Number 80, January 2010

ERCIM NEWS

European Research Consortium
for Informatics and Mathematics
www.ercim.eu

Special theme:

Digital Preservation

Also in this issue:

Keynote:
by Patricia Manson

Joint ERCIM Actions:
ACGT - Evolution of a Semantic Grid Infrastructure

R&D and Technology Transfer:
Impacts of an ICT Breakdown on the European Economy

ERCIM News is the magazine of ERCIM. Published quarterly, it reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology and Applied Mathematics. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of 9,000 copies. The printed version of ERCIM News has a production cost of €8 per copy. Subscription is currently available free of charge.

ERCIM News is published by ERCIM EEIG
BP 93, F-06902 Sophia Antipolis Cedex, France
Tel: +33 4 9238 5010, E-mail: contact@ercim.eu
Director: Jérôme Chailloux
ISSN 0926-4981

Contributions
Contributions must be submitted to the local editor of your country

Advertising
For current advertising rates and conditions, see
http://ercim-news.ercim.eu/ or contact peter.kunz@ercim.eu

ERCIM News online edition
The online edition is published at http://ercim-news.ercim.eu/

Subscription
Subscribe to ERCIM News by:
sending email to en-subscriptions@ercim.eu
or by filling out the form at the ERCIM News website:
http://ercim-news.ercim.eu/

Next issue:
April 2010, Special theme:
"Modelling and Simulation for Research and Industry".

ERCIM
European Research Consortium
for Informatics and Mathematics

Digital Preservation Research: An Evolving Landscape

Pat Manson
Head of Unit
Unit E3 "Cultural Heritage & Technology Enhanced Learning"
European Commission
Information Society and Media Directorate-General

Digital preservation research tackles the problems of keeping - preserving - digital content, particularly that which is born digital and, therefore, by definition does not exist in any other format. As early as the mid 1990s the European Commission recognised that this was an emerging and important issue and started funding pioneering research projects in digital preservation. At that time the challenge of managing digital content so that it could be accessed and used reliably in the future was one that was being confronted mainly by national libraries and archives, the key institutions with the mandate to keep publications and records for the future. They were at the sharp end of facing the problems posed by the new shifts towards electronic journals and towards electronic records.

Today the picture has changed and continues to change rapidly. In 2007, the International Data Corporation (IDC) estimated that the current size of the digital universe was 161 billion gigabytes or 161 exabytes and that this would increase sixfold by 2010. By 2008 it calculated it had already expanded to 281 exabytes and revised its four year estimate upwards from sixfold to tenfold. From a problem faced by archives and libraries, digital preservation is an issue affecting all domains which rely on digital data be it administrations, industry, research. Even as individuals we record all aspects of our lives and create and store our memories in digital form, through photos, blogs, e-mails etc.

One irony of the information age is that keeping information has become more complex than it was in the past. We not only have to save physical media and electronic files, we also need to make sure that they remain compatible with the hardware and software of the future.

So what does this mean for research? Of course, research is continuing to explore and develop solutions that will support libraries and archives, including audio-visual archives, in more efficient and cost-effective preservation, through automating workflows and decision making. However, there are new challenges for research arising out of: the increasing dependency on digital resources; the increasing volumes and complexity of digital resources; and the risks of losing digital resources or of having information that is no longer usable or understandable. At the same time, research needs to address the needs of organisations that are only now beginning to face the problems of keeping their digital content so that its authenticity and integrity can be maintained while ensuring that it can be transformed and used by new systems in the future.

European research is at the forefront of anticipating these challenges. Through FP6 and FP7 the objectives for the research have moved from a library/archive centric view to one that is increasingly focused on understanding the challenges posed by the nature of the digital content itself. This is leading our research projects to tackle new methods for web archiving ensuring the authenticity and integrity of the archived content which is characteristically distributed, dynamic and disappearing if not captured at the right point in time. The average life of a web page is less than that of the house fly. Scientific data (eg earth observation data) requires preservation systems that can handle significantly large volumes, document the original context, and curate the data so that it is usable and combinable in future uses. Often it is not an option to go back and re-capture these data and, for example, our models of climate change depend heavily on being able to understand and use data collected in the past often for different purposes. Digital objects are increasingly complex, combining text, image and embedded software. And that describes the objects of today without taking account of emerging softwares that may impact on their use in the future. At the same time, research needs to support the needs of organisations that are only now beginning to face the problems of keeping their digital content so that its authenticity and integrity can be maintained while ensuring that it can be transformed and used by new systems in the future.

As the volumes of information, the diversity of formats and types of digital object increase, digital preservation becomes a more pervasive issue and one which cannot be handled by the current approaches which rely heavily on human intervention. Research is needed on making the systems more intelligent. We need to accelerate the move from human monitoring and decision making to embedding reasoning and intelligence in the systems themselves.

For the research community, the challenge is also to build new cross-disciplinary teams that integrate computer science with library and archival science (and even with social and historical sciences). We need to ensure that future technological solutions for preservation are well founded and grounded in understanding what knowledge from the past and from today we need to keep for the future.

The views expressed in the article are the sole responsibility of the author and in no way represent the view of the European Commission and its services

2 **Editorial Information**

KEYNOTE

3 **Digital Preservation Research: An Evolving Landscape**
by Patricia Manson

JOINT ERCIM ACTION

6 **14th ERCIM Formal Methods for Industrial Critical Systems Workshop**
by María Alpuente, Byron Cook and Christophe Joubert

7 **ACGT - Evolution of a Semantic Grid Infrastructure**
by Alexander Hoppe and Manolis Tsiknakis

THE EUROPEAN SCENE

8 **The Future of CLOUD Computing – Report from EC CLOUD Computing Expert Group**
by Keith Jeffery

9 **The OpenAIRE Project - Open Access Infrastructure for Research in Europe**
by Donatella Castelli and Paolo Manghi

11 **Alliance for Permanent Access to the Records of Science**
by Keith Jeffery

SPECIAL THEME

Digital Preservation
coordinated by Ingeborg Solvberg, NTNU, Norway; and Andreas Rauber, Vienna Technical University/AARIT, Austria

Introduction to the Special Theme

10 **Digital Preservation**
by Ingeborg Solvberg and Andreas Rauber

13 **Drivers for Digital Preservation**
by Matthias Hemmje and Ruben Riestra

Frameworks and Systems

14 **The Planets Interoperability Framework**
by Ross King

15 **PROTAGE: Long-Term Digital Preservation Based on Intelligent Agents and Web Services**
by Xiaolong Jin, Jianmin Jiang, and Josep Lluis de la Rosa

17 **SHAMAN: Sustaining Heritage Access through Multivalent Archiving**
by José Borbinha

18 **HOPPLA - Archiving System for Small Institutions**
by Michael Greifeneder, Stephan Strodl, Petar Petrov and Andreas Rauber

20 **The CARA Approach for Long-Term Preservation and Exploitation of Medical Images and Reports**
by Hanan Bouzid, Mimouna Guenfoud, Andreas Jahnen, Pierre Plumer, Cédric Pruski and Frédéric Zucconi

21 **Designing a Trusted Distributed Long-Term Archive for Health Records**
by Frej Drejhammar

22 **Tackling the Problem of Complex Interaction Processes in Emulation and Migration Strategies**
by Klaus Rechert and Dirk von Suchodoletz

Tools

24 **Trustworthy Preservation Planning with Plato**
by Christoph Becker, Hannes Kulovits and Andreas Rauber

26 **Towards Document Process Preservation: Xerox Launches Document Process Modelling Technology 'Xeproc©'**
by Thierry Jacquin, Hervé Déjean, Jean-Pierre Chanod

27 **Cyclops: An Interface for Producing and Accessing Archives of Artistic Works**
by Nicolas Esposito, Bruno Bachimont and Erik Gebers

29 **Magnetic Tape Storage and the Growth of Archival Data**
by Jens Jelitto, Mark Lantz and Evangelos Eleftheriou

Testbeds

30 **The ESA Approach to Long-Term Data Preservation using CASPAR**
by Sergio Albani

32 **Digital Preservation of Interactive Multimedia Performances**
by Kia Ng

33 **The Planets Testbed: A Collaborative Research Environment for Digital Preservation**
by Brian Aitken and Andrew Lindley

OAIS and other standards

35 **Which Repositories are Worth their Salt?**
by David Giaretta

36 **Best Practices for an OAIS Implementation**
by Luigi Briguglio, Carlo Meghini,and David Giaretta

37 **Preserving the Past for the Future: Digital Technology for Film Archives**
by Arne Nowak

39 **Considering Software Preservation**
by Brian Matthews, Arif Shaon, Juan Bicarregui, Catherine Jones, Esther Conway and Jim Woodcock

40 **Electronic Records Management in Luxembourg: Challenges and Perspectives**
by Lucas Colet

Knowledge Management for Digital Preservation

41 **Knowledge Management for Digital Preservation**
by Yannis Tzitzikas and Vassilis Christophides

43 **Automating the Ingestion and Transformation of Embedded Metadata**
by Yannis Tzitzikas and Yannis Marketakis

44 **The Art of Preserving Digital Creativity in Planets**
by Andrew McHugh and Leonidas Konstantelos

45 **User-Centered Digital Preservation of Multimedia**
by Egon L. van den Broek, Frans van der Sluis and Theo E. Schouten

Surveys

47 **Communication and Preservation in Academic Research: Current Practices and Future Needs**
by Filip Kruse and Annette Balle Sørensen

48 **Preservation Planning: User Requirements for Digitally Preserved Materials**
by Annette Balle Sørensen and Filip Kruse

R&D AND TECHNOLOGY TRANSFER

50 **Five Steps to Green Desktop Computing**
by Howard Noble, Kang Tang, Daniel Curtis and Paul Jeffreys

52 **Ranking the Stars with MonetDB**
by Annette Kik and Milena Ivanova

53 **3D Reconstruction by Multimodal Data Fusion**
by Dmitry Chetverikov and Zsolt Jankó

55 **The Virtualization Gate Project**
by Edmond Boyer, Benjamin Petit and Bruno Raffin

56 **Providing Web Accessibility for the Visually Impaired**
by Barbara Leporini, M.Claudia Buzzi and Marina Buzzi

57 **ICASE Project: New Challenges in Computer-Based Assessment**
by Thibaud Latour and Sandrine Sarre

58 **Impacts of an ICT Breakdown on the European Economy**
by Fabio Bisogni, Simona Cavallini and Cristiano Proietti

EVENTS

60 **D4Science World User Meeting**
by Donatella Castelli, Marc Taconet and Virginie Viollier

60 **Announcements**

IN BRIEF

62 **InterLink Research Roadmaps Published**

62 **New book on Advanced Computational Methods**

63 **INRIA is Recruiting 45 Researchers**

63 **Christer Norström new CEO for SICS**

63 **EIT ICT Labs Wins Prestigious European Race for Excellence in Innovation**

14th ERCIM Formal Methods for Industrial Critical Systems Workshop

by María Alpuente, Byron Cook and Christophe Joubert

The 14th ERCIM Formal Methods for Industrial Critical Systems (FMICS) workshop was held in Eindhoven, The Netherlands, on 2-3 November 2009. It was part of FMweek, the first Formal Methods Week, which offered a choice of events in the area including TESTCOM/FATES (Conference on Testing of Communicating Systems and Workshop on Formal Approaches to Testing of Software); FACS (Formal Aspects of Component Software); PDMC (Parallel and Distributed Methods of verification); FM2009 (Symposium of Formal Methods Europe); CPA (Communicating Process Architectures); FAST (Formal Aspects of Security and Trust); FMCO (Formal Methods for Components and Objects); and the REFINE workshop.

The aim of the FMICS workshop series, organized annually by the ERCIM FMICS Working Group, is to provide a forum for researchers who are interested in the development and application of formal methods in industry. In particular, these workshops are intended to bring together scientists and engineers who are active in the area of formal methods and are interested in exchanging their experiences in the industrial usage of these methods. These workshops also strive to promote research and development for the improvement of formal methods and tools for industrial applications.
The topics chosen for FMICS 2009 included, but were not restricted to:

- design, specification, code generation and testing based on formal methods
- methods, techniques and tools to support automated analysis, certification, debugging, learning, optimization and transformation of complex, distributed, real-time and embedded systems
- verification and validation methods that address shortcomings of existing methods with respect to their industrial applicability (eg scalability and usability issues)
- tools for the development of formal design descriptions
- case studies and experience reports on industrial applications of formal methods, focusing on lessons learned or new research directions
- impact and costs of the adoption of formal methods
- application of formal methods in standardization and industrial forums.

In response to the call for papers, 24 contributions were submitted from sixteen countries. The Program Committee selected ten papers, basing this choice on their scientific quality, originality and relevance to the workshop. Each paper was reviewed by at least three program committee members or external referees. The workshop also included four invited contributions by Dino Distefano (Queen Mary, University of London, UK), Diego Latella (CNR/ISTI, Italy), Thierry Lecomte (ClearSy, France), and Ken McMillan (Cadence Berkeley Labs, USA), as well as six poster descriptions. The resulting program offered the participants a complete landscape of the recent advances in this area. On-site proceedings were published by Springer-Verlag as volume 5825 of Lecture Notes in Computer Science.

From left: Christophe Joubert, María Alpuente, Bárbara Vieira and Alessandro Fantechi.

Following a tradition established over the past few years, the European Association of Software Science and Technology (EASST) offered an award to the best FMICS paper. This year, the award was given to Bárbara Vieira from Universidade do Minho, Braga, Portugal for the paper 'Correctness With Respect to Reference Implementations', written together with José Bacelar Almeida, Manuel Barbosa and Jorge Sousa Pinto.

The award was presented by María Alpuente, PC co-chair of FMICS 2009, Christophe Joubert, workshop chair of FMICS 2009, and Alessandro Fantechi, FMICS Working Group coordinator since November 2008 (see photo).

Links:
http://users.dsic.upv.es/workshops/fmics2009/
http://www.inrialpes.fr/vasy/fmics
http://www.win.tue.nl/fmweek

Please contact:
Christophe Joubert
Universidad Politécnica de Valencia/SpaRCIM, Spain
E-mail: joubert@dsic.upv.es

ERCIM Innovation

ERCIM Innovation no. 2 has been published in November 2009. This magazine is part of ERCIM's strategy to foster ICT innovation for the benefit of the European economy and society. This issue showcases innovative developments in ERCIM member institutions and their spin-off companies with a view to finding commercial exploitation partners.

The magazine is available online at http://www.ercim.eu/publications/ercim-innovation. Printed copies can be requested from catherine.marchand@ercim.eu

ACGT - Evolution of a Semantic Grid Infrastructure

by Alexander Hoppe and Manolis Tsiknakis

ACGT (Advancing Clinico-Genomic clinical Trials on cancer: open grid services for improving medical knowledge discovery) is an Integrated Project funded by the 6th Framework Programme of the European Commission, under the action line 'Integrated biomedical information for better health'. The overall vision of the project is to become a pan-European voluntary network or grid connecting individuals and institutions, which will enable the sharing of data and tools. This will create a European Wide Web of cancer clinical research with the goal of speeding the delivery of innovative approaches for the prevention and treatment of cancer.

Life sciences are currently at the centre of an information revolution. The development of new techniques and tools is making possible the collection and organization of biological information at an unprecedented level of detail and in extremely large quantities. With respect to cancer research, the use of high-throughput technology has resulted in an explosion of information and knowledge about cancers and their treatment.. However, the lack of an open and shared information infrastructure is preventing clinical research institutions from being able to mine and analyse disparate data sources. Our inability to share technology and data that have been developed by different organizations is severely hampering the research process. As a result, very few cross-site studies and multicentre clinical trials are being performed.

The ultimate objective of ACGT is therefore the development of a semantic grid infrastructure facilitating a common platform for researchers, clinicians, biostatisticians and software developers that will (i) facilitate seamless and secure access to heterogeneous, distributed multilevel databases; (ii) provide a range of semantically rich reusable, open tools for the analysis of such integrated, multilevel clinico-genomic data in the context of discovery-driven (eScience) workflows; (iii) support the creation and management of dynamic virtual organizations (VOs); and (iv) do so in full compliance with existing ethical and legal regulations.

To achieve this high-level goal, the project is delivering a Master Ontology on Cancer – soon to be submitted for membership in the Open Biomedical Ontologies (OBO) foundry, and an innovative software mediation tool, the ACGT Mediator, used for hiding the complexity of query translation and data integration. One fundamental component is the implementation of a clinical trial management system, called ObTiMA, based on an ontology-driven software development process, providing uniform access to heterogeneous clinical trial, genetic and public biological databases. Another is the implementation of a range of tools for bioinformatics and biomedical data analysis and visualization, and grid-enabled knowledge discovery. The gridified version of the R package (GridR) enables the use of the complete ACGT platform and services by choosing the (familiar) R language.

The main outcome of the work to date is the development of a generic integration framework using semantic Web technology, standards and tools, which enables the integration of not only ACGT compliant services but also of other third-party bioinformatics services like BioMOBY.

The status of the two pilot trials is being continuously monitored. Clinical, imaging and genetic data are used for the evaluation and defining codes for the Oncosimulator in order to simulate tumour growth and response to treatment. A range of new clinico-genomic trials and scenarios were defined to use the ACGT platform. Keeping end-users in line with ethical and legal requirements demands special consideration. To address these needs, the CDP (Center for Data Protection) promoted at European level by the consortium was established on the legal side, and several security mechanisms were implemented (eg an innovative data anonymization and privacy enhancing tool) on the technical side.

Two large scenarios were developed and used to test the logical and execution architecture in general as well as separate tools and services. A consequent user-developer loop led to experiences gathered in building and demonstrating these end-to-end demonstrations. Training-evaluation sessions with end users were conducted to further increase robustness, reduce technical complexities in using the systems/services and improve the usability of the whole platform. As the ontology-based clinical trial management system for ACGT (ObTiMA) plays a crucial role from a clinical perspective the main focus was to steadily improve the functionality of ObTiMA including a user-friendly integration of the Master Ontology.

In 2009 ACGT established a formal collaboration with the European Organisation for Research and Treatment of Cancer (EORTC). EORTC's experience is an invaluable asset in achieving the project goals. ACGT was also made visible to end-user communities by being presented to important organizations in the field of oncology, such as the Breast International Group (BIG), and the International Society of Paediatric Oncology (SIOP) and by inviting external experts (such as an the European Clinical Research Infrastructures Network (ECRIN) representative) to an advisory board meeting, where highly positive feedback was received.

Mature demonstrations of all the technological components functioning as a whole are now running, and since the project has been extended to July 2010 we believe that by the end of the project we will deliver a solid platform for end-user communities.

The ACGT project is managed by ERCIM.

Link:
http://www.eu-acgt.org/

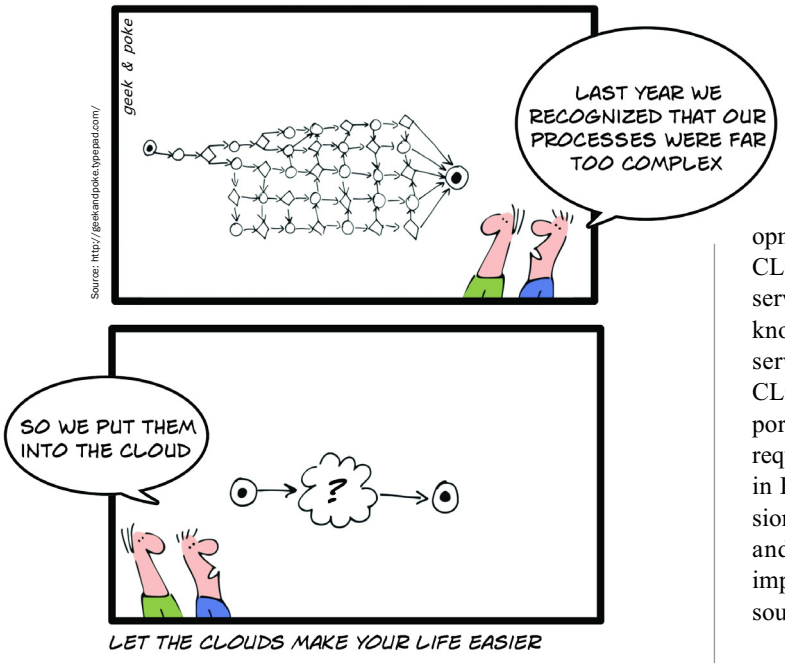
Please contact:
Manolis Tskinakis
FORTH, Greece
E-mail: tsiknaki@ics.forth.gr

The Future of CLOUD Computing – Report from EC CLOUD Computing Expert Group

by Keith Jeffery

The EC DG Information Society and Media, Software & Service Architectures and Infrastructures convened an expert group on CLOUD Computing in 2009. The group has produced a report that outlines the future directions of Cloud Computing research. The findings of the final report of the group of experts and the orientations for future research in cloud computing will be presented and discussed on 26 January 2010, in Brussels.

The group was moderated by Burkhard Neidecker-Lutz of SAP Research and the author representing ERCIM. The rapporteur is Lutz Schubert of HLRS at Stuttgart and Maria Tsakali, the official responsible from the European Commission. After several meetings of the group, one open meeting and much internet-based discussion a final report has been produced. It will be presented at an event organised by the EC DG Information Society and Media, Software & Service Architectures and Infrastructures on 26 January 2010 in Brussels http://cordis.europa.eu/fp7/ict/ssai/events-20100126-cloud-computing_en.html.



The report characterises different kinds of CLOUDs (both existing and future) and suggests the open research issues that need to be addressed. While recognising that CLOUDs are being used now, both privately within an organization and as a service external to an organization, the report indicates a much greater possible utilisation if the research issues (which address limitations of CLOUDs as seen from an end-user perspective) have solutions provided.

CLOUDs offer potentially unlimited scalability (both up and down) and virtualization of resources so system management issues are hidden from the end-user. CLOUDs can offer IaaS (infrastructure as a service), PaaS (platform as a service), AaaS (application(s) as a service) or a totally outsourced ICT capability. The use of CLOUD services can reduce greatly the carbon footprint of an organization due to the efficiencies of a datacenter environment and so can claim 'green ICT' credentials.

The major technological concerns which act as barriers to take-up of CLOUD computing centre on (a) security, trust and privacy; (b) lack of standardization and therefore supplier lock-in; (c) insufficient virtualization to provide real hiding of systems management (especially in resource sharing/failover); (d) data movement and management; (e) programming models to provide the required elasticity; (f) systems / services development methods.

There are also non-technological concerns, mainly (a) business / economic / cost models for CLOUD computing (including 'green ICT' aspects) that are robust and realistic; (b) legalistic issues concerning data processing in another country or multiple countries and/or using an outsourced service.

The report identifies three areas where Europe could become prominent in CLOUDs: (1) large companies – especially but not exclusively the telecommunications industry – could provide CLOUD services; (2) development by companies (especially SMEs) of services for the CLOUD environment leading to an open market in CLOUD services matching that in goods, services, human capital and knowledge; (3) provision of business model and legalistic services (including 'green ICT') to accompany the use of CLOUD computing. The report requests the EC (a) to support R&D in the technological aspects and (b) to set up the required governance framework for CLOUDs to be effective in Europe. Subsidiary recommendations include the provision of testbeds, joint collaboration groups across academia and industry, standardisation and open source reference implementation (rather like W3C) and the promotion of open source solutions.

So, what is all the fuss about? People ask if CLOUDs are not the same as one or many of: Cluster computing, GRIDs, Future Internet, the Internet of Things or SOA (Service Oriented Architecture). The answer is – as usual – yes and no! Basically CLOUDs provide a new perspective on ICT provision for an organisation. The CLOUDs technological solution (virtualization, resource sharing, elasticity) allows organisations to do their ICT differently.

Internal CLOUDs allow the organization to optimise ICT in one datacenter (almost certainly replicated – probably externally - for business continuity) and so increase server utilisation and allowing server hibernation or switch off in periods of low demand. This reduces capital expenditure and associated maintenance / systems administration. It also reduces energy consumption. Departments in the organization buy services provided in the CLOUD so having better cost-management of their ICT; similarly the ICT department is more efficient.

External CLOUDs allow the organisation to outsource some or all of its IT to another organisation providing the service (and probably providing such a service to several other organisations). This leaves the organisation to concentrate on its primary business and treat ICT as a utility service.

In both cases there is a shift in accounting for ICT from a CAPEX (Capital expenditure) dominated state to an OPEX (Operational expenditure) state. This means that an organisation has neither to reduce its liquid capital not take out expensive loans to procure ICT but can 'pay as you go'.

Interestingly, this concept links up with recent work on Data and Information Spaces where problems of integration across heterogeneity in data and information – not unlike in infrastructural ICT resources – are at least partially overcome by the 'pay as you go' philosophy.

CLOUDs provide a real opportunity for European business. However, for this to be realised there are research issues to be addressed both technological and non-technological. With the strong background in Europe in the relevant technologies and in dealing with legal and cultural heterogeneity the challenges surely will be met and the problems overcome.

Link:
http://cordis.europa.eu/fp7/ict/ssai/events-20100126-cloud-computing_en.html

Please contact:
Keith Jeffery
STFC, UK
E-mail: keith.jeffery@stfc.ac.uk

The OpenAIRE Project - Open Access Infrastructure for Research in Europe

by Donatella Castelli and Paolo Manghi

OpenAIRE will deliver "an electronic infrastructure and supporting mechanisms for the identification, deposition, access, and monitoring of FP7 and ERC funded articles", where the main supporting mechanism will be a European Helpdesk System. The infrastructure will be based on state-of-the-art software services of the D-Net Software Toolkit developed within the DRIVER and DRIVER-II projects and the Invenio digital repository software developed at CERN.

Although simple in conception, unrestricted availability of research publications (as well as scientific data) is still far from reality; the implementation of policies that promote Open Access to these important research products has proved to be challenging. For this reason, the recent European Commission Open Access Mandate pilot was followed by a Call soliciting pilot projects aimed at developing a software system addressing such issues in the context of published peer-reviewed articles reporting on outcomes of FP7 and European Research Council (ERC) projects in seven selected disciplines: energy, environment, health, cognitive systems/interaction/robotics, e-infrastructures, science in society, and socioeconomic sciences/humanities.

The OpenAIRE project was financed to meet the pilot requirements. Thematically, the project focuses on publications in the pilot seven disciplines and on research datasets in a subset of them: environment, health, cognitive systems/interaction/robotics, and socioeconomic sciences/humanities. Geographically, it has a definitive "European footprint" by covering the European Union in its entirety, engaging people and scientific repositories in almost all 27 member states and one associated state (Norway).

The project will deliver a technical infrastructure, through which Open Access publications and research data will be harvested, author-ingested, curated and fruitfully combined with EC project information, and a networking infrastructure, through which the EC Open Access mandate and OpenAIRE system will be disseminated across Europe and beyond.

OpenAIRE technical infrastructure

The technical infrastructure will be based on the D-Net software toolkit, developed within the DRIVER and DRIVER-II projects, and the Invenio digital repository software, developed at CERN. These already offer most of the desired functionality but will be enhanced and complemented with services developed within OpenAIRE to address critical requirements and issues that arise in the target environment and require further investigation.

The infrastructure will support the OpenAIRE Information Space, to be populated with bibliographic metadata records of Open Access publications funded in FP7/ERC projects. Such records will either be deposited directly by authors in the Invenio-powered repository (established to host so-called repository-orphan publications) or, after author or institution notification, harvested (and later curated) from the institutional and thematic repositories where they were originally ingested. Moreover, descriptive metadata about EC/ERC funded projects will be also ingested to be linked with the related articles and to enable further content analysis. In particular, monitoring tools and services will infer relevant information and statistics on FP7 and ERC funded research from articles, research data, project metadata and the relationships between them. Finally, OpenAIRE will develop a portal for easily searching, browsing and accessing its rich Information Space and statistic data.

OpenAIRE will work with several subject communities to explore the state of the art of research datasets management and their combination with research publications. Prototype services will be developed to demonstrate the feasibility of complex processes and structures and show the benefit for researchers in both depositing and re-using these combined information resource packages.

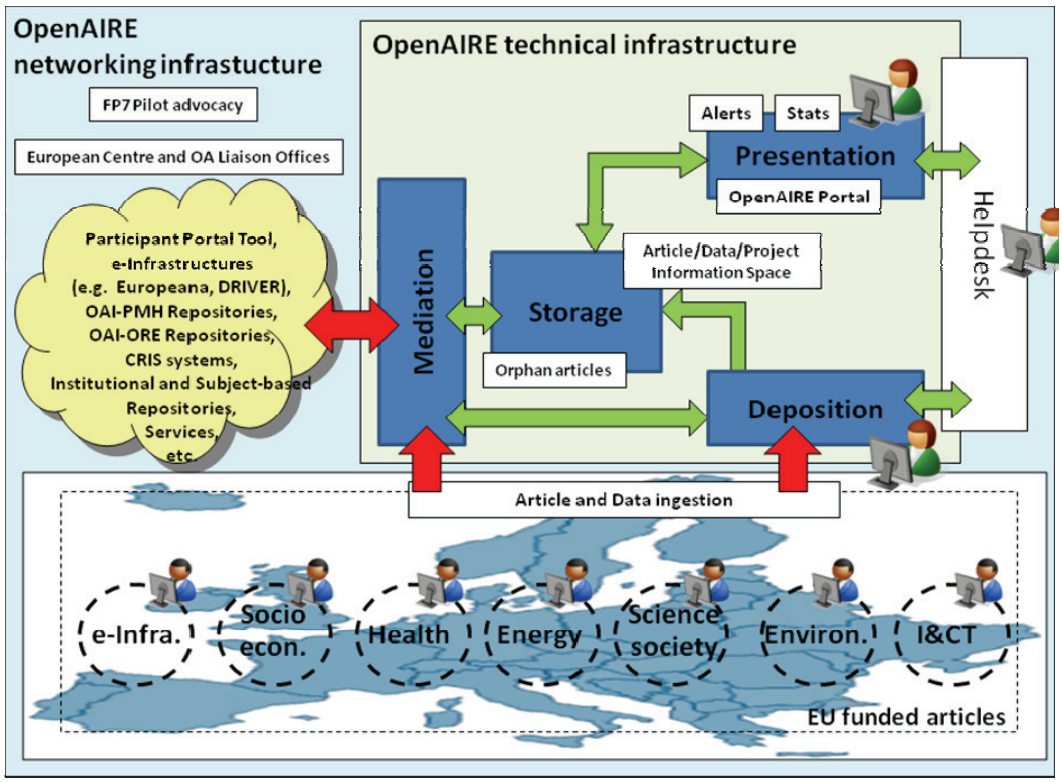


Figure 1: OpenAIRE infrastructure.

OpenAIRE networking infrastructure

The Open Access mandate of the European Commission has been an indispensable step towards free access to research results from Europe. Experiences with other Open Access mandates, however, show that acceptance and broad take-up by the scientific community critically depends on accompanying support mechanisms, as any activity beyond the actual research and publishing process is considered by researchers as administrative burden and essentially as a waste of time. Therefore, in addition to organizing advocacy, promotion, and training events, OpenAIRE will establish a networking infrastructure supporting structures and tools that enable article deposition to be carried out as easily and efficiently as possible, thereby ensuring that a critical mass of articles will be deposited. For this purpose, the project will deliver a European Helpdesk System, which will consist of a European Centre and national Open Access liaison offices in all but one EU member states and one associated state (Norway). The European Helpdesk System will be accessible online through the envisaged portal "OpenAIRE.eu", which will also provide access to the FP7 and ERC research articles. From this portal, links will go out to national Open Access support pages, such as "open-access.net" (Germany), "rcaap.pt" (Portugal), "recolecta.net" (Spain), etc. The portal will also link to the European Commission "Participants Portal", CORDIS, and other relevant trans-national Open Access initiatives and organisations, such as SPARC Europe and learned societies.

Conclusions

By facilitating access to scientific literature, OpenAIRE will allow European researchers to become more efficient in conducting their own investigations. Scientists and scholars from around the globe will be attracted to the OpenAIRE infrastructure and portal for their own work. A wealth of usage statistics and other impact metrics will become available to European policy makers and strategists, who will be able to use them to sharpen and refine policies on open access or other issues at the European level. Access to such statistics should also assist in identifying important scientific trends, key researchers in a particular domain, or other relevant variables, which may be helpful to better understand the European research landscape.

Links:

ftp://ftp.cordis.europa.eu/pub/fp7/docs/open-access-pilot_en.pdf
http://www.openaire.eu
http://www.driver-repository.eu/D-NET_release
http://cdsware.cern.ch/invenio/index.html

Please contact:
Donatella Castelli
ISTI-CNR, Italy
E-mail: donatella.castelli@isti.cnr.it

Paolo Manghi
ISTI-CNR, Italy
E-mail: paolo.manghi@isti.cnr.it

Alliance for Permanent Access to the Records of Science

by Keith Jeffery

The Alliance for Permanent Access to Records of Science (usually referred to by its acronym APA) has been set up by concerned persons representing organisations working on digital preservation/curation who believe that the record of science must be curated for future use. APA is a not-for-profit organization registered in the Netherlands.

Some information has only one chance of capture (eg state of the earth at any one time); other information is expensive to reproduce by collection/observation, experiment or simulation. APA membership includes national libraries, large data/computing centres; funding organizations, scholarly publishers and organizations concerned with digital preservation/curation. STFC is a member and indeed the author currently holds the chairmanship of the Executive Board. The organisation intends to provide the point of reference in

Andreas Rauber speaking about R&D and Technical Tools in Digital Preservation.

Europe (and wider) for all matters related to digital preservation/curation of the scientific record. Indeed, already APA has interacted strongly with and/or participated in groups such as e-IRG (the e-Infrastructure Reflection Group) associated with ESFRI (the European Strategy Forum on Research Infrastructures) and a member of ERC (European research Council) attended the first conference. At present APA is working on organising a workshop on metadata for curation.

APA organized its first annual conference in Budapest in November 2008 with a theme concerning business models for curation and preservation. The second was held in The Hague November 2009. At the second conference keynote talks were given on PARSE.Insight by Dave Giaretta of STFC and on GEANT and e-Infrastructure by Konstantinos Glinos of the EC. A talk on R&D and Technical Tools was given by Andreas Rauber (Vienna University of Technology and ERCIM vice president and Director for AARIT, and previous Cor Baayen Award Winner). With sessions on scientific community insights and cross-community insights the conference achieved its objective of both informing and stimulating common approaches.

The Conference Chair was Peter Tindemans who in many ways is the father of APA. His key summarising messages were that we needed better R&D leading to technology for preservation/curation but above all we needed R&D to generate sustainable business models and legal models for rights backed by policies (including research funders providing resources for curation) and that APA should communicate its expertise to the EC and to national governments as well as to other appropriate international bodies.

Link:
http://www.alliancepermanentaccess.eu

Please contact:
Wouter Spek
APA Executive Director
E-mail: Wouter.Spek@KB.nl

ERCIM "Alain Bensoussan" Fellowhsip Progtramme

ERCIM offers fellowships for PhD holders from all over the world. Fellowships are of 18 month duration, spent in two of the ERCIM member institutes (the duration might be increased to 24 months for the next round), or of 12 months duration spent in one ERCIM institute.

Next deadline for Applications: 30 April 2010

Conditions

Applicants must

- have obtained a PhD degree during the last eight years (prior to the application deadline) or be in the last year of the thesis work with an outstanding academic record
- be fluent in English
- be discharged or get deferment from military service
- have completed the PhD before starting the grant (a proof will be requested)
- the fellowship is restricted to two terms (one reselection is possible).

ERCIM encourages not only researchers from academic institutions to apply, but also scientists working in industry.

A detailed description of the programme, its topics and the online application form is available at: http://www.ercim.eu/activity/fellows

Introduction to the Special Theme

Digital Preservation

by Ingeborg Solvberg and Andreas Rauber

When digital representations of information objects first became available, they were seen as the solution to a myriad of problems relating to replication, distribution, ease of use and maintenance. Instead of filling up shelves and filing cabinets with documents, numeric data or fragile physical objects, the digital versions of these data promised to be space saving. They could also be copied and stored without loss or degradation - right up until the moment when the hardware and software environment required to interpret them became obsolete and they were suddenly lost (not degrading slowly, but in a very binary fashion, suddenly and completely lost).

Digital objects need specific viewer applications to be interpreted. These, in turn, need specific libraries installed on a specific operating system. The operating system runs on very specific hardware configurations for which drivers are provided. At the same time, digital objects are preserved on storage media, which themselves are fragile, as is the device technology needed to read them. All of these factors combine to ensure that digital objects are severely vulnerable to obsolescence: if any of the layers in the dependency tree is lost, the entire object ceases to be accessible and usable. On top of that, we find vulnerabilities regarding the interpretation of objects, documenting their provenance and limitations, ensuring that they are authentic and trustworthy.

Digital Preservation encompasses the activities that try to ensure that digital objects remain accessible and usable in an authentic way for long periods of time. Here, 'long term' can mean anything from a few years to decades and ultimately centuries: anything that is long enough to experience technological change that threatens the availability of a digital object.

In the last few years we have witnessed drastic changes in the field of digital preservation. First of all, the topic has moved from niche specialist discussions into broad mainstream awareness: public media are publishing and broadcasting detailed accounts of the challenges of keeping digital objects available for longer times than the increasingly short technology cycles allow. This has also led to a sudden recognition of the size of the stakeholder community, which includes not only data archives and heritage institutions (for a long time the sole drivers of this topic), but (e)-governments, industry, small and medium enterprises, and society at large, as our lives become increasingly shaped and based upon digital processes and artefacts.

The second drastic change we have witnessed recently is the fact that more solutions are becoming available and being deployed. This may be attributed to the provision of significant R&D funds for this massive challenge of keeping digital objects accessible and usable. Numerous research projects have been supported globally in recent years, with tools, guidelines and standards maturing to the point of production-level deployment. Still, the problem is far from being solved. While some areas are already quite well understood, for others we can only guess at the suitability of the solutions available; while for some we have solid tools available, for other areas we have only prototypes that hint at possible solutions or show the scale of the problem; while in some cases we have well-defined procedures in place, there are others where we have done little more than bought a few years of time in which to come up with better solutions.

This edition of ERCIM News brings together an exciting overview of the achievements and current activities in the field of Digital Preservation. It presents some of the most outstanding work in the field, ranging from issues of standardization and the development of models, via the creation of concrete solu-

tions, tools and frameworks to deploy them, to case studies and reports of evaluating and discussing the state of research and development in a number of case studies and testbed settings. We believe that this selection of articles provides an excellent overview of and entry point to this challenging domain, serving both as a starting point to identify solutions to one's own personal and institutional preservation needs as well as a good overview of the topics currently being addressed in research projects.

Please contact:
Andreas Rauber
TU Vienna, Austria, AARIT
E-mail: rauber@ifs.tuwien.ac.at

Ingeborg Solvberg
NTNU, Norway
E-mail: Ingeborg.Solvberg@idi.ntnu.no

Links:

European Initiatives:

Alliance for Permanent Access:
http://www.alliancepermanentaccess.eu

PLANETS - Preservation and Long-term Access through Networked Services: http://www.planets-project.eu

KEEP - Keeping Emulation Environments Portable:
http://www.keep-project.eu

PrestoPRIME - Keeping Audiovisual Content Alive:
http://www.prestoprime.eu

LiWA, Living Web Archives:
http://www.liwa-project.eu

PROTAGE - Preservation Organizations using Tools in Agent Environments: http://www.protage.eu

SHAMAN - Sustaining Heritage Access through Multivalent Archiving
http://www.shaman-ip.eu

3D-COFORM - Tools and Expertise for 3D Collection Formation
http://www.3d-coform.eu

Recently terminated projects with relevant results:

CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval:
http://www.casparpreserves.eu

DPE - DigitalPreservationEurope:
http://www.digitalpreservationeurope.eu

other Links:

Digital Preservation Videos on YouTube:
http://www.youtube.com/user/wepreserve

Drivers for Digital Preservation

by Matthias Hemmje and Ruben Riestra

Digital preservation (DP) is becoming a relevant issue for ensuring the future accessibility and usability of knowledge, information and data that only exist in digital formats, ie the so-called 'born-digital' content. The overwhelming expansion of this content is creating a continuously increasing spectrum of opportunities and threats to organizations exposed to the need to preserve such digital assets for decades or even centuries.

The current demand for DP solutions and services is mainly driven by institutions having a legal mandate to handle the preservation of society's collective memory, the so-called evant societal assets, eg the records of science or governmental bodies. However, it needs to be recognized that other businesses and even whole industries within globalizing economies are dealing with information resources which need to be preserved for decades. To achieve a more differentiated view of DP, we must define its various domains of application more explicitly.

Ongoing research in Europe has shown that because pioneers such as the memory institutions (MIs) have the only legal mandate to preserve the collective cultural heritage of society, they therefore carry the sole responsibility. That is why they have been the first to encounter the challenges and complex problems of preserving born-digital content without any kind of fallback solution like digitization from the original physical information carrier or storage media in case of failure. Scientific research institutions (SRIs) as well as to some degree the data centres which are supporting them have a similar problem, although it is not necessarily their task to preserve but only to collect and provide access to scientific data and content.

Aside from the MIs and SRIs, economically the most important application domains that are relevant to this problem are the many different types of information-dependent industrial sectors that generate business value from their specific application knowledge and at the same time cater for the various demands of our society. Such Businesses, Corporations, and any kind of economically acting Enterprises (BCEs) represent the third and most important application domain for DP technology. Many BCEs are not only involved in the production of physical services and goods, but increasingly take part in the processing of information and knowledge within so-called knowledge/value chains of today's 'knowledge economy'. For all of these, preservation of digital content is of crucial importance because digital information objects and knowledge resources are an essential part of their business processes and therefore one of their most important resources.

Despite the fact that DP is not a concept or term usually recognised by BCEs, explicit and very specific archives of digital information assets supporting access and reuse of information in the long term are strongly integrated into most enterprises' policies and data management practices. This means that businesses, companies and enterprises have been involved and have experience in the preservation of digital objects that can be considered digital assets, often in multiple cycles, over the last 30-50 years. In enterprises however, DP processes are proprietary and part of infrastructures that support IT

workflows to secure the optimization of operating expenses. As DP is a continuous process, it forms part of the integrated management of digital content and assets in such infrastructures (access, printing, reusing etc) and is difficult to separate from the work routine and from its operational context. In a business environment therefore, DP is the responsibility of data and asset management or legal compliance. In other words, it is driven by IT and legal departments, and is not part of the corporate mission as is the case for MIs. The common feature between MIs and BCEs is the record-keeping aspect. BCEs are safekeeping their organization's records under the broad definitions of a Unified Information Management. Furthermore, it can be recognized that in recent years, BCEs have taken on a more specific, new, but still evolving meaning that refers to the storage and preservation of the organization's digital information or knowledge assets, either for compliance or as a source of revenue.

In summary, if the global development and spread of the Information Society and its knowledge-based economies continues with its current or even accelerating speed, it will further increase the need to preserve born-digital content. Certainly the future demand for DP technology, infrastructure, tools and solutions is likely to accelerate: this process already has significant momentum.

Please contact:
Matthias L. Hemmje
University of Hagen, Germany
E-mail:
Matthias.Hemmje@FernUni-Hagen.de

The Planets Interoperability Framework

by Ross King

The Planets Interoperability Framework is a software infrastructure for the preservation of digital documents, and was developed as part of the European Integrated Project Planets. It provides the technical environment that governs the integration of the Planets end-user applications with preservation services and data repositories. Since most of the institutions that will be interested in the Planets applications already have some kind of archiving system in place, archival storage is not part of what Planets delivers. Instead, our approach is to provide a framework and services that can be integrated with existing systems. The design of the framework was driven by the requirements of logical preservation in libraries and archives, including demands for a robust and extensible infrastructure for the characterization and migration of digital documents.

The Planets Interoperability Framework (henceforth referred to as the IF) provides a service-based infrastructure that leverages a number of standards and open source tools. The core of the IF implementation is based on the Java Enterprise Edition (Java EE 5) standard, which among other things provides a framework for the efficient implementation of Web services and Web applications. The IF installation package includes a pre-configured JBoss application server that provides common services like single-sign-on, user management, authorization and authentication. This application server provides the container for web-based preservation applications (the Planets Testbed application and the Planets Preservation Planning Tool, PLATO). In addition, a number of commonly required software components and their associated APIs are bundled with the IF; for example, a component for user management, a component for service registration and discovery, and a component for executing preservation workflows. Preservation action tools are deployed as Web services are hosted as a distributed network.

The Interoperability Framework enforces a set of standard Web service profiles for preservation services and a common model for the digital objects on which preservation actions are carried out. The interfaces define atomic preservation actions such as Identify, Characterize, Compare, Modify, Migrate, and View. Preservation tools that are provided using these interfaces can be easily registered with a Planets IF instance and immediately used within Planets workflows. A preservation workflow typically consists of a sequence of parameterized preservation actions, carried out in a specific order, in which the output parameters of one action are validly mapped to the input parameters of the following action. An example of a preservation workflow could be: for a given file, first identify a file format, then validate the document against the format, then determine a number of significant characteristics, then migrate it to a new format, then characterize the new file, then compare with the original. Each of these steps involves a number of different services within the Planets architecture; hence, the orchestration of these services is required. In general, the IF allows one to formally and technically describe preservation processes through a workflow template system. The aim of this approach is to shield the user from the complexity of the underlying architecture and implementation issues, allowing non-experts (i.e. librarians and archivists) to create and execute preservation workflows.

Thus the Workflow Execution Engine (WEE) provides an essential component within the IF service environment. We surveyed service orchestration approaches and experimented with WS-BPEL (Web Service Business Process Execution Language). WS-BPEL is an XML-based workflow description language for SOAP-based Web services. Within the IF, experimental preservation workflows were implemented using WS-BPEL v2.0 definitions and the JBPM (JBoss Business Process Management module) as a Workflow Execution Engine. The Eclipse BPEL Visual Designer served as a graphical interface for designing and visualizing the process flow. However, work in this direction was hindered by two difficulties; first, that the BPEL language is quite powerful but also low-level and hence complex;

and second, at the time we conducted the experiments, BPEL related-tools proved to be not yet mature. Both points turned out to be a major hindrance for implementing preservation workflows by non-BPEL experts.

Consequently, we chose to implement a much simplified, custom workflow description language and corresponding execution engine. The Planets WEE is based on a high-level application programming interface (API) and a corresponding template mechanism. This allows workflow developers to build abstract workflow definitions from Java components and serialize them into XML document. The Java components may act upon a preservation service or provide utility functions such as metadata manipulations. We implemented a template-repository service that allows users to choose from various abstract workflow scenarios (templates). Using the WEE, selected workflow templates can be dynamically configured and executed based on simple XML descriptions, which also can be generated from a visual workflow design tool.

In the fourth and final year of the project, we have explored different options for improving the scalability of preservation workflows. First, we have improved the IF architecture to allowing clustering of the application server and database layer. Second, we have demonstrated the use of Planets services with open source workflow engines like Taverna and Triana. Finally, we have performed experiments with the IF making use of data-intensive computations using the Amazon utility cloud infrastructure (AWS) .

To summarize, the Planets Interoperability Framework provides the glue that holds together the Planets user applications and preservation services. It enforces a technical contract (the service interfaces) and semantic interoperability (the digital object model) between the various services of a preservation workflow and provides a number of commonly required software components. Making use of the Planets IF, workflow templates, and preservation tool suite can save an organization effort, time, and money by basing preservation workflows on existing best practices, or by re-using existing preservation patterns.

Links:
http://www.planets-project.eu

Farquhar, A., Hockx-Yu, H. “Planets: Integrated Services for Digital Preservation.” International Journal of Digital Curation, Vol. 2, No. 2 (2007): http://www.ijdc.net/index.php/ijdc/article/view/45/31

Rainer Schmidt, Christian Sadilek, Ross King. “A Workflow System for Data Processing on Virtual Resources.” International Journal on Advances in Software, IARIA, ISSN 1942-2628, Vol. 2, No.2&3 (2009): http://www.iariajournals.org/software/tocv2n23.html

Please contact:
Ross King
AIT Austrian Institute of Technology GmbH/AARIT, Austria
Tel: +43 50550-4271
E-mail: ross.king@ait.ac.at

PROTAGE: Long-Term Digital Preservation Based on Intelligent Agents and Web Services

by Xiaolong Jin, Jianmin Jiang, and Josep Lluis de la Rosa

Digital objects have emerged as the primary means by which people create, disseminate and exchange information. The huge volume of digital information being constantly produced means there is now a pressing demand for long-term preservation. In this article, we briefly introduce a European FP7 Research Programme funded project, PROTAGE, which will investigate new technology for computerizing long-term digital preservation based on intelligent software agents and Web services.

In recent decades, a rapidly increasing amount of information has existed in digital form, with much now being ‘born digital’. Digital objects are now the primary means by which people create, disseminate and exchange information, and they are changing the ways in which people work, live and play. Compared to traditional non-digital information such as paper documents, audio tapes etc, digital information has many appealing advantages. For example, digital information can be made available to a greater number of users; it requires less space for storage; and it is much easier to search and retrieve. As a result, it can be readily reused to create new or adjusted information.

As the volume of digital information is growing at an explosive speed, there is a pressing demand for digital objects to be transferred from various IT systems (eg PCs, laptops, PDAs) to digital repositories, libraries and archives for long-term preservation. However, due to rapid changes and ongoing development in hardware and software systems and in the ICT infrastructure, long-term archiving of digital objects is a highly complicated task. Moreover, the diversity in the size and complexity of digital objects implies that modern digital preservation systems must be highly scalable and adaptable to various types of digital objects, as well as their input, storage and access. However, existing strategies for digital preservation are labour intensive and often require specialist skills. To meet our pressing preservation demands, it is necessary to find new levels of automation and self-reliance in preservation solutions. For these reasons, long-term digital preservation has been attracting significant research and development efforts from a variety of communities with a stake in digital preservation.

PROTAGE (PReservation Organization using Tools in AGent Environments), funded by the European FP7 Research Programme, aims to employ the promising technology of intelligent software agents and Web services to computerize long-term digital preservation. This

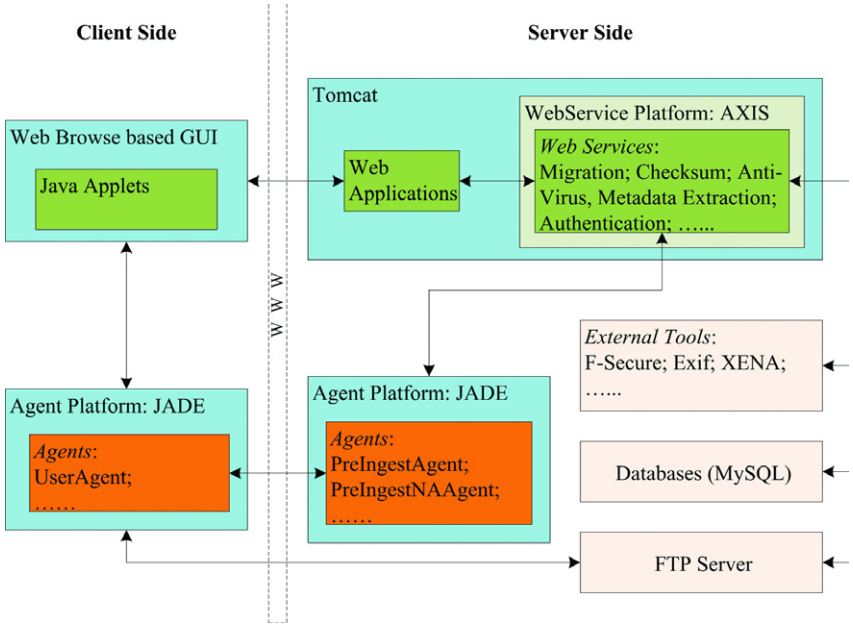


Figure 1: The architecture of the first PROTAGE prototype.

project commenced in November 2007 and is expected to finish in October 2010. Seven partners from six different European countries are involved in the PROTAGE project, namely, the National Archives of Sweden, Lulea University of Technology (Sweden), the National Archives of Estonia, Fraunhofer Gesellschaft (Germany), the University of Bradford (UK), EASY Innova SL (Spain), and Giunti Labs Srl (Italy).

PROTAGE intends to make digital preservation simple and automated, such that end users can readily preserve their digital objects while reducing the cost and increasing the capacity of preservation. PROTAGE will develop flexible and extensible software agent tools and Web services for long-term digital preservation and access, which are able to cooperate and be integrated with new or existing preservation systems. The PROTAGE system can be used to automate the submission of digital objects and their transfer between repositories, and monitor the digital preservation process. More specifically, the PROTAGE system along with its software agent tools and Web services will:

- enable digital-content creators to produce and publish digital objects in a preservation-compatible manner
- provide digital repositories with the means for further automating the preservation processes and
- facilitate seamless interoperation between content creators, libraries and archives, and end-users throughout Europe.

In general, the objectives of the PROTAGE project can be summarized as follows:

- to research the potential of intelligent software agents and Web services to support the automation of digital preservation tasks
- to demonstrate the technical feasibility of software agents and Web services by means of the PROTAGE prototype
- to analyse how the PROTAGE system can be implemented in various organizational environments
- to explore the possible integration of PROTAGE solutions with other or existing digital preservation environments
- to explore synergies with other RTD activities dealing with digital preservation.

The PROTAGE project adopts an iterative and incremental process to implement the targeted system, which begins with identification and analysis of user needs and functional requirement analysis, followed by technical specifications, implementation and system testing. In particular, upon analysis of user needs, thirteen typical scenarios have been identified, covering the overall process of long-term digital preservation. The first prototype of the PROTAGE system, which concentrates on transferring digital objects between repositories, was implemented and released in March 2009. Figure 1 presents the architecture of the first prototype, while Figure 2 shows one of its screenshots. The second prototype, which focuses on monitoring related issues, is under development.

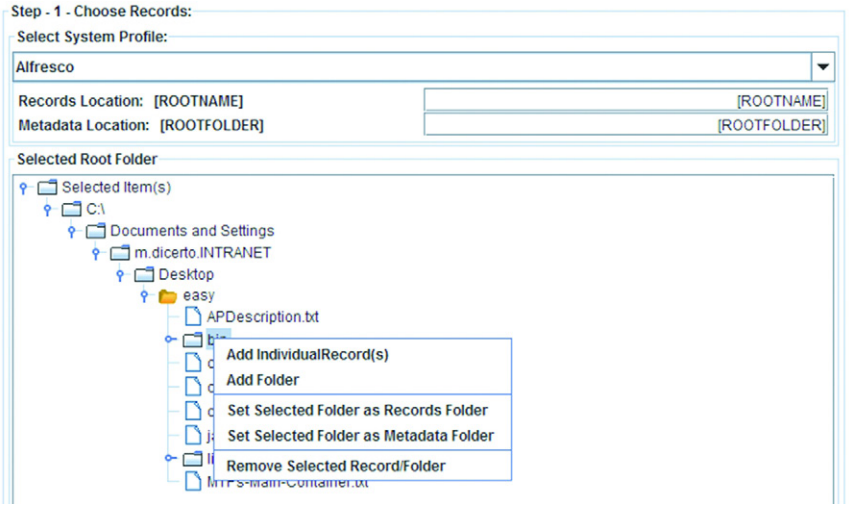


Figure 2: A screenshot of the first PROTAGE prototype.

Links:

http://www.protage.eu/index.html
http://cordis.europa.eu/fp7/home_en.html
http://www.protage.eu/Video_Prototype_1.html

Please contact:

Xiaolong Jin
School of Computing, Informatics and Media, University of Bradford, UK
Tel: +44 1274 234070
E-mail: x.jin@brad.ac.uk

Jianmin Jiang
School of Computing, Informatics and Media, University of Bradford, UK
Tel: +44 1274 233695
E-mail: j.jiang1@brad.ac.uk

Josep Lluis de la Rosa
Rensselaer Polytechnic Institute, USA, & University of Girona, Spain
E-mail: peplluis@cs.rpi.edu

SHAMAN: Sustaining Heritage Access through Multivalent Archiving

by José Borbinha

The SHAMAN project (Sustaining Heritage Access through Multivalent Archiving) will develop a next-generation digital preservation framework. Furthermore, it involves developing the relevant preservation tools for analysing, ingesting, managing, accessing and reusing information objects and data across libraries, archives or any other deployment scenario in which the SHAMAN 'Theory of Preservation' proves to be relevant.

The SHAMAN Theory of Preservation makes assertions about the ability to maintain the context, arrangement and management of information objects and the preservation environment itself, while taking into consideration:

- authenticity (the provenance of objects)
- 'respect du fonds' (the arrangement of objects)
- integrity (the management of objects)
- chain of custody (the ownership of objects)
- context of production (preservation, access and reuse).

These assertions also require that the functions performed by preservation processes remain consistent over time. We note that for these principles to apply, a theory of preservation must also make assertions about the information context managed by the preservation environment. Special attention must be given to the reuse of information objects across distributed repositories, as well as to securing the authenticity and integrity of the objects through time. These requirements led us to the present project.

Three prototypes will support the testing and validation of the results. These Integration & Demonstrator Subprojects (ISPs) cover real cases in memory institutions (ISP1), industrial design and engineering (ISP2) and scientific application domains in scenarios of e-Science (ISP3).

To achieve these goals, SHAMAN is focusing its research on: integrating data grid, digital library and persistent archive technology; developing support for context representation and annotation, with deep linguistic analysis and corresponding semantics; and modelling of preservation processes. In the end, SHAMAN is also expected to deliver a

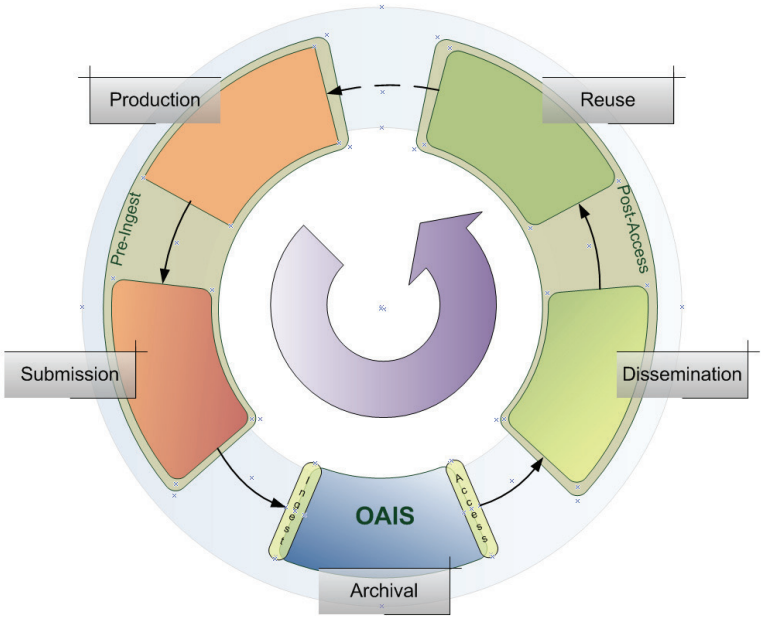


Figure 1: The informal SHAMAN context (which is supporting the work in progress towards the final definition of the SHAMAN Reference Architecture).

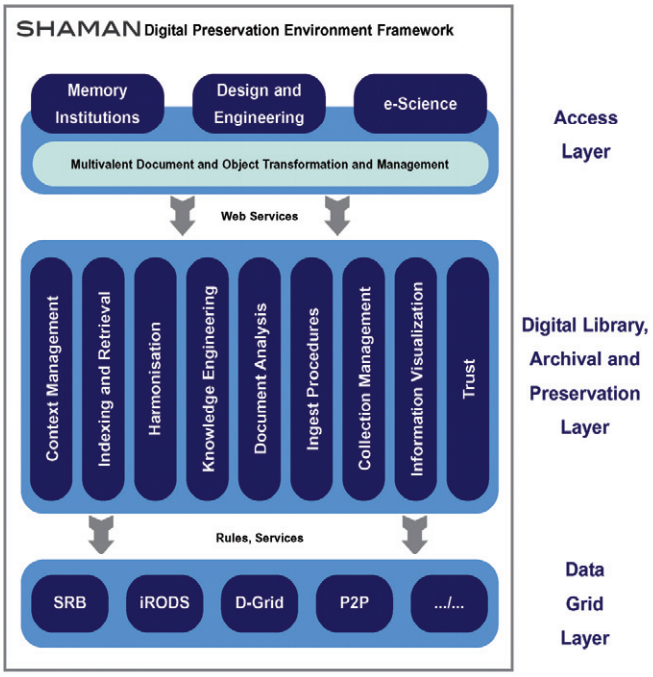


Figure 2: The informal and generic conceptual view for the SHAMAN architectural framework.

reference architecture for the design and development of solutions for digital preservation in distributed scenarios.

Until now, the project has been busy performing state-of-the-art analyses (such as reviewing OAIS), better understanding the real usage scenarios, defining solution architectures, and developing the first set of demonstrators. In first half of the project, the focus was on the analysis and development of the ISP1 scenarios, and the analysis of the ISP2 scenarios. These results will be presented at the review to be held in early 2010, and will be publicly disseminated after that.

The third year will focus on the revision of the ISP1 scenarios, the implementations of the ISP2 scenarios, and the analysis of the ISP3 scenarios. The final year will focus on the implementation of the ISP3 scenarios, the revision of the other two and the consolidation of the results. The final definition of the SHAMAN reference architecture will be an especially significant result expected for this term.

The consortium comprises a well-balanced group from academia, research labs, industry and intended final users. The full list of partners is available on the project Web site. SHAMAN also has strong connections to the United States, including collaboration with the Data Intensive Cyber Environments (DICE) research group, which leads the development of the open-source iRODS (Integrated Rule-Oriented Data System). The group is based at the DICE Center at the University of North Carolina at Chapel Hill, and the Institute for Neural Computation at the University of California, San Diego. We expect iRODS to play a fundamental role in SHAMAN, with its openness and flexibility supporting our vision and proposed strategies.

SHAMAN is a Large Integrated Project co-financed by the European Union within the 7th Framework Programme. It will run from 2008 to 2011.

Links:
http://shaman-ip.eu/
https://www.irods.org/

Please contact:
José Borbinha
INESC-ID, Portugal
E-mail: jlb@ist.utl.pt

HOPPLA - Archiving System for Small Institutions

by Michael Greifeneder, Stephan Strodl, Petar Petrov and Andreas Rauber

Hoppla is an archiving solution that combines back-up and fully automated migration services for data collections in small office environments. The system allows user-friendly handling of services and outsources digital preservation expertise.

Small companies are often hardly aware of changes in their technological environment. This can have serious effects on their long-term ability to access and use their highly valuable digital assets. In some countries, the law requires that business transactions remain available and auditable for up to seven years. Moreover, essential assets such as construction plans, manuals, production process protocols or customer correspondence need to be at hand for even longer periods of time, in case of maintenance issues, lawsuits or for business value. To avoid the physical loss of data, companies implement various backup solutions and strategies. Although the bitstream preservation problem is not entirely solved, there exists many years of practical experience in the industry, with data being constantly migrated to current storage media types, and duplicate copies held to preserve bitstreams over years.

A much more pressing problem is logical preservation. The interpretation of a bitstream depends on the environment of hardware platforms, operating systems, software applications and data formats. Even small changes in this environment can cause problems in opening important files. There is no guarantee that a construction plan for part of an aircraft, stored in an application-specific format, can be rendered again in five, ten or twenty years. Logical preservation requires continuous activity to keep digital assets accessible and useable.

Digital preservation is mainly driven by memory institutions like libraries, museums and archives, which have a focus on preserving scientific and cultural heritage, as well as dedicated resources to care for their digital assets.

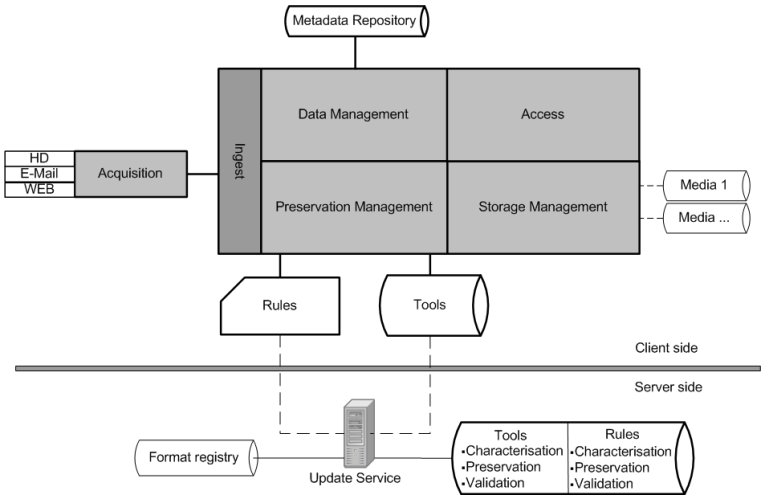
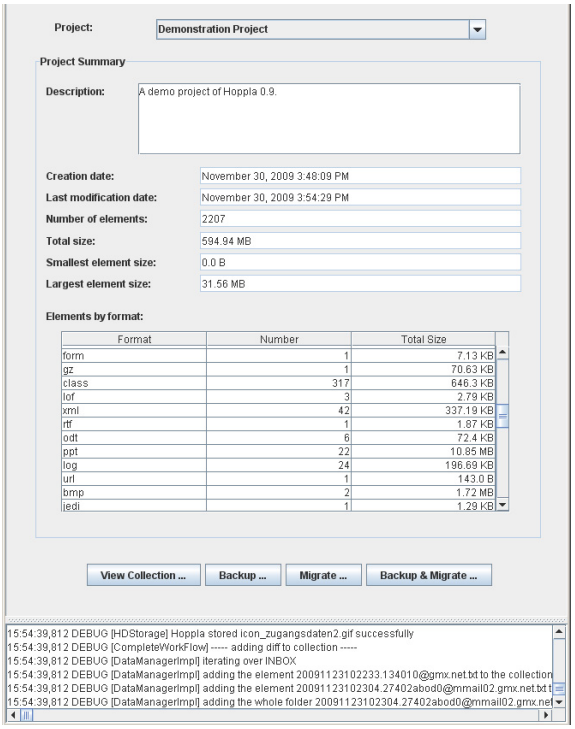
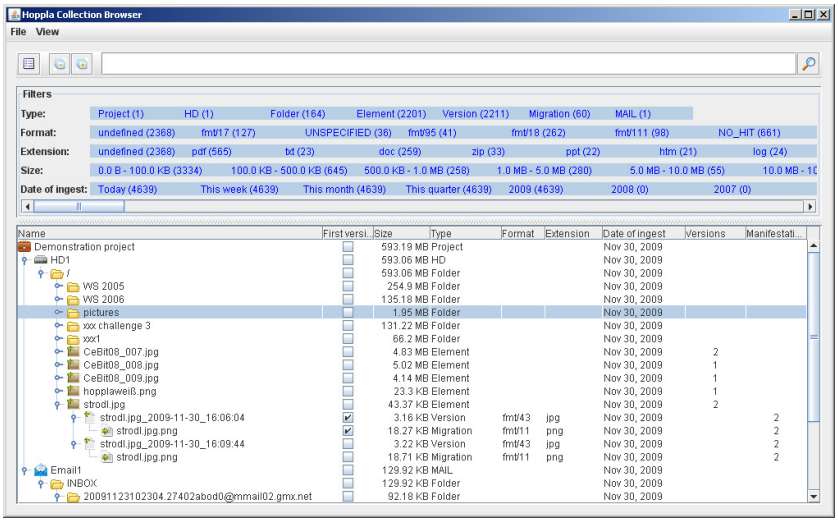


Figure 1: Architecture of the Hoppla system.





Figures 2 and 3: Hoppla screenshots.

Enterprises whose core business is not data curation are going to have an increased demand for knowledge and expertise in logical preservation solutions to keep their data accessible. Long-term preservation tools and services are developed for professional environments to be used by highly qualified employees in this area. In order to operate in more distant domains, automated systems and convenient ways to outsource digital preservation expertise are required.

With Hoppla, we are currently developing a solution that combines back-up and fully automated migration services for data collections in small institutions and small and home offices. The system builds on a service model similar to current firewall and antivirus software packages, providing user-friendly handling of services and an automated update service, and hiding the technical complexity of the software.

A central update service provides preservation rules and services to local Hoppla instances. The archiving system ingests data from a number of sources such as data carriers, email repositories and online storage locations. The user can define filter criteria for the collection, such as location, size and content type.

Metadata relating to the objects in the collection are important for later searching, retrieval and preservation. Hoppla collects available metadata from the source systems and additionally extracts metadata from the objects. In order to provide the system with appropriate preservation rules and tools, a collection profile is provided to the Web update service. For privacy reasons, the user can define the level of detail provided to the service by the profile. According to the received preservation rules and tools the objects in the collection are migrated.

Hoppla performs verification checks of the migration activity, and the archiving system supports versioning of objects. The storage module manages multiple backups of objects across different media. Hoppla can use offline and online storage media in both write-once as well as rewritable forms like DVDs, hard disks or cloud services for storage via plug-in infrastructure.

Missing in-house knowledge and expertise in digital preservation and data management of small institutions will be replaced by external expertise via a Web update service. Expert groups will provide guidelines and rules for the migration of endangered objects. We utilize this knowledge in an automated process and keep digital collections in an accessible form. The Web update service provides preservation rules and the relevant tools to the client side for migration and implements the preservation planning of the archiving system.

Hoppla is a research prototype development with a special focus on modular design and clearly defined interfaces between modules (Figure 2). This allows integration with existing solutions (eg as storage services), exchange of modules and easy enhancement for further acquisition sources, storage media or online back-ends. The architecture of the Hoppla archiving system is highly influenced by the reference model for an Open Archival Information System (OAIS), which has been widely accepted as a key standard reference model for long-term archival systems. As auditing and certification are becoming important issues for data storage, the system provides full documentation of all actions in the archive.

It assists in the fulfilment of software-related criteria of the TRAC (Trustworthy Repositories Audit & Certification) checklist by the OCLC/RLG Programs and National Archives and Records Administration (NARA).

Links:
http://www.ifs.tuwien.ac.at/dp/hoppla

Please contact:
Michael Greifeneder, Stephan Strodl, Petar Petrov and Andreas Rauber
Vienna Universtiy of Technology/AARIT
E-mail: {greifeneder, strodl, petrov, rauber}@ifs.tuwien.ac.at

The CARA Approach for Long-Term Preservation and Exploitation of Medical Images and Reports

by Hanan Bouzid, Mimouna Guenfoud, Andreas Jahnen, Pierre Plumer, Cédric Pruski and Frédéric Zucconi

Storing images and medical reports for long periods of time is becoming a challenge for medical institutions. Technology developed to collect detailed images from patients' bodies is producing an increasing quantity of large files that need to be stored for many years and to remain accessible when requested. This context has driven CRP Henri Tudor and the Health Ministry of Luxembourg to collaborate in the definition and implementation of a common project, named CARA (Carnet Radiologique). The main objective of this project is to set up IT solutions to improve the quality of the information available in the National Electronic Medical Record (EMR) without excessively increasing the size of this database; and to provide appropriated services for health professionals to access, manage and use, according to strict security policies, the content of the EMR.

Improving the quality of health care and reducing unnecessary costs are two major challenges for the health community. Reducing redundancy in exams due to misinformation about existing health data, especially in radiology where there is a risk of overexposure to radiation, is at the heart of the research carried out in the CARA (Carnet Radiologique) project developed at the Resource Center for Healthcare Technologies (CR SANTEC) of the Centre de Recherche Public (CRP) Henri Tudor in Luxembourg.

To achieve the initial objectives of the project, we face a set of organizational, technical, semantic and legal problems (see Figure 1). First, by virtue of the huge amount of data produced by radiology, appropriate storage space is required. In addition, we must be able to guarantee that the data can be kept for a long, predefined period of time. Second, problems exist in the data integration phase, mainly due to heterogeneous aspects of both the data (from a structural and semantic point of view) and the environment from which the data comes (eg hospitals or private practices). This also implies security issues, since users providing the data must be identified, the access to the data must be controlled, and the integrity and confidentiality of the data must be preserved. Finally, the exploitation of the preserved data is also problematic. The fast and accurate retrieval of rele-

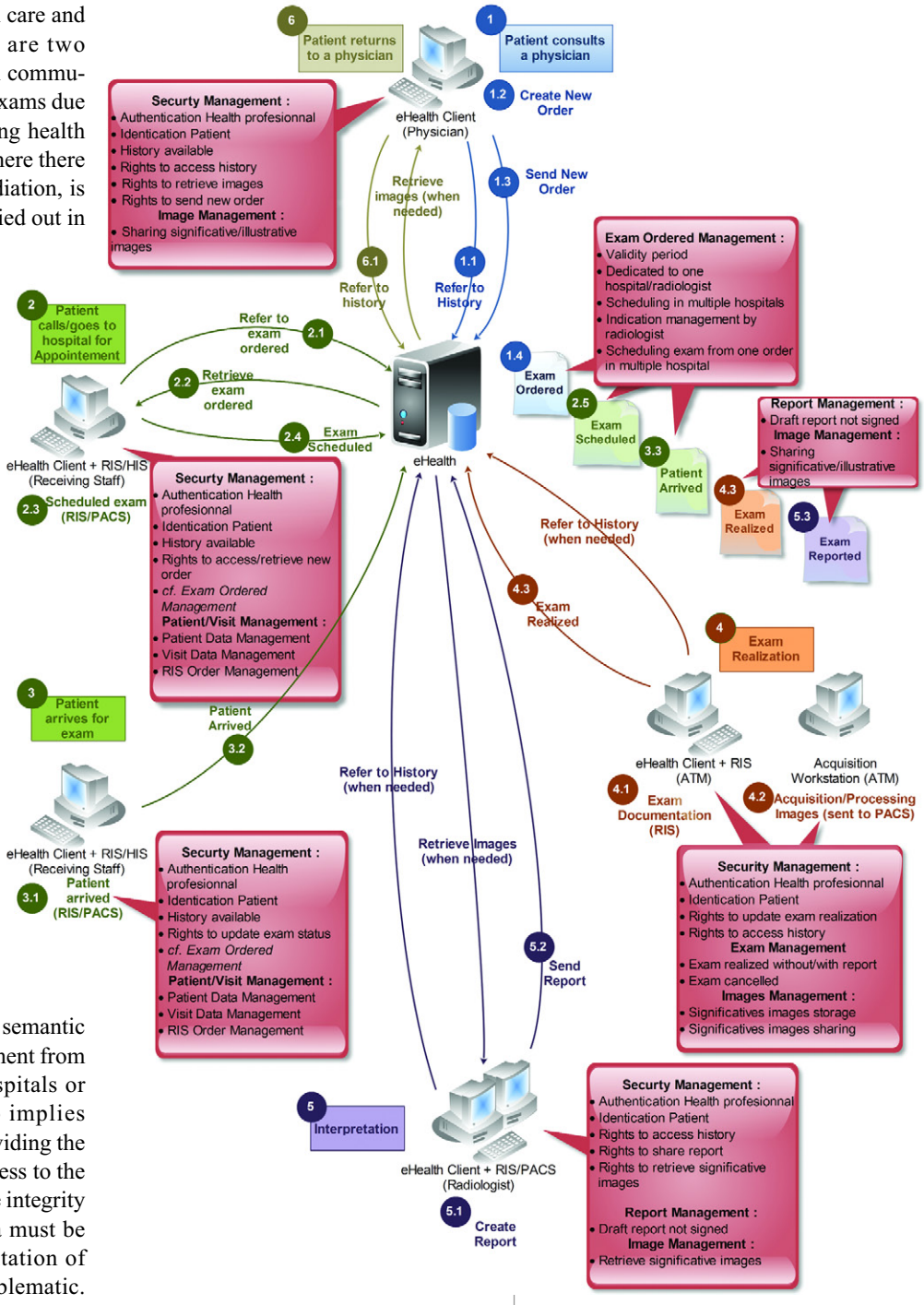


Figure 1: The CARA approach.

vant results from the huge quantity of stored information is particularly crucial, since doctors base critical diagnoses on the retrieved data.

Our research addresses all these problems. In order to improve data storage, we must first define the concept of relevant data. Only some of the images produced can be usefully exploited by doctors in making a diagnosis: such images can be considered relevant and must be identified and stored in an adapted structure. In order to facilitate data integration, we plan to define a common structure for medical documents such as prescriptions and reports, which will simplify both the capture of the medical information by doctors and the integration and indexation of the data (including images) by the system. The use of existing standards like HL7 will also speed up the integration of the data by providing a common framework for addressing technical interoperability during the data exchange between health actors.

A particularly challenging aspect of our research addresses data exploitation issues. It is necessary that data retrieval be efficient (with respect to the relevance of results and the speed of the search process): through the use of dedicated ontologies, Semantic Web technology has shown great promise. To this end we will investigate how we can use semantic technology to optimize the exploitation (and indexation) of the stored data.

CARA is a pilot project, and will pave the way for the development of a general health platform in Luxembourg. This general framework will integrate several health domains that require long-term data preservation and exploitation, such as laboratory exam results, pharmaceutical data and the management of patients' complete electronic health records.

Links:
eSanté-CARA web site :
http://www.santec.lu/project/esante/cara/start

Please contact:
Andreas Jahnen,
CR SANTEC, CRP Henri Tudor,
Luxembourg
Tel: +352 425991294
E-mail: Andreas.jahnen@tudor.lu

Designing a Trusted Distributed Long-Term Archive for Health Records

by Frej Drejhammar

Long-term archiving of Electronic Health Records (EHRs) is a complex task with a number of different requirements. This article describes how these requirements shape the implementation the DIGHT distributed EHR database.

The aim of the DIGHT (Distributed Information store for Global Healthcare Technology) project is to build a scalable and highly reliable information store for the Electronic Health Records (EHRs) of the citizens of India. The project partners are the Swedish Institute of Computer Science (SICS) and the Indian Centre for Development of Advanced Computing (C-DAC). SICS is responsible for developing a trusted and reliable distributed long-term archive for EHRs, while C-DAC is working on the standardization of EHRs and front-end software. The DIGHT system is a federated system where the participating entities are medical service providers (hospitals, clinics etc). These entities provide computing resources for their own use and contracts with other participants for sharing data and access to geographically dispersed data storage.

EHR data must be preserved for at least the citizen's lifetime. For demographical, genealogical and other research purposes it may be desirable to preserve EHRs for even longer time periods. Long-term archiving of EHRs can be approached from a number of directions. We have legal requirements, requirements on reliability, organizational aspects and software maintenance requirements to consider. In this article we will describe how these aspects constrain and shape the design of the DIGHT distributed EHR database.

To ensure the availability of EHR data and robustness in the face of unexpected occurrences such as sabotage, fires and natural catastrophes, EHR data must be replicated to avoid data loss. On the other hand, physical replication is expensive as it requires network connectivity and the maintenance of computing equipment at several locations. As there are legal requirements on keeping and archiving health records, we have designed the system such that all data have an explicit owner. The DIGHT implementation uses the owner information to guarantee that the data is permanently replicated on storage nodes controlled by the owner. The explicit ownership of data motivates a participant to pay for the upkeep of replicas as it is the only way it can fulfill its legal requirements.

Other legal requirements such as protection of patient confidentiality mandate the use of strong cryptography to protect against information disclosure. A way to guarantee patient confidentiality would be to let the patient carry the key needed to decrypt his or her EHRs. However, a patient might lose their key or be admitted to a hospital in a state where they cannot produce the key, making this method impractical. Such a scheme would also hinder research that uses patient data. To solve this problem, DIGHT uses trusted hardware to secure disk storage, and a public key infrastructure-based authorization policy and authentication system to control access to EHRs. By necessity a healthcare professional assigned to, for example, an emergency ward is allowed

SICS is responsible for developing a reliable distributed long-term archive for the health records of India's 1,2 billion citizens. Photo: iStockphoto.

access to any patient's EHR. To control abuse of the confidence placed in them by this policy, the DIGHT design uses secure logs to audit access. The log typically stores a request to access the patient's data signed by the healthcare professional's private key. Likewise new EHRs created by a healthcare professional are signed by his or her private key and time-stamped by the system. The time stamp and signature are important if, for example, malpractice is suspected, since created EHRs cannot be manipulated without detection.

If the DIGHT system is successful, its design choices will probably influence the software that handles EHRs for centuries to come, as converting to a newer incompatible system will probably not be economically feasible. The DIGHT database is designed from the beginning to support gradual upgrades as new requirements evolve.

Over time, the types of data stored in the database will inevitably change. The database supports a generic explicitly typed data format. The type information allows old data objects to be upgraded for use by newer software, and can also be used to temporarily downgrade data to allow old software to access newly created entries. To ensure data integrity and confidentiality in the face of improved cryptographic attacks, the system is designed with mechanisms for upgrading and re-certifying stored data.

From a software maintenance perspective, we support gradual updates by structuring the system as a set of cooperating services communicating over documented platform- and implementation-neutral interfaces. This allows us to upgrade and replace parts of the system, while maintaining its availability, to accommodate new storage technologies and upgrade legacy hardware and software. For example, a storage node is upgraded online using bootstrap and handover protocols that ensure the new node has replicated the old node's data before it takes over from the old node.

To summarize, the guiding principles behind our design of the DIGHT distributed EHR database are to make legal and operational responsibilities coincide; to support gradual online updates of software and hardware; to choose a flexible data model which can be extended; and to avoid vendor and technology lock-in by using open and documented interfaces.

Link:
http://dight.sics.se/

Please contact:
Frej Drejhammar
SICS, Sweden
Tel: +46 8 633 1617
E-mail: Frej@sics.se

Tackling the Problem of Complex Interaction Processes in Emulation and Migration Strategies

by Klaus Rechert and Dirk von Suchodoletz

In order to be handled, viewed or executed, digital objects require software environments. Most of these environments were designed with human interaction in mind, and this represents a major challenge for organizations wishing to use these now obsolete environments to handle huge numbers of objects in non-interactive ways for migration or in emulation.

Archiving and preservation organizations already have a large quantity of digital objects of various types created with a wide range of different GUI-oriented tools. At some point, new computer environments become unable to open or execute the original format of these objects. As a consequence, these organizations will need to convert the objects to a current, sustainable file format or would like to set up and emulate the original environment. Due to the scale of typical collections, the only financially and organizationally feasible way to support both actions and achieve both goals is by automated procedures. A major challenge for deploying automated processes is the availability of suitable tools. In most cases, a digital object is best viewed using the application with which it was created or in its original environment. Most of these applications were programmed to be handled in an interactive manner, and

little effort was put into automation for tasks such as batch handling of large numbers of files. A particular issue from the viewpoint of a digital archive manager is that spreadsheet, product design, audio/video or word processing programs cannot execute basic tasks such as the opening and saving of a file in another format as an unattended and fully automated task.

The attempt to later add such functions on to an application whose lifespan has already ended is in many cases simply impossible, since the source code and the required knowledge are no longer available. It is also becoming increasingly difficult to find staff able to operate the rising number of obsolete user environments.

Traditionally, so-called macro-recorders have been employed to help users automate interactive tasks to a certain degree. These are specialized tools or functions of an application or the user interface of an operating system that capture sequences of actions carried out; eg create a new file, open the address database and select an address, copy some text and save or print the file for serial letters. However, this functionality is not standardized, differs in its usability and features and might not be present in certain ancient environments at all.

Given these problems, the Planets Working Group (Preservation and Long-term Access through Networked Services) at Albert Ludwig University has suggested a different approach. The authors, together with a group of students, are exploring the option of handling typical repetitive tasks within specially wrapped hardware emulators. We hope to gain a perspective that is abstract enough to handle quite different tasks on applications in a reliable manner regarding defined in and outputs. The proposed method uses an operating system and application-independent interactive workflow for the migration or execution of digital objects using an emulated environment.

The approach is to interactively record a particular workflow once, such as installing a specific printer driver for PDF output, loading an old Word Perfect document in its original environment and converting it by printing into a PDF file. Such a recording can serve as the base for a deeper analysis and the generation of a machine script for the then completely automated repetition. An interactive workflow is defined as an ordered list of actions which are passed on to the emulated environment through a defined interface like the well-known Virtual Network Computing (VNC). These events may be mouse movements or keystrokes, and each is linked with a precondition and an expected outcome which can be observed as a state of the emulated environment. Until this effect is seen, the next event cannot occur. To link events with special preconditions and outcomes is necessary, since a workflow depends on the level of capacity of the emulation environment: programs will take different amounts of time to run depending on the load of the hosting machine, the size of the object being handled or the number of blocks already cached in memory. In the interactive case, this can occur, for example, through visual control by the user. For an automated run, the definition of expected states and a reliable verification is indispensable. We hope to produce time-independent action files which abstract in a machine way from written installation guidelines.

During the recording of a given workflow, the archivist is supported by an interface to a software archive for storing all additional necessary components like applications, operating systems, codecs, font sets and hardware drivers for the emulated machine. This additional service would extend the Planets framework, offering the possibility of interactively ingesting the software into the archive and enriching it with sufficient metadata. Such a supporting service would help to resolve the software dependencies originating with the object.

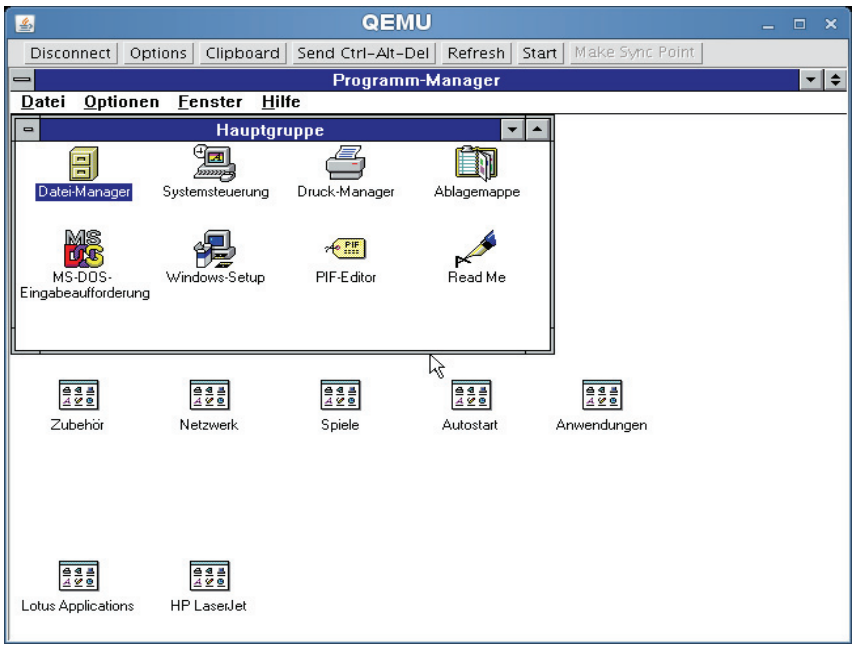


Figure 1: The Java VNC interface for the archivist to record workflows running interactively within the open-source processor emulator QEMU.

Please contact:
Dirk von Suchodoletz
University Freiburg, Germany
E-mail:
dirk.von.suchodoletz@rz.uni-freiburg.de

Klaus Rechert
University Freiburg, Germany
E-mail:
klaus.rechert@rz.uni-freiburg.de.

Trustworthy Preservation Planning with Plato

by Christoph Becker, Hannes Kulovits and Andreas Rauber

Digital content is short-lived, yet may prove to have value in the future. How can we keep it alive? Finding the right action to enable future access to our cultural heritage in a transparent way is the task of Plato.

The rapid changes in technology in today's information landscape have considerably shortened the lifespan of digital objects. While analogue objects such as photographs or books directly represent the content, digital objects are useless without the technical environment for which they were designed. In contrast to a book, word-processor documents cannot be read, a simulation cannot be re-run and re-evaluated, and sensor data cannot be interpreted without the right hardware, software and documentation environment. Digital objects are under threat at several levels: media failure, file format and tool obsolescence, or the loss of necessary metadata. Especially for born-digital material this often means that the contained information is lost completely. Digital preservation has become a pressing challenge for any kind of IT-related operation.

Given that a digital object needs the correct environment in order to function, we can either recreate the original environment (emulation) or transform the object to work in different environments (migration). A growing number of tools performing migration and emulation are available today, with each having particular strengths and weaknesses. Often there is no optimal solution. On the other hand, requirements vary across institutions and domains, and for each setting, very specific constraints apply. The process of evaluating potential solutions against specific requirements and building a plan for preserving a given set of objects is called preservation planning. Preservation planning is the centerpiece of the reference model for an Open Archival Information System (OAIS, ISO Standard 14721:2003, see link below). So far, it is a mainly manual, sometimes ad-hoc process with little or no tool support.

The planning tool Plato, developed as part of the Planets project (Preservation and Long-term Access through Networked Services) by the Digital Preservation lab at the Vienna University of Technology, is a publicly available Web-based decision support tool accessing a distributed architecture of preservation services. It implements a solid planning process and integrates a controlled environment for experimentation and automated measurements of outcomes. This enables trustworthy, evidence-based decisions to be made, as required by the Trustworthy Repositories Audit & Certification Criteria (TRAC, currently under evaluation for ISO standardization).

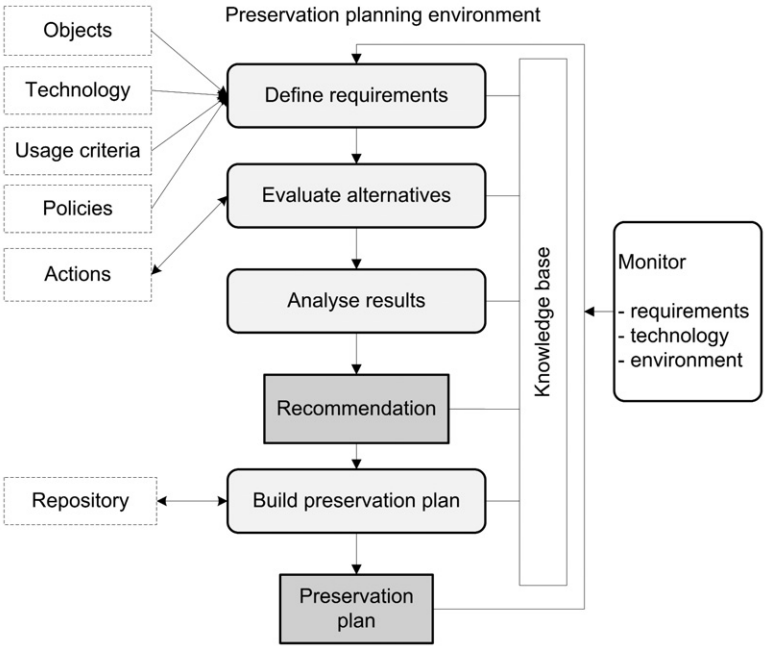


Figure 1: Preservation planning environment.

Preservation Planning

To ensure digital content remains accessible to and authentic for future users, a plan must be created that takes into account legal and technical constraints such as storage space, infrastructure and delivery, copyright issues, costs, user needs and object characteristics.

A preservation plan defines a series of preservation actions to be taken by a responsible institution due to an identified risk for a given set of digital objects or records (called a collection). The Preservation Plan takes into account the preservation policies, legal obligations, organizational and technical constraints, user requirements and preservation goals and describes the preservation context, the evaluated preservation strategies and the resulting decision for one strategy, including the reasoning for the decision. It also specifies a series of steps or actions (called a preservation action plan) along with responsibilities and rules and conditions for execution on the collection. Provided that the actions and their deployment as well as the technical environment allow it, this action plan is an executable workflow definition, such as a Planets workflow (see article on page 14).

The four-phase high-level workflow shown below can further be divided into fourteen steps. Evaluation of candidate actions uses controlled experiments and increasingly automated measurements.

Potential migration and emulation tools are applied to sample content and evaluated according to a hierarchy of require-

Figure 2: Visualization of results.

ments, based on Utility Analysis. A service-oriented framework greatly automates experiments and allows users to leverage various publicly available Web service registries that provide access to potential preservation action tools. Quality-aware services measure execution parameters and quality of the action tools, removing this burden from the experimenter.

The result of using the tool is a complete preservation plan that can be deployed and executed.
Current and future work includes:

- *Repository integration:* we are working on an integration of Plato with leading digital repository systems such as ePrints, RODA and other Fedora-based solutions to add preservation planning functionality to these systems.
- *Monitoring:* continuous monitoring of repository operation is essential and should include monitoring preservation plans.
- *Proactive recommendation:* by building recommender technology, we want to further increase the level of proactive planning in Plato.
- *Deployment:* Plato is being evaluated and used by several institutions to assist in planning long-term preservation (including the British Library, the Royal Library of Denmark and the Bavarian State Library), with further case studies focusing specifically on non-heritage application domains such as the medical sector (medical imaging), e-Government, production processes and scientific data sets.
- *Compliance validation:* with both service provision and trust gaining importance in the handling of digital content, full integration and validation within operational procedures are being evaluated in the context of respective international standardization initiatives.

Plato is publicly available free of charge at the project Web site.

Links:
Plato Project:
http://www.ifs.tuwien.ac.at/dp/plato

ISO-14721:2003: OAIS, Blue-Book:
http://public.ccsds.org/publications/archive/650x0b1.pdf

Planets Project:
http://www.planets-project.eu

Digital Preservation Lab, Department of Software Technology and Interactive Systems, Vienna University of Technology:
http://www.ifs.tuwien.ac.at/dp

Trusted Repositories Audit and Certification Checklist:
http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Full preservation plan definition:
http://www.ifs.tuwien.ac.at/dp/plato/docs/plan-template.pdf

Please contact:
Christoph Becker
Technical University Vienna/AARIT, Austria
Tel: +43 1 58801 18818
E-mail: becker@ifs.tuwien.ac.at

Towards Document Process Preservation: Xerox Launches Document Process Modelling Technology 'Xeproc©'

by Thierry Jacquin, Hervé Déjean, Jean-Pierre Chanod

Developed at the Xerox Research Centre Europe in the context of the EU Integrated Project SHAMAN, Xeproc© technology lets you define and design document processes while producing an abstract representation that is independent of the implementation. These representations capture the intent behind the workflow and can be preserved for reuse in future unknown infrastructures. Xeproc© is available under Eclipse Public Licence.

Xeproc© technology can be used to build a wide range of applications based on document processing, including transformation, extraction, indexing and navigation. It can be easily integrated with more global business processes and customized to match specific requirements and infrastructures. In the spirit of service-oriented architecture (SOA), Xeproc© embeds references to services and documents and provides loose coupling not only to services but also to data resources, with respect to both their location and format.

Xeproc© was developed in the context of the Integrated Project SHAMAN (Sustaining Heritage Access through Multivalent Archiving), co-funded by the European Union within the FP7 Framework. SHAMAN aims at developing a long-term digital preservation framework and tools to analyse, ingest, manage, access and reuse digital objects.

More specifically, within the context of SHAMAN and digital preservation, Xeproc© models XML pipelines and XML validation checkpoints. These capture the intent behind the workflow irrespective of the implementation at a given point in time. These abstract representations are preserved, so that the Xeproc© models can be seen as independent specifications to be instantiated and deployed over time and as technology evolves. These logical and persistent descriptions, when associated with the accurate components, are interpreted or translated into any SOA orchestration language to produce logically structured documents (typically XML). These make explicit how the source document content is logically and semantically organized.

Available on Eclipse 3.5.1 under the Eclipse Public License, Xeproc© combines a domain-specific language (DSL), an associated graphic designer and extension APIs (application programming interfaces).

The Xeproc© DSL: extensible, easy to use and focussed

The Xeproc© Domain-Specific Language (DSL) is used to describe the document process you want to design. It specifies a chain of processing steps, which may point to components such as document services or project-specific resources. All components take a document as input and generate another document as output.

To take full advantage of Xeproc©, the designer links processing steps with validation resources. While validations are traditionally exploited just before deployment, the Xeproc© Designer is conceived in such a way that they are exploited throughout the design process. Thanks to a continuous monitoring mechanism, validations not only verify but also specify, and lead the

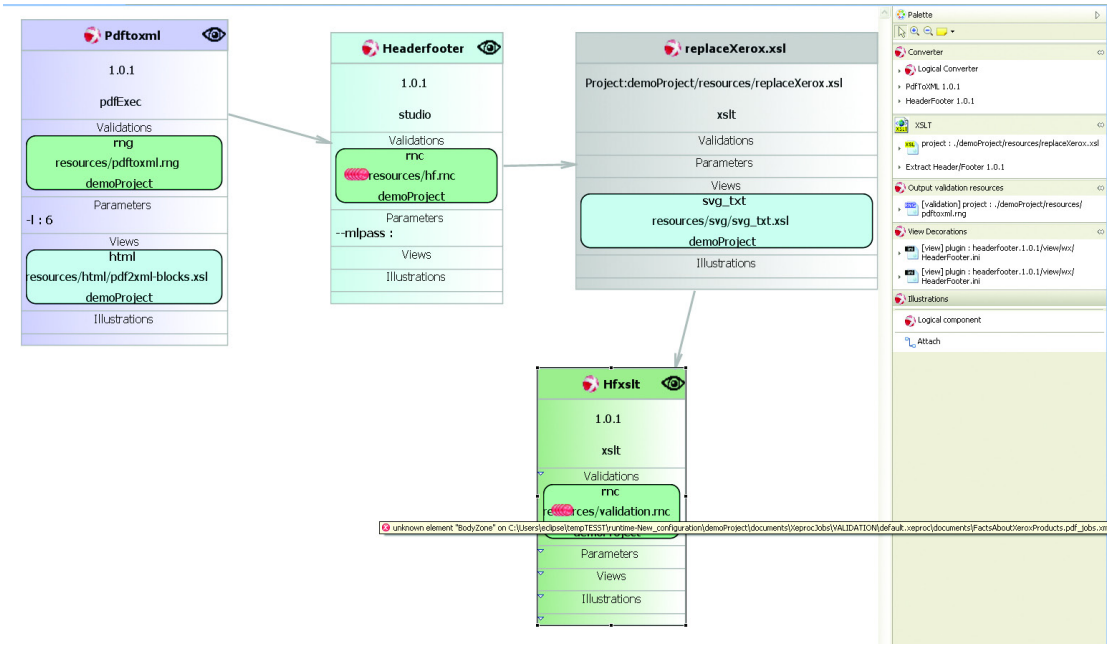


Figure 1: representation of a designed process with Xeproc©.

design process from the specification to instantiation.

In addition, processing steps can be linked to visualization specifications, highlighting selected outputs. These views, which are captured on demand and throughout the entire monitoring of the process, make it easier to identify and pinpoint errors, undertake corrections or consult the relevant experts.

The Xeproc© DSL is open enough to support any document format, validation syntax and resource location.

The Xeproc© Graphic Designer

The Xeproc© Graphic Designer is a user-friendly Eclipse plug-in editor which allows the user to manipulate abstract representations of objects relevant to the Xeproc© application domain.

The Designer provides an intuitive representation of underlying Xeproc© models and the ability to draw, rearrange and tune document-processing chains. This is achieved by combining project-specific resources (processing components, validations and views) with generic document services organized in a palette. The processing elements are represented as boxes, intermediate documents as arrows and validation constraints and views as icons on boxes.

The Designer was generated from the Xeproc© model using the EMF/GMF (Eclipse Modelling Framework and Graphical Modelling Framework) technologies provided by Eclipse. Model-Driven Architecture methodologies supported by the Object Management Group were applied.

Example scenario

A document transformation project will typically create an Eclipse project, share it amongst all the technical partners and initialize it with the reference resources such as documents, requirements and schemas to be validated. The process designer will consider the context and customize the palette of components with those considered useful from a site update. From there (s)he will start the building process and may drag and drop from the component palette or from the project workspace, quickly drawing specific logical and persistent pipelines for document analysis and transformation.

Extension APIs

The extension APIs allow the palette to be enlarged with components and associated validations and views published as document services. Such resources need to stipulate the interpretation engine responsible for injecting the input documents.

The extension APIs also allow you to extend your Xeproc© Designer with new resource processors, be they component, validation or view engines. The uploaded resources will then declare the processor type required at runtime. The Xeproc© Designer will dynamically delegate the operation to the right processor if plugged in.

This coupling of resource with interpreter makes it possible to realize amazing combinations, including WSDL/SoapClient, main.java/JVM, python.py/python.exe, XSLT Stylesheet/XSLT processor, or any other combination one may care to imagine. Mechanisms are provided to plug processing resources and processing engines into the Xeproc© Designer.

Links:
http://www.xrce.xerox.com/Xeproc
http://shaman-ip.eu/shaman/

Please contact:
Jean-Pierre Chanod
Xerox Research Centre Europe, France
Tel: +33 (0)4 76 61 50 75
E-mail: Jean-Pierre.Chanod@xrce.xerox.com

Cyclops: An Interface for Producing and Accessing Archives of Artistic Works

by Nicolas Esposito, Bruno Bachimont and Erik Gebers

Within the scope of the EU project CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval, started in 2006) and the OAIS standard (Open Archival Information System), our team (CNRS/Université de Technologie de Compiègne) is focusing on the long-term preservation of artistic resources. The aim was to propose a framework which preserves access to these resources and maintains their intelligibility over the long term. The more precise objectives were to aid both the study of artistic productions and new performances of these works. Our contributions led us to design a tool for producing and accessing archives: Cyclops.

During the project, our team worked with a number of institutions involved in artistic productions: INA (Institut National de l'Audiovisuel), IRCAM (Institut de Recherche et Coordination Acoustique/Musique), University of Leeds, and CIANT (International Centre for Art and New Technologies). From the point of view of preservation, their archives are complex objects. Indeed, works such as electroacoustic music or multimedia installations are based on hardware and software that can quickly become obsolete, and they involve non-digital components which need to be described/digitized in order to preserve the intelligibility of the whole. So, the question is: how to achieve long-term preservation with the relevant designated community (ie the users of these archives, including composers, musical assistants, archivists and musicologists)?

Since the specialty of our team is knowledge engineering, we started by modelling activities. We proposed a

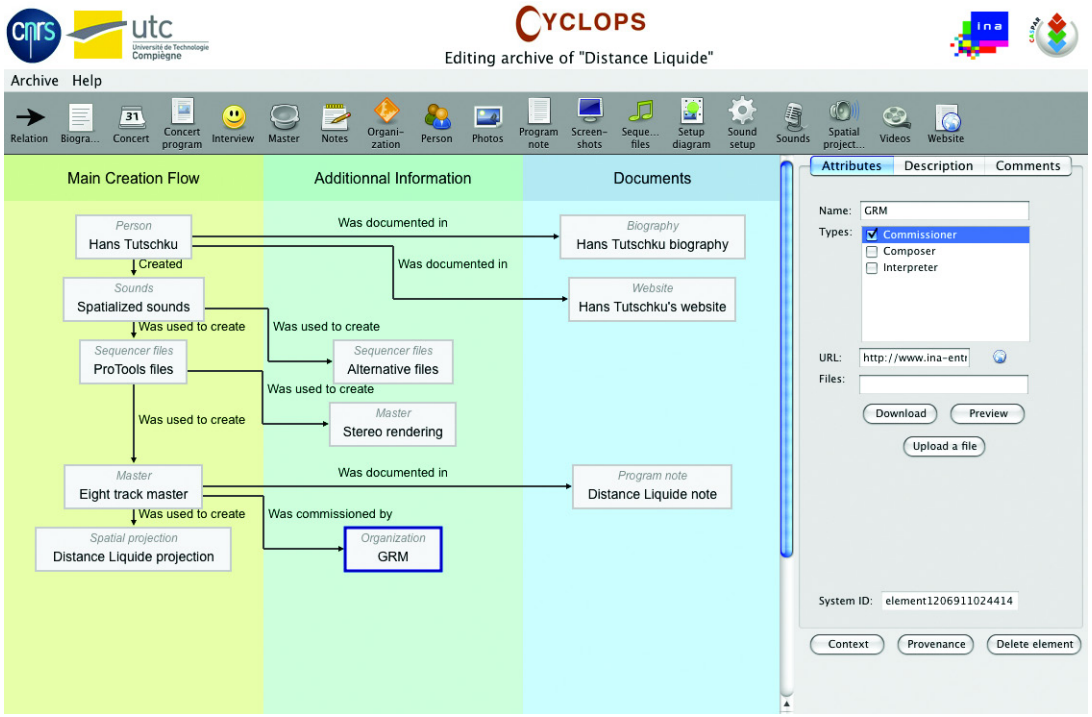


Figure 1: Screenshot of Cyclops.

way to structure the archives using the CIDOC CRM ontology (International Committee for Documentation – Conceptual Reference Model, ISO 21127) in order to describe the life cycle of the works. It allows users to account for each resource according to its relationship with all the others. It makes the history of the works explicit by answering this question at each step of the production flow: who did what?

We also provided a model for the acousmatic works from INA (electroacoustic music) and a methodology with which to archive artistic productions. We validated the model and the methodology for several acousmatic works, with other partners using a similar approach: eg the University of Leeds with interactive multimedia performances. However, since the members of the designated communities are not knowledge engineers, we proposed an interface for the CASPAR system that meant they need not handle the CIDOC CRM ontology. We named this interface Cyclops.

The Cyclops tool allows archive producers to describe the life cycle of an artistic work. When Cyclops is installed in an institution, it is associated with a set of concepts, relations and types that come from its domain. Users manipulate these terms (which are common for them) within a graphical representation of the life cycle (see Figure 1).

Some members of these designated communities have successfully produced archives using Cyclops. It was also used for the scenarios of the artistic testbed, for instance when the implicit knowledge of the production team is changing. The tool provided concrete elements related to some of the key concepts of the OAIS standard, especially concerning representation information, context and provenance. Moreover, demonstrations of interaction with other CASPAR components were made, especially with the DRM component (Digital Rights Management) and for the IRCAM scenarios (electroacoustic music).

Cyclops is also a tool for accessing the archives. A dedicated browsing mode allows users to search and access the archives. Reading the life cycle of an artistic work and referring to attached documents is helpful in understanding it, but is not sufficient for the long term. This is why we implemented features related to preservation through access. Each element of a life cycle can be commented on and thus documented, thanks to the contributions of the users. If the designated community accesses an archive actively, this archive is still alive.

Cyclops is a Web application and is open source. It uses the following technologies: XUL, JavaScript, SVG, HTML, CSS, XML, PHP, MySQL. While it is planned that Cyclops will be used on top of a CASPAR system, it additionally provides a stand-alone mode using its own Web server to store the files.

Now that the CASPAR project is completed, further uses will be found for Cyclops. The tool will continue to be used by some partners on top of the CASPAR system. Some artists who are interested in the approach will also use it through its standalone mode. Furthermore, we can establish new collaborations to adapt Cyclops to other institutions.

This work was partially supported by the EU project CASPAR (FP6-2005-IST-033572). The partners were ACS, Asemantics, CIANT, CNR, CNRS/UTC, Engineering, ESA, FORTH, IBM Haifa, INA, IRCAM, Metaware, STFC (project coordinator: David Giaretta), UNESCO, University of Glasgow, University of Leeds and University of Urbino.

Links:
http://www.utc.fr/caspar/
http://www.casparpreserves.eu/

Please contact:
Nicolas Esposito, Bruno Bachimont
Université de Technologie de Compiègne, France
Tel: +33 344 23 44 23
E-mail: {nicolas.esposito, bruno.bachimont}@utc.fr

Magnetic Tape Storage and the Growth of Archival Data

by Jens Jelitto, Mark Lantz and Evangelos Eleftheriou

The volumes of digital data being produced are growing at an ever increas-ing pace. According to an International Data Corporation study for 2007, 264 exabytes of data were created. In the future, this staggering volume of data is projected to grow at a 57% annual growth rate, faster than the ex-pected growth of storage capacity. Moreover, new regulatory requirements mean that a larger fraction of this data will have to be preserved. All of this translates into a growing need for cost-effective digital archives.

While Hard Disk Drive (HDD) technology has made significant progress over the years, so has magnetic tape recording, such that tape still remains the least ex-pensive long-term archiving medium. Current tape technology achieves a storage density of about 1 Gb/in2 and a cartridge capacity on the order of a terabyte. An analysis of the limits of current tape technology suggests that tape areal density can be further pushed by two orders of magnitude, leading to cartridge capacities in excess of 100 terabytes. This makes tape a very attractive technology for data archiving with a sustainable roadmap for the next ten to twenty years, well be-yond the anticipated scaling limits of HDD technology.

It is clear that tape will never become the primary storage medium for average computer users. HDDs are much better suited to this purpose, with access times of a few milliseconds, storage densities of 300-400 Gb/in2 and capacities of up to a terabyte.

However, for long-term archiving, backup and disaster recovery, there are con-siderable advantages to using tape:

- *energy savings:* once data is recorded, the medium is passive; it sits in a rack and no power is needed
- *security:* once the data is recorded and the cartridge removed from the drive, the data is inaccessible until the cartridge is reinstalled. This means that the data cannot be corrupted by a virus while it is offline. Security is further enhanced by drive-level encryption
- *lifetime:* because the medium is passive, it is extremely reliable with a long lifetime. Some tapes have been in use for forty years
- *reliability:* tape media is removable and interchangeable, meaning that unlike HDDs, mechanical failure of a drive does not lead to data loss, because a cartridge can simply be mounted in another drive.

All these factors contribute to the major net advantage:

- *cost:* savings estimates of the total operating cost of tape backup relative to HDDs range from factors of three to twenty-three, even if the latest developments, such as data deduplication, are taken into account. In archival applications, where deduplication can not be used effectively, cost savings can be even higher.

Today's archival tapes have a storage capacity of about 1 Gb/in2. A recent study indicates that improvements in technology may increase this density to 100 Gb/in2 without a fundamental change in the tape recording paradigm. There are five main technologies involved:

1. *Media:* the main challenge for tape systems is that the tape medium is flexible while HDDs are rigid. HDD heads 'fly' over the media while those for tape sys-tems are in contact with the tape. Current tapes are based on metal particles, but promising research is underway into new tape media such as barium-ferrite, which provides a smoother surface and improved signal quality.

2. *Heads:* HDDs currently use very sophisticated head technology, based on tunneling magnetoresistive (TMR) sensors. It is expected that tape heads will also move to using TMR, which has increased sensitivity leading to an im-proved signal-to-noise ratio for detection.

3. *Transport and track-following control:* the spacing between adjacent tracks today is around 10 µm. The target for the future is to reduce this to the order of 0.2 µm, which requires much better control of the lateral positioning of the tape head in the sub-micrometer range as well as very tight tape speed and tension control.

4. *Signal processing:* signal processing plays a crucial role in reliably retriev-

Figure 1: IBM System Storage TS3500 Tape Library, a highly scalable, automated tape library for mainframe and open systems backup and archiving in midrange to enterprise environments with a capacity of up to 45 PB (with 3:1 compression).

ing the recorded digital information. High areal recording densities pose signifi-cant challenges in terms of "write" and "read" operations. The main challenge is to ensure highly reliable operation of both these signal-processing func-tions, including adaptive equalization as well as gain and timing control, de-spite significant reductions in the available signal to noise ratio. To achieve the envisaged linear recording densities leading to multi-terabyte tape sys-tems we are investigating novel advanced noise-predictive detection schemes that take into account the special statistical properties of the noise process.

5. *Error protection:* to guarantee an uncorrectable bit error probability of less than 1 x 10-17, redundancy is added to protect the data with error-correcting codes (ECC). In the Linear Tape Open (LTO-4) standard, the overhead amounts to 27%. We investigate ways to reduce the overhead without sacri-ficing performance. For instance, currently data is coded first for error correc-tion and then for modulation constraints. In the future, a reversal of this order called 'reverse concatenation' could lead to a gain in efficiency and enable more powerful iterative data detection and decoding schemes.

Tape will remain a vital storage medium for the foreseeable future, as there is a good chance of reaching a density of 100 Gb/in2 on tape. This will help solve the huge problem of preserving our ever-growing mountain of data in an economical and ecological manner. IBM is working on advanced technologies to keep tape storage systems as attractive for long-term data storage in the future as they have been in the past.

Links:

http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf

http://www.research.ibm.com/journal/rd/524/argumedo.pdf

http://www.clipper.com/research/TCG2008009.pdf

http://www.ultrium.com/pdf/Tape%20Fallacies%20Commentary%20Final.pdf

Please contact:

Jens Jelitto, Mark Lantz, Evangelos Eleftheriou
IBM Research GmbH, Zurich Research Laboratory, Switzerland
E-mail: jje@zurich.ibm.com, mla@zurich.ibm.com, ele@zurich.ibm.com

The ESA Approach to Long-Term Data Preservation using CASPAR

by Sergio Albani

Long-term preservation of earth science data in the European Space Agency has been studied using a framework constituted by components developed in the EU CASPAR project.

Earth Observation (EO) Space Missions provide global coverage of the Earth across both space and time, generating on a routine and continuous basis huge amounts of data (from a variety of sensors) that must be acquired, processed, elaborated, appraised and archived by dedicated systems. ESA-ESRIN, the European Space Agency Centre for Earth Observation, is the largest European EO data provider and operates as the reference European centre for EO payload data exploitation. The long-term preservation of both EO data and the ability to discover, access and process them is clearly a fundamental issue and a major challenge at all levels (programmatic, technological and operational). The need to address this challenge is one of the reasons why ESA participates in several EU-funded projects in addition to conducting an internal research program.

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval), an Integrated Project co-financed by the EU within the Sixth Framework Programme (Priority IST-2005-2.5.10, 'Access to and preservation of cultural and scientific resources'), has built a framework to support the end-to-end preservation life cycle for digital information, based on the OAIS reference model, with a strong focus on the preservation of the knowledge associated with the data. Three testbeds have been established to validate CASPAR solutions in the domains of Cultural Heritage, Contemporary Performing Arts and Earth Science. ESA plays the role of both user and data/infrastructure provider in the Earth Science domain and has built a number of dedicated scenarios for testing purposes.

The main objective of the ESA scientific testbed is the preservation of the ability to process data over different levels, ie the ability to generate higher-level products (using auxiliary data and suitable processors) starting from raw satellite-acquired data. ESA's first demonstrator focused on GOME (Global Ozone Monitoring Experiment, a sensor on board ESA ERS-2 satellite) data; specifically the ability to produce Level 1C data (fully calibrated) from Level 1B data (raw signals plus calibration parameters). The ESA testbed can demonstrate the preservation of this GOME processing chain with respect to changes of the operating system or compilers/libraries/drivers which affect the ability to run the GOME Data Processor.

The preservation scenario is the following: a complete and OAIS-compliant GOME L1 processing dataset has been ingested into the ESA CASPAR System (a framework developed in ESA-ESRIN using only CASPAR components by Advanced Computer Systems ACS SpA, technical partner for the testbed implementation). At a certain point an external event affects the

ability to run the processor (eg a library or the operating system changes) and a new L1B->L1C processor has to be developed/ingested to preserve the ability to process data from L1B to L1C. These changes must be catered for by ensuring correct information flow through the ESA CASPAR System, the system administrators and the users.

The ESA testbed is divided into four phases:

1. *ESA CASPAR System setup:* a basic EO ontology has been developed based on a specialization of the ISO 21127:2006 CIDOC-CRM ontology, representing relationships and dependencies of GOME data and OAIS representation information stored on the CASPAR system. (Representation information maps a data object onto more meaningful concepts, eg the ASCII definition that describes how a sequence of bits – the data object – is mapped onto a symbol.)

2. *Ingestion of data and related representation information*: the ingestion process allows the data producer to ingest an OAIS compliant dataset composed of GOME L1B data, the L1B->L1C processor and the representation information including all knowledge related to the GOME data and processor.

3. *Data access, browsing, searching and retrieval:* CASPAR is able to provide a user asking for L1C data not only with the related L1B data plus the processor needed to generate them, but also with all the information needed to perform this process, depending on the user's needs and knowledge (ie different Designated Communities will retrieve different representation information during the same search-and-retrieve session to fill their specific knowledge gap).

4. *Software processing preservation (upgrade):* the preservation phase can be summarized as follows.
   - an external event affects the processor (eg a library or the operating system changes) and an alert sent through CASPAR by informed users is forwarded to the system administrator
   - the system administrator uses the ontology to identify which modules need to be updated
   - the system administrator is able to retrieve, download and work on the source code of the processor to deliver a new version of the processor
   - the new processor, with appropriate (updated) OAIS preservation description information and representation information, is reingested into the CASPAR system
   - an alert mechanism notifies the users that a new version of the processor is available
   - the new processor can be directly used to generate Level 1C products, meaning the scientific capabilities of users are maintained.

The above scenario has been implemented in ESA-ESRIN through a Web-based interface and has demonstrated the effectiveness of the CASPAR preservation framework in the Earth Science domain. The ESA CASPAR System is available (for further enhancement and testing) to users and data owners/providers interested in a practical approach to preservation using CASPAR solutions.

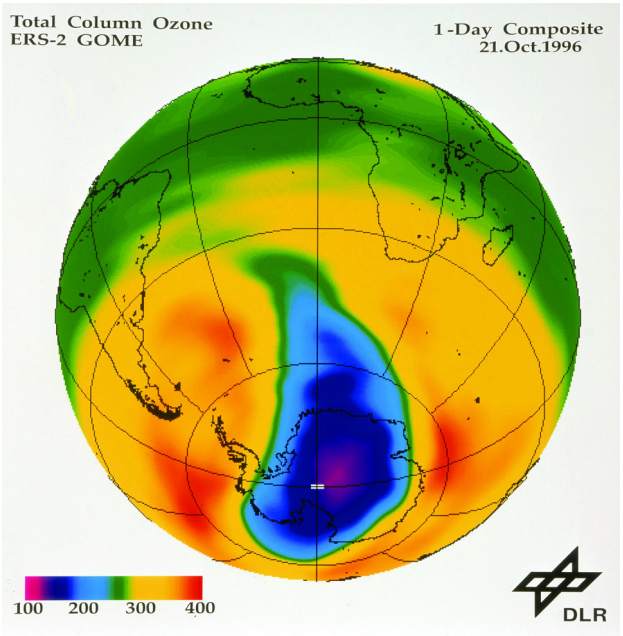


Figure 1: opening of the ozone hole during austral spring. This image was produced from GOME derived data. Image: ESA

The need to preserve and link Earth Science tools and data has become more evident recently and the ESA-ACS team is confident that the CASPAR solutions will be increasingly adopted in the years to come.

Links:
http://www.casparpreserves.eu
http://www.esa.int
http://www.acsys.it

Please contact:
Sergio Albani
ACS c/o ESA-ESRIN, Italy
Tel: +39 06 94180561
E-mail: sergio.albani@acsys.it

Digital Preservation of Interactive Multimedia Performances

by Kia Ng

Digital media and technology are becoming increasingly important for the performing arts, particularly with regard to technology-enhanced performance. Digital preservation is now an urgent consideration for performing arts in many aspects, such as ensuring possible future re-performance and analysis, and the preservation of intangible heritage (beyond the usual audio-visual recording) and plying/gesture styles.

Digital preservation aims to ensure the intelligibility of digital information at any given time in the near or distant future. Digital preservation must address changes that inevitably occur in hardware or software, in organizational or legal environments, as well as in the designated community, ie the users of the preserved information. In order to be preserved against these changes, digital information must be enriched with metadata or 'representation information', which can be used for the interpretation of the original information. In addition, representation information needs to be connected to the knowledge base of the designated community. Ontologies offer a means of organizing and representing the semantics of this knowledge base.

This article presents a brief overview of the contemporary arts testbed of the CASPAR EC IST project, with particular focus on the digital preservation of Interactive Multimedia Performances (IMP). The CASPAR framework is based on the Open Archival Information System (OAIS) reference model, providing a consistent set of concepts, terminology and a framework for the development of archival information systems and related standards.

An IMP involves one or more performers who interact with a computer-based multimedia system, making use of multimedia content that may either be pre-prepared or generated in real time, including music, manipulated sound, video and graphics.

The architecture of the proposed archival system allows multiple multimedia systems to be preserved, such as the Music via Motion (MvM) software for transdomain mapping of motion and sound (see Links), the i-Maestro 3D Augmented Mirror (see Figure 2, Links), and many others.

Due to multiple dependencies, the preservation of an IMP requires a robust representation and association with digital resources. This can be achieved with an ontological approach that describes an IMP and its internal relationships to support the preservation process. With the CASPAR framework we use entities and properties defined in the CIDOC-CRM and FRBRoo ontologies for the description of an IMP.

CIDOC-CRM was originally designed to describe cultural heritage collections in museum archives. A harmonization effort has also been carried out to align the Functional Requirements for

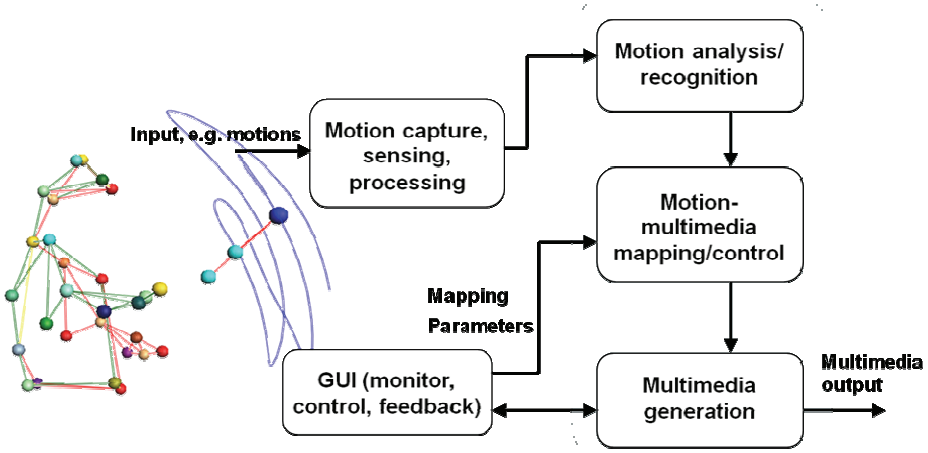


Figure 1: Main processes of an IMP.

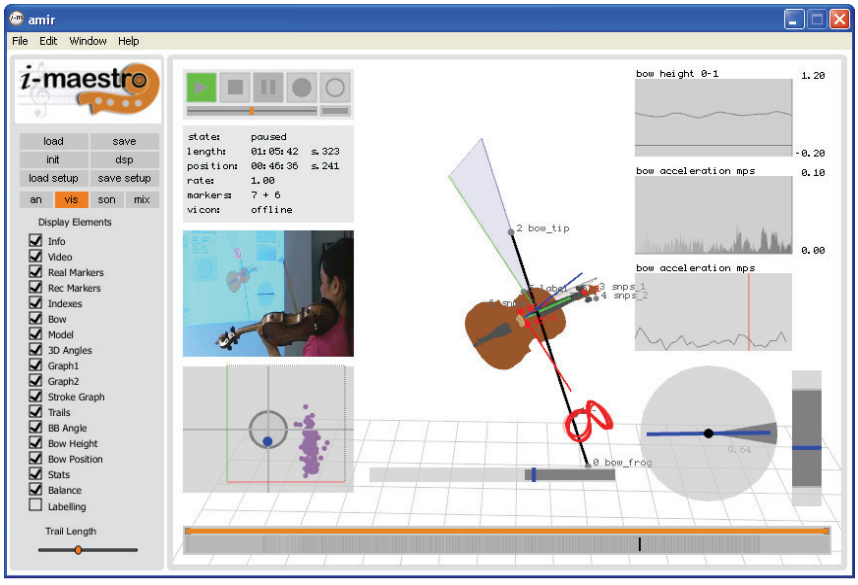


Figure 2: A screen snapshot of the i-Maestro 3D Augmented Mirror (capturing and analysing audio, video, 3D motion data, pressure and balance).

Bibliographic Records (FRBR) to CIDOC-CRM for describing artistic contents. The result is an object-oriented version of FRBR, named FRBRoo. The concepts and relations of the FRBRoo are directly mapped to CIDOC-CRM.

For the testbed, a Web-based archival system called the ICSRiM IMP Archival System has been developed. It is based on the OAIS model and integrates the CASPAR components by extending their functionality for the efficient preservation of the IMPs. The interface of the prototype communicates with a repository containing both the IMPs and the metadata necessary to preserve the IMPs. Both the ingestion and retrieval procedures are based on the representation information of the IMP. The representation information is generated in RDF/XML format with the use of the CASPAR Cyclops tool (see Links) and captures all the appropriate metadata of the IMP to enhance virtualization and future re-performance of the IMP.

For the ingestion procedure, the archival system parses the representation information of an IMP, detects the defined modules, and then guides the user to ingest the corresponding digital content. After the ingestion of the package, the representation information is updated with information regarding the location of the IMP files.

For retrieval, the system calls the CASPAR FindingAids service, which performs RQL queries on the representation information and returns the related objects to the user.

The captured representation information plays a significant role in providing access to an IMP, as it encapsulates the semantic information needed for the reproduction and comprehension of an IMP. Thus, the user has the ability to retrieve the digital files of an IMP along with comments and additional information on the IMP for re-performance.

The Web-based IMP archival prototype has been successfully validated with a group of IMP producers and artists. It is now being piloted at the University of Leeds Interdisciplinary Centre for Scientific Research in Music (ICSRiM). Further works include setting automated responses to changes that may affect the preserved IMP. For example, upon the release of a new software version, the metadata and information packages should be updated with the new information.

Links:
CASPAR: http://www.casparpreserves.eu

CASPAR Cyclops tool:
http://www.utc.fr/caspar/wiki/pmwiki.php?n=Main.Proto

i-Maestro: http://www.i-maestro.org

University of Leeds Interdisciplinary Centre for Scientific Research in Music (ICSRiM):
http://www.icsrim.org.uk,
http://www.leeds.ac.uk/icsrim

Music via Motion (MvM):
http://www.leeds.ac.uk/icsrim/mvm

Body movement to create music, BBC News, Technology, URL:
http://news.bbc.co.uk/1/hi/technology/3873481.stm

3-D movement captured to conduct music, Reporter, issue 500, University of Leeds, URL:
http://reporter.leeds.ac.uk/500/s14.htm

Please contact:
Kia Ng
ICSRiM – University of Leeds
Tel: +44 113 3432583
E-mail: k.c.ng@leeds.ac.uk,
kia@icsrim.org.uk

The Planets Testbed: A Collaborative Research Environment for Digital Preservation

by Brian Aitken and Andrew Lindley

The digital objects that are so fundamental to 21st-century life may have a precarious future due to the rapid pace of technological change. Digital preservation specialists have proposed an almost overwhelming variety of preservation actions and tools that may help to mitigate this risk, but there is a lack of empirical evidence to help librarians, archivists and non-specialists to make informed decisions about the most applicable and effective preservation tools. The Planets project (Preservation and Long-term Access through Networked Services) has developed a digital preservation Testbed that aims to provide such an evidence base.

The Planets Testbed is a freely available and easy-to-use controlled environment in which users can experience and compare different preservation tools and approaches through their Web browser. The Planets approach of a service-oriented architecture makes preservation tools available on heterogeneous platforms and in well-defined and controlled surroundings. The tools are given a Web service wrapper which exposes certain aspects of their functionality through a standardized vocabulary and therefore allows users to access them through the Testbed's Web-based interface, as shown in Figure 1.

Through the Testbed, users can design and execute a variety of experiments, such as migration, emulation and executable preservation plan experiments. The focus of a migration experiment may be to analyse the performance and trustworthiness of tools that transform digital objects from one format (such as obsolete word processor files) into more up-to-date or preservable formats (such as PDF/A). The focus of an emulation experiment may be to investigate how effectively and accurately an obsolete digital object is perceived within an emulated hardware and software environment.

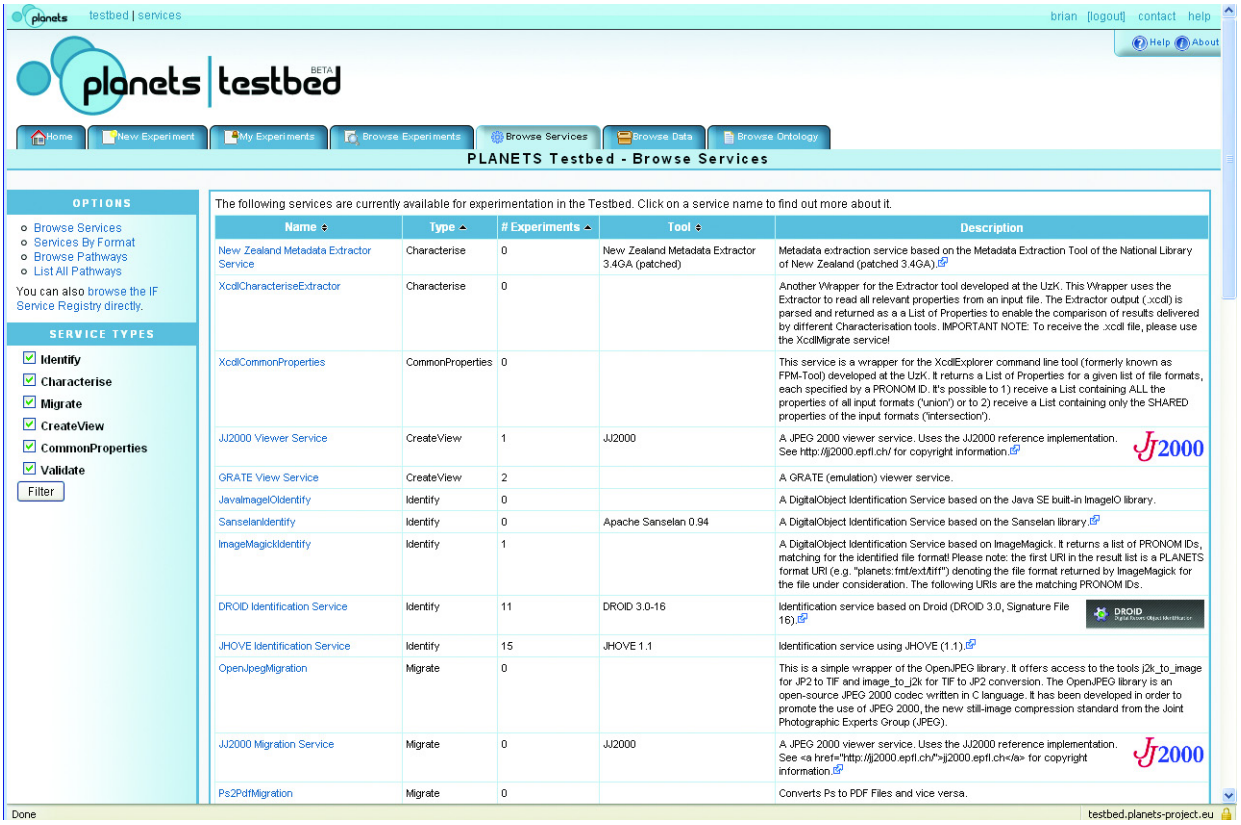


Figure 1: The testbed interface, listing services.

In order to perform such experiments, users can either provide their own data by uploading it to a dedicated FTP area or choose content from several large corpora of test files provided by the Testbed. These corpora cover a variety of popular and important file formats, and include edge-cases such as malformed PDFs and GIF files that have experienced bit rot. The Testbed not only provides access to this vast array of sample data, it also allows the exploration of corpora annotations that have been documented using the Extensible Characterization Definition Language (XCDL) developed by Planets, making them ideal control files for experimentation.

One of the principal aims of the Testbed is to create a shared knowledge base of digital preservation tool performance both on aggregated comparative measurements and on an individual experiment level. For this reason, experiment details, input files and outcomes are made available to all Testbed users. Furthermore, the Testbed facilitates the reproducibility of experiments: users can rerun any experiment to prove the validity of the results, or even adapt an existing experiment to specific requirements.

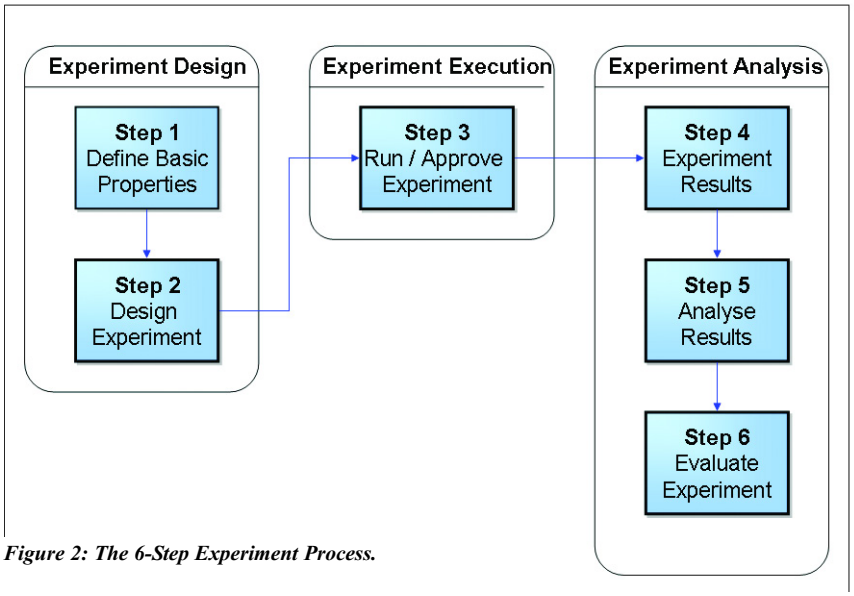


Figure 2: The 6-Step Experiment Process.

Testbed experiments follow a six-step process that is simple to use and flexible, as shown in Figure 2. At Step 1 the basic properties for the experiment are defined, including the overall experiment aims and objectives, contact details and references. In Step 2, the user formulates the design of the experiment, which includes selecting an experiment type and workflow, choosing the required preservation services and 'fine-tuning' parameters of the tool. Experiment input files are also selected at this stage. Step 3 executes the experiment workflow, gathers statistics relating to service execution and stores all results as output data, interim results and tool-specific log information within the Testbed.

At Step 4 the results of the experiment execution are displayed. Input and output files are listed (together with file

properties such as name, size and a thumbnail if appropriate) and can be opened or downloaded if required. Additionally, this page displays records of every operation that was conducted on each digital object, even in the case of a failure.

In Step 5 the results can be analysed. In order to assess the effectiveness of a preservation tool we need to investigate how digital objects that have undergone a preservation action differ from their original form. A variety of characterization and identification services that can automatically extract digital object properties are offered at this stage, together with options for manually recording properties. By comparing the significant characteristics in the original objects with the post-preservation action objects (ie migrated files or perceived objects within an emulated environment) it is possible to gain an understanding of the effectiveness of a tool. Finally, in Step 6 the user can provide an overall evaluation of their experiment, stating how well the requirements were met along with any other factors the user wishes to document.

By supplying a controlled environment in which users can experiment with preservation tools and share outcomes with others, the Planets Testbed aims to enhance the knowledge of preservation approaches and help users to make informed decisions about which tools and approaches are the most useful for particular tasks. For more information or a Testbed instruction pack, please contact helpdesktb@planets-project.eu.

Planets is partially supported by the European Community under the Information Society Technologies (IST) Programme of the 6th FP for RTD - Project IST-033789. Development of the Testbed will continue until May 2010.

Links:
http://www.planets-project.eu
https://testbed.planets-project.eu/testbed/

Please contact:
Brian Aitken
Humanities Advanced Technology and Information Institute (HATII)
University of Glasgow
Tel: +44 141 330 3392
E-mail: b.aitken@hatii.arts.gla.ac.uk

Andrew Lindley
Future Networks and Services
AIT Austrian Institute of Technology GmbH
Tel: +43 50550 4272
E-mail: andrew.lindley@ait.ac.at

Which Repositories are Worth their Salt?

by David Giaretta

...getting closer to an international system for audit and certification of the trustability of digital repositories.

The Preserving Digital Information report of the Task Force on Archiving of Digital Information (Garrett & Waters, 1996) declared, "A critical component of digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital collections", and "A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information". The issue of certification, and how to evaluate trust into the future (as opposed to a relatively temporary trust which may be more simply tested) is a request that has been repeated in many subsequent studies and workshops. The Open Archival Information System (OAIS) reference model provides many of the concepts on which such evaluations can be based.

Development of an ISO Accreditation and Certification process

The development of OAIS was hosted by the Consultative Committee for Space Data Systems (CCSDS) and approved by ISO as ISO 14721. OAIS contained a roadmap which listed a number of possible follow-on standards, some of which have already become ISO standards, after development within CCSDS.

The need for a standard for certification of archives was included in that list and the RLG/NARA work which produced TRAC (2007) was the first step in that process. The next step was to bring the output of the RLG/NARA Working Group back into CCSDS. This has now been done: the Digital Repository Audit and Certification (RAC) Working Group has been created, the CCSDS details are available from http://cwe.ccsds.org/moims/default.aspx#_MOIMS-RAC, and the working documents are available from http://wiki.digitalrepositoryauditandcertification.org. Both are open to everybody. The openness of the development process is particularly important and the latter site contains the notes from the weekly virtual meetings as well as the live working version of the draft standards.

The first of the two basic standards, Audit and Certification of Trustworthy Digital Repositories (RAC, 2009), has been submitted to the standardization process and the second, Requirements for bodies providing Audit and Certification of Trusted Digital Repositories (drafts available on the RAC wiki) is near submission.

Besides developing the metrics, the Working Group has also been designing a strategy for creating the accreditation and certification process. In addition to the 'central' accreditation body there will be an eventual need for a network of local accreditation and certification bodies.

Conclusions

It has long been recognized that there is a need for a way to judge the extent to which an archive can be trusted to preserve digitally encoded information. On the one hand, funders of such archives need some formal certification process to provide assurance that their funding is well spent and that their important digital holdings will continue

to be usable and understandable into the future. On the other hand it is probably also true that many who manage such archives would want some less formal process.

Considerable work has been carried out on the second of these aims, namely peer or informal certification. The RAC Working Group is close, at the time of writing, to taking important steps towards the first aim (formal ISO certification). Difficult organizational issues still need to be addressed but there is a clear roadmap for doing this. Even if all this is put in place, the take-up of the process and its impact on, for example, determining the funding of digital repositories is far from guaranteed. However in order to make progress, the RAC Working Group believes that the effort must be made.

Please contact:
David Giaretta
STFC, Rutherford Appleton Lab, UK
E-mail: david.giaretta@stfc.ac.uk

Links:

Garrett, J. & Waters, D, (Eds). (1996). Preserving Digital Information, Report of the Task Force on Archiving of Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group. Retrieved from: http://www.ifla.org/documents/libraries/net/tfadi-fr.pdf

National Science Foundation Cyberinfrastructure Council (NSF, 2007), Cyberinfrastructure Vision for 21st Century Discovery: http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

Open Archival Information System (OAIS) – Reference Model, ISO 14721:2003, (2003): http://public.ccsds.org/publications/archive/650x0b1.pdf

RLG-OCLC, (2002), Report on Trusted Digital Repositories: Attributes and Responsibilities: http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

TRAC, (2007), Trustworthy Repositories Audit & Certification: Criteria and Checklist: http://www.crl.edu/PDF/trac.pdf

RAC wiki: http://wiki.digitalrepositoryauditandcertification.org

RAC, (2009) Audit and Certification of Trustworthy Digital Repositories: http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206520R1/Overview.aspx

Best Practices for an OAIS Implementation

by Luigi Briguglio, Carlo Meghini,and David Giaretta

We introduce a more concrete and detailed intermediate level for long-term digital preservation between the conceptual model of the Open Archival Information System (OAIS) and the real world of the certified archives.

Digital information innervates modern civilization, and yet it is extremely vulnerable. Precious digital information created and stored all over the world is becoming inaccessible at a very fast pace. At the same time, paper documentation is increasingly being converted into electronic format, with even more now being 'born digital'. The availability of these electronic resources must be guaranteed for the future. Such reasons justify the need to acquire, preserve and maintain digital resources, so that the information contained in them may be always accessible and usable.

Much effort and activity has been dedicated towards digital preservation in recent years, but it still represents an open issue for research and development. Models, methodologies, approaches and prototypes have been proposed, and often solutions have been conceived and driven by the requirements of specific libraries and archives. However, the growing community of experts now recognizes and widely accepts the standard OAIS (Open Archival Information System; ISO:14721:2003) as the primary reference in this domain. Although OAIS itself is primarily a conceptual model and aims to standardize key concepts and mandatory responsibilities for any long-term archive, this reference model does not specify a design or an implementation. Hence, building a system that supports an OAIS-compliant archive requires the introduction of an intermediate level of detailed models and specifications. Following this, the implementation and prototyping should be evaluated (better if in different scenarios and domains) in order to validate the completeness and correctness of the models and specifications.

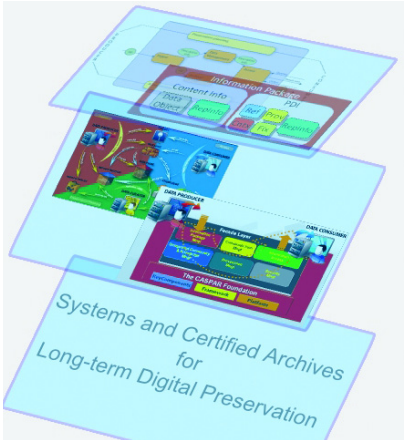


Figure 1: From OAIS to Certified Archives: the intermediate step provided by CASPAR.

Such a process has already occurred during the European Integrated Project CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval), which based modelling and specification activity on the OAIS reference model. In three years, CASPAR has addressed impor-

tant and still open research issues in digital preservation by:

- extending the information model and investigating fundamental related aspects such as representation information (RepInfo), preservation description information (PDI), Designated Community knowledge base characterization and evolution, authenticity concepts and tools, digital rights management and rights evolution
- proposing revisions to OAIS
- identifying, designing and prototyping key components which deal with the main concepts of the OAIS information model and represent best practices for many important digital preservation issues
- raising awareness in a variety of other communities (ie cultural, artistic and scientific) of digital preservation issues and the importance of the OAIS standard
- undertaking impact analysis on the validation methodology for the preservation activities.

The CASPAR approach has advanced the state of the art in digital preservation by enhancing and extending the OAIS standard, and this is its most important strength. In addition, by using the OAIS reference model, CASPAR has identified, specified and prototyped a framework and key components which could be applied in the digital preservation process and represent a valuable outcome to be reused for best practices, patterns, methodologies and references in the domain (for further details see the OSS community 'Digital Preservation Services' on SourceForge). It may reasonably be asserted that CASPAR has added an intermediate step between the abstract conceptual model provided by OAIS and the concrete world of digital preservation.

Links:
CASPAR:
http://www.casparpreserves.eu

Development Community:
http://developers.casparpreserves.eu

Please contact:
Luigi Briguglio
Engineering Ingegneria Informatica S.p.A, Italy
E-mail: luigi.briguglio@eng.it

Carlo Meghini
ISTI-CNR, Italy
E-mail: carlo.meghini@ISTI.CNR.IT

David Giaretta
STFC, Rutherford Appleton Lab, UK
E-mail: david.giaretta@stfc.ac.uk

Preserving the Past for the Future: Digital Technology for Film Archives

by Arne Nowak

The use of digital technology in filmmaking is on the increase. With this trend, the question arises of how to preserve for the future the enormous amounts of image data being produced. The Fraunhofer Institute for Integrated Circuits IIS together with leading European film archives has developed methods and formats for the long-term preservation of digital films and to provide easier access in a plethora of formats. Not only digitally produced movies benefit from this approach: it can also be used to make the inventories of film archives available to a wider audience.

Photographic film is known to have great longevity. Studies from various sources estimate the potential lifespan of film to be between several tens and some hundreds of years (Figure 1). Even films that are stored in non-optimal conditions have a good chance of surviving. This is not the case with digital images, however, and this is now an issue of major importance. In the mid-1990s, special effects and post-production began to move to digital, and the same change is currently taking place for onset image acquisition, with many film cameras having been replaced by digital versions. While in principle digital data can last forever, the media currently used to store the data often has a lifespan of only a few years. Additionally, there is the risk of obsolescence of media formats, leading to a high probability that even if a magnetic data tape is in excellent condition, it may be hard to find a drive and suitable software with which to read the data into a computer system. The same is unfortunately true regarding data formats. Most formats used in production are specifically developed for this purpose and are not well suited to long-term archiving. Proprietary formats are also used, and these may become unreadable if the manufacturer ceases supporting them or goes bankrupt.

The Fraunhofer Institute for Integrated Circuits IIS has worked together with its partners within the EU co-funded project EDCINE on the development of system concepts and formats to overcome these problems. The concept is based on the asset store approach of the Open Archival Information System (OAIS) reference model, where images and sound are stored together with descriptive data in the one place. JPEG2000 and MXF formats were chosen for encoding and packaging because they represent well-documented and open available standards that are used widely in the film industry. This ensures long-term support for and the usability of the archived data.

In the described system two main formats are used. The Master Archive Package (MAP) that contains images in their original resolution and colour bit depth with compression that uses only mathematically lossless methods. With resolutions currently at 4K (4096x2160) or even higher and compression ratios in the region of 2:1 this results in amounts of data to large to be handled in day-to-day usage. A second format was therefore introduced: the Intermediate Access Package (IAP). Here, a restricted range of resolutions

Figure 1: Film cans in a film archive.

and very high-quality, visually lossless compression is used to bring the data down to a reasonably usable volume.

The MAPs are normally stored on offline media such as magnetic tapes, while the IAPs are stored on hard disks and are used on a daily basis to create several different so-called dissemination formats. These are then delivered as requested to the users and customers of the archive. Each user thus receives the material in the format that is most useful for his or her particular use case (Figure 2). The main advantage of this approach is that it is much less complex than directly supporting a continuously growing number of original source formats. In addition, the supported dissemination formats can be changed over time and as appropriate to the target audience of a particular archive.

To enforce the idea of open and freely accessible standards for archiving, Fraunhofer IIS is active in the JPEG committee and has, during the course of the EDCINE project, developed three profiles for JPEG2000 that describe how to use this format in the context of audio-visual archiving. With the help of presentations at international conferences and a series of workshops for film archives, the idea of using JPEG2000 for this use case has been spreading in the archiving world and is now widely recognized.

During the EDCINE project, a prototype of a digital film archive system was developed and used to publicly demonstrate and evaluate the system concept. Fraunhofer IIS is now beginning to implement the concept in pilot projects with several leading European archives.

The EDCine Enhanced Digital Cinema project was funded by the European Commission within the 6th Framework Programme FP6/2004/IST/4.1, contract no. 038454 EDCine.

Link:
http://www.iis.fraunhofer.de/EN/bf/bv/cinema/index_edcine.jsp

Please contact:
Arne Nowak
Fraunhofer Institute for Integrated Circuits, Germany
Tel: +49 9131 776 5154
E-mail: arne.nowak@iis.fraunhofer.de

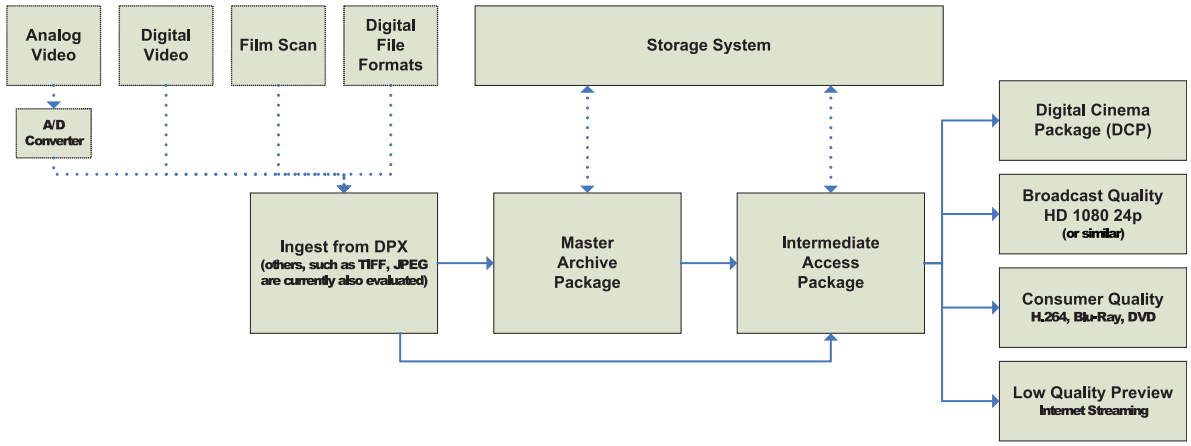


Figure 2: Overview of a digital film archive system.

Considering Software Preservation

by Brian Matthews, Arif Shaon, Juan Bicarregui, Catherine Jones, Esther Conway and Jim Woodcock

Software is a class of electronic object which is by its very nature digital, and the preservation of software is often a vital prerequisite to the preservation of other electronic objects. However, software has many characteristics that make preserving it substantially more challenging than for many other types of digital object. Software is inherently complex, normally composed of a very large number of highly interdependent components and often forbiddingly opaque for people other than those who were directly involved in its development. Software is also highly sensitive to its operating environment, with the typical software artefact depending on a large number of other items including compilers, runtime environments, operating systems, documentation and even the hardware platform with its built-in software stack. Preserving a piece of software thus involves preserving much of its context as well.

Handling these challenges is a major barrier to the preservation of software, so much so that the preservation of software is often seen as a secondary activity, less critical than the preservation of the data it manipulates. This is despite the fact that in many cases, such data becomes unusable without the software to handle it; and recreating software from partial information can be a near-impossible task.

Software preservation is thus a relatively underexplored topic and there is little practical experience in the field of software preservation as such. The e-Science Centre, Science and Technology Facilities Council (STFC) has undertaken preliminary studies sponsored by the UK Joint Information Systems Committee (JISC) into the Significant Properties of Software for preservation (2007), and subsequently in a development project on Tools and Guidelines for Preserving and Accessing Software Research Outputs (2007-09). Given the relative immaturity of the field, the studies became explorations of the notion of software preservation, looking at the stakeholders and motivations behind software preservation as much as identifying methods and technology.

Different communities have different motivations for preserving software, such as libraries and archives, managers of data archives who have a need to preserve associated software, and software developers who themselves maintain and reuse software over the long term. We therefore considered different approaches to software preservation, ranging from a strong emphasis on preserving software executables directly, which uses hardware preservation and emulation, to an emphasis on preserving the essential behaviour of software in a new context via migration and porting. The choice of preservation approach depends on the nature of the software artefacts available, the extent to which the original operating environment of the software can also be preserved or reproduced, and legal restrictions such as software licensing.

The project identified some concepts useful for a software preservation methodology, discussing the stages of retrieval, reconstruction and replay which need to be passed through to reproduce a usable performance of a software product. We defined a notion of adequacy of preservation, an aspect of the authenticity of preservation which tests the future performance of software against specified preservation properties once it has been reconstructed into a working version in the new environment. We developed a new software preservation framework to categorise software components and identified the properties required to ensure adequate preservation. This software preservation framework uses concepts from the Open Archival Information System (OAIS) information model; indeed the framework can be seen as a specialization of OAIS for the case of software.

We tested the software preservation framework in collaboration with the British Atmospheric Data Centre (BADC). This included assessing the overall efficiency of the framework against a variety of BADC software, specifically in terms of its relevance (to the software) and sufficiency (of the information recorded) for long-

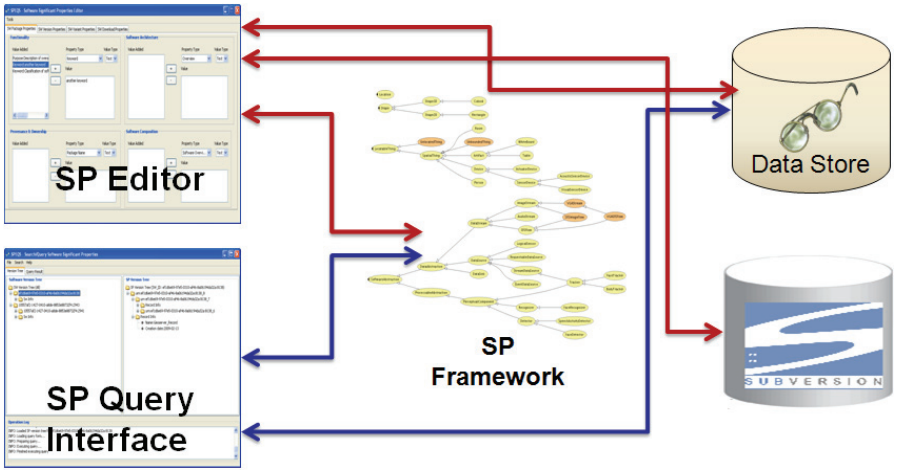


Figure 1: Using the SPEQS tool to capture preservation properties of software.

term preservation of the software. The cost-effectiveness of the framework must be considered within the context of the BADC's approach to accommodating changes in the technological environment to ensure effective long-term software maintenance and reuse.

Software engineering best practice shares many of the concerns of software preservation in producing quality software that can be maintained and reused in the future, such as providing version control, dependency analysis and good documentation. Software preservation could be integrated into the software life cycle to systematically capture those properties required for preservation and an adequate replay of the software. We have provided a Java-based tool called Significant Properties Editing and Querying for Software (SPEQS), developed in view of our analysis of the BADC use case (Figure 1). SPEQS demonstrates the feasibility of capturing preservation properties identified in the software preservation framework and integrating these with the software development life cycle, to aid in long-term preservation.

Link:
British Atmospheric Data Centre:
http://badc.nerc.ac.uk/home/index.html

Please contact:
Brian Matthews
e-Science Centre, Science and Technology Facilities Council, UK
E- mail: brian.matthews@stfc.ac.uk

Electronic Records Management in Luxembourg: Challenges and Perspectives

by Lucas Colet

In order to comply with regulations or for management purposes, it is increasingly required that information be stored for long periods of time. Moreover, to retain their legal value, such records need to have the following properties: authenticity, reliability, integrity and usability. But how can these characteristics be guaranteed in an electronic environment? This problem is of major concern in Luxembourg, where a large part of the economy consists of service companies with a strong technological background.

To solve this problem, it is necessary to use Electronic Records Management Systems (ERMS) that satisfy the requirements related to good records management. However, these requirements are complex and not uniformly defined in the literature. It is thus a real challenge to manage electronic records and create an ERMS.

ERMS activities can be structured as illustrated in Figure 1. The workflow starts when a document meets the requirements of the capture policy: it is then captured by the ERMS and becomes a record. The workflow ends when a record reaches the end of its legal or internal use duration and needs to be deleted (if the law gives such instruction or if the organization no longer needs it). Otherwise, it can be archived, but has lost its legal value.

While the workflow of the ERMS is important, how can we guarantee that the probing value and the characteristics of the record are preserved during the whole process? Many standards have been defined in attempts to deal with this issue.

The ISO 15489 standard was published in 2001. Its objective is to define the scope and benefits of records management (RM), to provide guidance in determining the responsibilities of organizations for records and records policies, procedures, systems and processes, and to provide guidance in the design and implementation of a records system. However it does not include the management of archival records within archival institutions. This standard is generic, and does not specifically address the electronic question.

MoReq2 is a model prepared in 2008 for the European Commission. It can be seen as an electronic and operational approach to the ISO 15489 standard. The main objective is to describe a way to create specific software for electronic records management. This model does not deal sufficiently with organizational aspects. The DoD 5015.02-STD from the US Department of Defense is quite similar.

The ISO 14721 defines a model (Open Archival Information System; OAIS) that targets long-term archiving of electronic information. The main goal of OAIS is to ensure that the Designated Community is still able to understand the content of the archived documents without any additional explanation. OAIS does not cover the entire process of RM, because it deals with archived documents rather than records.

The NF Z42-013 standard provides specifications for technical and organizational measures for capturing, storing and retrieving electronic records in an information system, to assure integrity and preservation during the whole record's lifetime. This standard is general in the conception of the system.

To meet the requirements of the Luxembourgish activities, a combination of these standards would be necessary.

NormaFi-IT (*Normalisation, Finance & Information Technology*) is a joint R&D project started in 2009 between ILNAS (Institut luxembourgeois de la normalisation, de l'accréditation, de la sécurité et qualité des produits et services), the Luxembourgish organization responsible for norms and accreditation, and CRP Henri Tudor which aims at investigating the state of the art of electronic records management.

From this perspective, a Working Group has been set up to define the present and

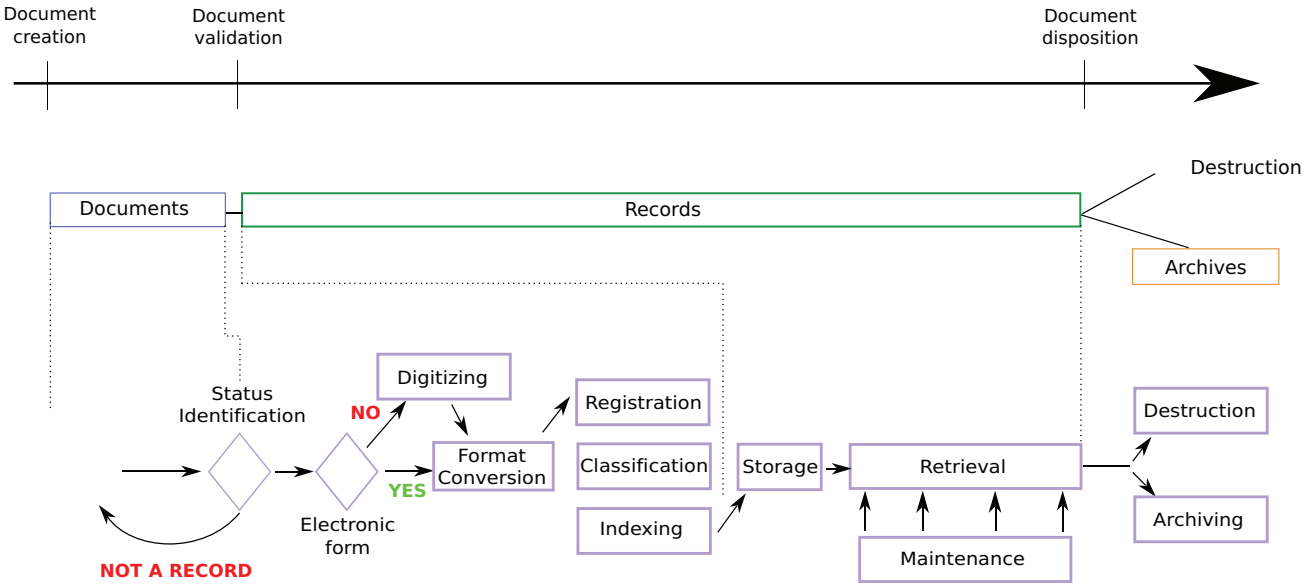


Figure 1: Workflow of the ERMS.

future needs and objectives of electronic records management in Luxembourg, and has already broadened the vision of the field. Indeed, while the ISO 302XX series (Management System for Records) is still under development, the need for a standard (or at least guidelines) is urgent for many sectors of the Luxemburgish economy.

Any future guidelines or standards might be complementary with the Luxembourgish PSF/S (*Professionnel du Secteur Financier/Support*", an organisation managing key business process or data for a financial organisation) status, which allows financial professionals to trust certified third-party service providers for their outsourcing. A major step will be the introduction of the third-party archiving status, and the description of accreditation methods. Moreover, players in the process need to be identified, and their roles and obligations clearly defined.

Furthermore, NormaFi-IT will rely on future legislation on dematerialization in Luxembourg, which will potentially create equivalence between paper-based and electronic-based (born digital or otherwise) probing value for audits, for administration or in front of a court. The creation of a new legal and normative framework will need the consensus of all concerned players.

Finally, a new research study will start to allow private individuals to take advantage of ERMS, in administration, e-invoicing and other areas.

Please contact:
Lucas Colet
CRP Henri Tudor, Luxembourg
Tel: +352 42 59 91 - 421
E-mail: lucas.colet@tudor.lu

Knowledge Management for Digital Preservation

by Yannis Tzitzikas and Vassilis Christophides

We can preserve the bits, but what about the knowledge encoded in them? Modern societies and economies are increasingly dependent on a deluge of information that is only available in digital form. The preservation of this information in an unstable and rapidly evolving technological (and social) environment is a challenging problem of great importance. The CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project built a pioneering framework to support the end-to-end preservation 'life cycle' for scientific, artistic and cultural information based on existing and emerging standards. CASPAR aimed to preserve not simply the bits of digital objects but also the information and knowledge that is encoded in these objects.

The key contributions of CASPAR regarding knowledge management revolve around four main topics:

(a) Intelligibility. Since it is hard to define explicitly what information or what knowledge is, it is equally difficult to claim that a particular approach, methodology or technique can indeed preserve information and knowledge. To tackle this issue and for preserving the meaning of digital objects, CASPAR formalized the notion of intelligibility in an OAIS-compliant manner and provided guidelines, methodologies and components that can aid humans in preserving information and knowledge. Specifically, it formalized the notion of

intelligibility and an intelligibility gap through the notion of dependency. This perspective allows us to answer questions relating to (a) what kind of (and how much) representation information we need, (b) how this depends on the designated community, and (c) what kind of automation we can offer (regarding packaging and dissemination). Apart from developing formal and conceptual models, CASPAR has developed tools and applied them to real data.

(b) Semantic Web evolution management. Evolution is a key concept, because preservation is a dynamic process; the world evolves, software and hardware evolve, metadata schemas evolve, digital objects evolve, community knowledge evolves. This change poses several challenging requirements for semantic Web repositories, including bulk metadata updates, versioning metadata and ontologies, ontology evolution, comparison operators and change log management. CASPAR has developed an advanced platform for semantic Web management (SWKM) for tackling the above requirements.

(c) Provenance modelling and querying. There is a need for a comprehensive and extensible conceptual framework that will allow provenance information to be integrated, exchanged and exploited within or across digital archives. CASPAR extended the ISO standard CIDOC CRM, defining CIDOC CRM Digital to explicitly model digital objects, and showed how it can be employed for provenance queries.

(d) Automating the ingestion of metadata. The creation and maintenance of metadata is a laborious task whose benefits may only be obvious in the longer term. There is a need for tools that automate as much as possible the creation and curation of preservation metadata. CASPAR developed PreScan, a tool for automating the ingestion phase. It can bind together automatically extracted embedded metadata with manually provided metadata and dependency management services (recall Intelligibility), and transforms the metadata according to CIDOC CRM Digital.

For each of the aforementioned topics, CASPAR has provided guidelines, methodologies and components. Regarding software infrastructure and components, Figure 1 depicts some of the key software components. SWKM (Semantic Web Knowledge Middleware) is middleware for semantic Web data management, GapManager (API and WebApp) offers dependency management services related to intelligibility, PreScan is a tool for automating the extraction, transformation and ingestion of metadata, while a number of other tools for inserting, editing, browsing, searching and visualizing semantic Web data have been designed and developed. The detailed results of this research have been published at DEXA'07, ECDL'07, ISWC'07, ESWC'08, ECAI'08, MEDES'09 and ISWC'09.

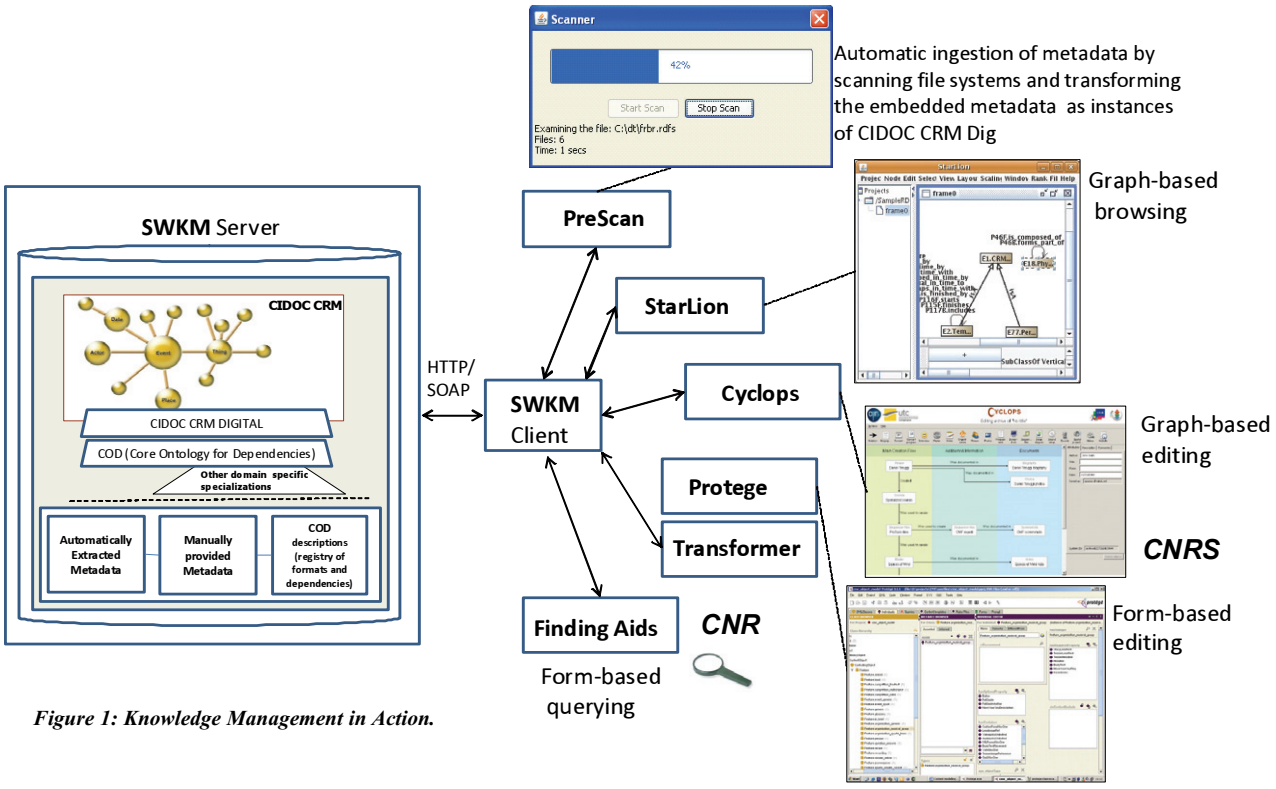


Figure 1: Knowledge Management in Action.

Links:
CASPAR project:
http://www.casparpreserves.eu/
GapManager:
http://athena.ics.forth.gr:9090/Applications/GapManager/
PreScan:
http://www.ics.forth.gr/prescan
SWKM:
http://athena.ics.forth.gr:9090/SWKM/
StarLion:
http://www.ics.forth.gr/~tzitzik/starlion/

Please contact:
Yannis Tzitzikas
FORTH-ICS, Greece
E-mail: tzitzik@ics.forth.gr

Automating the Ingestion and Transformation of Embedded Metadata

by Yannis Tzitzikas and Yannis Marketakis

Can we create automatically and at no cost ontology-based metadata repositories? Work carried out in the context of the CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project has been tackling this challenge.

The majority of preservation approaches rely on metadata. However the creation and maintenance of metadata is a laborious task that does not pay off immediately. For this reason there is a need for tools that automate as much as possible the ingestion and management of metadata. Our objective is to bypass the strict (often manual) ingestion process while at the same remaining compatible with it. According to the traditional approach, the ingestion phase starts with assigning identifiers to the objects and then extracting or creating metadata for these objects. These metadata can be expressed using various metadata schemas and formats, and usually the updating or movement of metadata records is prohibited. We want to relax these constraints and automate the process of metadata extraction and ingestion. Automation is crucial for the preservation of emergent systems and structures, like file systems, which are much more complex and dynamic than traditional digital archives.

In general, metadata can be stored either internally, ie in the same file with the data, or externally, ie data and metadata are stored in separate places or systems. The former are called embedded and the latter detached. PreservationScanner (PreScan for short) is a tool that we have developed for automating the extraction, transformation and maintenance of embedded metadata. PreScan is quite similar in spirit to the crawlers of Web search engines (WSE). For the problem at hand, we have to scan file systems, extract the embedded metadata from files of various types and build a metadata repository. In contrast to WSE crawlers, we have to (a) support more advanced extraction services, (b) allow the manual enrichment of metadata, (c) use more expressive frameworks for representing metadata (ie SW languages), (d) associate the extracted metadata with other sources of knowledge (eg registries of format types), and (e) offer rescanning services that do not start from scratch but exploit the previous status of the repository in order to preserve the human-provided metadata.

PreScan starts like an antivirus program by scanning files from a specific folder and continues transitively to its subfolders. For every file it encounters it identifies the file format and extracts the embedded metadata. PreScan has a modular design and can work with several format identifiers and metadata extractors. Currently it uses JHOVE (JSTOR/Harvard Object Validation Environment) for this purpose. PreScan is capable of transforming the extracted metadata into ontological metadata expressed in RDF. Currently the extracted metadata are transformed to descriptions according to CIDOC CRM (CIDOC Conceptual Reference Model) Digital ontology. The user has the option to enrich the metadata of a file by providing additional information.

Periodic scans are supported too. Here PreScan identifies the files that have been renamed or moved to another location by comparing (through hash functions) the contents of files that have vanished (files that existed at the previous scan but are not there now) with new files encountered during the current scan. It suggests these matches to the user who in turn approves the correct ones (this is critical for preserving the human-provided metadata of files that have changed location).

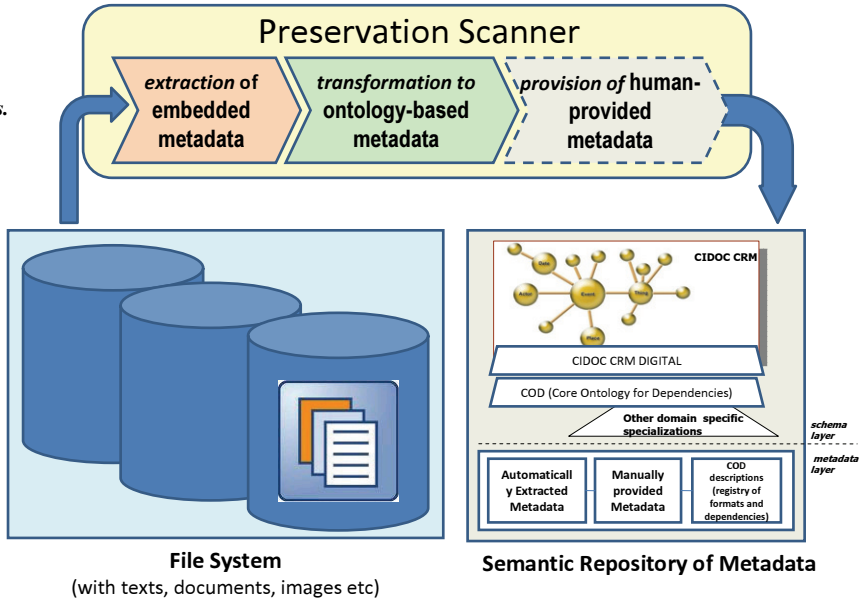


Figure 1: Creating automatically ontology-based metadata repositories.

PreScan currently recognizes and extracts the embedded metadata from twelve file types (from which we get around 150 attributes in total), and it takes around ten hours to scan, extract and transform the metadata of a hundred thousand files.

Regarding the metadata repository, several (not mutually exclusive) options are supported: (a) all metadata records are stored in a folder specified by the user, (b) each metadata record is stored in the same folder as the scanned file, and (c) the contents of the metadata records are stored in a semantic Web knowledge base (specifically at SWKM (Semantic Web Knowledge Middleware)). The latter choice allows these metadata to be linked with other sources of knowledge (eg from registries). Furthermore, it offers declarative query and update services, which are important for building obsolescence risk detection services, notification services, and services relating to the intelligibility of digital objects.

This work has been done in the context of the CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project.

Links:
PreScan: http://www.ics.forth.gr/prescan
CASPAR: http://www.casparpreserves.eu/
Related Publication:
Y. Marketakis, M. Tzanakis and Y. Tzitzikas, "PreScan: Towards Automating the Preservation of Digital Objects", ACM Conference on Management of Emergent Digital EcoSystems, MEDES'2009, Lyon, France, Oct. 2009

Please contact:
Yannis Tzitzikas
FORTH-ICS, Greece
E-mail: tzitzik@ics.forth.gr

The Art of Preserving Digital Creativity in Planets

by Andrew McHugh and Leonidas Konstantelos

While characterizing digital art materials for long-term preservation is laced with considerable complexity, it offers insights applicable to the preservation of all kinds of contemporary and emerging materials. The Planets project (Preservation and Long-term Access through Networked Services) is leading research into emerging characterization approaches that will safeguard the availability of diverse digital experiences.

As part of its contribution to the Planets project, the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow is creating vocabularies and information structures for adequately characterizing the value of digital art. Value is encompassed in those qualities that must be understood and captured in order to ensure that art works' sensory, emotional, mental and spiritual resonance remain. Facets of interactivity, modularity and temporality associated with digital art present critical questions that the preservation community must acknowledge; HATII and Planets' intention is to highlight risks and approaches that may be applicable to a wider, more generic range of materials. A definitive ontological model for characterizing the many relevant facets of new media is being conceived; because digital art materials exhibit fundamental multidimensionality, validating the preservation of creative experience demands the explication of more than just file characteristics. Understanding relationships between objects also implies an understanding of their respective functional qualities. By aiming to solve some of the more difficult issues of digital persistence within this notably challenging domain, we hope to highlight future research directions that could accommodate the increasingly complex digital infrastructures of tomorrow.

Figure 1: Lynn Hershman Leeson's 'Synthia'. In this work the mannerisms of an animated character rendered onscreen are influenced by live stock market data.

Art communicates simultaneously on sensory, emotional, mental and spiritual levels. For digital art, these levels of impact and our comprehension of value are based not just on tangible characteristics of the piece in question, but on many additional contextual factors that may be permanent or transitory, localized or global and either physical or conceptual. Furthermore, those qualities considered more intrinsic to works may be difficult to characterize. Contemporary art typically establishes, encourages and demands greater levels of dialogue than the traditional fruits of creativity. Whereas paintings or sculptures are largely consumed in a passive manner by audiences, digitally equipped installations (most obviously net art) promote a high degree of often distributed user involvement. Meaning is less than self-evident; unlike more traditional art where the materials used are largely subservient to the implicit message, it is commonplace within contem-

porary works for specific component materials to have tremendous implications for the overall meaning. These issues are shared by digital materials more generally – they regularly exhibit complexity of interpretation, consumption and application in excess of those physical materials with implicit, unambiguous usefulness. Of critical importance is the extent to which information and the associated means of representation or experience are tightly or loosely coupled. Numerous logical and physical layers must exist to support the presentation and understanding of digital information: this is in contrast to analogue information, which exists largely atomically. More layers introduce more complex dependencies between those layers; any preservation action (to alter the format of a digital image component for example) can have implications far in excess of the intended extent of the intervention. From the artist's perspective, complexity creates opportunities for variation of behaviour and performance. While this contributes to, rather than detracts from, the significance and impact of the creative expression, it introduces difficulties for those seeking to characterize and preserve that which is definitive in and around a digital work.

Further complications arise from the often modular nature of contemporary installations, whereby components operate based on inputs from discrete linked systems. Lynn Hershman Leeson's 'Synthia' is a good example. In this work the mannerisms of an animated character rendered onscreen are influenced by live stock market data. Partly contextual, partly intrinsic, the flow of data must be made persistent for the piece to be correctly exhibited. We see similar phenomena within the digital context more generally; applications and file formats are increasingly networked, and are more and more reliant on decentralized services. How we deal with the preservation challenges associated with maintaining third-party services or user contributions is particularly challenging. Web archiving appears trivial when dealing with simple networks of linked, static Web pages. When the relationships between scripts, users, Web services, databases and rights management systems become more intricate and integral, preservation becomes less akin to photocopying and more like performing organ transplant surgery, with all of the risks that digital materials will be 'rejected' within their anticipated preservation environment.

From the conservator's perspective, documentation takes on a critical role. In those cases where art relies on bespoke, deteriorating materials, externally managed and originating services or a critical mass of community involvement there may be no way to ensure its availability. Nevertheless, the maintenance of appropriate linkable and navigable documentation can assist conservation and preservation strategies, most notably offering opportunities to characterize value and formulate priorities for individual examples. This can then inform the selection of subsequent conservation or restoration strategies, and ensure their consistency with the spirit of the piece.

This work was funded by the Planets (IST-2006-033789) Project, funded by the European Commission's IS&T 6th Framework Programme.

Links:
http://www.planets-project.eu/
http://www.hatii.arts.gla.ac.uk/

Please contact:
Andrew McHugh
University of Glasgow, UK
Tel: +44 141 330 2675
E-mail: a.mchugh@hatii.arts.gla.ac.uk

User-Centered Digital Preservation of Multimedia

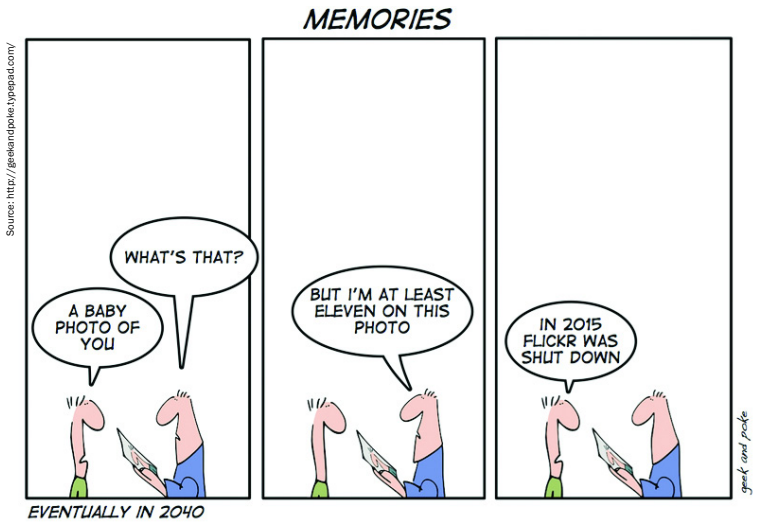
by Egon L. van den Broek, Frans van der Sluis and Theo E. Schouten

Everything expressed by humans in whatever form, arouses emotions in every one, who witnesses that expression. Those emotions are dependent on the witness and vary over time. For instance, an expression like "I'm now going to smoke a cigar in my office" uttered today brings about other emotions than 10 years ago. To really preserve (digital multimedia) expressions, the different kinds of emotions it arouses have to be preserved.

Traditionally, digital preservation (DP) is approached from engineering rather than a user perspective. Consequently, definitions such as these have been proposed: "Digital preservation combines policies, strategies and actions that ensure access to digital content over time" (ALA, 2007). This article approaches DP of multimedia from an end-user perspective. More specifically, we suggest that users' most important association with virtually all media be taken into account and, as such, that we introduce a new dimension to DP: emotion. This article discusses how this new dimension can be integrated in traditional frameworks as used for DP; see also Figure 1.

Let us start with denoting our perspective on knowledge representation (KR), which is founded on its traditional definition. In line with AIM (1993), we argue that KR can play five distinct roles, each crucial. A KR can be (i) a substitute for the object itself; (ii) a set of ontological commitments; (iii) a fragmentary theory of intelligent reasoning; (iv) a medium for pragmatically efficient computation; and (v) a medium of human expression. This perspective on KR is already more user-centred than the ALA (2007) definition of DP.

Through traditional KR, a range of information can be captured, including multimedia; eg SMIL (2008). Regrettably, we suffer from an information overload and multimedia retrieval techniques are not as good as sometimes thought, relying on the extraction



of low level features. While more complex compounds of these low-level features enable the definition of high-level features (eg objects), the mapping of such features, either low- or high-level, on semantics is still an unsolved problem. Consequently, (semi-automatic) annotation of multimedia data is still the best solution.

As stated, we want to take multimedia KR one step further. As is now generally acknowledged by the community of artificial intelligence, emotions play a crucial role in understanding human intelligence, creating artificial intelligence, and the interaction between entities (eg human and computer). We adopt this notion and, in addition, state that it is of importance to include emotions when aiming for DP, as it is a primary form of human expression and consequently human communication; cf AIM (1993). For example, laypersons can benefit from an enriched representation of an abstract painting that describes its emotional expression.

Recently, the W3C launched the first working draft of the Emotion Markup Language (EmotionML). EmotionML is "a 'plug-in' language suitable for use in three different areas: (1) manual annotation of data; (2) automatic recognition of emotion-related states from user behaviour; and (3) generation of emotion-related system behaviour." Each of these three areas is of interest for DP, as we will explain next.

EmotionML gives concrete possibilities for capturing the emotional communication of digital art; manual (or semi-automatic) annotations can preserve a much richer representation than can be obtained solely through traditional KR means. The manual and semi-automatic annotations of the emotional expression of art is, to some extent, currently possible; see also Figure 1.

Automatic recognition of emotions - 'affective computing' - goes well beyond the scope of traditional DP. Nevertheless, it can be exploited to automatically annotate DP; see also Figure 1. Although not yet mature and struggling with various problems, affective computing is currently moving into its subsequent stage of development (Guid, 2009). Through speech signals, computer vision, movement recordings and biosignals, users' affective states can be determined (to a certain extent). EmotionML can help in automatically capturing and preserving the different

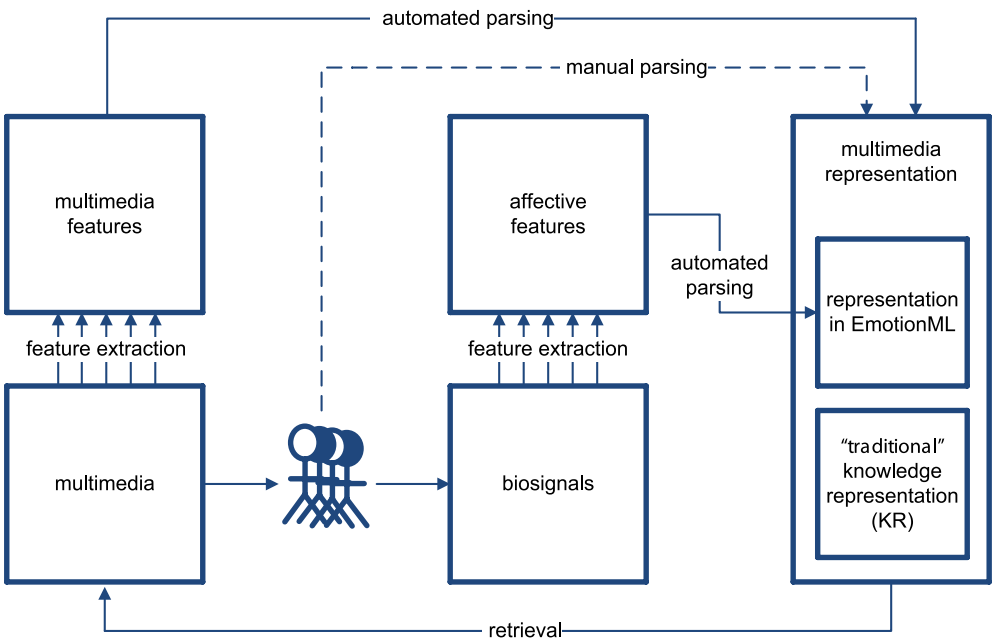


Figure 1: A scheme for the process of user-centred digital preservation of multimedia. The process starts with either an annotation of the user, which enables digital preservation, or a user's query that requires digital preserved multimedia. These users can be both the same or different persons. The boxes denote information sources. Arrows denote either core (solid lines) or optional (dashed line) processes. Further, please note that this scheme is simplified; eg fusion and classification processes are omitted.

emotional experiences, and through that various perspectives on the emotional expression.

The further DP develops, the more important the area of 'emotion-related system behaviour' will become. With DP, not only the storage of the KRs is crucial: access to the system (including its user interface) and the retrieval of the KR is of the utmost importance. In addition, non-specialists will increasingly need to be able to access the systems, to fully experience the "replay" from a wanted perspective. Affective computing can aid this interaction, as is generally acknowledged in the usability and human-computer interaction communities.

Let us consider the example of an abstract painting. The emotion will depend on the painting, the viewer, and the context (eg time). EmotionML can help in capturing the emotional expression of the painting; eg through (low-level) multimedia feature extraction. In further iterations, EmotionML can support capturing and preserving the emotional experiences through affective computing and, thereby, make emotional preservation automatic and incorporating different perspectives. However, an open issue remains in the user's perspective of the painting's expression; possible awareness for this perspective can be supplied using emotion-related system behaviour.

Taken together, fully automated generation of KRs of multimedia is beyond science's current reach. Nevertheless, we introduce a new dimension: emotion. The same holds for this new dimension as for multimedia analysis in general: semi-automatically is the best we can do. Nevertheless, developments in multimedia analysis, affective computing and in understanding humans continue to gain in speed. So, it is a matter of time before enriched digital preservation of multimedia, including its affective annotation, will evolve from theory to practice.

Links:
ALA (2007):
http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm
AIM (1993):
http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html
EmotionML (2009):
http://www.w3.org/TR/emotionml/
SMIL (2008):
http://www.w3.org/AudioVideo/
Guid (2009):
http://emotion-research.net/acii/acii2009/guidelines-for-affective-signal-processing-from-lab-to-life/

Please contact:
Egon L. van den Broek
Human Media Interaction, University of Twente, The Netherlands
Karakter, Radboud University Medical Center Nijmegen, The Netherlands
Tel: +43 1 956 1530; +31 24 355 8120
E-mail: vandenbroek@acm.org
http://www.human-centeredcomputing.com

Communication and Preservation in Academic Research: Current Practices and Future Needs

by Filip Kruse and Annette Balle Sørensen

Today's researchers work in hybrid environments that require, in addition to traditional, analogue methods of working, the use of an increasing amount of digital material and communication tools. This is forcing us to change our understanding of how researchers communicate, how they are connected in networks, and which parts of their research activities need to be preserved. But how should we rethink these issues? Researchers' current practice and requirements are analysed and future consequences reflected upon in a questionnaire-based survey from Aarhus University in Denmark.

According to most researchers, intermediate research results such as drafts, preliminary findings or datasets are important to preserve. Likewise, access to such results should not be restricted to the actual researchers involved. Previous research activities and professional networks are also essential to the majority of researchers, for the generation of new ideas as well as for the research process in general. Finally, communication with the network is of central importance for the entire research process including generation of new ideas as well as initiation and completion of projects.

Almost all the researchers' networks are cross-institutional and international. E-mail is the preferred form of communication with the network. Researchers generally prefer information or data in digital form rather than printed, but researchers from the arts and humanities are split down the middle on this question.

The findings of the survey highlight the importance of e-mail communication, both as a medium in itself and as an element in maintaining researchers' professional networks. The issue of preservation thus carries a double meaning, i.e. both preservation of the communication of research results from the initial idea to the final results, and preservation of the network. The first stresses the influence of preservation and dissemination on the creative flow of thoughts and ideas. The second focuses on social interactions and their role in the formative processes of the network.

Though the importance of e-mail communication is obvious, researchers from the arts and humanities and the social sciences rate its importance slightly lower than those from the health and natural sciences. The importance of e-mail communication does not imply that all e-mails should be preserved. On the other hand most researchers state that they do need to preserve more research data or information. The message from the research community is quite clear: e-mail is important, but not every e-mail related to research should be preserved; they are already critically

sorted. Facilities are needed to preserve more research data and information, as well as to ensure that data remains accessible: more than two thirds of the researchers, mostly from the natural and health sciences, have experienced problems in accessing old digital data.

In the 'good old days', one might expect that research was mainly an individual activity with the final results of the research being the only item worthy of preservation. In contrast, our findings clearly indicate that both professional networks and previous research activities are very important for the majority of researchers. The importance of networks lies in communication: networks act as forums for the exchange of ideas, and for reflection and discussion. In this context the importance of preserving intermediate research results and making them accessible to other interested parties becomes clear. Research thus seems to be developing into a more collaborative and cooperative activity in which the networks play an important role. Almost all the researchers' networks are cross-national and cross-organizational. Communication is digital – via e-mail – but is frequently supplemented by face-to-face communication.

The planning of future preservation activities must take into account the obvious need to preserve intermediate research results as well as final publications. At the same time, the consequences of the changed social organization of research activities must be considered. This points to a greater focus on researchers' networks as spheres of communication, and thus to new items for future preservation.

The Planets DT/7 Work Package group, from left: Filip Kruse, Annette Balle Sørensen, Jørn Thøgersen, Bart Ballaux, and John W. Pattenden-Fail.

This survey was carried out in November/December 2008 as part of the Planets (Preservation and Long-term Access via NETworked Services) DT/7 work package. The Web-based questionnaire was mailed individually to all researchers at Aarhus University. The survey population included 2722 researchers from the five faculties: theology, humanities, social sciences, natural sciences, and health sciences, and the questionnaire was completed by 404 researchers corresponding to a total average response rate of 14.8%.The survey was preceded by a series of qualitative analyses (interviews and data probes) carried out by HATII, University of Glasgow, UK, The Nationaal Archief of the Netherlands, The Hague, NL, and State and University Library, Aarhus, DK.

The survey was carried out by the authors and Jørn Thøgersen, also of the State and University Library. The questionnaire was prepared jointly by the authors, John W. Pattenden-Fail of HATII, University of Glasgow, UK, Bart Ballaux of The Nationaal Archief of the Netherlands, The Hague, NL , and Jørn Thøgersen.

Link:
The full report with appendices and the questionnaire deployed is available at: http://www.planets-project.eu/docs/reports/Planets_DT7-D4_Questionnaire_Report.pdf

Please contact:
Filip Kruse
State and University Library, Denmark
Tel: +45 8946 2241
E-mail: fkr@statsbiblioteket.dk

Annette Balle Sørensen
State and University Library, Denmark
Tel: +45 8946 2154
E-mail: abs@statsbiblioteket.dk

Preservation Planning: User Requirements for Digitally Preserved Materials

by Annette Balle Sørensen and Filip Kruse

Libraries and archives carry the responsibility of capturing and preserving 'representative samples of society', covering both cultural and scientific production. In the digital world this obligation has extended to include not only diverse physical outputs (books, journals, music records, newspapers etc), but also digital preservation of both analogue and digitally born materials. Here we describe a subproject of a larger user study targeted to identify user requirements for the digital preservation of documents, records and data sets. The central message to be communicated from this study is that requirements reflect usage type rather than user type.

A preliminary model of general user requirements for preservation planning was developed based on qualitative studies, including cultural data probes and contextual design adjusted for this specific model (Snow et al., D-Lib Magazine, 14, (5/6) 2008). These initial studies focused on users of archives (in the Netherlands), users of data centres (in Scotland), and users of libraries (in Denmark), and assembled the identified general requirements into a first-round preservation planning model (see link below). In order to refine and improve this model, including identification of

more specific requirements, a final qualitative approach was chosen, in which differences among user groups and collections were explored, instead of identifying common themes.

It might be anticipated that users of archives, data centres, and libraries would produce very different requirements for a preservation planning tool. However, our study suggests that this is in fact not the case. It is the type of usage, rather than the type of user that determines the specific requirements. For example, a scientist using a data centre of natural science information may have little concern for appearance, since the data may just be tables of numbers. This is contrary to an archival user, who is often dealing with scanned or digitized information, and who may be very concerned with the completeness or resolution of archival documents. An archivist or preservation officer at either of these organizations would take these intents into consideration when selecting requirements for preservation planning. Based on our findings of different types of usage, we propose a series of questions to be asked during the preservation planning process that, depending on the answers, will alter the priorities of the requirements. These questions are not focused directly on libraries, archives and data centres, as those boundaries are somewhat artificial; any type of usage can certainly occur in any type of institutional setting.

We have found that the following six central questions pay regard to different usage needs and scenarios:

- Is the content digital-born?
- Is this content likely to be represented in paper/analogue format?
- Is the appearance of this content relevant?
- Should the content be searchable?
- Should it be possible to alter or edit a personal copy of the content?
- Should it be possible to verify the provenance of the content?

By combining these simple questions with a tree of abstract requirements, the requirements can be weighted to indicate their relevance, on the basis of which a decision support tool can be designed.

Archive users were presented with experimental sets of records, including three representations of an original Word Perfect 5.1 document and two representations of an e-mail, along with different representation methods for metadata. The users were asked to reflect and comment upon the representations during the 1½ hour session and to indicate, among other things, their preference for a specific representation, the reasons for this, and their preferences for availability of metadata.

Users of data centres in the initial phase of the study provided actual sample documents representing a cross-section of their daily work. The samples were subsequently manipulated, creating a series of variations to emulate potential migrations that may occur during preservation actions. These variations were presented to the participants at the second interview, where the alterations were discussed and commented on.

Users of libraries and their collections, represented by a group of active researchers, were invited to a three-hour workshop on the theme 'original content and potentially migrated or altered content'. They were asked to comment on and discuss cases from three different collections using their individual professional experiences as the basis for their reflections. The collections targeted were: the Danish national collection of newspapers, represented by different digital and paper versions of an article in Politiken (a national newspaper), examples from the digitization of the collected works of Søren Kierkegaard, and examples from the Web Archive (a comprehensive archive of Danish web pages).

This article draws upon the 'Report on usage models for libraries, archives and data centres: final results' by John W. Pattenden-Fail (HATII, University of Glasgow, UK), Bart Ballaux (The Nationaal Archief of the Netherlands, The Hague, NL), Laura Molloy (HATII, University of Glasgow, UK), Jørn Thøgersen (State and University Library, Aarhus, DK), Filip Kruse (State and University Library, Aarhus, DK), and Annette Balle Sørensen (State and University Library, Aarhus, DK) [unpublished] under the auspices of PLANETS (Preservation and Long-term Access via NETworked Services).

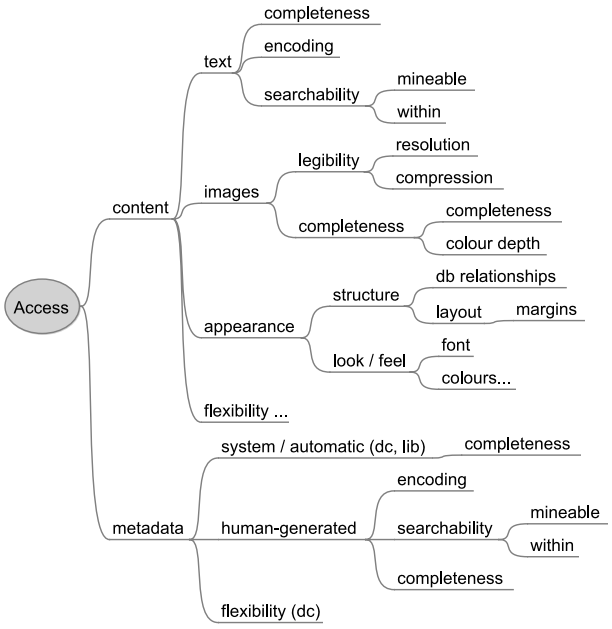


Figure 1: Preservation planning model. Three slightly different approaches were taken. In previous studies, affinities between user groups were important; in this study however, the focus was on specific clarification of existing requirements, discovering new requirements and differences in usage across the three target areas.

Links:

http://www.planets-project.eu/
http://www.planets-project.eu/publications/?search[0]=9
(Report on Usage Models for Libraries, Archives and Data Centres)

Please contact:
Annette Balle Sørensen
State and University Library, Denmark
Tel: +45 8946 2154
E-mail: abs@statsbiblioteket.dk

Filip Kruse
State and University Library, Denmark
Tel: +45 8946 2241
E-mail: fkr@statsbiblioteket.dk

Five Steps to Green Desktop Computing

by Howard Noble, Kang Tang, Daniel Curtis and Paul Jeffreys

It is estimated that in the UK, about half of the 1.47 million desktop computers owned by further and higher education institutions are left on all the time. If power management practices are improved annual savings would be in the order of £64 million and 285 million kg of CO_2 equivalents. Tools and techniques are readily available to achieve these reductions. This paper outlines the approach taken by staff at the University of Oxford who have developed a five step approach to 'green' desktop computing.

Most desktop computers can be configured to go into S3 (sleep/standby) and S4 (hibernate) power saving modes automatically. This has a substantial effect on power usage, typically reducing power consumption from ~80 Watts to below 5 Watts. Monitors can also be configured to go into sleep mode to reduce power consumption when not in use.

There are good reasons why computer power saving is often disabled, as it potentially:

- prevents remote access to computers by people and third-party services, eg for backup or from a conference
- interferes with operating services eg network drives
- prevents immediate access when users return
- causes usability issues that require support.

A full discussion of these issues is given by Lisa Hopkins et al on the 'PowerDown' Web site (see Links) and also the British Computer Society (http://www.bcs.org/server.php?show=conWebDoc.28412)..

The easiest, safest and most reliable way to reduce energy consumption, across all types of operating systems and hardware, is to switch off idle computers, but the issues listed above need to be addressed in order for this to be effective. Furthermore, organizing the switching off of computers can be a challenge in itself. Sometimes it is difficult for an automatic process to know when a computer is truly idle, eg it may be running a background service for a researcher overnight; at other times, users may forget to close down applications and switch computers off manually.

Figure 1 shows the number of computers powered in a typical unit at the University of Oxford over several weeks in August 2009. Office computers that are switched on have an IP address which is visible on the network. Address Request Protocol (ARP) queries can be used to gather and plot the number of devices with IP addresses; this list can be filtered to separate desktop computers from virtual machines, printers, switches and other network devices (ARP provides higher accuracy in identifying devices than normal scanning protocols such as Ping). It is apparent from the graph that while some users switch off their computers outside the working day, there are a substantial number of idle computers left powered.

The University of Oxford has developed two enterprise services to encourage and facilitate power management of desktop computers. The first provides monitoring statistics as shown in Figure 1, and makes the information accessible to users. This is an instance of an important class of tool that provides environmental data in a form that encourages behavioural change; users

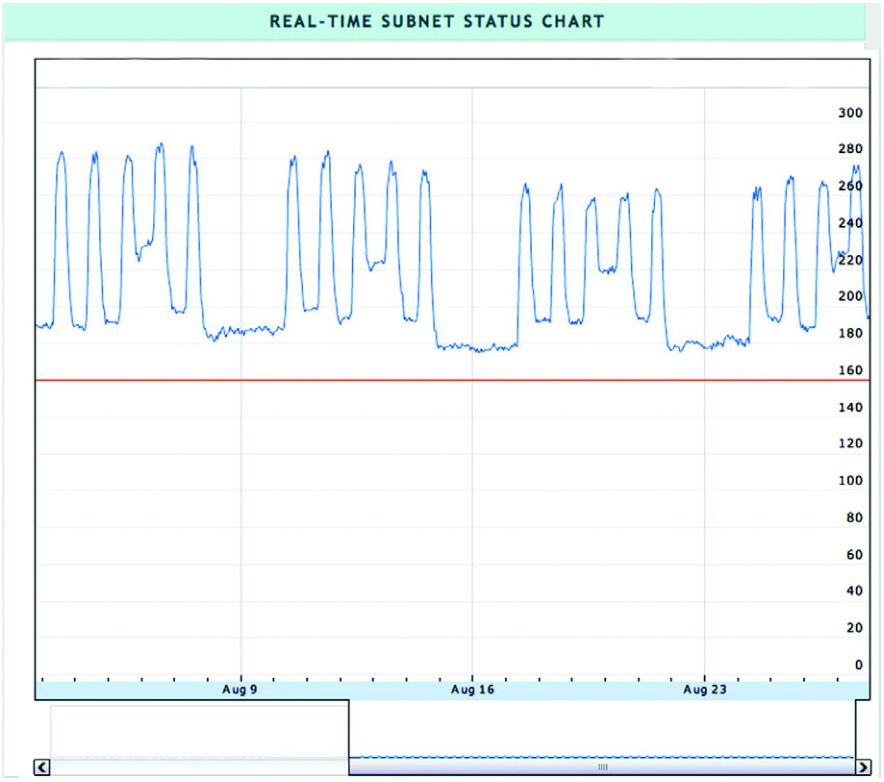


Figure 1: Number of computers powered in a typical unit in the university in August 2009. The x-axis measures weeks and the y-axis the number of desktop machines visible on the network. The five-day working week is apparent, as is some regular activity which occurs overnight between Wednesday and Thursday.

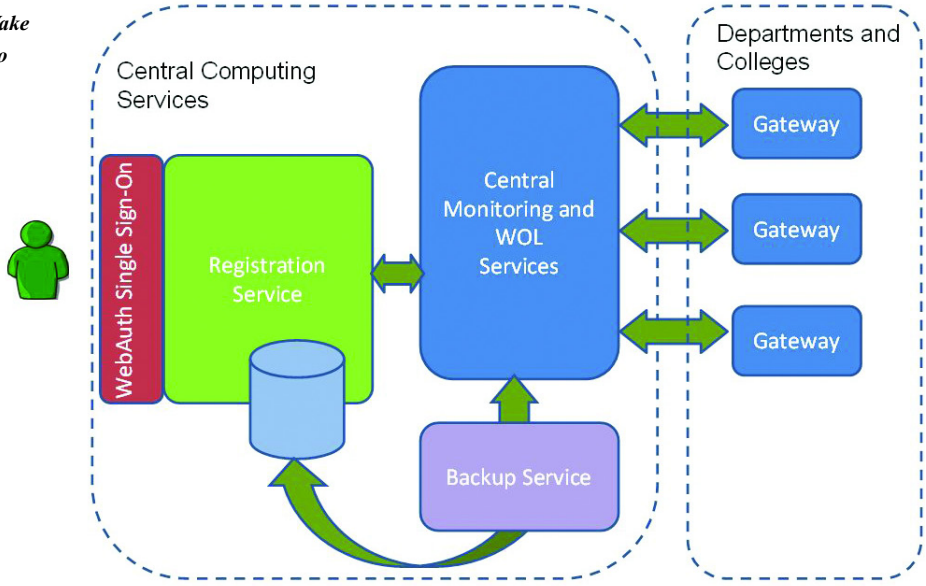


Figure 2: Schematic representation of Wake on LAN infrastructure. The Gateway also hosts the monitoring shown in Figure 1.

will be able to track the direct benefits from improved power management.

The second University enterprise service, Wake on LAN (WOL), enables a computer to be switched on remotely. Assuming the BIOS and network card settings support 'wake on LAN' and the feature is enabled, a computer can be switched on from the S3, S4 and also S5 (off) states just before third-party services require access, eg for a scheduled back-up, for a remote user, or automatically before a user returns to the office, in order to reduce the time for start-up.

Figure 2 is a schematic representation of the WOL infrastructure. Units across the university can 'plug' into both enterprise services by installing a local gateway server. Full instructions are provided at http://www.oucs.ox.ac.uk/greenit/wol.xml. Software is freely available under an open-source licence and further information can be requested from greenit@oucs.ox.ac.uk.

A service to automatically switch off computers matched with WOL would offer the essential tools required for power management.

With the enterprise monitoring and WOL tools in place, the goal is to encourage units across the University to engage with green desktop computing practices. To this end, a five-step process is being promoted:

- Estimate how much electricity your desktop computers consume
- Research what other groups have done
- Implement tools to reduce IT-related electricity consumption
- Communicate effectively to help people to 'do their bit'
- Share your experiences with others.

The practical steps to be taken by an individual unit are straightforward. First, the unit invests around twenty minutes estimating costs and CO_2 emissions for desktop computers and consequent potential savings. Second, it invests 1-4 hours installing a gateway server on a spare computer, VM, a mini-ITX, or a small and very cheap device such as a Sheeva plug (a Sheeva plug is a very low-cost plug computer that can be used as the server).

The unit then waits a few days for monitoring data to become available, and decides on a course of action. If all computers are consistently switched off overnight, there is little to be gained in a power management sense, but monitoring should continue. (Note that WOL could still offer benefits to users in terms of ensuring computers are switched on before the start of work.) If it is found that a significant number of computers are left switched on when idle, the unit should promulgate the monitoring statistics and request that users switch off their desktop computers when not in use.

If after this step a significant number of computers are still left switched on, and for those users who are unable to switch off their computers for the reasons discussed above, then there are two things the unit can do. The first is to offer to provide the WOL service in order that desktop computers can be switched on when required; the second is to offer to install tools such as PowerDown that can automatically turn computers off. Some units are considering investing in a commercial solution that makes automatic power-saving modes (S3: sleep/standby) easier to implement. Such services will be used in conjunction with WOL, but there are licence costs and some effort is required to set up and maintain services.

Following these actions, the unit provides a clear definition of its power management policy, eg switch office computers off when idle, log out from shared computers, disable screen savers and enable 'sleep' on monitors. The final step is to write up and publish the unit's approach in the form of a case study.

More information about the Five Steps to Green Office Computing at the University of Oxford is available from the links below.

Links:
http://www.oucs.ox.ac.uk/greenit/oucs.xml

http://www.oucs.ox.ac.uk/greenit/desktop.xml

PowerDown:
http://www.liv.ac.uk/csd/greenit/powerdown/index.htm

Please contact:
Howard Noble
University of Oxford, UK
Tel: +44 1865 273211
E-mail: howard.noble@oucs.ox.ac.uk

Ranking the Stars with MonetDB

by Annette Kik and Milena Ivanova

The international Sloan Sky Server project has put together a huge scientific database of information on celestial bodies. Querying these vast amounts of data is a great challenge. Using the MonetDB database system, Milena Ivanova and her fellow researchers at CWI in the Netherlands have implemented the first open-source solution. The MonetDB/SkyServer project now provides a valuable experimentation platform for developing new techniques for scientific data management.

The Sloan Digital Sky Survey (SDSS) started in 2000. Its aim is to map a quarter of the sky and to obtain observations of 100 million objects, including galaxies, nebulas and quasars. The SkyServer application gives public access to these data through a Web site. Both researchers and school children can now easily learn more about temperature, mass or the chemical composition of objects like the Whirlpool Galaxy and Owl Nebula. The survey is immense. In 2008, the sky object catalogue alone contained four terabytes (4000 GB) of information. The data are organized in a relational database, containing tables with millions of rows and hundreds of columns.

Figure 1: The famous Whirlpool Galaxy is one of the many objects in the SDSS database. SDSS acts as a well-documented benchmark for scientific database management. (Picture: The Sloan Digital Sky Survey.)

Figure 2: The Sloan Telescope, a 2.5-meter telescope at Apache Point Observatory, did all SDSS imaging and spectroscopy. (Picture: The Sloan Digital Sky Survey.)

These vast amounts of data stress the capabilities of most database management systems (DBMSs), with efficient querying being a particular problem. The architectures of most modern DBMSs are based on an original design that is now three decades old, and was originally intended for business applications (eg bank transactions) having many small and frequently updated records. Scientific databases, however, have large records, with data that stay unchanged once they have been put in the database. They require a different type of database management.

The original SDSS Skyserver, based on Microsoft SQL Server, was the first to bridge the gap between databases and astronomy. It became a successful showcase of scalable database support for scientific applications. There were several other attempts to port the complete SkyServer application to other commercial and open-source systems, but they did not succeed. That is, until the MonetDB solution was implemented.

MonetDB is an innovative open-source database system that has been under development at CWI for over a decade. MonetDB has several advantages, such as efficient data access patterns, flexibility with changing workloads, reduced storage needs, and run-time query optimization. It is a column-store database system. Where other systems organize data in rows, MonetDB reads and stores columns. This approach minimizes the data flow from disk through memory into the CPU caches since only the columns relevant for processing have to be fetched from disk. This can be especially favourable in data analysis applications that need to efficiently retrieve and process large portions of stored data, as in this real-life astronomy application.

The team of CWI researchers - Milena Ivanova, Martin Kersten, Niels Nes, Arjen de Rijke and Romulo Goncalves - intended to test the maturity of this column-store technology by working on a new version of the SkyServer database that is both scalable to growing amounts of data and more efficient to query. The researchers considered MonetDB the best candidate to act as an experimentation platform, since it enables experimentation at all levels of a DBMS architecture.

To make the new SkyServer version, the team members optimized MonetDB for

scientific data. They improved its scalability through partitioning and distribution, and made it more efficient. The first functional prototype of MonetDB/SkyServer went live in 2006. It was a 1% subset of the archive, called 'Personal SkyServer', having a size of 1.5GB. The large vendor-specific database schema and its extensive use of a specific SQL functionality required a significant engineering effort. The initial performance was competitive with the reference platform, MS SQL Server 2005. This demonstrated the benefit of column-stored database techniques for scientific database management. The full-size version went live at the end of 2008 – a major achievement.

The team is currently investigating a number of techniques to increase the system's efficiency, such as parallel load, interleaving of column I/O with query processing, exploitation of commonalities in query batches, and self-organizing indexing schemes like 'crackers'.

Crackers, developed by other members of the MonetDB team, Stratos Idreos, Stefan Manegold, and Martin Kersten, are methods that 'crack' the database into smaller pieces based on querying.

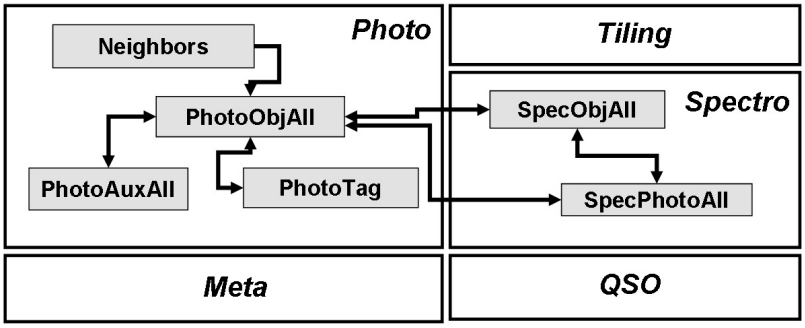


Figure 3: The database schema of SkyServer. The photometric data are stored in the photo-section. The PhotoObjAll table has 454 columns and over 585 million rows. (Picture: The Sloan Digital Sky Survey.)

A cracker structure converges quickly towards a partial index for fast access. Instead of ordering the data in a fixed manner, this is performed dynamically during query processing.

This project is a good example of one of the CWI key research themes: the data explosion. The current explosion in the amount of digital data confronts science and society with new questions. How can concise and relevant information be extracted from this flood of data? There is great need for models, methods and techniques to control it. The MonetDB/SkyServer project contributes to this objective.

The MonetDB SkyServer project was funded by the Dutch Bsik BRICKS programme, NWO Focus and MultimediaN. The MonetDB platform was developed in the Bsik programme MultimediaN.

Links:
SDSS (Sloan Digital Sky Survey/SkyServer): http://cas.sdss.org
MonetDB: http://monetdb.cwi.nl

Please contact:
Milena Ivanova
CWI, The Netherlands
Tel: +31 205924317
E-mail: Milena.Ivanova@cwi.nl

3D Reconstruction by Multimodal Data Fusion

by Dmitry Chetverikov and Zsolt Jankó

At the Geometric Modelling and Computer Vision laboratory of SZTAKI, a new method for 3D reconstruction by fusing multimodal data obtained using a laser scanner, a camera and illumination sources is being developed. The system processes and fuses geometric, pictorial and photometric data using genetic algorithms and efficient methods of computer vision.

Building photorealistic 3D models of real-world objects is a fundamental problem in computer vision and computer graphics. To construct such models, both precise geometry and detailed surface textures are required. Textures allow one to obtain visual effects that are essential for high-quality rendering. Photorealism is enhanced by adding surface roughness in the form of so-called 3D texture, represented by a bump map. Different techniques exist for reconstructing the object surface and building photorealistic 3D models. Although the geometry can be measured by various methods of computer vision, laser scanners are usually used for precise measurements. However, most laser scanners do not provide texture and colour information, or if they do, the data is not accurate enough.

Our primary goal is to create a system that uses only a PC, an affordable laser scanner and a commercial uncalibrated digital camera. The camera can be used freely and independently from the scanner. No other equipment (special illumination, calibrated set-up etc) is used. Neither are any specially trained personnel required to operate the system: after training, a computer user with minimal engineering skills will be able to use it.

The 3D reconstruction system developed in our laboratory receives as input two datasets of diverse origin: a number of partial measurements (3D point sets) of the object's surface made by a hand-held laser scanner, and a collection of good-quality images of the object acquired independently by a digital camera using a number of illumination sources. The partial surface measurements overlap and cover the entire surface of the object; however, their relative orientations are unknown since they are obtained in different, unregistered coordinate systems. A specially designed genetic algorithm automatically pre-aligns the surfaces and esti-

mates their overlap. A precise and robust iterative algorithm developed in our laboratory is then applied to the roughly aligned surfaces to obtain a precise registration. Finally, a complete geometric model is created by triangulating the integrated point set.

The geometric model is precise, but lacks texture and colour information. The latter is provided by the other dataset, the collection of digital images. The task of precise fusion of the geometric and visual data is not trivial, since the pictures are taken freely from different viewpoints and with varying zoom. The data fusion problem is formulated as photo-consistency optimization, which amounts to minimizing a cost function with numerous variables represented by the internal and the external parameters of the camera. Another dedicated genetic algorithm is used to minimize this cost function.

When the image-to-surface registration problem is solved, we still face the problem of seamless blending of multiple textures, that is, images of a surface patch appearing in different views. This problem is solved by a surface-flattening surface algorithm that gives a 2D parameterization of the model. Using a measure of visibility as weight, we blend the textures providing a seamless solution that preserves details. The process of image-to-surface registration and texture merging is illustrated in Figure 1. A measured surface and a textured model are shown in Figure 2.

Finally, photometric data is added to provide a bump map reflecting the surface roughness. We use a photometric stereo technique developed in our lab to refine surface geometry. For this, a number of images are taken from the same viewpoint under varying illumination. The surface is assumed to be diffuse, but the lighting properties are unknown. The initial sparse 3D mesh obtained by the 3D scanner is exploited to calibrate light sources and then to recover surface normals. Figure 3 provides an example of surface refinement by photometric stereo.

All major components of the reconstruction software, including data registration, surface flattening and photometric stereo, have been developed at SZTAKI and are novel. Sample results of 3D reconstruction by our system are available on the Web page of the Geometric Modelling and Computer Vision laboratory (see Links). We are currently involved in a related national project IRIS: 'Integrated Research for Innovative technological Solutions in static and dynamic 3D object and scene reconstruction', funded by the National Office for Research and Technology of Hungary.

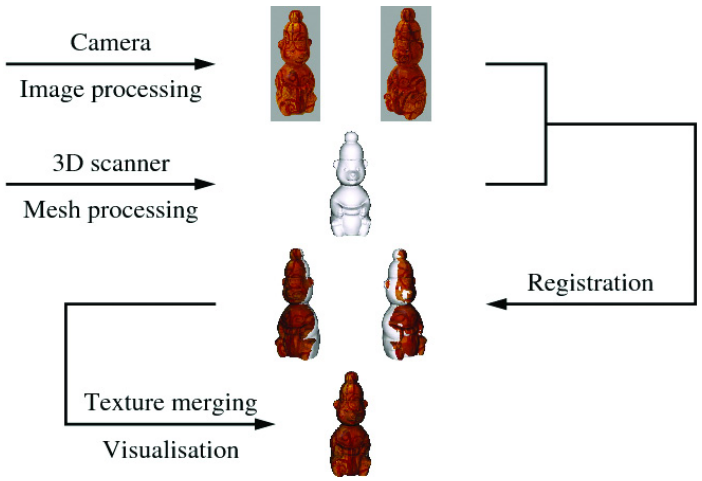


Figure 1: Image-to-surface registration and texture merging.

Figure 2: left: surface measured by laser scanner; right: textured model.

Figure 3: Refining the surface model by photometric stereo.

Links:
http://vision.sztaki.hu/
http://vision.sztaki.hu/iris-nkth/index.php

Please contact:
Dmitry Chetverikov
SZTAKI, Hungary
Tel: +36 1 2796161
E-mail: csetverikov@sztaki.hu

The Virtualization Gate Project

by Edmond Boyer, Benjamin Petit and Bruno Raffin

The Vgate project introduces a new type of immersive environment that allows full-body immersion and interaction with virtual worlds. The project is a joint initiative between computer scientists from research teams in computer vision, parallel computing and computer graphics at the INRIA Grenoble Rhône-Alpes, and the 4D View Solutions company.

A Virtualization gate virtualizes users into graphical objects that capture both their shape and their appearance in real time. These graphical objects can be plugged into any virtual reality application including immersive and interactive applications where users can see and act on virtual worlds. Geographically distant users can also be immersed into a common virtual environment for collaborative applications.

Images from the VGate project.

Traditional immersive solutions developed in the virtual reality community are generally based on advanced display technology such as head-mounted displays (HMDs) and immersive multi-projector environments like Caves. Though they provide an impressive sense of immersion, users tend to be limited both in their interactions with virtual objects, and in their presence (eg appearance) in the 3D world. The main reason for this lies in the perception capabilities of the environment. Interactions usually rely on sensors that provide local information on position or velocity, for instance, but not full-body or appearance information. In contrast, the Vgate environment relies on cameras that provide both geometric and photometric information on users' body shapes. The contribution of Vgate is therefore to associate multi-camera 3D modelling, physical simulation and tracked HMDs for immersion and interaction purposes.

Hardware Setup

Vgate uses several video cameras and 3D modelling tools to build a graphical model of the observed shape in real time (about 20 frames per second). This model is fed into a physical simulation where it becomes a solid object that can act upon other objects. User-centred visualization is provided through a head-mounted display that is tracked with an infrared positioning system (Cyclope). Computations are distributed on a PC cluster to enable real-time execution.

Software Components

Vgate uses the following software:

- Computer Vision: silhouette-based models are computed in real time from the video streams. They are represented with meshes onto which the acquired images are back-projected to produce photorealistic models (http://www.4Dviews.com).
- Simulation: the Simulation Open Framework Architecture (SOFA) runs the physical simulation. It allows objects of very different natures to interact, including rigid bodies, deformable objects and fluids (http://sofa-framework.org).
- Middleware: the Vgate application is developed on top of the FlowVR library, a middleware dedicated to high-performance interactive applications. It enforces modular programming through a hierarchical component model that leverages software engineering issues while enabling efficient execution on parallel architectures (http://flowvr.sourceforge.net/).

Link:
http://vgate.inrialpes.fr/

Please contact:
Edmond Boyer
Grenoble Universities and INRIA
Grenoble, France
E-mail: Edmond.Boyer@inrialpes.fr

Providing Web Accessibility for the Visually Impaired

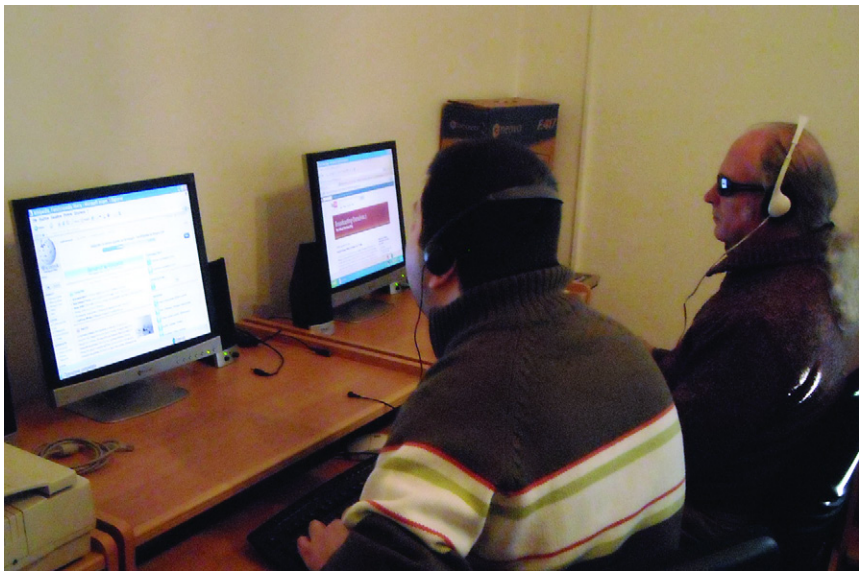
by Barbara Leporini, M.Claudia Buzzi and Marina Buzzi

Web accessibility means ensuring that online content, services or applications can be accessed and used by everyone, including those with special needs. Usability, on the other hand, is a multidimensional concept that depends on the application, the user context and on the goal itself, and its aim is to provide a fully satisfactory user experience. Although closely related, accessibility and usability are frequently addressed as two separate issues. Nevertheless, it is very important to apply them synergistically from the earliest phases of design in order to guarantee satisfactory interaction for users with disabilities.

Several general accessibility and usability guidelines have been proposed in the literature. One of the more authoritative sources is the WAI group (Web Accessibility Initiative) of the World Wide Web Consortium (W3C), which defines accessibility guidelines for Web content, authoring tools and user agent design. The W3C Web Content Accessibility Guidelines (WCAG) are general principles for making Web content more accessible and usable for people with disabilities. However these general criteria need to be broken down with explicit and detailed guidelines that can be assessed simply and rapidly in order to be concretely applied. The context of use as well as the desired goal must be carefully considered (eg the goals of e-commerce are different from those of social networks).

We are working on simplifying interactions with Web content and services for visually impaired people, without developing an alternative Web site. The objective is to offer blind users an interface that is easy to use, simple to understand and quick to navigate. By intervening at the level of the interface source code, a satisfactory interaction can be guaranteed for everyone while maintaining a Web site's original graphic appearance (and thereby avoiding any negative impact on the sighted community). An additional goal is to promote the adoption of W3C standard guidelines and technologies such as WCAG 2.0 and WAI-ARIA (Accessible Rich Internet Applications), and to contribute to their diffusion and application.

Blind people usually explore the Web using a screen reader and voice synthesizer. The screen reader is an assistive technology that interprets the user interface (UI), announcing its content sequentially as it appears in the (X)HTML source code. In addition to the text, the screen reader announces the interface control elements and non-textual components, such as links, images and window objects, that are embedded in the page content. These elements are important for helping a blind user to figure out the page structure, but if the layout is too complex, the actual reading process can require considerable cognitive effort.



Investigating the screen-reader usability of Wikis.

Lack of visual perception prevents the user from having an interface overview at a glance, meaning blind people can spend much time navigating without finding relevant content. Serialization also makes the reading time-consuming and annoying when part of the interface (menu, navigation bar etc) is repeated on every Web site page. As a consequence, blind users often prefer to navigate from link to link with the tab key, or to explore content row by row via arrow keys. This can lead to:

1. Lack of context. When navigating via the Tab key the user can access only small portions of text and may lose the overall page context; thus it may be necessary to reiterate the reading process.
2. Difficulty understanding UI elements. Screen reader commands allow movement via links, tables and buttons; these should be context-independent and self-explanatory.
3. Difficulty working with form control elements. The order in which the user visits form elements (eg via the tab key) should reflect the logical order for the user compiling the form. If possible, the user should be able to jump to a group of homogenous elements.

Our aim is to enhance Web user interaction, in particular with respect to collaborative and cooperative aspects. In the last two years our team has been investigating the screen-reader usability of Wikis (Wikipedia editing page), eCommerce systems (eBay transactions) and Learning Management Systems (Moodle Demo Course). Appropriate design allows users to make additions to UI elements to clarify

the page structure and to navigate more quickly via the keyboard (alternative descriptions for images, summaries for tables etc). Similarly, adding specific tags or attributes such as heading elements or hidden labels can contribute by creating more usable content. Moreover, due to the serialization process, the arrangement of the content (ie the location of UI sections/elements in the source code) is crucial to making it clear and comprehensible.

Our methodology starts by analysing the Web UI (inspection via screen reader) and defining specific design guidelines aimed at removing accessibility and usability barriers. We then develop a prototypal UI conforming to the criteria identified, and test it with blind users.

To build an accessible UI we use the WAI-ARIA suite which, by defining roles, states and properties for UI elements, can greatly enhance interaction for the blind. Adding semantics to (X)HTML objects, ARIA makes dynamic Web content more accessible to differently-abled people since changes in the UI can be captured by the assistive technology and communicated to the user. Furthermore, UI logical sections may be marked as ARIA regions, specifying standard XHTML landmarks (main, navigation, search etc) or defining customized regions. In this way the user is able to get a page overview (the list of UI regions), to move around a specific region and also to jump from one region to the next. Last, if appropriately tagged, a UI element may be silently ignored by the screen reader (eg a table used as layout). In this way the user has more control over the interface and over the amount of text announced by the screen reader, giving them a better interaction experience.

Much positive feedback from users, coupled with the fact that Web accessibility is still generally poor, encourage us to continue in this direction. We have recently started to analyse online social networks like Facebook, to evaluate whether they actually offer the visually impaired an opportunity for active social participation. In future work, we will consider more specific collaborative environments and groupware applications that can be used for distance meetings or classes.

Links:
http://hci.isti.cnr.it/accessibility/
http://usability.iit.cnr.it/
http://www.w3.org/WAI/
http://www.w3.org/TR/WCAG20/

Please contact:
Barbara Leporini
ISTI-CNR, Italy
E-mail: Barbara.Leporini@isti.cnr.it

Claudia Buzzi, Marina Buzzi
IIT-CNR, Italy
E-mail: Claudia.Buzzi@iit.cnr.it,
Marina.Buzzi@iit.cnr.it

ICASE Project: New Challenges in Computer-Based Assessment

by Thibaud Latour and Sandrine Sarre

The ICASE project addresses issues related to computer-based assessment practices, analysed from the perspectives of IT, psychometrics and assessment, and is intended to open up new research strands.

ICASE is a research, development and innovation project undertaken by the Public Research Centre (CRP) Henri Tudor in Luxembourg. ICASE (Intelligent Computer-Assisted Skill Evaluation) is part of a project portfolio called the ESIS (Educational Systems and International Surveys) programme, which supports education policy makers in lifelong learning, teaching and assessment activities by offering products and services. The project started in 2008 and will end in 2011. The initial presentation made at the International Conference of Psychology (ICP 2008) has demonstrated interest in a cross-disciplinary approach. As a strategic project, it aims at setting up a long-term roadmap for the development of computer-based assessment (CBA) followed by the development of assessment tool prototypes and empirical investigations in the field.

CRP Henri Tudor is gaining international expertise and a strong reputation with TAO (Testing Assisté par Ordinateur), a testing platform created in 2002 in cooperation with the University of Luxembourg. TAO is used in the Programme for International Student Assessment (PISA), and in the Programme for the International Assessment for Adult Competencies (PIACC) international surveys for the OECD, which is led by ACER (Australian Council for Educational Research) and ETS (Educational Testing Services). Through this, hundreds of testing items translated into more than 20 languages are managed and supported.

From a general point of view, assessment addresses a wide range of dimensions (political, economic, social, educational, psychological, linguistic and cultural). Each of these endorses characteristics that may vary considerably depending on national, organizational or individual contexts and requirements. The complexity of this area has increased significantly with the integration of new forms of assessment, tools and methodologies that have dramatically changed the landscape of skill assessment.

The participation of the TAO team in many meetings, workshops and seminars in the domain of CBA has led to the identification of five challenging research areas addressed by different work packages (WP). The exploitation of advanced results (WP1) is mainly based on data mining and is used to analyse candidates' logs and extract hidden patterns from chronometric and behavioural data. Behavioural patterns dedicated to cheating detection are also

considered. Security issues (WP2) concerning particularly high-stake and large-scale tests are still complex (eg brain dump, cheating, item protection etc), and so is identity management; only a few solutions have been provided so far. New forms of testing (WP3) encompass a wide range of mature ubiquitous technologies used for checking how to perform formative assessments, collaborative assessments (eg with serious games) as well as situational skill assessments (eg with mobile devices). Assessment of business-related skills (WP4) is challenging because, in a workplace context, assessments mostly lack scientific validity and consume too much time and money for both test takers and producers. Last but not least, management of e-assessment resources (WP5) covers metadata management and the storage, annotation, search, retrieval and exchange of e-testing resources within a Web-based distributed community.

The main expected deliverables include new software, methods, tools and standards that will significantly improve the quality of assessment practices and spread new efficient ways of assessing the new skills of the 21st century (eg collaborative assessment, portfolio assessment etc).

Amongst the successive phases planned for achieving these goals, the prototyping phase is undoubtedly of the utmost importance. It consists of designing and developing services with pragmatic features. These new components will be directly integrated into the TAO platform and evaluated in real-world applications. Scientific publications, roadmaps and technological developments will enable the CRP to define innovative projects in the future.

Topics addressed by ICASE.

The centre plans to enlarge its network of international RDI collaborations by creating and reinforcing synergies at the scientific level with leading players worldwide. ICASE is therefore providing doctoral and post-doctoral opportunities and expects to include a wide variety of stakeholders affected by the future challenges of CBA, whatever their role (psychometricians, educators etc) or organizational setting (testing centres, research institutes etc).

Finally, our future activities will not only permit us to study how assessment practices in teaching and learning evolve towards formative assessments (eg supporting test takers in knowledge construction), but they will also allow us to better understand and integrate CBA into education in general. The ICASE project represents a starting point for the development of new business and research opportunities at the service of the European and international (e)-assessment community.

Link:
http://www.tao.lu

Please contact:
Thibaud Latour
Tel: + 352 42 59 91 327
E-mail: thibaud.latour@tudor.lu

Sandrine Sarre
Tel: + 352 42 59 91 825
E-mail: sandrine.sarre@tudor.lu

Impacts of an ICT Breakdown on the European Economy

by Fabio Bisogni, Simona Cavallini and Cristiano Proietti

What would happen to your economy in the case of ICT breakdowns? Can the economic impacts of ICT shortages be assessed? Are the effects of ICT breakdowns the same in all European countries? The VIS research project attempts to answer these questions by investigating the vulnerability of European industrial systems.

In 2007, the Italian FORMIT Foundation was awarded a grant for the research project 'The Vulnerability of Information Systems and inter-sectorial economic and social impacts – VIS' within the framework of the programme 'Prevention Preparedness and Consequence Management of Terrorism and other related Risks' of the European Council. This project analyses the economies of all the 27 member states and provides results at national and European level.

VIS collects information on the vulnerability of economic sectors, which is provided to policy makers responsible for guaranteeing prevention and/or ensuring preparedness for potentially dangerous events. In this framework, VIS deals specifically with the potential damage that can be caused by an unexpected ICT shortage.

The VIS project has developed a model and an econometric computer-based tool (VIS-S) for this purpose. The model identifies the condition of equilibrium by measuring the weight (or contribution) of ICT in the production activity of the industrial sectors in a given economy in absence of shocks. Technically, the VIS model is a computational general equilibrium model where a supply side (production activities of enterprises in a given economic sector) and a demand side (final consumption activities of individuals and intermediate consumption of enterprises of the given economic sector) define the condition of equilibrium. A VIS-S simulation breaks this equilibrium by introducing into the model a shock caused by an ICT shortage and calculates the consequences at sector level, taking into account interdependencies and cascading effects. In particular, the tool can simulate different intensities of shortage, different breakdown lengths of time and different recovery times, thus computing the sectorial effects on five specific variables:

- labour deviation (expresses the increase or diminution of employment)
- output loss (calculates quantity of production loss)
- production value loss (calculates the monetary value of the production loss)
- price deviation (captures the variation in prices necessary to restore the condition of equilibrium)
- welfare loss (synthesizes the damages according to a more social perspective).

The output of the simulation is a ranking of the sectors most likely to be affected by an unexpected ICT breakdown.

Table 1 gives an example of a simulation that can be generated by VIS-S. Output loss variations are compared in the EU27 area and for four member countries, showing the effects on the top 20 sectors of a hypothetical 10% ICT breakdown (sectors are ranked following the NACE classification provided by EURISTAT. This classification breaks down European economies into specific economic sectors. The digit (1,2,3 and 4) indicates the partition level of the country economy. In this case, the research refers to the NACE 2-digit classification). The simulation considers the effect after one day and a five days

Percentage output deviation, NACE 2-digit
Comparison of percentages (Top 20 sectors in EU27)
Impact after 1 day - 10% ICT breakdowns (recovery 50% loss in 5 days)

NACE code	Economic sector	EU27	Ireland	Italy	Romania	Spain
K72	Computer and related activities	**-10,36%**	**-9,80%**	**-11,01%**	**-11,17%**	**-10,13%**
J67	Activities auxiliary to financial intermediation	**-0,63%**	-0,20%	-1,20%	**-0,68%**	-0,15%
J65	Financial intermediation, except insurance and pension funding	**-0,57%**	-0,24%	-1,25%	**-0,62%**	-0,20%
J66	Insurance and pension funding	-0,50%	-0,18%	-1,08%	-0,56%	-0,12%
I64	Post and telecommunications	-0,48%	-0,21%	**-1,38%**	-0,41%	**-0,23%**
K71	Renting of machinery and equipment	-0,44%	-0,21%	-1,05%	-0,35%	-0,09%
K73	Research and development	-0,43%	-0,17%	**-1,87%**	-0,28%	**-0,29%**
DL30	Manufacture of office machinery and computers	-0,40%	-0,18%	-0,66%	-0,42%	-0,14%
K74	Other business activities	-0,40%	-0,18%	-1,07%	-0,32%	-0,10%
DA16	Manufacture of tobacco products	-0,36%	-0,28%	-0,62%	-0,26%	-0,09%
K70	Real estate activities	-0,36%	-0,21%	-0,96%	-0,25%	-0,07%
G51	Wholesale trade and commission trade	-0,36%	-0,22%	-0,86%	-0,33%	-0,09%
CA11	Extraction of crude petroleum and natural gas	-0,35%	**-0,71%**	-0,76%	-0,23%	-0,08%
I62	Air transport	-0,35%	-0,24%	-1,09%	-0,23%	-0,09%
I63	Supporting and auxiliary transport activities	-0,35%	-0,20%	-1,14%	-0,23%	-0,10%
O90	Sewage and refuse disposal, sanitation and similar activities	-0,33%	**-0,57%**	-0,66%	-0,21%	-0,06%
E41	Collection, purification and distribution of water	-0,33%	-0,24%	-0,77%	-0,22%	-0,10%
DA15	Manufacture of food products and beverages	-0,33%	-0,23%	-0,78%	-0,24%	-0,10%
E40	Electricity, gas, steam and hot water supply	-0,33%	-0,22%	-0,74%	-0,21%	-0,12%
L75	Public administration and defence; compulsory social security	-0,32%	-0,23%	-0,58%	-0,28%	-0,12%

Top 3 sectors in red.
Source: VIS Project Research

Table 1: Output loss deviation in the top 20 sectors In EU27, Ireland, Italy, Romania and Spain.

recovery time in order to re-establish a 50% of the production loss capacity (to better clarify,if the system affected by the ICT shortage loses the amount X of its production capability, the simulation considers 5 days as the time necessary to recover the half of X. The first column provides a ranking of the 20 most affected sectors in Europe, while the other columns show the same sectors for a sample of four European countries. The top three ICT vulnerable sectors per country are highlighted in red.

The output loss variable is of great interest because it provides information about the reaction of selected economies to the ICT shock in terms of production quantity loss. As shown in the table, there is a close similarity between the Italian and Spanish results, especially concerning the top positions. However, the values differ, with the Italian values being significantly higher than the Spanish ones. This implies a major maturity of the Italian economy in terms of ICT use and pervasiveness. Another interesting element highlighted by the table is the similarity of the Romanian system with the data for EU27 in terms of sector ranking. In contrast, the role of ICT in the Irish economy has weights and modalities completely different from the others areas investigated.

The simulation of the effects on output loss is just one of the examples of the VIS-S functionality. The potentially many scenarios (in terms of percent of ICT damages) and the possible different perspectives (obtained, for instance, by analysing the effects through the different variables) can provide strategic information to support policy makers and stakeholders (such as IT managers, business continuity managers, organizations for citizen security etc) in the analysis of decisions and design of interventions to reduce the socio-economic impacts of ICT breakdowns.

Stakeholders such as national civil defence agencies, those involved in prevention and consequence management of security risks, national and international associations for the protection of critical infrastructures, national and international organizations for business impact analysis, and national and European policy makers, can test VIS-S via customized simulations on the dedicated session of the project Web site.

Links:
http://www.formit.it/vis
http://ec.europa.eu/justice_home/funding/cips/funding_cips_en.htm
http://ec.europa.eu/eurostat/

Please contact:
Fabio Bisogni, Simona Cavallini and Cristiano Proietti
FORMIT, Italy
Tel: +39 065165001
E-mail: {f.bisogni, s.cavallini, c.proietti}@formit.org

D4Science World User Meeting

by Donatella Castelli, Marc Taconet and Virginie Viollier

D4Science (DIstributed colLaboratories Infrastructure on Grid ENabled Technology 4 Science) is a two-year project which started in January 2008 and is co-funded by the European Commission's 7th Framework Programme for Research and Technological Development. D4Science follows the path initiated by GÉANT, EGEE, and DILIGENT aimed at creating networking, grid-based, and data-centric e-Infrastructures.

The D4Science e-Infrastructure is designed for demanding science and serves virtual organizations affiliated to the fields of Environmental Monitoring and Fisheries and Aquaculture Resources Management. Communities are led by three international project participants: the Food and Agriculture Organization of the United Nations, the WorldFish Center and the European Space Agency.

The Food and Agriculture Organization of the United Nations (FAO) hosted the first D4Science World User Meeting in Rome, Italy, 25-26 November 2009.

The objectives of the meeting were to:
- Follow up on opportunities and innovative projects implementing advanced e-Infrastructures for multi-disciplinary scientific communities;
- Share experiences, best practices and discuss the most recent advances in e-Infrastructures with emphasis on how to exploit possible synergies;
- Showcase the latest experiences in building advanced applications operating on distributed heterogeneous sources, like dynamic monitoring progress reports, and the most recent results in harmonisation and combination of distributed structured data sources;
- Identify the perspectives of the user communities and discuss the road map for D4Science-II (second phase of the D4Science project) Virtual Research Environments.

The D4Science World User meeting hosted an international audience of more than 80 experts from about 20 countries, plus distinguished guest speakers from the scientific and user communities whose in-depth knowledge of the subject and experience in the field represented an added value to the meeting and to the D4Science communities.

The audience was very diverse:
- Scientists and information managers from the User Communities related to D4Science who expect to make use of Virtual Research Environments in their daily work;
- Professionals from the information technology domain interested in, or actually taking part in development of e-infrastructures for their user communities;
- European Commission representatives and members of large International projects;
- D4Science, External Advisory Board, D4Science-II members;
- Policy and decision makers in the various above-mentioned fields.

The event provided the opportunity for participants to contribute to lively discussions and to share experiences. Representatives from international projects were invited to present and lead talks on how to build synergies. User dedicated sessions aimed at understanding how an e-infrastructure such as D4Science could support their scientific needs. Representatives from D4Science and D4Science-II presented their perspectives for the future and for possible collaboration.

Links:
http://www.d4science.eu/
http://www.d4science.eu/worldusermeeting

Please contact:
Donatella Castelli, ISTI-CNR, Italy
E-mail: donatella.castelli@isti.cnr.it

iPres 02010

Call for Papers

7th International Conference onPreservation of Digital Objects

Vienna, Austria, 19-24 September

The Austrian National Library and the Vienna University of Technology are pleased to host the International Conference on Preservation of Digital Objects (iPRES2010) in Vienna in September 2010. iPRES2010 will be the seventh in the series of annual international conferences that bring together researchers and practitioners from around the world to explore the latest trends, innovations, and practices in preserving our digital heritage.

Digital Preservation and Curation is evolving from a niche activity to an established practice and research field that involves various disciplines and communities. iPRES2010 will re-emphasise that preserving our scientific and cultural digital heritage requires integration of activities and research across institutional and disciplinary boundaries to adequately address the challenges in digital preservation. iPRES2010 will further strengthen the link between digital preservation research and practitioners in memory institutions and scientific data centres.

Submissions

iPRES2010 will adopt a two-track scheme, focussing on research papers reporting on novel, previously unpublished work, as well as case studies and best practice reports. The conference programme will be designed to encourage interaction between these areas, rather than seeing them as separated fields. Furthermore, iPRES2010 will offer a set of tutorials on the Sunday preceding the conference, as well as focussed workshops following the main conference.

Submissions are invited for full and short papers, demos/posters, panels, workshops, and tutorials. All contributions will be reviewed by members of the Programme Committee. A detailed call for papers is available at the iPRES2010 homepage.

Topics (include but not limited to):
- Theoretical, Formal and Conceptual Models of Information and Preservation
- Trusted Repositories: Risk Analysis, Planning, Audit and Certification
- Scalability and Automation
- Metadata Issues for Preservation Processes
- Business Models and Cost Estimation
- Personal Archiving
- Innovation in Digital Preservation: Novel Approaches and Scenarios
- Training and Education
- Domain-specific Challenges: Web, GIS, Primary/Scientific/Sensor Data, Governmental & Medical Records
- Case Studies and Best Practice Reports: Systems, Workflows, Use Cases

Important Dates:
- 18 March 2010:Workshop Submission
- 9 April 2010: Workshop Notification of Acceptance
- 5 May 2010: Paper/Tutorial/Panel Submission
- 18 June 2010:Paper/Tutorial/Panel Notification of Acceptance
- 11 July 2010: Submission of final versions.

Conference Organisation
- General Chairs: Andreas Rauber, VUT, Austria; Max Kaiser, ONB, Austria
- Programme Chairs: Rebecca Guenther, Library of Congress, US; Panos Constantopoulos, Athens University of Economics and Business, Greece; Digital Curation Unit, Greece
- Panel Chair: Heike Neuroth, Göttingen State and University Library, Germany
- Tutorial Chair: Shigeo Sugimoto, University of Tsukuba, Japan
- Workshop Chair: Perry Willett (California Digital Library, US); John Kunze, University of California, US
- Publicity Chairs: Priscilla Caplan, University of Florida, US; Joy Davidson, University of Glasgow, Scotland
- Local Organising Chair: Johann Stockinger, Austrian Computer Society, Austria

More information:
http://www.ifs.tuwien.ac.at/dp/ipres2010

Conference on Trust in the Information Society

León, Spain 10-11 February 2010

Under the auspices of the Spanish EU Presidency the European Commission Information Society & Media Directorate-General organises jointly with INTECO (The National Institute of Communication Technologies), and eSEC-AETIC (Spanish Technology Platform for Security, Trust and Dependability) a Conference on Trust in the Information Society on 10-11 February, in Léon, Spain. The event will focus on the findings of the Advisory Board of Research and Innovation for Security, Privacy and Trustworthiness in the Information Society (RISEPTIS). High level experts from Public Administrations, Industry and Research will speak on the challenges in RTD and policy to ensure an Information Society which will be secure and trustworthy.

Programme
Wednesday, 10 February 2010:
- Opening Session
  Minister or State Secretary of Spanish Government
  Zoran Stančič Dep. Director General Information Society and Media, EC; Local political representative
- RISEPTIS report "Trust in the Information Society"
  (The RISEPTIS Report is available for download at http://www.think-trust.eu/)
  George Metakides (Chairman RISEPTIS, University of Patras)
- Trust in Digital Life - an Industry View
  Participants: Willem Jonker (Philips); Krishna Ksheerabdhi (Gemalto); Luís Fernando Álvarez-Gascón (GMV); Laila Gide (THALES); Jerry Fishenden (Microsoft)
- Trustworthy Networked Service & Computing environments
  Chair: Willie Donnelly, (WIT, THINK-Trust),
  Participants: Michel Riguidel (ENST); Volkmar Lotz (SAP); José Maria Cavanillas (Atos); Simon Foley (University College Cork)
- A European Framework for e-Identity management
  Chair: Kai Rannenberg (Goethe Universität Frankfurt),
  Participants: Reinhard Posch (CIO Austria); Kim Cameron (Microsoft); Speaker from Indra (to be determined); Jan Camenisch (IBM).

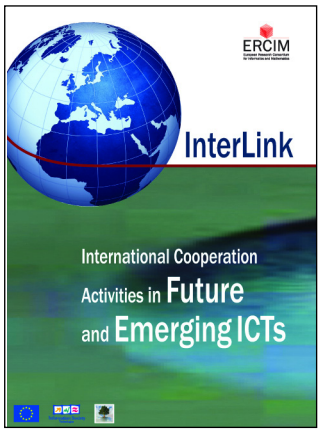
Thursday, 11 February 2010:
- Technology development and the EU Legal framework of Data Protection and Privacy
  Chair: Udo Helmbrechts (ENISA),
  Particpants: Peter Hustinx (EDPS); Mireille Hildebrandt (Vrije Universiteit Brussel); Michelle Chibba (Director Policy, IPC Toronto, CA); Jos Dumortier (KUL)
- International Cooperation on Trust and Security
  Chair: Neeraj Suri (Technische Universität Darmstadt) ,
  Participants: Tai Znati (Director NSF, USA); Malcolm Crompton (IISPartners, AU); Priscila Solis Barreto (University of Brasilia); Jan Eloff (SAP Research CEC Pretoria/SAP Meraka UTD).

More information: https://trustworthyict.inteco.es/

InterLink Research Roadmaps Published

The Coordination Action InterLink (International Cooperation Activities in Future And Emerging ICTs), coordinated by ERCIM and ICS-FORTH, and funded by the Future and Emerging Technologies (FET) Programme of the European Commission, has elaborated research roadmaps for international collaboration in the domains of Software-Intensive Systems and New Computing Paradigms; Ambient Computing and Communication Environments and Intelligent and Cognitive Systems.

To attract and foster trans-disciplinary research excellence, research programmes involving international cooperation need to be defined around new grand challenges and/or key technological issues that have major economic importance or are derived from major societal drivers. Such programmes should explore visionary research themes, demanding breakthroughs in basic research and engineering in key technologies and investigating radically new uses for technology.



InterLink booklet summarising the results of the research roadmaps for international collaboration in the domains of Software-Intensive Systems and New Computing Paradigms; Ambient Computing and Communication Environments and Intelligent and Cognitive Systems.

The main goals of InterLink were to:

- bring together internationally renowned scientists and highlight the latest advances in their areas
- facilitate the exchange of experiences and discussion of the latest progress and findings in challenging research problems relevant to the selected thematic areas
- collectively identify new research topics
- link European research communities to the best research carried out in other developed countries in the respective research fields
- enable European researchers to access knowledge, skills and technology available outside the EU
- provide a critical assessment of the advantages and disadvantages of different kinds of international collaboration
- promote European solutions and knowledge worldwide and influence the way in which science and technology evolve internationally
- build new international strategic alliances, wherever this may be of benefit to European efforts
- influence the design of new research programmes to be funded by the EC, and also by other funding agencies worldwide
- broadly disseminate the findings of InterLink at a European and international level.

InterLink has addressed three thematic areas carefully selected based on the need to address the evolution of the Information Society in the next ten to fifteen years and the challenges this imposes on computing, software engineering, cognition and intelligence:

1. Software intensive systems and new computing paradigms
2. Ambient computing and communication environments
3. Intelligent and cognitive system.

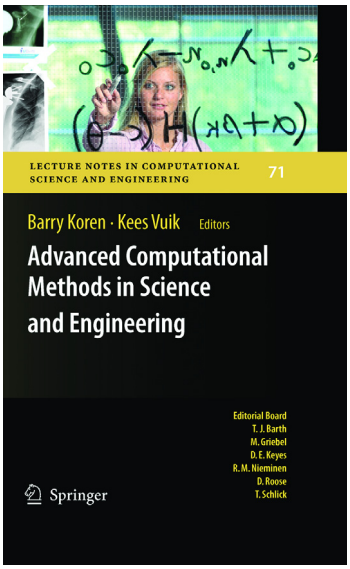
For each thematic area, a Working Groupwas established, and these worked in a coordinated fashion. They had a scientifically and geographically balanced participation, involving experts, mainly from the academic and research sectors, representing various research practices and innovation strategies, from Europe, North America, Australia, Asia and the Far East.

The final versions of the roadmaps are available at the InterLink project Web site. A booklet summarising the results is currently being produced and will also be available on the project Web site.

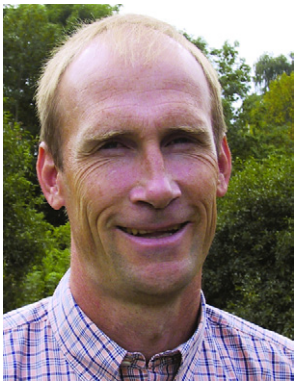
More information:
http://interlink.ics.forth.gr/

New book on Advanced Computational Methods

In November the book Advanced Computational Methods in Science and Engineering was published by Springer, edited by Barry Koren and Kees Vuik. The aim of the book is to show the state-of-the-art in computational science and engineering. It deals with fast and accurate numerical algorithms, model-order reduction, grid computing, immersed-boundary methods, and specific computational methods for simulating a wide variety of challenging problems. Examples of these problems are fluid-structure interaction, turbulent flames, bone-fracture healing, micro-electro-mechanical systems, failure of composite materials, storm surges, and particulate flows. The authors of the chapters are all specialists from the separate disciplines.



Barry Koren and Kees Vuik (eds.), Advanced Computational Methods in Science and Engineering, Lecture Notes in Computational Science and Engineering, Vol. 71, Springer (2009). 498 p., Hardcover. ISBN: 978-3-642-03343-8.



Christer Norström new CEO for SICS

Christer Norström has been appointed new CEO for the Swedish Institute of Computer Science, SICS.

Christer Norström leaves his position as a Professor and vice president at Mälardalen University to take over the leadership of SICS on February 1st, as Staffan Truvé moves on to take the chair of the Board.

Before Mälardalen University, Christer has a background at ABB as well as with small innovation based startup enterprise. He has a thorough experience of leading and developing research organizations, and is, with this mixed background in academy and industry, very well suited to lead an applied research institute like SICS.

"I am very pleased to welcome Christer as CEO and looking forward to working with him in my future role as chairman" says Staffan Truvé. "Christer's broad experience and documented leading talent makes him an ideal new CEO for SICS".

"I am very honoured to be appointed CEO for such a well renowned national research institute as SICS" says Christer Norström. "It is a stimulating challenge to continue the development of SICS to add even better value to Swedish industry".

As the new CEO of SICS Christer Norström will join the ERCIM Board of Directors.

INRIA is Recruiting 45 Researchers

In 2010, the Institute is recruiting 45 scientists for its eight research centres spread across France.

INRIA is opening a competitive selection process to recruit 18 senior research scientists, eight experienced research scientists and 19 Junior research scientists. Positions are offered in its five major research fields:

- Applied mathematics, computation and simulation
- Algorithmics, programming, software and architecture
- Networks, systems and services, distributed computing
- Perception, cognition, interaction
- Computational Sciences for Biology, Medicine and the Environment.

More information:
http://www.inria.fr/actualites/2009/concours-chercheurs.en.html

EIT ICT Labs Wins Prestigious European Race for Excellence in Innovation

Turn Europe into the global leader in ICT innovation - this is the mission for "EIT ICT Labs", the new Knowledge and Innovation Community (KIC) selected in December 2009 in a tough competition by the European Institute of Innovation and Technology (EIT). The EIT is a new independent community body which was set up to address Europe's innovation gap. The aim of EIT is to rapidly emerge as a key driver of EU sustainable growth and competitiveness through the stimulation of world-leading innovation.

By highlighting the "Future Information Society" addressed by EIT ICT Labs, the EIT recognizes the fact that 80 % of new developments in the key economic sectors of Europe are based on ICT. EIT ICT Labs aims at radical transformation of Europe into a knowledge society with an unprecedented proliferation of internet-based services and will establish a new partnership between business and academia based on trust, transparency and mobility of ideas and people. The consortium connects world leading companies, globally renowned research institutes – including the ERCIM members Fraunhofer Gesellschaft, INRIA, SICS and VTT – and top-ranked universities all dedicated to speeding up innovation to address grand challenges of our society. Committed to an efficient open innovation model, EIT ICT Labs will generate faster transformation of ideas and ICT technologies into real products, services and business, boosting Europe's future competitiveness in all sectors of society.

EIT ICT Labs builds on five co-location centres – Berlin, Eindhoven, Helsinki, Paris, and Stockholm – to build a world-class network of innovation hotspots. These main hotspots are complemented with our extensive network of local and international innovation partners. EIT ICT Labs inspires creative students, researchers, and business people to embrace a risk taking and entrepreneurial attitude and help them to identify new business opportunities. EIT ICT Labs catalyzes the creation of strong ventures and help them to grow to become the future world leaders in the ICT arena. EIT ICT Labs will make Europe the preferred place for ICT innovation and attract top talent, R&D units of large companies and investors from all over the world.

"Becoming a KIC is a tremendous opportunity for us to make Europe the global leader in ICT Innovation" says Magnus Madfors, Acting CEO of EIT ICT Labs. "We have the team, the experience, and we are ready to start building a world class innovation ecosystem, turning the potential of the Future Information Society into benefits for the citizens of Europe and the world."

EIT ICT Labs is one of the first three KICs launched by EIT. The two other KICs are: Climate change mitigation and adaptation (Climate-KIC) and Sustainable energy (KIC InnoEnergy)

More information: http://www.eitictlabs.eu/
http://eit.europa.eu/

ERCIM
European Research Consortium for Informatics and Mathematics

ERCIM – the European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development, in information technology and applied mathematics. Its national member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.

ERCIM is the European Host of the World Wide Web Consortium.

AARIT
Austrian Association for Research in IT
c/o Österreichische Computer Gesellschaft
Wollzeile 1-3, A-1010 Wien, Austria
http://www.aarit.at/

CNR
Consiglio Nazionale delle Ricerche, ISTI-CNR
Area della Ricerca CNR di Pisa,
Via G. Moruzzi 1, 56124 Pisa, Italy
http://www.isti.cnr.it/

Crcim
Czech Research Consortium
for Informatics and Mathematics
FI MU, Botanicka 68a, CZ-602 00 Brno, Czech Republic
http://www.utia.cas.cz/CRCIM/home.html

CWI
Centrum Wiskunde & Informatica
Centrum Wiskunde & Informatica
Science Park 123,
NL-1098 XG Amsterdam, The Netherlands
http://www.cwi.nl/

danaim
Danish Research Association for Informatics and Mathematics
c/o Aalborg University,
Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark
http://www.danaim.dk/

Fonds National de la Recherche Luxembourg
Fonds National de la Recherche
6, rue Antoine de Saint-Exupéry, B.P. 1777
L-1017 Luxembourg-Kirchberg
http://www.fnr.lu/

FWO
Egmontstraat 5
B-1000 Brussels, Belgium
http://www.fwo.be/

FNRS
rue d'Egmont 5
B-1000 Brussels, Belgium
http://www.fnrs.be/

FORTH
Foundation for Research and Technology – Hellas
Institute of Computer Science
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
http://www.ics.forth.gr/

Fraunhofer IUK-TECHNOLOGIE
Fraunhofer ICT Group
Friedrichstr. 60
10117 Berlin, Germany
http://www.iuk.fraunhofer.de/

INRIA
Institut National de Recherche en Informatique
et en Automatique
B.P. 105, F-78153 Le Chesnay, France
http://www.inria.fr/

IUA
Irish Universities Association
Cumann Ollscoileanna Éireann
Irish Universities Association
c/o School of Computing, Dublin City University
Glasnevin, Dublin 9, Ireland
http://ercim.computing.dcu.ie/

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and
Electrical Engineering, N 7491 Trondheim, Norway
http://www.ntnu.no/

PEG
PORTUGUESE ERCIM GROUPING
Portuguese ERCIM Grouping
c/o INESC Porto, Campus da FEUP,
Rua Dr. Roberto Frias, nº 378,
4200-465 Porto, Portugal

Universitas Varsoviensis
PLERCIM
Universitas Wratislaviensis
Polish Research Consortium for Informatics and Mathematics
Wydział Matematyki, Informatyki i Mechaniki,
Uniwersytetu Warszawskiego, ul. Banacha 2, 02-097 Warszawa, Poland
http://www.plercim.pl/

Science & Technology Facilities Council
Science and Technology Facilities Council,
Rutherford Appleton Laboratory
Harwell Science and Innovation Campus
Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom
http://www.scitech.ac.uk/

SpaRCIM
Spanish Research Consortium for Informatics and Mathematics
Spanish Research Consortium for Informatics and Mathematics,
D3301, Facultad de Informática, Universidad Politécnica de Madrid,
Campus de Montegancedo s/n,
28660 Boadilla del Monte, Madrid, Spain,
http://www.sparcim.es/

Swedish Institute of Computer Science
SICS
Swedish Institute of Computer Science
Box 1263,
SE-164 29 Kista, Sweden
http://www.sics.se/

SARIT
Swiss Association for Research in Information Technology
c/o Professor Daniel Thalmann, EPFL-VRlab,
CH-1015 Lausanne, Switzerland
http://www.sarit.ch/

MTA SZTAKI
Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutató Intézet
P.O. Box 63, H-1518 Budapest, Hungary
http://www.sztaki.hu/

VTT
Technical Research Centre of Finland
PO Box 1000
FIN-02044 VTT, Finland
http://www.vtt.fi/

Order Form

If you wish to subscribe to ERCIM News ***free of charge*** *or if you know of a colleague who would like to receive regular copies of ERCIM News, please fill in this form and we will add you/them to the mailing list.*

Send, fax or email this form to:

ERCIM NEWS
2004 route des Lucioles
BP 93
F-06902 Sophia Antipolis Cedex
Fax: +33 4 9238 5011
E-mail: office@ercim.eu

Data from this form will be held on a computer database. By giving your email address, you allow ERCIM to send you email

I wish to subscribe to the

☐ ***printed edition*** ☐ ***online edition (email required)***

Name:

Organisation/Company:

Address:

Postal Code:

City:

Country

E-mail:

You can also subscribe to ERCIM News and order back copies by filling out the form at the ERCIM Web site at http://ercim-news.ercim.eu/