

Major Preservation Projects under the 6th Framework Program

Vittore Casarosa

ISTI-CNR, Pisa, Italy and HATII at University of Glasgow
casarosa@isti.cnr.it

1 Introduction

Preservation of traditional (physical) objects is (was ?) mainly a management function, whose objective is “to ensure that information survives (unaltered) in usable form for as long as it is wanted”. Preservation of digital information may be different, due to the presence of (rapidly changing) technology needed to access the information. Therefore, while traditionally preserving things meant keeping them unchanged, the ubiquitous digital environment has fundamentally changed the concepts of preservation requirements. If we hold on to digital information without modifications, accessing the information will become increasingly difficult, if not impossible.”

The problem in preservation therefore is not that of just “maintaining the bits” (which is relatively easy), but is the one of ensuring that the digital information can be used (rendered) at any time in the future.

The 6th Framework Program of the European Commission has funded three major projects dealing with the preservation of digital objects, namely DPE (Digital Preservation Europe), Planets (Preservation and Long-term Access through NETworked Services) and CASPAR (Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval).

DPE, which ended in March 2009, was a Coordinated Action fostering collaboration and synergies between many existing national and international initiatives across Europe. It was addressing the need to improve coordination, cooperation and consistency in practices and activities for the effective preservation of digital materials.

Planets and CASPAR are two large Integrated Projects, both addressing the preservation issues in order to provide a “complete solution” to memory institutions. Although sometimes the two projects can be seen as competing, a deeper examination of their approaches reveals that most often they are complementary. In the following we will briefly highlight those two projects.

2 PLANETS

Planets is a four-year project co-funded by the European Commission under the 6th Framework Programme (IST-033789). It started in June 2006 and will end in May 2010, with a total budget of 14 million Euro, about half of which contributed by FP6. The project's goal is to provide long-term access to digital scientific and cultural assets. Planets Partners are:

Libraries The British Library The National Library of the Netherlands Austrian National Library The Royal Library, Denmark State and University Library, Denmark	Archives The National Archives of the Netherlands The National Archives of England, Wales and the United Kingdom Swiss Federal Archives
Universities University at Cologne University of Freiburg HATII at the University of Glasgow Vienna University of Technology	Technology companies ARC Seibersdorf research GmbH IBM Netherlands B.V. Microsoft Research Limited Tessella Support Services Plc

Planets activities are focusing on five major technical areas, plus the Dissemination, Take-up and Training activities, which are customary for European projects, as sketched in Fig. 1.

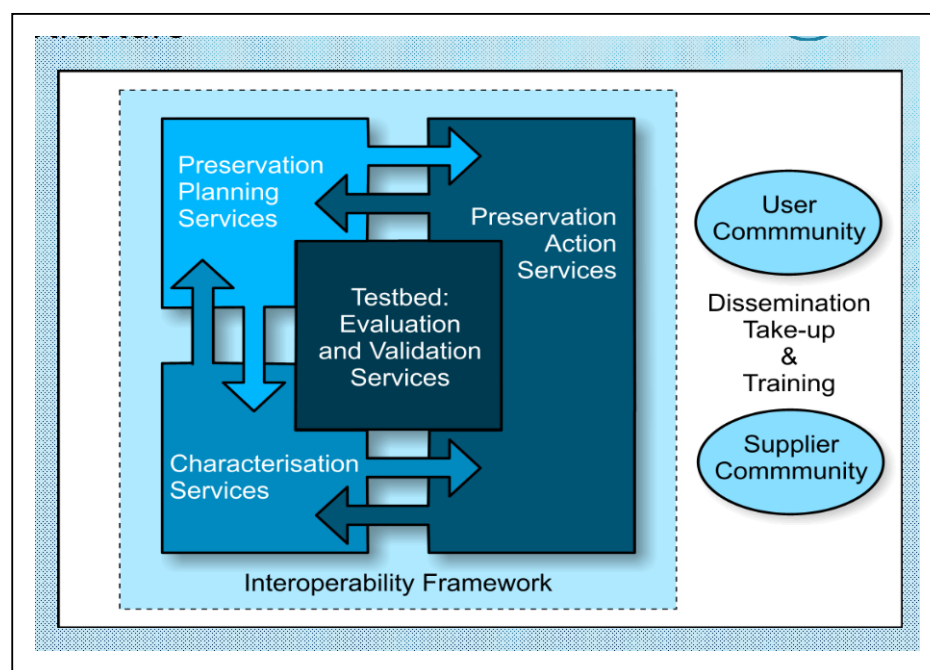


Figure 1

Preservation Planning is a set of services that enables an organisation to identify risks to its digital collection, come up with a set of alternative treatment plans to address the problem, evaluate the plans to determine the most effective one, and then execute the plan with as much automation as possible. In addition to technical information about the collection(s) to be preserved, the planning services require information also about the organisation's policies, goals, and constraints; the environments in which end-users interact with the content, and the goals that the end-users have. The preservation planner uses two key sets of Planets services: content characterisation and preservation action.

Content characterisation services identify and extract features of the content in the collection that are relevant to the planning process. Examples include basic technical features such as the pixel-depth, height, width of images, rendered features such as the number of pages in a document, and other aspects such as whether a database uses stored procedures. Planets has developed two XML based languages, namely XCEL to specify how to extract properties from a digital object, and XCDL, to describe the properties of a digital object. A tool to compare the properties of two digital objects is also available, as an aid in checking that the migration process does not change them.

Preservation action services determine what can be done. For example, one preservation action might take a specific type of digital object and convert it to a new format; another might produce an emulation environment that enables users to interact with the original digital object using the original software application; another might provide a viewing tool that provides limited interaction with a digital object, but can be readily executed on a user's portable device.

Although some preservation plans will be simple, others will comprise complex multi-step work-flows that involve extracting content from a repository, characterising it, using the results to select one or more services to treat, transform, or encapsulate the content, and then either returning the result to the repository with a detailed record of treatment, or providing a capability that can be used in a delivery environment so that end-users can get appropriate access. The Preservation actions will rely on a Registry of available tools and services, either developed by Planets or provided by different sources, such as open-source projects, commercial vendors, and third-party service providers.

The Interoperability framework enables the intimate relationship between preservation planning, preservation actions, and content characterisation as well as extensibility with a plug-in approach. It provides essential shared services such as auditing and security, as well as an extensibility mechanism, by which composite services and workflows can be defined by combining existing services available in the Registry.

The Testbed application provides a controlled environment that enables researchers and practitioners to conduct experiments and gather data on a scientific basis. They will be able to define, schedule and execute an experiment, collecting data to analyze the

result or to reproduce it later. The data collected in the testbed can provide information about the strengths, weakness, and properties of transformation tools on different types of content, can provide a basis for selecting better preservation plans, can enable the users to compare different approaches to preservation planning.

Complete information about Planets can be found at the project web site:

<http://www.planets-project.eu/>

3 GASPAR

CASPAR is 42 months project co-funded by the European Commission under the 6th Framework Programme (IST- 033572). It started in April 2006 and will end in September 2010, with a total budget of 15 million Euro, of which about 9 million Euro are contributed by FP6. The project's main goals are to implement, extend, and validate the OAIS reference model and to enhance the technology for capturing preservation related information for digital objects. Planets Partners are:

Academic HATII at the University of Glasgow, UK University of Urbino, Italy University of Leeds, UK	Research council Science and Technology Facilities Council, UK Consiglio Nazionale delle Ricerche – Institute of Information Science and Technologies, Italy Centre National de la Recherche Scientifique, France
Industrial Engineering Ingegneria Informatica S.p.A., Italy IBM Haifa Research Laboratory, Israel	International Organizations European Space Agency, ESRIN, Italy UNESCO, France
Small and Medium Enterprises Advanced Computer Systems S.p.A., Italy Asemantics S.r.l., Italy Metaware S.p.A., Italy	National Organizations Institut National de l'Audiovisuel, France Foundation for Research and Technology Hellas, Greece Institut de Recherche et Coordination Acoustique/Musique, France International Centre for Art and New Technologies, Czech Republic

The CASPAR research and development activities have been focused on the implementation of the OAIS reference model (Open Archival Information System, ISO:14721:2002), and on testing this architecture on three application domains, namely cultural data, contemporary performing arts and scientific data. An OAIS is an archive

consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated user community. The information being maintained has been deemed to need indefinite preservation, even if the OAIS itself is not permanent.

An overall pictorial view of the CASPAR Framework is depicted in Fig. 2.

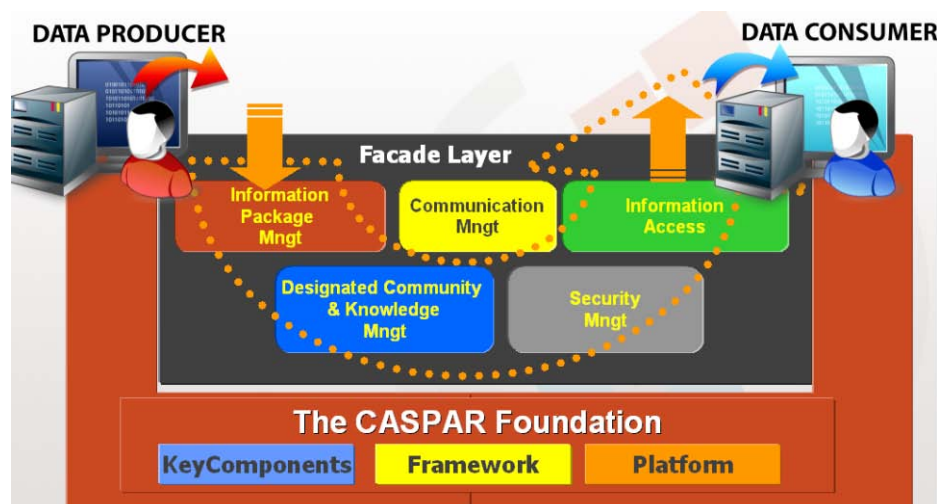


Figure 2

As shown in Figure 2, the CASPAR Foundation provides OAIS-compliant services with many features for supporting Data Producer and Consumer in their preservation activities. From an overall point of view, it is possible to group the features together in 5 main blocks:

1. Information Package Management
2. Information Access
3. Designated Community and Knowledge Management
4. Communication Management
5. Security Management.

Based on the OAIS Reference and Functional Models, CASPAR has defined the basic infrastructure for supporting those main features and providing digital preservation services, called the CASPAR Foundation, which is composed by 11 Key Components. The CASPAR Framework guarantees portability and interoperability (i.e. compliance to WS-I open standard) with existing systems and platforms.

The Key components are:

registry (REG), knowledge (KM), orchestration (POM), representation information (REPINF), preservation datastore (PDS), data access and security (DAMS), digital rights (DRM), finding aids (FIND), virtualisation (VIRT), packaging (PACK) and

authenticity (AUTH). The overall architecture of the CASPAR Foundation is depicted in Fig. 3.

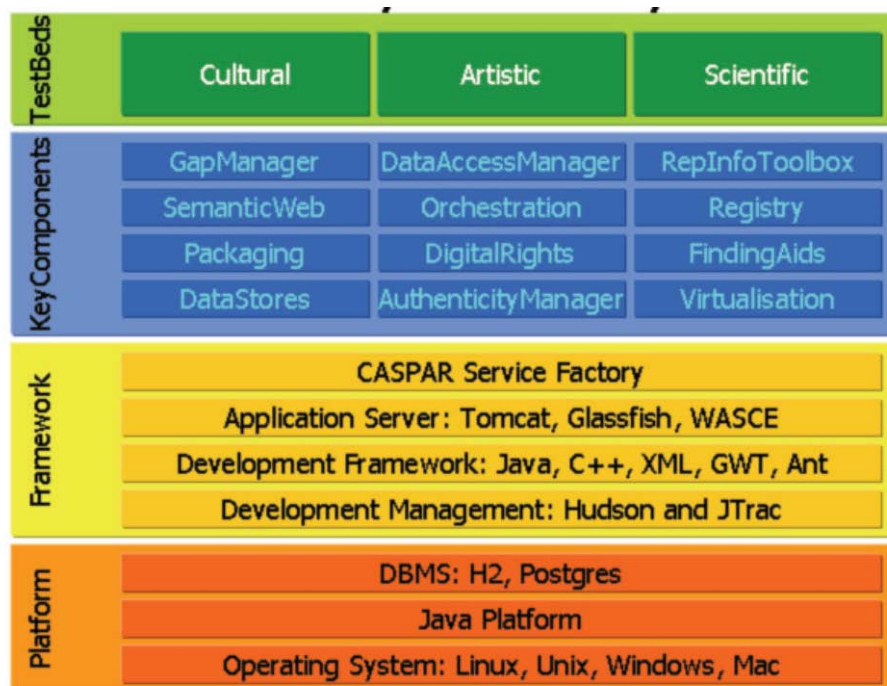


Figure 3

On top of the key components, CASPAR is testing and demonstrating preservation scenarios and strategies in order to validate its conceptual model and architectural solutions within different testbed domains: cultural, artistic and scientific.

Cultural data testbed

This testbed focuses on the preservation of all data necessary to document, visualise and model heritage sites and will provide a valuable resource to assist conservation experts in restoring the associated site while keeping its original integrity. The documentation, visualisation, and modelling of natural and cultural heritage sites is a complex task that requires large amounts of data and information. The objective of this testbed is to test the solutions developed throughout CASPAR, including virtual digital objects, spanning between processing and display.

Contemporary art testbed

This testbed is focused on the implementation for contemporary arts: contemporary music, performing arts, and other forms of technology-enhanced arts like video games. These objects also include dynamic interactive digital objects oriented towards presentation and replay. Contemporary music, as well as performing arts and video games produce very complex objects including specific hardware, instructions, and

specific equipment. These electronic models often contain highly complex extensions such as interaction devices and instructions. Interaction devices include specific sensors, instruments, and equipment while instructions often include score, software components, and audio files. The objective of this testbed is to study these specific extensions, their relationship with the generic models, procedures, and services, and to start the analysis and set-up of both a generic and specific infrastructure.

Science data testbed

This testbed is aimed at satisfying the Earth Science community requirements, by developing the necessary specific services and prototyping an Earth Observation science data preservation. It also aims at the preservation life cycle of scientific data and the preparation of specific tools that will allow visualisation and navigation of the complex metadata associated with science data complex objects. In terms of preservation, these are often very high volume, complex digital objects, oriented towards processing.

Complete information about CASPAR can be found at the project web site:

<http://www.casparpreserves.eu/>