



European quarterly preservation digest

The Second Digital Preservation Challenge deadline extended to 31 July 2008	2
DigitalPreservationEurope (DPE) Research and Industrial Exchange (DPEX) call for applications	3
DRAMBORA	4
SHAMAN (Sustaining Access through Multivalent Heritage ArchiviNg).....	5
CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)	7
Planets (Preservation and Access through Networked Services)	9
LiWA – Living Web Archives.....	13

DigitalPreservationEurope (DPE) Research and Industrial Exchange (DPEX) call for applications

DigitalPreservationEurope are pleased to announce that the first DPE Research and Industrial Exchange has been successfully completed. After careful consideration of all applications received the independent review committee awarded the first DPE Exchange to a young researcher working in digital preservation to visit the Vienna University of Technology, Austria, in October 2007. A report detailing the activities and results of this exchange will shortly be made available on the DPE website.

Exchanges should typically last for four weeks and the DPEX support of up to 3500 Euros per exchange can be used to partially meet the costs of accommodation, subsistence, and travel. Unfortunately, DPEX support cannot be used to meet salary costs and exchanges must involve participants from and institutions located in EU Member States.

The deadline for applications is the 1st of each month. The application process can be completed online and full programme details (including terms and conditions) are available at <http://www.digitalpreservationeurope.eu/exchange>

Further information can be found on the DPE webpage
<http://www.digitalpreservationeurope.eu/exchange>

The DPEX team is happy to provide support and guidance to anyone considering applying for an exchange, questions or comments should be sent to:
DPEX@digitalpreservationeurope.eu

DRAMBORA

Subject-based and institutional digital repositories are increasingly being hailed as the preferred means for safeguarding the future accessibility of digital information. Repositories will not only store digital materials, but also act to limit the threats that are posed to their authenticity, integrity and understandability over time. Consequently, they are being increasingly relied upon to facilitate curation for a wide range of digital content, and there is considerable interest in identifying where successes and shortcomings exist.

The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) is a methodology for undertaking repository self assessment which adopts a bottom-up approach that takes risk and risk management as its principle means for determining repository success and charting improvement.

From January 2008 DRAMBORA will be available as a fully online tool, offering an intuitive form based interface, peer-comparison features, sophisticated reporting mechanisms and maturity tracking. Both DPE and the UK's Digital Curation Centre will continue to develop DRAMBORA to reflect closely the expectations of those working within the repository community that require mechanisms to validate their efforts, assist their repository's development, and prepare them for external assessments.

For further information please visit
www.repositoryaudit.eu

SHAMAN (Sustaining Access through Multivalent Heritage ArchiviNg)

The first round of the FP7-ICT programme has funded three projects addressing digital preservation:

Name	Focus	Consortium
SHAMAN	Data Grid, Federated Digital Libraries, Persistent Data Archives and Multivalent Architecture. Test-beds: Documents in Memory Institutions and Governmental Collections, Objects in Industrial Design and Engineering, eScience	Universities and research organizations in EU and US, industry and government
LiWA	Web archiving: fidelity, coherence and interpretability, transforming pure snapshot into living web archive	Universities and research, new media archiving
PROTAGE	Explore software agent technologies to automate preservation processes (self-preserving objects)	Archives and universities, research, ICT

SHAMAN (Sustaining Access through Multivalent Heritage ArchiviNg) is Large-scale Integrating Project¹. The aim of SHAMAN is to develop the framework for the next generation of long term (more than one century) digital preservation systems and tools. It includes the definition of a SHAMAN theory of preservation integrating the analysis, ingestion, management, access to and reuse of information objects across distributed repositories. The data preservation capabilities offered will secure the authenticity and integrity of data objects through time. It has 18 academic and industrial partners from 9 European countries, and will run for four years. The project budget is €13.3m, of which about €8.4m is being contributed by the European Union.

The development work will be structured around four core components whose objectives can be described as follows:

- to establish an open distributed resource management infrastructure framework enabling grid-based resource integration (fig. 1), reflecting, refining and extending the OAIS model and taking advantage of the latest state of the art in virtualisation and distribution technologies from the fields of GRID computing, Federated Digital Libraries, and Persistent Archives;
- to develop and integrate technologies to support contextual and multivalent archival and preservation processes which are adapted and significantly extended from the fields of content and document Management and Information Systems;

¹ SHAMAN website will be available soon. More details on the project can be found at ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/shaman_en.pdf

- to develop and integrate technologies to support semantic constraint-based collection management to target one of the key challenges in automating one class of digital preservation core functions;
- to support the managing of future requirements by securing interoperability with future environments and maintaining essential properties of the preserved content.
- Three prototypical applications will support trialling and validation in the following domains: 1) scientific publishing in libraries and documents in governmental (parliamentary) archives, 2) digital objects used (i.e. CAD) in industrial design and engineering and 3) data resources used in e-Science applications.

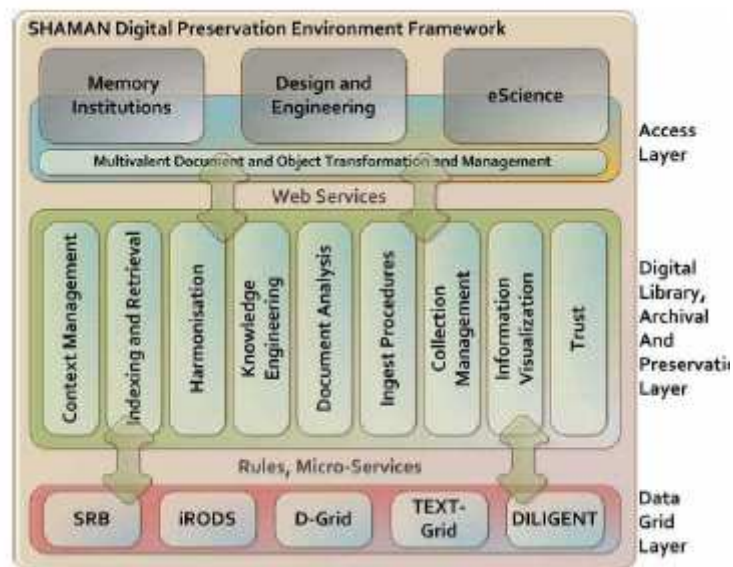


Figure 1. SHAMAN Conceptual Architecture & Grid Utilization

SHAMAN's dissemination and exploitation plans aiming at actively foster outreach and take-up of results will be tailored according to the specific needs of the scientific / academic world and of industry users. SHAMAN's work will be naturally coordinated with other digital preservation European projects (CASPAR, DPE, PLANETS)² as well as initiatives at national (DGrid, Germany)³ and international level (NDIIPP/NSF, US)⁴.

² <http://www.wepreserve.eu/>

³ http://gks06.fzk.de/slides/D-Grid_Gentzsch.pdf

⁴ <http://www.loc.gov/today/pr/2008/08-004.html>

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)

CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) Project has made significant advances in its research and development, releasing a number of key documents, now available on the CASPAR website at <http://www.casparpreserves.eu/publications/deliverables>).

These documents include the **CASPAR Conceptual Model**, which builds on the main concepts of the OAIS Reference Model, refining them to provide the basis for an implementation of the key preservation components to answer fundamental questions. These questions include

- how can one preserve any digital object irrespective of whether it is only to be rendered, for example a simple document or image, or processed and analysed, for example a complex piece of scientific data?
- how can the artefacts produced as part of the preservation process, such as Representation Information, Fixity etc, themselves be preserved?
- how can one have a common support infrastructure for preservation?
- how can such a preservation infrastructure itself be preservable?
- how can one facilitate the sharing of the effort and costs needed in the long term preservation of all types of digitally encoded information.

Another milestone for the project was the publication of the **CASPAR Architecture**. The architecture systematically presents the basic components implementing the CASPAR Conceptual Model. It is a loosely coupled, highly asynchronous architecture which allows a variety of ways of deployment, with no single point of failure. An interface is presented for each component and the relevant technologies for its implementation are discussed, with a particular focus on the preservability of the components.

Significant progress has also been made in the aim of putting knowledge at the heart of preservation. This research has focused on defining the key functionality of a component for identifying “gaps” in the Representation Information network which need to be filled in to make a digital object fully understandable by the underlying Designated Community. The result is the **Gap Manager**, based on a Semantic Web technologies, a key component of the **CASPAR Knowledge Manager**,

The ontology-based preservable **Digital Rights Management** system will be released. This is integrated with the CIDOC CRM core ontology, which has been extended with digital rights concepts. This work aims at using knowledge modelling techniques to formalise the right management problem in the preservation domain.

Also, the conceptual modelling of managing **Authenticity** will be made available. The model is novel in that it does not propose yet another metadata schema for

authenticity metadata, but aims at capturing the dynamic aspects of the process of defining and assessing authenticity of digital objects.

Finally a generalised **Information Discovery service** for managing descriptive information associated with the digital objects to be preserved will be prototyped.

In addition, CASPAR is collaborating with diverse EU funded initiatives in digital preservation. One of the most recent highlights is the WePreserve platform highlighting activities of CASPAR, DPE and PLANETS. This is designed to provide a common entry point to digital preservation projects and provide common services. CASPAR is also taking part in a stand at CEBIT 2008 in Hannover, Germany on the 4th-9th of March jointly with these projects as well as Nestor. We will showcase the latest digital preservation research and technology developments and their relevance to industry.

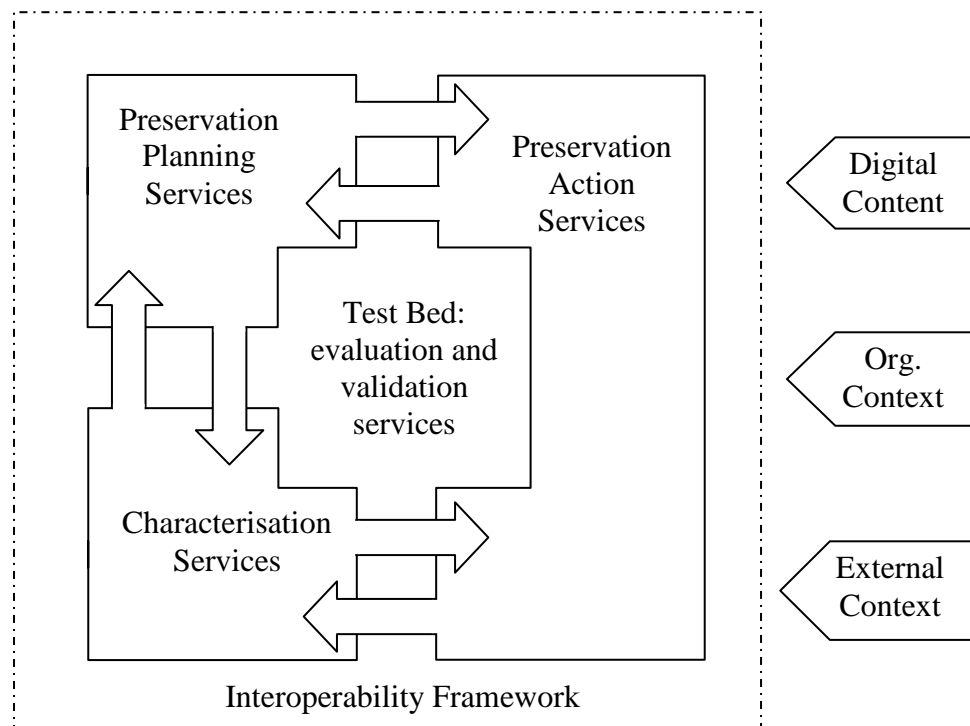
The **CASPAR Preservation User Community** continues to grow and incorporate members from around the world. The 181 members are kept up to date on project initiatives and populate the project website at www.casparpreserves.eu. Community members will also have access to the Training Programme which is being finalised and will be implemented shortly! Visit the CASPAR website for more project developments!

info@casparpreserves.eu

Planets (Preservation and Access through Networked Services)

Planets is a four-year pan-European project funded by the European Commission under its Framework Programme 6. The project was set up in June 2006 to help organisations in Europe address challenges with preserving digital information in Europe. From mid-2010, Planets will provide a common framework that will underpin digital preservation activity, plus downloadable software to help libraries, archives and other institutions to preserve their digital content for the long-term. The project brings together 16 national libraries, archives, research institutions and leading technology companies in Europe.

Planets (Preservation and Access through Networked Services) has reached the half-way point in the four-year project and is making good progress towards reaching its end-goals. The team has now made first and even second releases of the core Planets software components.



Preservation Planning

In April 2008, Planets presented its preservation planning tool, Plato, for the first time. Plato will help users of Planets to systematically plan and identify strategies for preserving digital content in their organisations. The tool is now being updated to

incorporate new functionality that will help define organisations' needs at the start of the planning process, and confirm a preferred plan at the end of the process.

Planets is also developing a model which will assist organisations in creating preservation policies and provide them with a framework to embed preservation planning into day-to-day activity. This work is leading to clearer definitions of what a preservation plan is, what it includes and to arrive at a shared understanding.

Understanding Digital Objects

The process of preserving digital objects requires a language which makes it possible to describe those objects' essential characteristics. Planets has developed two languages to do this. The Extensible Characteristics Description Language (XCDL) makes it possible to describe characteristics of digital objects. Typical characteristics include colour depth of an image, number of images in a text document, metadata in a file, or the font type of text. The Extensible Characteristics Extraction Language (XCEL) makes it possible to define how these characteristics can be extracted from the object. Using a software application called an interpreter, Planets can automatically execute the XCEL definitions to extract XCDL characteristics from a file.

Planets has now extended its characterisation work by developing a Comparator. This will allow the characteristics of two files to be automatically compared and so make it possible to identify the changes that have occurred during a migration process. A design is in place for a second tool which will allow organisations to profile their collections automatically.

Preservation Actions

Specific actions must be performed on digital objects if they are to continue to be accessible. Typically these actions fall into two categories. Migration involves taking the digital object and converting it into a new format so that it can still be read. Emulation involves providing a tool which replicates the operating environment, in which the software was created, so that users can continue to interact with digital content.

To support these actions, Planets has completed desk research and surveys with prospective users in Denmark and the United Kingdom to identify the file formats most commonly found in digital archives. This has led to the creation of an inventory containing 133 file formats. The project has also surveyed tools that exist to characterise and preserve each file format and conducted a gap analysis so that it can identify where tools do not exist or are insufficient. A blueprint has been defined for those that will need to be created, including emulation tools. Planets has been working to integrate second versions of emulation software Dioscuri and Universal Virtual Computer.

Work has been carried out with organisations such as the British Library, National Library of the Netherlands, IBM, National Archives of the Netherlands and University of Freiburg to consider the impact of such emulation tools. The work has indicated

that emulation can be a viable approach to preserving digital objects although further investigation is required.

Learning What Works

The Testbed component of the project will allow users of Planets to identify through experiment the most appropriate tools and strategies for preserving particular content based on its specific characteristics. The Planets Testbed was demonstrated for the first time in April 2008. Subsequently, plans have been established for building the Testbed corpora. These groups of reference documents will act as benchmark content, making it possible to ascertain how far Testbed results are reliable and can be reproduced against well-known sets of digital content. Planets aims to open up the Testbed to external researchers and institutions in Spring 2009.

Planets Framework

The Planets Interoperability Framework provides shared functions and integrates the Planets tools and services into an easily managed preservation system. It allows plug-in of third-party tools and services. The first release of the Interoperability Framework has been available to Planets partners since September 2007. A second release is expected in Summer 2008. The project is also working on guidelines for 'wrapping' preservation tools for integration into Planets.

Sharing the Results

Communication, outreach and training are important components of the project's work. Planets has presented work at various conferences and in scientific papers and articles. Copies of many of these, and Planets newsletters, are available from the Planets website.

The project's work on Characterisation was presented at the 'What to Preserve: Significant Properties of Digital Objects' workshop in London in April 2008, attracting 140 delegates. A joint stand with CASPAR and DPE and Nestor at CeBIT – the world's biggest ICT exhibition - in Hannover, Germany attracted 1,000 visitors including Bern Neumann, the German Minister of State to the Federal Chancellor and Federal Government Commissioner for Culture.

Planets held its first tutorial 'Planning the Future with Planets: a preservation planning tutorial' in Vienna in April 2008, which incorporated a demonstration of the Planets Testbed and hands-on experience of Plato. Following the success of this event, 'Digital Preservation: Principles, Examples and Planets' organised jointly with the Digital Preservation Coalition, will take place on 29 July at the British Library Conference Centre. For further details see <http://www.planets-project.eu/events>.

Planets will also take part in WePreserve's Training in Prague on 13-17 October and the WePreserve Annual Conference in Nice on 29&30 September. Another date for the diary is the Fifth International Conference for Preservation of Digital Objects, iPRES 2008, the world's longest-running conference on digital preservation. IPRES 2008 will be held at the British Library Conference Centre in London on 29-30 September 2008. For more information, see www.bl.uk/ipres2008.



For more information about Planets, publications, newsletters, training and outreach events , and to sign up to receive news about Planets, please visit: www.planets-project.eu. Planets may be contacted at info@planets-project.eu .

LiWA – Living Web Archives

The Web today plays a crucial role in our information society: it provides information and services for seemingly all domains, it reflects all types of events, opinions, and developments within society, science, politics, environment, business, etc. Due to the central role the World Wide Web plays in today's life, its continuous growth, and its change rate, adequate Web archiving has become a cultural necessity in preserving knowledge. Consequently a strongly grown interest in Web archiving library and archival organizations as well as emerging industrial services can be observed.

However, web preservation is a very challenging task. In addition to the “usual” challenges of digital preservation (media decay, technological obsolescence, authenticity and integrity issues, etc.), web preservation has its own unique difficulties:

- distribution and temporal properties of online content, with unpredictable aspects such as transient unavailability,
- rapidly evolving publishing and encoding technologies, which challenge the ability to capture web content in an authentic and meaningful way that guarantees long-term preservation and interpretability,
- the huge number of actors (organizations and individuals) contributing to the web, and the wide variety of needs that web content preservation will have to serve.

A first generation of Web archiving technology has been built by pioneers in the domain like the Royal Library of Sweden and the Internet Archive based on existing search technology. It is now time to develop the next generation of Web archiving technology, which is able to create high-quality Web archives overcoming the limitations of the previous generation. The aim of the European funded project LiWA (IST FP7 216267) is create innovative methods and services for Web content capture, preservation, analysis and enrichment.

The LiWA project, started in February 2008 and led by the L3S Research Center, brings together a consortium of highly qualified researchers (L3S, Max Planck Society, Hungary Academy of Science), industrial users (European Archive, Hanzo Archives), and archiving organizations (Sound and Vision, National Library of the Czech Republic, Moravian Library). It is the intention of the project partners to turn Web archives from pure Web page storages into “living Web archives” within the next three years. Such living archives will be capable of handling a variety of content types and dealing with evolution as well as improving long-term content usability. For creating Living Web Archives, the LiWA project will address R&D challenges in the three areas: Archive Fidelity, Archive coherence and Archive interpretability:

- **Archive Fidelity:** development of effective approaches and methods for capturing all types of Web content including the Hidden and Social Web content,

for detecting capturing traps as well as for filtering out Web spam and other types of noise in the Web capturing process.

- **Archive Coherence:** development of methods for dealing with issues of temporal Web archive construction, for identifying, analysing and repairing temporal gaps as well as methods for enabling consistent Web archive federation for fostering synergies between Web archiving stakeholders;
- **Archive Interpretability:** development of methods for ensuring the accessibility, and long-term usability of Web archives especially taking into account evolution in terminology and conceptualization of a domain;

The results of the project will be demonstrated within two application scenarios namely “Streaming Archive” and “Social Web Archive”. The Streaming Archive application will showcase the building of an audio-visual Web archive and how audio and video broadcast related web information can be preserved. The Social Web application will demonstrate how web archives can capture the dynamics and the different types of user interaction of the social web.

The main expected outcome of the LiWA project is a set of innovative methods and services for capturing, preserving, analysing and enriching web content, which implement the vision of a Living Web Archive as described above. As a further outcome of the project, these services will lead to an integrated reference web archiving solution, based on LiWA results and on open source developments and operated by the project partner European Archive as well in as in the industrial-grade commercial solution developed by Hanzo Archives Ltd.

More information about the project can be found on the LiWA homepage: <http://www.liwa-project.eu/>