

At the Event

Progress towards Addressing Digital Preservation Challenges

Planets, CASPAR and DPE are projects funded by the European Union to address digital preservation challenges under its Framework Programme 6 (FP6). [Helen Hockx-Yu](#) reports on the 2nd Planets, CASPAR and DPE annual conference, held on 5-6 September 2007 in Lisbon, Portugal.

Main Contents	Section Menu	Email Ariadne	Search Ariadne
-------------------------------	------------------------------	-------------------------------	--------------------------------

Introduction

Digital preservation has become an area of strategic importance for the European Union in recent years. This has been reflected in the investment of €17 million in co-funding three major digital preservation projects under call 5 of its Framework Programme 6 in September 2005. Planets (Preservation and Long-term Access through NETworked Services) [1], CASPAR (Artistic and Scientific knowledge for Preservation, Access and Retrieval) [2] and DPE (DigitalPreservationEurope) [3] are all co-ordinated by British organisations: Planets by the British Library, CASPAR by the Science and Technology Facilities Council (formerly CCLRC) and DPE by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow. This is an indication of the UK's leading position in digital preservation research and development. Planets and CASPAR are both termed 'integrated projects' with new knowledge as their primary deliverable. DPE is a 'co-ordination action' project, funded to foster collaboration and synergies between existing national initiatives and to improve co-ordination, co-operation and consistency in current digital preservation activities. All three projects commenced in the first half of 2006.

The purpose of the 2nd joint annual conference was to report the projects' progress to date but also to place the latter in the context of the wider international digital preservation and curation landscape. Furthermore, it was intended to provide a forum for networking and bridge-building, with collaboration as the key objective. The conference was presented under 'wePreserve', the umbrella for PLANETS, CASPAR

and DPE's synergistic activities. A Web site with the same name has also been set up to deliver a collaborative Web platform shared by the projects [4].

Day One

Setting the Scene

Seamus Ross, Director of HATII, opened the conference on behalf of DPE. HATII is not only the lead partner in DPE but also a partner in both Planets and CASPAR. Seamus welcomed the participants and thanked the European Commission for supporting the conference.

The opening presentation was given by **Carlos Oliveira**, Deputy Head of the Cultural Heritage and Technology Enhanced Learning Unit, Directorate General Information Society and Media, European Commission (EC). Carlos emphasised the importance of digital preservation in knowledge economy and explained that the purpose of public investment was to support emergence of a coherent policy framework covering the organisational, economic, legal and technological aspects of digital preservation. He then gave an overview of the EC's activities in digital preservation, including initiatives at policy level and various strands of work funded under the Framework Programmes. Carlos also presented an action plan for digitisation, online access and digital preservation which included ambitious goals to develop national strategies for long-term preservation and deposit with quantitative and qualitative targets, standards for digital preservation and a legislative framework supporting digital preservation. There is an increased level of funding in the new Framework Programme 7, which is intended to fund areas which complement the current portfolio of projects and explore possibilities offered by new ICT (Information Communications Technology) to consider new approaches to digital preservation. Carlos concluded the presentation by flagging up some areas of challenge for digital preservation, such as analysis, identification and spread of good practices, self-assessment and certification, models for long-term sustainability and expansion beyond the 'knowing community'.

Herbert van de Sompel of the US Los Alamos National Laboratory (LANL) presented the aDORe Project, supported by the Library of Congress as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP). aDORe is a standards-based, repository federation architecture which has been implemented at the LANL for local storage of digital assets. The flexible, highly modular architecture facilitates the presentation of autonomous distributed repositories as a single logical repository. aDORE is not a digital preservation solution itself but the interoperability it provides is an enabler and eases the digital preservation tasks across distributed, heterogeneous repositories.

Ross Harvey of the Charles Sturt University, Australia, who is currently a visiting research fellow at the Digital Curation Centre (DCC) in the UK, presented an international overview of digital preservation activities to analyse trends and directions. Ross started with an exploration of the definition of 'digital preservation' and related terms. He introduced Nancy McGovern's definition for digital preservation as a three-legged stool: the organisational leg (the 'what') and the technological leg ('the how') need to be backed by the resources leg (the 'how much')

and co-ordinated to develop compliant and feasible digital preservation strategies [5]. He pointed out the strong research focus in European initiatives, as well as the theme of dissemination and collaboration. Skills development and links with the industry have also been seen in a number of major projects. The US initiatives cover a wide range of activities but much is happening in the areas of electronic records and digital repositories. Australia and New Zealand were given as examples to illustrate the development in other countries, where practice varies considerably. Ross concluded the presentation by summarising the trends in digital preservation in 2007. There is an ongoing strong emphasis on 'community', dissemination, testing and evaluation, and toolkit development. The new emphases are standards, public policy and strategy development, skills identification and development, and links with the ICT industry.

Helen Hockx-Yu of the British Library gave a general introduction to the Planets project. She presented the aims and objectives of the project and outlined Planets' approach to digital preservation. Involving 16 national libraries and archives, educational institutions and technology companies, Planets is developing practical digital preservation solutions to meet the needs of libraries and archives. Its work on automation of digital preservation processes should also provide useful answers for the wider digital preservation community. Helen reported on Planets' progress to date and referred to a number of prototype tools and services which were going to be presented at the conference by other Planets colleagues. She also highlighted the key deliverables by November 2008, explicitly demonstrating what could be expected of Planets in the near future.

David Giaretta of the Science and Technology Facilities Council (STFC) introduced the CASPAR Project, which aims to develop techniques and services to meet the challenge of understanding and using digitally encoded information in the future when software, systems, and everyday knowledge will have changed. The CASPAR consortium includes data holders, research universities, national organisations and a number of small and medium enterprises (SMEs) and industrial partners. CASPAR focuses on scientific, cultural and artistic data which are complex and need to be rendered as well as processed. David argued that the key to address the challenge is Representation Information, which is used to convert the bit sequences within a digital object into more meaningful information. One of the objectives for CASPAR is to design the infrastructure for capturing and storing various types of representation information, taking into account the knowledge base of the designated community.

Collaboration and Co-ordination

Echoing Ross Harvery's observation on the strong emphasis on collaboration and co-ordination, **Luigi Briguglio** of Engineering Ingegneria Informatica presented the Alliance for Permanent Access, a membership initiative formed by major stakeholders from the world of science, including national and pan-European research organisations, research support organisations such as national libraries and publishers, and funding organisations for research. The primary goal for the Alliance is to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information. Luigi explained in detail the proposed mechanisms to achieve this goal, including collaboration within and between communities, development of general policies/practices and sustainable business models for preservation, as well as a research and development programme focusing

on prototyping and testing of new knowledge. The Alliance will work closely with national governments and the European Union (EU) to strengthen strategies, policies and their implementation.

Maurizio Lunghi of Fondazione Rinascimento Digitale presented DPE's work in benchmarking competence centres. 'Competence centre' is the concept used by the DPE to review the current international landscape with regard to the availability and provision of digital curation and preservation expertise in the EU and beyond. Maurizio explained in detail the '7C's' benchmarking model DPE developed to assess competence centre sources and models, including: Capacity, Context, Credibility, Commitment, Certification, Competition and Communication. Based on the results of its assessment using the '7C's' benchmarking model, DPE recommends a federated approach to the provision of support and guidance for digital preservation. Maurizio concluded the presentation by outlining the benefits of establishing federated competence centres which should help nurture strong community relationships from a range of disparate stakeholders.

Planets, CASPAR and DPE all have a mandate to disseminate their findings and to provide dedicated training. In addition there are other organisations such as the DCC which provide training in digital preservation. The increasing number of training events being offered can potentially cause confusion amongst participants. A natural course of action is to join forces and collaborate. **Joy Davidson** of the DCC gave a presentation entitled 'Collaboration in Training Provision', in which she outlined the importance and benefits of collaboration, both for training providers and participants. Joy explained that collaboration can take many formats, ranging from simply sharing information to more formal collaboration and the delivery of joint events. She then reported on the effort of DCC, Planets, CASPAR and DPE in establishing an international curation and preservation training roadmap to ensure a coherent approach to training development and provision. A workshop took place in March 2007, involving key European stakeholders in digital preservation training. Issues discussed at the workshop included target audiences, training options, methods and themes, promotion and branding and business models. There was a consensus among participants that collaboration is extremely important, but that there is a serious lack of dedicated time available to participants. Joy argued that any follow-up meetings and activities must be carefully scheduled to maximise the potential benefits while minimising the time required to participate and plan joint events. She ended the presentation by announcing that the first DPE/Planets/nestor collaborative training event would be held in Vilnius, Lithuania on 1-5 October 2007.

Day Two

Having set the scene and covered the wider context, the conference proceeded on day two to provide insight into different aspects of Planets and CASPAR.

Curation Challenges

Esther Conway of STFC gave a presentation on the complexity of data and the related digital preservation problems. In the context of specifying user requirements and scenarios, CASPAR had surveyed a number of data archives to understand the

practices related to data sources, data access and use, and rights issues. The intention was also to understand aspects of changes which might affect the preservability of the information encoded in bit sequences. Having faced a large variety of proprietary data formats, Esther argued that there is room for standardising the representation of data and that SAFE (Standard Archive Format for Europe) is a positive step in this direction. SAFE has been designed to act as a common format for archiving and conveying data within the European Space Agency Earth Observation archiving facilities. She then gave detailed examples of the changes that affect our ability to understand data, such as changes in hardware, software and environment and termination of organisational support. In addition, retirement of key personnel, changes in copyright ownership or legal restrictions could all affect the knowledge base required to understand data.

Jérôme Barthélemy of IRCAM (Institut de Recherche et Coordination Acoustique/Musique), France presented the Planets Performing Arts Testbed, where work has been undertaken to identify requirements and preservation scenarios for electro-acoustic music and allied fields. Jérôme explained that musical work in digital form suffers from accelerated obsolescence due to some of its unique properties. Some work created as recently as the 1980s and 1990s has already been lost. He then focused on the preservation of interactive multimedia performances and analysed a complex set of components of such performances, ranging from people, documents and musical instruments to mapping and content-generation applications, multimedia outputs and supporting applications for processing and rendering. Preserving individual components is already a challenge. More daunting still is assembling the components in a logical and temporal order, while preserving the knowledge about the performances' internal logical and temporal relationships over time. Using case studies, it was possible to derive a set of detailed preservation requirements by identifying the scenarios in which changes occur and how they affect any component of a performance.

Birte Christensen-Dalsgaard of the State and University Library, Denmark, presented the ongoing study of users within Planets. The purpose was to understand how the digital revolution affects the way in which the research community functions and how users access and employ digital collections. Answers to these questions will influence preservation strategies and quality measures, and consequently the preservation tools and services Planets is developing. Birte introduced in detail the methodology for the user study, the techniques used to observe users and analyse data. In order to anticipate future trends, interviews have been held with a number of futurologists, to tap into their anticipatory thinking about the changes in storing and using knowledge in the context of libraries.

Preservation Models

David Giarretta presented the CASPAR conceptual model, which is guided by the OAIS Reference Model. David opened the presentation by stressing the importance of information and saying that the ultimate goal for digital preservation is to ensure that information to be preserved is independently understandable to (and usable by) the designated community. He explained that there is a strong focus on representation information within CASPAR, which is vital to the interpretation of data objects. An added bonus is that a piece of representation information can be associated with a

large number of different digital objects, hence sharing the burden of preservation. David presented the key preservation components of CASPAR, describing in detail the information flow in a couple of scenarios in which representation information is created or retrieved from one or more registry/Repository of representation information, in case the representation information packaged with the data is not sufficient. He also explained the role of virtualisation within CASPAR, which is a technique of isolating dependencies on hardware, software and environment. Virtualisation creates external interfaces that hide an underlying implementation. The benefits for preservation arise from hiding the specific, changing technologies from the higher level applications which use them. David concluded the talk by saying the conceptual model had led to the CASPAR architecture which is broadly applicable and useful both for preservation and for interoperability.

Adam Farquhar of the British Library gave a presentation entitled 'Planets: Integrated services for digital preservation' and detailed the types of problems Planets is addressing as well as the rationale for project partners' involvement in the project. Adam started by examining the scale of the problem and concluded that losing digital information costs money and hurts everyone. All partners within Planets have vested interests in digital preservation and in the success of the project. For the national libraries and archives involved in Planets, preservation and access over the long term is their primary mission. Planets, for example, is expected to provide the technology component of the British Library's digital preservation solution. For researchers, digital preservation touches upon complex disciplinary issues and has a potentially huge impact on a broad spectrum of society. For the technology companies, this is an opportunity to introduce innovative services and products and to increase competitiveness. Adam then presented the Planets architecture, explaining the functions of the key components, including preservation planning, preservation action, preservation characterisation, the testbed and the interoperability framework, which integrate the different tools and services to provide one easily managed digital preservation system. In the context of two scenarios, Adam demonstrated how Planets methods, tools, and services can help organisations diagnose and treat obsolescence problems with their digital objects. He hoped that Planets' high levels of automation and scalable components will reduce the costs and improve the quality of digital preservation.

Solutions or Snakeoil

There is a Planets testbed and there are CASPAR 'testbeds'. Attentive readers will have noticed that one is singular and the other is plural. Some may have thought that both have something to do with validation. The details are hard to tell just from the names. There have been questions from outside the projects as to how the two differ. This was fortunately to a great extent clarified by two presentations, one on the CASPAR testbeds by David Giaretta and one on the Planets testbed by **Max Kaizer** of the Austrian National Library. Both presentations contained a great deal of detail and included a look into the future.

The 'CASPAR testbeds' seems to be the collective name for the following three aspects of work:

1. A set of proposed metrics which can be used to validate digital preservation tools and techniques
2. A methodology for simulating the effect upon the usability of digital information caused by changes over time in hardware, software, environment and the knowledge base of the designated communities.
3. The application of the metrics and the methodology to a variety of digital objects from the domains of science, cultural heritage and contemporary performing arts.

The Planets testbed is a software system which provides a controlled environment for experimentation which enables benchmarking of preservation tools, services and strategies. Its role within Planets is two-fold:

1. Test and validate the technical solutions and approaches developed in PLANETS, more specifically:
 1. provide a controlled hardware and software environment for testing and evaluating preservation action (migration, emulation) and characterisation tools and services
 2. record data from experiments in registries for further analysis and comparison
 3. assist the validation of the effectiveness of different digital preservation plans
2. Assess the suitability of the approaches across 'real life' scenarios in various organisations:
 1. analyse applicability of the outcomes of PLANETS in existing workflows and organisational contexts
 2. evaluate their efficiency in providing practicable solutions for organisations engaged in digital preservation

At a later stage testbed services will be offered to organisations outside Planets so that they can test preservation tools and services against benchmark content and validate preservation plans against organisational policies and content profiles. The first release of the Planets testbed is expected in early 2008.

Key Components for Preservation Infrastructures from CASPAR

Adam Farquhar chaired this session which included three presentations from CASPAR, providing a focused view of the various aspects of the project at detailed and technical level. **Luigi Briguglio** presented the CASPAR architecture. He explained the iterative and traceable process used to develop the CASPAR architecture and reported on its current status. The architecture is based on the conceptual model and will go through a number of stages to specify eventually key components interfaces. Although the main focus is on representation information, the overall architecture also includes key components such as data access and the security manager, digital rights manager, authenticity manager and visualisation manager. Luigi also provided a useful mapping of the CASPAR architecture to the OAIS functional model. A second presentation from Luigi focused on the concept of intelligibility of digital objects, which can be defined as the capacity to be correctly understood. Luigi argued that the intelligibility of digital objects is something that must be preserved along with the objects in addition to the bit sequences. Based on a

model developed by Yannis Tzitzikas, it is possible to formalise the intelligibility of digital objects, using the notions of modules and dependencies. This can also be mapped onto the representation information requirements with the OAIS and formally model the community's knowledge and the gaps within it.

David Giaretta has touched upon representation information in a number of previous presentations. His presentation in this session was dedicated to it. He analysed in detail and provided examples of different types of representation information. He then focused on the Registry / Repository for representation information, explaining how it can be used and how the link is maintained between the data and the representation information. David also talked about the various desired properties of the Registry / Repository itself, such as its trustworthiness, extensibility and distributed nature.

Planets Integrated Preservation Services

David Giaretta chaired this session which included three presentations from Planets, providing a focused view of a number of key components at detailed and technical level. **Christoph Becker** of Vienna University of Technology presented a methodology for specifying preservation plans, which allows explicit definition of preservation requirements and offers a systematic way to compare candidate preservation strategies. The ultimate goal is to make an informed and accountable selection of the preservation strategy which is most appropriate to the organisation. The methodology has been implemented in the context of a number of case studies. Christoph provided a sneak preview of the software tool called Planets Preservation Planning Tool (Plato), which is being developed to implement the methodology and automate the preservation planning process. Plato (including decision support and risk assessment modules) is expected to be released in August 2008.

Adrian Brown of the National Archive of the UK presented the content characterisation work. Tools and services are being developed within Planets to characterise the significant properties of digital objects, which are necessary to support the development of preservation plans and validate preservation actions (evaluating change). The aims and objectives of this strand of work are to define methodologies for describing significant properties, to develop tools and services for automating measurement and comparison of these properties and to make recommendations on improving the preservation characteristics of digital object types. Adrian then reported on achievements to date and provided technical details on the characterisation registry, the Extensible Characterisation Description Language (XCDL) and the Extensible Characterisation Extraction Language (XCEL), and the registry-driven characterisation tool framework. Adrian ended the presentation by sharing with the participants the planned year 2 activities.

Jeffrey van der Hoeven of the National Library of the Netherlands gave a presentation on emulation, a digital preservation technique which adapts the computer environment to render the digital object authentically. There has been some level of scepticism around emulation due to its technical complexity and high initial costs. Emulation has never been applied in an operational digital archiving environment. Work within Planets on emulation continues from the Dioscuri project, funded by the National Library and National Archive of the Netherlands in 2004, in recognition of the need for emulation, especially for rendering complex digital objects in the future

without affecting their authenticity and integrity. Jeffrey offered a detailed description of the project's achievements. The Dioscuri emulator has been designed for durability and flexibility. It is built on top of a virtual layer, called a virtual machine (VM), which reduces the dependency of the emulator on the actual hardware and software it runs on. It is also highly component-based. Each component, called a module, imitates the functionality of a particular hardware component (i.e. processor, memory, hard disk, etc.). By combining various modules any computer emulation can be created. Jeffrey explained the areas where improvements are needed and presented a plan to take things forward within Planets. Jeffrey ended the presentation by presenting a diagram showing how emulation tools and services fit with other Planets tools and services.

Conclusion

It has been two intensive yet fascinating days. The participants went home with answers and expectations as well as further questions no doubt. Even for people directly involved in the projects, it has been a great opportunity to understand each other's work and discuss many digital preservation related issues. One could not help remarking upon the strategy behind the funding of three very different but highly complimentary projects. Planets focuses on the immediate problems of libraries and archives and aims to offer solutions here and now. CASPAR deals with more complex scientific and artistic data and provides a valuable insight into more generic systems for digital preservation. DPE is the glue which binds the synergies of PLANETS, CASPAR and ongoing national initiatives.

References

1. Planets <http://www.planets-project.eu/>
2. CASPAR: <http://www.casparpreserves.eu/>
3. DPE: <http://www.digitalpreservationeurope.eu/>
4. wePreserve: <http://www.wepreserve.eu>
5. McGovern, Nancy (2007) 'A Digital Decade: Where Have We Been and Where Are We Going in Digital Preservation?' *RLG DigiNews* v11 no1: http://www.rlg.org/en/page.php?Page_ID=21033#article3

Author Details

Helen Hockx-Yu

Planets Project Manager
The British Library

Email: Helen.hockx-yu@bl.uk

Web site: <http://www.bl.uk/>

[Return to top](#)

Article Title: "Progress towards Addressing Digital Preservation Challenges"

Author: Helen Hockx-Yu

Publication Date: 30-October-2007 Publication: Ariadne Issue 53

Originating URL: <http://www.ariadne.ac.uk/issue53/fp6-2007-rpt/>
[Copyright and citation information](#) File last modified: Monday, 05-Nov-2007
09:59:34 UTC

Main Contents	Section Menu	Email Ariadne	Search Ariadne
-------------------------------	------------------------------	-------------------------------	--------------------------------

Ariadne is published every three months by [UKOLN](#). UKOLN is funded by [MLA](#) the Museums, Libraries and Archives Council, the [Joint Information Systems Committee \(JISC\)](#) of the Higher Education Funding Councils, as well as by project funding from the JISC and the [European Union](#). UKOLN also receives support from the [University of Bath](#) where it is based. Material referred to on this page is [copyright Ariadne \(University of Bath\) and original authors](#).