SIXTH FRAMEWORK PROGRAMME
PRIORITY IST-2005-2.5.10
Access to and preservation of cultural and scientific resources

Contract for:

INTEGRATED PROJECT

Annex I - "Description of Work"

Project acronym:	**CASPAR**
Project full title:	**Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval**
Proposal/Contract no.:	033572
Related to other Contract no.:	
Date of preparation of Annex I:	10 March 2006

TABLE OF CONTENTS

1. Project Summary

CASPAR will address the growing challenge facing society of a deluge of intrinsically fragile digital information, upon which it is increasingly dependent, by building a pioneering framework to support the end-to-end preservation "lifecycle" for scientific, artistic and cultural information, based on existing and emerging standards.

The ambitious challenge to build up a common preservation framework for heterogeneous data and a variety of innovative applications will be achieved through the following high level objectives:

- **to establish the foundation methodology applicable to a very wide range of preservation issues**.

  The guiding principle of **CASPAR** is the application of the OAIS Reference Model

- **to research, develop and integrate advanced components** to be used in a wide range of preservation activities. These components will be the building blocks of the **CASPAR** Framework

- **to create the CASPAR framework**: the software platform that enables the building of services and applications that can be adapted to multiple areas

The **CASPAR** consortium will demonstrate the validity of the **CASPAR** framework through **heterogeneous testbeds, covering a wide range of disciplines from science to culture to contemporary arts and multi-media**, providing a reliable common infrastructure which can be used or replicated in many more areas.

The **CASPAR** consortium will seek to guarantee the *future evolution* of **CASPAR**. This ambitious goal will be pursued through:

- the building of the **CASPAR preservation user community** creating consensus around the initiative and gathering a critical mass of potential users

- **embedding the CASPAR framework** and components within key memory organisations, both national and international.

To achieve this, CASPAR brings together a consortium covering important digital holdings, with the appropriate extensive scientific, cultural and creative expertise, together with commercial partners, and world leaders in the field of information preservation.

2. Project objectives

The **CASPAR** challenge is to achieve 4 main goals that can be stated as follows:

Goal 1: build a pioneering preservation environment, based on a full use of the OAIS Reference Model[1] and building in the latest developments in knowledge technologies;

Goal 2: demonstrate its ability to handle the preservation of the digital resources of many user communities

Goal 3: advance the current state of the art in digital preservation

Goal 4: development of technological solutions **supporting the emergence of an offer of systems and services for preservation of digital resources**

Expanding these goals into more specific objectives, **CASPAR** will:

1. Implement, extend and validate the OAIS reference model
2. Enhance the techniques for capturing Representation Information and other preservation related information for content objects
3. Design virtualisation services supporting the preservation of digital resources over the long term, despite changes in the underlying computing (hardware and software) and storage systems, and the Designated Communities.
4. Integrate as standard features of **CASPAR**, digital rights management, authentication and accreditation
5. Research more sophisticated access to and use of preserved digital resources including intuitive query and browsing mechanisms
6. Develop case studies demonstrating the validity of the **CASPAR** approach to the preservation of digital resources across different user communities and assessing the conditions for a successful replication
7. Actively contribute to the relevant standardisation activities in areas addressed by **CASPAR**.
8. Raise awareness about the critical importance of the preservation of digital resources among the relevant user-communities and facilitate the emergence of a more diverse offer of systems and services for preservation of digital resources.

We propose a number of methods by which the above may be verified, and for each of these there are a number of specific metrics, associated Work Package(s) and the milestones (month number) at which these metrics can be evaluated.

The metrics are shown in Table 1, and they supplement the project deliverables which include reports produced by the consortium. **Reviews are to be conducted by an independent group of internationally recognised experts (External Review**

Committee), and/or EU representatives. Each of these metrics are to be evaluated at a number of points in order to monitor the development of capabilities, and will be linked to Key Performance Indicators.

Table 1 CASPAR Measurable Objectives

OBJECTIVE	PROGRESS ASSESSMENT METHODOLOGY	WP	AUDIT TIME(S)
DIGITAL PRESERVATION METRICS			
1) Demonstrate a sound theoretical basis for the approach taken, including			
a. the compatibility with the OAIS Reference Model and related standards – all of which have been peer reviewed extensively in the standards process itself and also by practitioners of digital preservation in a great number of areas	Review of **CASPAR** architecture and OAIS components with the External Review Committee – showing the relationship to OAIS and the relevance to digital preservation.	WP1200 WP1300	M6 M12
b. the development of a peer-reviewed theoretical under-pinning for the specifics of the **CASPAR** approach	Paper(s) submitted to peer-reviewed journal, and presented at international conference(s)	WP1200	M18 M36
2) Provide a practical demonstration by means of what may be regarded as "accelerated lifetime" tests. These should involve demonstrating the ability of the Framework and digital information to survive:			
a. environment (including software, hardware) changes	Demonstration to the External Review Committee of usability of a variety of digitally encoded information despite changes in hardware and software of user systems, and such processes as format migration for, for example, digital science data,	WP4100, WP4200, WP4300 WP4400	M18 M27 M30

	documents and music		
b. changes in the Designated Communities and their Knowledge Bases	Demonstration to the External Review Committee of usability of a variety of digitally encoded information by users of different disciplines	WP4100, WP4200, WP4300 WP4400	M24 M30
3) Show improved trustworthiness of repositories			
	Use (draft) certification standard at various stages of application of **CASPAR** techniques to a number of repositories. The repositories should improve in trustworthiness.	WP4100	M12 M24 M36
4) Review of evaluation methodology for:			
a. **CASPAR** Framework evaluation	External Review Committee report on proposed validation techniques for **CASPAR** Framework	WP4200	M12 M18
b. **CASPAR** Testbed evaluation	External Review Committee report on proposed validation techniques for methodology for validation of the Testbeds	WP4100	M12 M24 M36
5) Demonstrate the applicability of the CASPAR virtualisation techniques for digital information in the following areas (all of which also include normal documents of various kinds), and their maintainability over time.			
a. Science, Cultural Heritage, Contemporary Arts	Demonstrate to External Review Committee some specific examples virtualisation of data from a number of disciplines and show usability.	WP4200, WP4300, WP4400	M8 M18 M24 M30
b. Digital Rights Management	Demonstrate to External Review Committee the use of the DRM	WP2300	M30

	description for a variety of types of data and rights		
c. Preservation Description Information, including			
i. Provenance	Demonstrate to External Review Committee the capture of provenance information with change of ownership, events, processing etc and can allow for evolution of these over time.	WP2100	M24
ii. Persistent identification	Demonstrate to External Review Committee the use of a system of Persistent Identifiers together with an analysis of the socio-economic survivability of the infrastructure supporting this identifier.	WP2100 WP2200 WP2300	M18 M30
iii. Authenticity	Demonstrate to External Review Committee the system for proving authenticity of a digital object, together with an analysis of the maintainability of this authenticity system over time and also the issues arising from the transformation of the digital encoding of the information.	WP2300	M15 M30
PROJECT IMPACT METRICS			
Show contribution to standards	Production of evidence of contributions to the development of some or all of the standards in section 5.1	WP1200	M24 M36
Show adoption of **CASPAR** results in the preservation community	Demonstrate effective adoption of the **CASPAR** by users involved in test beds and number of entities external to the project, interested in the **CASPAR** framework and endorsing it in some way		M30 M38
Show that effectiveness of dissemination processes	Indicate a number of providers of tools and services of digital preservation interested in **CASPAR** framework and developing complementary services or tools	WP5200 WP5300	M36

3. Participant list

Partic. Role*	Partic. No.	Participant name	Participant short name	Country	Date enter Proj.**	Date exit Proj.**
CO	1	Council for the Central Laboratory of the Research Councils -	CCLRC	UK	1	42
CR	2	European Space Agency, ESRIN -	ESA	IT	1	42
CR	3	University of Glasgow HATII	UG	UK	1	42
CR	4	Universita' degli studi di Urbino, Istbal (Istituto di studi per la tutela dei beni archivistici e librari)	UU	IT	1	42
CR	5	UNESCO	UNESCO	FR	1	42
CR	6	Advanced Computer Systems ACS S.p.A. of Rome	ACS	IT	1	42
CR	7	Asemantics S. r. l.	ASMX	IT	1	42
CR	8	IBM Haifa Research Laboratory	IBM	ISR	1	42
CR	9	Consiglio Nazionale delle Ricerche (Istituto della Scienza e delle Tecnologie della Informazione)	CNR	IT	1	42
CR	10	Metaware	MW	IT	1	42
CR	11	Institut National de l'Audiovisuel	INA	FR	1	42
CR	12	University of Leeds	UNIVLEEDS	UK	1	42
CR	13	Engineering – Ingegneria Informatica	ENG	IT	1	42
CR	14	Foundation for Research and Technology – Hellas	FORTH	GR	1	42
CR	15	Centre National de la Recherche Scientifique	CNRS	FR	1	42
CR	16	l'Institut de Recherche et	IRCAM	FR	1	42

		Coordination Acoustique/Musique				
CR	**17**	**International Centre for Art and New Technology in Prague**	**CIANT**	**CZ**	**1**	**42**

*CO = Coordinator

CR = Contractor

** Normally insert "month 1 (start of project)" and "month n (end of project)"

These columns are needed for possible later contract revisions caused by joining/leaving participants

4. Relevance to the objectives of the IST 2.5.10 activity objectives

Relevance to the General Objectives

CASPAR will produce a coherent preservation methodology, an implementation of a support infrastructure, a validation methodology with case studies in science, cultural heritage and performing arts, with replicable testbeds. **CASPAR**'s broad scope fits the **Lisbon Strategy**[2] which includes the aim of making Europe "the most competitive and dynamic knowledge-based economy in the world" and the ***e-Europe Action Plan***[3] for accessing and preserving Europe's cultural, scientific, social and industrial digital heritage. Commissioner Reding's repeatedly[4] stated need for **enlarging the internal market for electronic communications and digital services** is contributed to by **CASPAR's** development of standards, methodologies and technologies devoted to the preservation (and re-use) of born-digital and digitised objects more generally, including for example those from e-Government, e-Business and e-Health.

The project will result in standards, policies, strategies, guidance and tools for different types of users, within a **multi-industry perspective**, to manage digital content preservation in support of specific business lines.

By focusing on knowledge, **CASPAR** contributes to **semantic-based interoperability,** which is a fundamental technological stream addressed by IST. Simultaneously, **CASPAR** participates in the **e-inclusion** vision for the European Information Society, one of the priorities running through the ***e-Europe 2005 Action Plan***.

Relevance to specific directives

The rights protection criteria and, specifically, DRM-related issues are controversial topics within the debate related to several European Directives such as the "***European Copyright Directive***" (EUCD[5]), the Directive on the "***Enforcement Of Intellectual Property Rights***"[6] and the previous regulatory initiative[7] on the "***Legal Protection of Databases***". **CASPAR** will devote specific resources to the definition and development of Digital Rights Management tools that are easily (maybe self-) adaptable to the evolution of technologies for storing, accessing, representing and rendering long-lived digital information. The directive on the "***Re-use of public sector information***"[8] will also be relevant for **CASPAR**.

Relevance to Strategic Objective 2.5.10

To improve Preservation Methodologies and standards

The **CASPAR** paradigm provides a framework and key components for the preservation of complex information, pushing the current state of the art and building on current, and anticipating new, standards. The validation of the adopted approach will open the door to further developments and services.

To move from data to knowledge

CASPAR intends to contribute specifically to this IST-addressed field by focusing on **preserving knowledge** for future archive intelligibility and information system/services interoperability**.** Preserving information and knowledge – not just "the bits" – allows the keeping of archives alive through time.

This approach locates **CASPAR** in the overall IST vision. Moving from data to information to knowledge is a pervasive IST objective, which has been addressed in depth in the ISTAG[9] reports and is considered the basic step towards the coordination of the research activities within the European Research Area (ERA).

To develop sizeable test-beds that present complementary requirements for preservation

CASPAR develops three test-beds that instantiate the generic framework functionalities into real domains, which present complementary requirements in terms of preservation. (1) very-high volume, complex digital objects, oriented towards processing: **Scientific testbed** (2) dynamic interactive digital objects, oriented towards presentation and replay: **Contemporary Arts testbed** (3) virtual digital objects, spanning between processing and display, and able to overlap with each: **Cultural Data Testbed**.

Knowledge-driven tools will be produced to demonstrate the availability and accessibility over time of multi-sourced and multi-formatted resources in the scientific, artistic and cultural fields.

To ensure the co-evolution of technology and applications so that technology advances are exploitable in innovative products and services.

The **CASPAR** consortium places heavy emphasis on the usability of the services it will develop. Even more importantly, the **CASPAR** consortium believes that the preservation infrastructure components the project will produce can form the basis of preservation services, but would require further developments in security and robustness before a commercial service could be provided. The audit and certification processes will underpin this market by means of suitable metrics, as will the cost modelling that forms part of **CASPAR**.

Strengthening the European Research Area (ERA).

Viewing preservation as "Interoperability with the future" allows one to see how **CASPAR** will play a fundamental role in contemporaneous interoperability where the divide is not one of ***time***, but rather one of ***discipline and expertise***. This interoperability is precisely what is needed to allow new and unexpected links to be made between disparate pieces of scientific or cultural resources.

- **To exploit new opportunities and respond to emerging needs.** The fragility of digital information, combined with its fundamental importance to our increasingly

connected and interdependent society, is making digital preservation come to the fore in the minds of funders, legislators and commercial leaders. **CASPAR**, and related IPs, will allow more fruitful discussions with other countries, notably the USA. **CASPAR** can enable European companies to provide key digital preservation services.

- **Encouraging funders to support preservation**. By allowing funders to see their information holdings as assets that will not wither away but can be put to repeated use, **CASPAR** will encourage them to put funds into preservation, supported by an infrastructure to encourage sharing of the preservation load and a detailed estimation of costs.

- **Growing the market in preservation**. By playing a major role in the audit and certification standard[10] **CASPAR** is in a good position to also lead European efforts towards an accreditation mechanism and audit and certification services. Having such a service in place will underpin the supply of commercial preservation services and help grow the European market.

- **Facilitate societal cohesiveness and legislative effectiveness**. By allowing easier access to ***information*** rather than just data, for example e-Government records, scientific datasets and cultural digital assets, while at the same time providing the citizen with the wherewithal to verify the integrity and authenticity of that information, **CASPAR** will promote openness and trust of the digital within Europe.

- **Support an information rich society.** As World markets change and Europe moves increasingly to a post-industrial era, the importance of access to and use of information becomes a key societal asset. Familiarity with and confidence in our scientific and cultural heritage, including performing arts, is vital. However, the tools for supporting this in our schools and universities are fragmented. The capabilities that **CASPAR** will provide will enable the greater usability of information for society at large.

State of the Art

The current state of the art is inadequate for the task at hand, which is to preserve digital information for generations to come. Europe needs to begin to tackle the hard issues of preservation, and **CASPAR** can be the instrument for this process.

Current testbeds (such as the National Archives of the Netherlands testbed[11]) begin to touch on some of the factors that affect preservation strategies. However, the focus so far has been on document-type data and there is a neglect of the detailed information content, especially of complex and interrelated objects. Moreover, the approach is difficult to cross-calibrate and reproduce; it is a framework that provides some numerical outputs in particular implementations but it does not answer the hard questions about whether the information will really be preserved – and usable – especially when one considers that there may be multiple migrations and transformations. One can imagine information being distorted and lost through a series of "***Chinese Whispers***" even though at each exchange the information has apparently been passed correctly.

The view of preservation as an issue only of media and formats is marginally supportable as long as one considers only what are essentially opaque formats for

documents and images which must be rendered by specific, often proprietary, software in order then to be examined by human intelligence. But so much digital information is inherently non-textual and is not simply displayed but is processed for each access. Indeed, in many cases the digital objects that are used are freshly created (from underlying information) for each user. Each object may be linked directly or indirectly to many more, whether locally or half a world away. There are of course many projects and products that claim to be "self describing" or "common" or "universal" (e.g. Universal Virtual Computer – UVC). However, close examination suggests these are like brand names for selling the product (like soap powder) rather than describing real capabilities.

Other panaceas on offer include transcription to XML in one way or another. Yet this is to overload what XML has to offer while at the same time crippling what can be done with the information at the present time. Much of the behaviour that is fundamental to digital information is embodied in the associated software, and that behaviour cannot be captured in an XML tag or associated schema, and surely one cannot re-implement all software to deal with the transcribed data, thereby risking crippling the current usage.

Although the OAIS methodology makes knowledge one key factor for preservation, so far very little use has been made of results in Knowledge Representation within the preservation field. **CASPAR** will overcome this limitation by acting on both the practical and the theoretical fronts. From a practical point of view, it will employ Semantic Web tools and techniques to address the "Semantic" type of Representation Information defined by OAIS, thus aligning preservation with the state of the art in knowledge management. From a theoretical point of view, **CASPAR** will research more powerful representation formalisms for the Preservation Description Information, notably provenance and context, building on the results and methodologies from Knowledge Representation, and in particular from Description Logics.[12]

The current state-of-the-art Digital Rights Management (DRM) systems are mainly focused on proprietary mechanisms for copy protection of digital artefacts; powerful content provider lobbies have put significant political and economic power in the legal framework surrounding copyright laws. Proprietary schemes exist for restricting usage of digital materials. Many efforts also exist in the declaration of Digital Rights with more or less well adopted formats. The ISO, OASIS, the MPEG Consortium and several other standardisation bodies have all defined XML-based formats for digital rights declaration, of which ORDL, XrML[13] and MPEG-REL[14] constitute the most widely accepted. All these formats try to model the entities involved in the Digital Rights Management domain as a set of principles participating in the process, a set of rights over a material and a set of conditions under which such rights are granted. The challenge here is to extend DRM over much longer timescales so that it can be relied upon into the dimly foreseen future.

CASPAR will use a novel approach for extending the management of digital rights over longer timescales by adding DRM-related concepts, such as right assertion and enforcement, to the preservation information model; but also by using current Semantic Web technologies to keep track of the evolution of the legal framework defining rights and procedures linked to digital objects production and distribution. Such a model will allow the evolution of licenses stored along with preserved objects, so that they could cover new distribution means and also new regulation policies governing copyright and, in particular, fair-use scenarios.

Much of what is being done elsewhere is useful to a greater or lesser extent and can be used within **CASPAR's** broader framework. It is not possible to pretend that **CASPAR**

will solve all problems, nor to hope to re-invent everything from scratch. The project's aim is to show how the different approaches can work together – in those cases where they actually do provide adequate, albeit limited, preservation mechanisms. More importantly, we can provide the key missing components and framework, described later, which can make this a practical possibility.

State of other projects

Open Archival Information Systems (OAIS) Reference model

The Open Archival Information Systems (OAIS) Reference Model is an ISO standard (ISO:14721:2002) of fundamental importance to this call as it addresses the long-term preservation of ***information***. The standard provides a consistent set of concepts and terminology for comparing and contrasting archives, and a framework for the development of related standards.

INTERPARES and SDSC projects -

InterPARES 1 and InterPARES 2 projects have been developing an analysis of the entire preservation chain in the heterogeneous domains of the scientific data/records and the cultural and government records. That work is based on the OAIS model and focuses on the identification of the activities, constraints and requirements for preserving authentic electronic records.

Grid technology – EGEE

EGEE (Enabling Grids for E-science, see http://public.eu-egee.org/), is a pan European Grid infrastructure that is deploying a state of the art operational Grid infrastructure in Europe with the aim of providing a reliable service capable of attracting new resources, applications and user communities.

Grid technology – DILIGENT

Diligent (Digital Library Infrastructure on Grid ENabled Technology, see http://www.diligentproject.org/) that aims to promote the advancement of a Knowledge e-Infrastructure in Europe by building a Digital Library (DL) demonstration testbed on top of EGEE focusing on two application areas - environmental eScience and cultural heritage.

CHRONOPOLIS

The CHRONOPOLIS project hopes for funding from the NSF at the level of $1M/year for 5 years. CHRONOPOLIS has the concept of a distributed, secure, replicated data store, supplemented by a research arm.

ADAR – ESA Advance Data ARchive System

The ESA ADAR R&T project (2002-2004) aimed to develop a logical model for a generic Earth Observation Payload Data Ground Segment (PDGS), based on the Open Archival Information System (OAIS). ESA ADAR plans include the HARM (Historical Archive Rationalization and Management) initiative, aiming at defining a rationalization strategy for the ESA's distributed long term EO archives.

Digital Curation Centre

The Digital Curation Centre (DCC) is a UK project with the aims of providing a resource to enhance digital data repositories; providing digital curation services;

conducting leading research in relation to digital curation; helping to establish standards for digital curation; creating testbeds for digital curation; and creating a network of interested parties.

DSpace

DSpace is a digital repository system developed jointly by MIT Libraries and Hewlett-Packard (HP) which captures, stores, indexes, preserves and redistributes an organization's research material in a variety of digital formats. An interface between SRB and DSpace is being developed which will allow ***storage virtualisation***, however DSpace does not support the higher levels of content virtualisation which we propose in **CASPAR**, nor do the preservation aspects go much beyond bit preservation.

Fedora

Fedora is a general purpose repository service developed jointly by The University of Virginia Library and Cornell University. The Fedora project aims to provide open-source repository software that can serve as the foundation for many types of information management systems. The Fedora Object Model which allows the description of object properties for management and tracking, and also identifies internal components of a complex object. However virtualisation of the contents is not available in Fedora.

BRICKS

The BRICKS Integrated Project (IP) aims at establishing the organisational and technological foundations of a Digital Library - a networked system of services over globally available collections of multimedia digital documents, covering Digital Museums, Digital Archives and other kinds of digital memory systems. BRICKS is based on a open, peer-to-peer architecture, Web services, with a well-defined and documented API.

Current state of the art of artistic preservation

Digital contents for performance art objects can consist of a setup enabling a cultural performance: content appears while playing with the devices or making the setup work. There is no agreed state of the art, but there is instead a number of innovative but heterogeneous ways to deal with digital preservation of cultural contents, including attempts to maintain hardware and software systems, emulate systems, migrate to new systems and finally “virtualise” systems by describing them.

5. Potential Impact

The expected impact of the project in various domains are:

- Research : advance the state of the art e.g. the application of OAIS, virtualisation etc
- Standards : development of and extensions to standards in many areas
- Policy developments : facilitating the understanding and widespread application of digital preservation
- Industrial : supporting emerging digital preservation systems and services

5.1 Contributions to standards

CASPAR is strategically focused on current and emerging standards and aims to influence these emerging standards to best support the preservation process. Examples of the areas **CASPAR** will influence include:

1. **The evolution of the OAIS Reference Model:** OAIS is due for its period review in 2007. It is expected that the experiences gained from **CASPAR** will allow us to make some significant enhancements to the Reference Model.

2. **Authenticity of Digital Information: CASPAR** will be implementing some of the concepts from InterPARES and will be working closely with that project to extend the theoretical work with lessons from the testbeds.

3. **Extensions to the Object Storage Device (OSD) standard to support an implementation of the OAIS concepts:** The initial analysis indicates that we will be extending some of the current industry standard interfaces with specific reference to digital preservation.

4. **Contribution to the definition of the national guidelines and recommendations for preservation:** All the European States have developed – within e-government policies and according to the 1999 EU specific directive about electronic signatures – specific legislation for transforming the traditional paper documents/records into digital resources. The **CASPAR** consortium can play a role in many national situations, to assist in the development of realistic schemes for the preservation of governmental records.

5. **Certification standard for trusted repositories:** will work with the draft standard produced by the RLG Task Force on Trusted Digital Repositories leading to a full international standard .

6. **Support the promotion of the ESA proposed SAFE[15] archive format standards for all Earth Observation missions:** ESA is developing a standard for an archival format for Earth Observation data. **CASPAR** expects to work closely with ESA to refine and enhance this standard and to promulgate the use of the standard in a broader context.

7. **Contributions to AIP-type packaging techniques**: the Archival Information Package (AIP) is a fundamental part of the OAIS approach. There are a number of candidate implementations, e.g. METS[16] and XDFU[17], but none of these are

complete at the time of writing. **CASPAR** will evaluate these alternatives and contribute to the completion of the most appropriate techniques.

8. **Contributions to OGC[18] effort for the geospatial data:** To support some of the virtualisation processes, **CASPAR** will extend some of the Geographical Information Systems (GIS) standards in order to facilitate interoperability in the Earth Observation part of the Science testbed.

9. **Digital Libraries evolution on GRID (GGF, SRB) Persistent Archives GGF: CASPAR** already has close connections with the ***Persistent Archives*** Working Group of the Global Grid Forum and will naturally contribute to that work.

10. **Contribution to the MoReq[19] guidelines with reference to preservation issues: CASPAR** could contribute specific representations by integrating the MoReq recommendation with the OAIS analysis and implementation developed in the project.

5.2 Contribution to policy developments

CASPAR will provide a fundamentally new paradigm for dealing with the deluge of digital information. Lessons learned from temporal interoperability may be applied to contemporaneous interoperability, where the separation is one of discipline rather than of time. In this way **CASPAR** can supply the kind of tools and procedures which are called for by ***i2010 – A European Information Society for growth and employment***[20] initiative which promotes, amongst other things, the knowledge society. In particular **CASPAR** addresses several of the basic challenges identified for ***Preserving Digital Content*** in the document ***i2010: Digital Libraries***[21], namely:

- ***Financial challenges***: addressed by the Cost Modelling work in **CASPAR**
- ***Organisational challenges***: addressed by **CASPAR** training and the **CASPAR** framework allowing the sharing of the effort of preservation.
- ***Technical challenges***: **CASPAR** will advance our understanding of many of the fundamental issues of digital preservation; the components and framework should improve the cost-effectiveness of digital preservation efforts.

CASPAR will:

- build the autonomic system to support the European body of shared digital scientific, cultural heritage and performing arts, and be extensible to many other areas.
- encourage those holding digital information to see themselves as part of a real chain of custody of evidence.
- provide the first systematic and integrated framework for preserving, disseminating and adapting contemporary artistic works based on interactive media.

CASPAR partners are based in some of the major European stakeholder institutions in digital preservation in science (ESA, CCLRC), cultural heritage (UNESCO) and performing arts (INA, IRCAM, and CIANT). This facilitates the take-up of the **CASPAR** results and testbeds. Some of these institutions are represented on the Task Force for Permanent Access[22] and are involved with the proposed Alliance of stakeholders which will foster the production of a European-wide infrastructure for

digital preservation. The expectation therefore is that **CASPAR** results will assist in the formation of this infrastructure.

5.3 Risk assessment and related communication strategy

The risk associated with ignoring the need to preserve digital information has been widely and increasingly recognised by information professionals. The memory of society at large and of individuals is held in fragile digital objects. The preservation of the "bits" is not fundamentally hard. Preservation of the information that those bits encode is much more difficult. Failure to address the issues risks leaving society unable to deal with the deluge of information on which it depends.

The public awareness of these issues needs to be raised, including those issues which affect their personal and family data such as digital photographs and other digital recordings, which have very different preservation demands to, for example, paper copies of photographs which can still be used even after sitting on a shelf untouched for decades. (see also Section 6.7).

6. Outline implementation plan for the full duration of the project

6A Activities

6.1 Research, technological development and innovation activities

Overview

The OAIS Reference Model provides the intellectual core of the **CASPAR** approach to preservation.

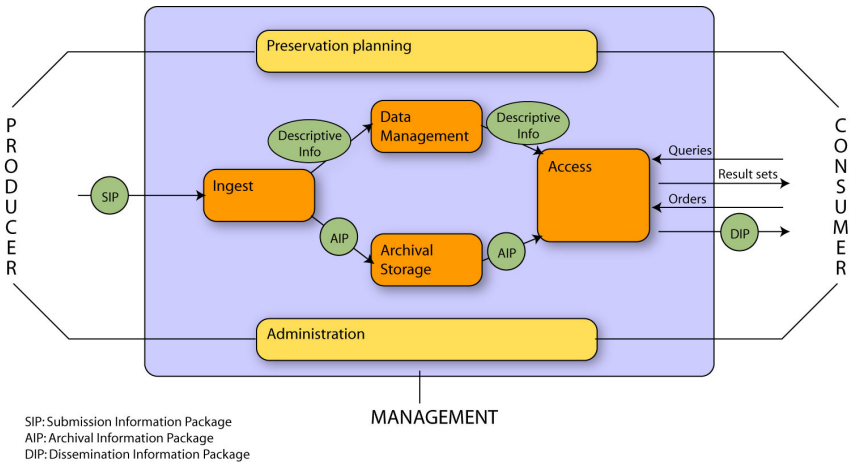


Figure 1 OAIS Functional model

From a functional point of view, the OAIS **Functional Model** (Figure 1) introduces the Ingest, Access, Archival Storage, Data Management and Preservation Planning OAIS components. This functional model is the one to which most projects refer when they claim OAIS compliance, although experience[23] shows that the least attention is normally given to Preservation Planning. All these components are part of the **CASPAR** Architecture.

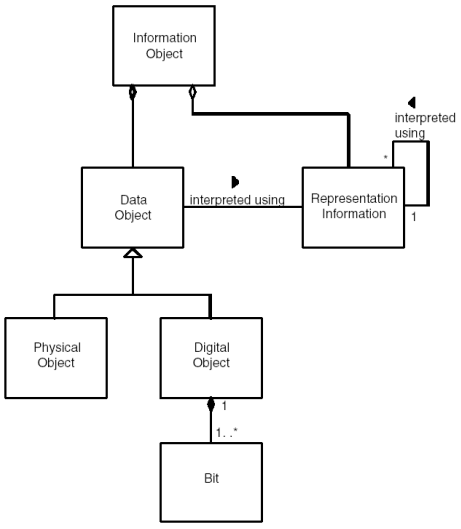


Figure 2: OAIS Information Model

From a conceptual modelling point of view, the OAIS **Information Model** distinguishes the **Data Object** from the more important **Information Object,** which plays a central role in the **CASPAR** approach, as the UML[24] class diagram in Figure 2 illustrates. This UML diagram shows that:

- An Information Object is made up of a Data Object and Representation Information.
- A Data Object can be either a Physical Object or a Digital Object. An example of the former is a piece of paper or a rock sample.
- A Digital Object is made up of one or more Bits.
- A Data Object is interpreted using

Representation Information.

- Representation Information is itself interpreted using further Representation Information.

The concept of Representation Information represents everything that is needed to produce an Information Object from the Data Object. This includes: the structure and semantics of the data object, for example from associated software to relevant standards, as well as any other information that may help in this respect (see Figure 3)

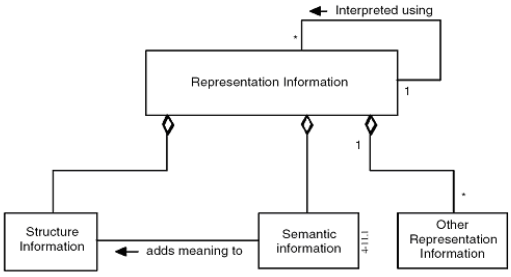


Figure 3: OAIS Representation Information

The Representation Information itself can be digital, and will itself have further Representation Information to allow it to be understood. This is the significance of the loop from Representation Information back to itself.

To collect the **complete** set of Representation Information, including the complete recursion is, almost certainly, an impossible task.

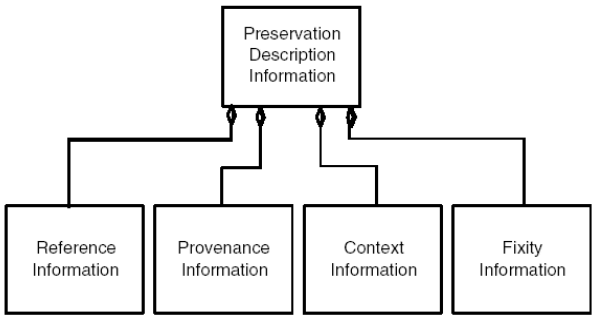


Figure 4 OAIS Preservation Description Information (PDI)

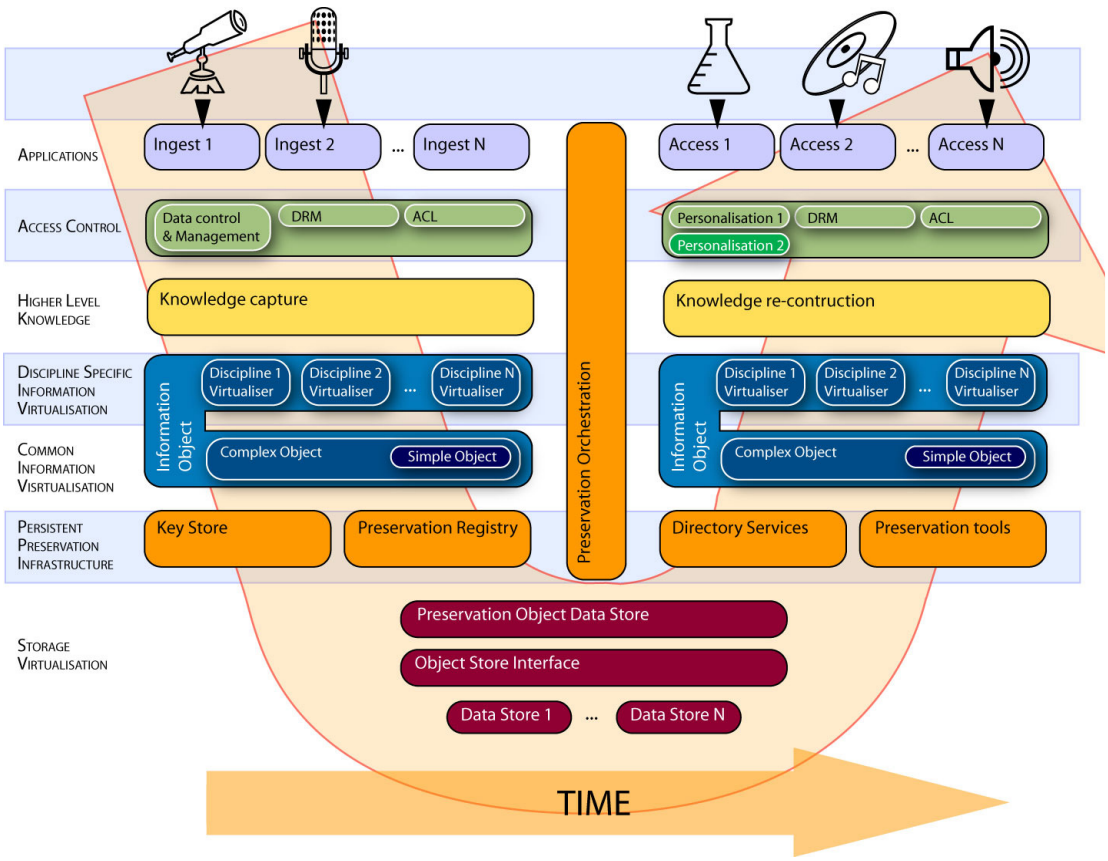
In order to make it a practical proposition for preservation, as opposed to simply contemporaneous use, the **Representation Information** required is linked to the **Knowledge Base** of the **Designated Community**. In other words, one must take account of the user community for whom the information is being preserved. Thus, as the user community changes or, for example, as that Community's normally available software changes, additional Representation Information must be supplied to ensure that the primary information content is understandable and usable.

Another important concept from OAIS is the Archival Information Package, which is a logical packaging together of all the information required for the long-term preservation of the primary information content. This (logically) contains not just the Representation Information but also the Preservation Description Information (PDI), which includes Fixity, Provenance, Context and Reference of the primary information (see Figure 4.)

Armed with these concepts and others from OAIS, the **CASPAR** consortium has performed an initial analysis and constructed the preservation workflow shown in Figure 5, and the important components described later.

These are the OAIS concepts that we wish to support – and there are many additional concepts arising from our initial analysis. What **CASPAR** will produce, as a deliverable, is software which can support these concepts in such as way that will encourage the sharing of effort over time and have itself built-in avoidance of obsolescence by making heavy use of virtualisation at all levels, using sophisticated knowledge management techniques.

Figure 5: CASPAR Workflow



The project identifies several major building, and the keystone to this edifice is the Information Object and its preservation.

CASPAR is built as a layered architecture that provides the foundation to ingest, preserve and access an Information Object. From the bottom up, the base is the **Storage Virtualisation** layer, which consists of the **Preservation Object Datastore** blocks; this is the foundation block for storing objects (both metadata objects and content data objects) in a persistent manner. Next in the stack is the **Persistent Preservation Infrastructure**; this layer contains all the necessary ancillary blocks (such as registries and directories) that are required in order to obtain the Representation Information of a given Information Object. The **Information Virtualisation** layer is responsible for creating (during Ingest) and using (during Access) a unified, standardised view of the Information Object to the application layer; the ability to present a unified Information Object (whether simple or complex) builds upon the layers below it. Next there is the layer dealing with the creation (during Ingest) and the usage (during Access) of the **Knowledge** required for understanding and using the object. The data **Access and Control** layer undertakes the operations for collecting the data to be preserved upon Ingestion. An important observation is that the Information Object is a virtual entity that is not physically maintained; in fact, it may be constructed on demand, based on the domain application that ingests or stores it. Representing an Information Object may involve access to many pieces of data, entries and software, which must be mitigated by data control and management mechanisms like access control (ACL), **Digital Rights Management** (DRM) and personalisation processes. Upon Access there may be **Personalisation**, aimed at capturing specific user preferences and using them to provide

a different experience for each different class of users. Finally, it is the **Preservation Orchestration** that provides a set of tools and software needed to control the lifecycle of data in this architecture.

The components to be developed are shown in Figure 6.

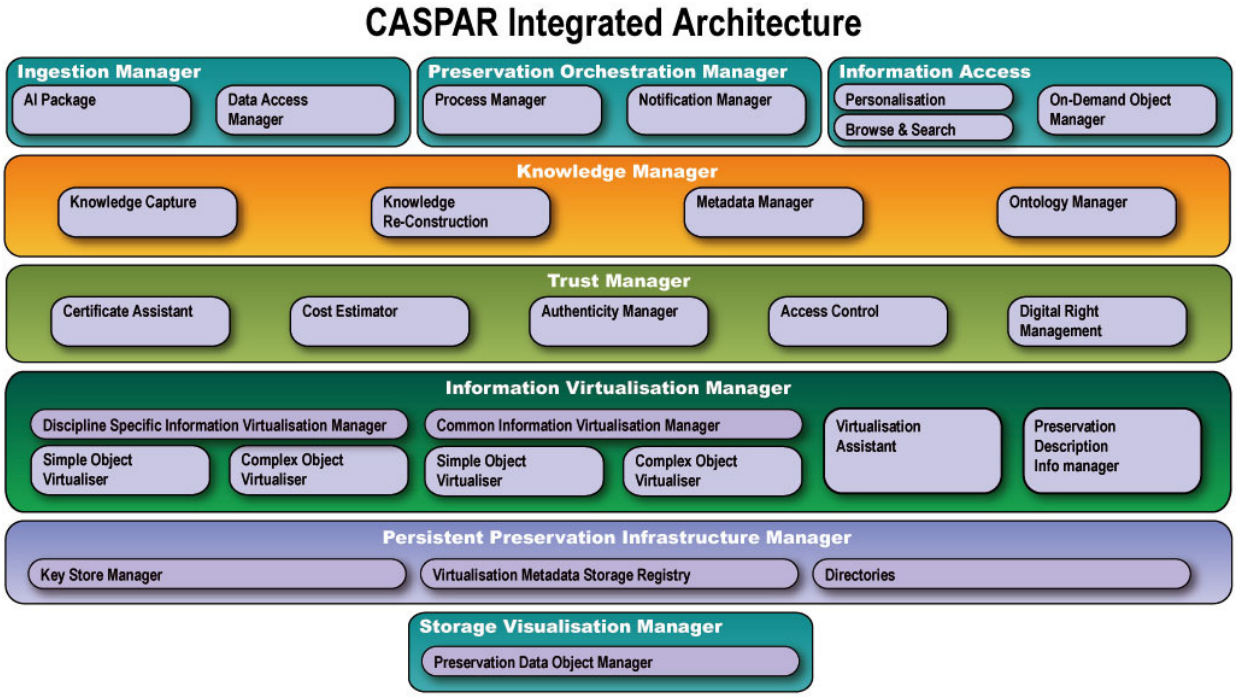


Figure 6: CASPAR Integrated Architecture

Research and Development approach

The first R&D activity of **CASPAR** is a broad analysis that lays the foundations for the rest of the project. The guiding principle of **CASPAR** is the application of the OAIS Reference Model, which is the only ISO standard addressing the long-term preservation of information. It is an abstract standard that defines fundamental concepts and approaches to information preservation.

The sequence of tasks is:

1. Review current methodologies used to preserve digital information; knowledge management, storage and preservation technologies that are being used and are coming to the market; the techniques that are being used for virtualisation of digital objects. This task, in close collaboration with the testbed requirements work, will gather use cases and user requirements, through extensive contacts with the user community, particularly those involved with setting up and running digital archives, as well as other research projects. It is worth, at this point listing some of the types of users which **CASPAR** will have immediate contact with:
    1) Data creators of all kinds
    2) Professionals dealing with funding decisions about digital preservation
    3) Operational people managing preservation activities
    4) Data owners, data holders and data curators

5) Preservation service providers

6) End-users of digital information.

2. Develop a specific **CASPAR** methodology, which provides the roadmap for the Project.

3. Review and refine the overall architecture, which guides how the various parts of the **CASPAR** Framework work together.

Overall, an **iterative approach** will be taken. For example the Methodology and Architecture are to be produced in a number of phases in order to allow them to be improved and refined through feedback from showing prototypes to **real users** and **real data holders** and holding discussions with other projects at the forefront of this field. In addition to the named deliverables, the project will produce a number of intermediate documents and studies, the majority of which will be available on the project Web site. It will however be constantly borne in mind that **CASPAR** is not simply developing software for the **here and now**, and for current users.

In many ways our **real users are not yet born** and the **target archive systems not yet conceived**; we will have to use novel and imaginative techniques to take these requirements into account, and in particular use knowledge management technologies to allow our infrastructure to be sufficiently adaptive.

The work will involve many ICT areas relevant to preservation including:

- Semantic Web: to put "knowledge" at the centre of preservation processes, to support the evolution of concepts over time
- Distributed systems: to support a distributed preservation infrastructure with no single point of failure
- Storage technologies: to leverage the latest ideas of virtualisation of storage – so that we are looking at digital objects not simply files
- Virtualisation techniques: for scientific, cultural and arts objects which can then be instantiated for future generations for their knowledge base and hardware and software environments

Some specific examples of areas we anticipate advancing the state of the art are:

- Tools and techniques for creating Representation Information for a range of generally useful complex objects, together with discipline specific examples.
- Digital Rights management which can be evolved with changes in recognised rights and legal frameworks.
- Future-proofed authenticity and provenance support.
- Brokering and notification services for information preservation, covering the OAIS Preservation Planning functional entity and much more.

The Framework is the software platform that enables the building of services and applications that can adapt to multiple areas of application and, in particular, to the three testbeds envisaged in **CASPAR** (see below). In addition to the functionalities developed in the Components-level R&D block, the Framework includes the integration/customisation of a number of basic but essential features to build the **CASPAR** testbeds, such as: security, authentication, accounting, authorisation,

monitoring, as well as the technologies to build systems that reflect a Service Oriented Architecture (SOA)[25] style.

In order to ensure the usability of the Framework, user input will be vital from several classes of users and examination of a wide variety of existing archive systems, both within and beyond the **CASPAR** partners.

Overall integration and functional testing will be performed at ESA/ESRIN.

Validation

Validation of what the consortium is doing and how it is being done will be an ongoing process supplemented by major, external, independent, reviews.

The "here and now" validation can be performed as for any other software system in terms of functional and performance tests. The practical implementations, embedded in a number of operational organisations in 3 different disciplines including ESA/ESRIN, CCLRC (science data), UNESCO (cultural heritage data), INA, IRCAM and CIANT (performing arts/multi-media data) will form the basis of the validation regime.

The more fundamental validation techniques on behalf of the "not yet born" are summarised in Table 1. These address the methodology and design which will be produced, and ways of standing in proxy of those future users. The latter will use the diversity to which the **CASPAR** consortium has access – internally and externally - to "mimic" the changes in hardware and software over time, and also to mimic the changes in the Designated Community for any particular dataset over time.

The University of Urbino and the European Space Agency (ESA) will play central roles within the consortium to monitor success, the former from the point of view of users, and the latter from the point of view of operation projects.

External monitoring will be undertaken by a small group of internationally recognised experts who are independent of the **CASPAR** consortium. The group will be funded by **CASPAR** – funds have been set aside under the University of Urbino, with Mariella Guercio – to ensure significant and long term engagement. The reports from this review group will be published by the project.

Additional external monitoring and assessment will come from the large number of users and user groups to which the **CASPAR** Consortium has access, including the large and diverse scientific users of ESA and CCLRC, the many distributed data holders connected with the Cultural Heritage sites linked via UNESCO, and the large and varied artists and performers linked to INA, IRCAM and CIANT. There will be many routes for such assessment, including user discussion groups/forums/focussed workshops, and users giving feedback on prototypes and giving their views via the Key Performance Indicator tool.

An additional metric is based on the developing audit and certification standards for trusted repositories, to assess improvements in preservation trustworthiness. This will be done in an incremental way to monitor improvements in trustworthiness of archives as the **CASPAR** methodology is applied.

Knowledge dissemination

The results and products from **CASPAR** will be widely disseminated and several of the key metrics address this.

The scientific knowledge arising from the R&D activities will be fully exposed to outside review through a policy of open access for documents and source code. In addition there will be formal peer review through reputable journals and conferences.

Embedding the testbeds in operational systems will support the longevity of those testbeds and also encourage the uptake of the **CASPAR** offerings. As previously mentioned, there is every reason to believe that **CASPAR** results will be incorporated into the European wide framework which will be created by the Task Force on Permanent Access Alliance, and the ***European Digital Information Infrastructure for Preservation and Access*** (EDIIPA) project which has been submitted to the ESFRI Roadmap process. These strategic European wide initiatives will be supplemented by a diffusion into individual organisations associated with the **CASPAR** consortium members.

By recognising at the start that **CASPAR** will not solve all problems, we adopt an open architecture which can encourage the mobilise providers of services and tools for digital preservation that can contribute to enrich the core functions of **CASPAR**

6.2 Demonstration activities

The Demonstration activities will be targeted to set up tools which will allow the widest audience to see and try out the tools and services which **CASPAR** is developing.

These demonstrations, including Web-based access to a number of components, should be simple yet at the same time powerful, based on interactive, appealing and very usable interfaces, and exploiting real, scientifically and culturally interesting contents.

Demonstrations of the three **CASPAR** testbeds will be provided, to show how the **CASPAR** technological framework can support the preservation methodologies and different needs. This kind of demonstration is intended to offer a wide spectrum of example scenarios.

The demonstration activities start in the second phase of the project, as soon as the required services are ready to be validated and submitted to the market sector.

6.3 Training activities

The **CASPAR** project will be delivering an approach to preservation of digital information based on OAIS. It will provide a significant number of components, both generic as well as innovative methods and technologies specific to cultural heritage, artistic and scientific communities. Training should be provided for this wide range of components and systems.

A caveat must be inserted at this point, namely that **CASPAR** funding will not cover the full Training activities described here. It is hoped that we can work with other Integrated Projects and Coordination Actions funded by the EU to deliver the full Training required.

CASPAR will approach the overall activity in four stages, of which three are concentrated in the first eighteen months and one of which begins during the first eighteen months and runs beyond the end of the project. The four phases are: definition of knowledge and skills development needs and opportunities, creation of a training

plan and identifying the educational methods that will be used, development of the training programmes, and delivery of the training programmes.

For each of the four core training domains, **CASPAR** will identify an approach to training that is responsive to the learning and teaching needs of the particular group we are training, is offered in the most appropriate training environment (e.g. lecture hall, seminar room, laboratory or workplace) and uses the most effective methods (large or small group teaching, one-to-one tutorials, self-directed learning, remote learning), and matches the level of detail and intensiveness to the ways in which the participants are likely to use the knowledge and skills acquired during the training.

The initial review of training needs and opportunities will enable **CASPAR** to establish where training should be offered and who will be encouraged to take up the training services. To optimise the training, **CASPAR** will then move on to define the training programme and the different educational methods that will be employed to deliver the training. For each training opportunity, specific courses will be defined and for each of these a number of key questions will be addressed.

A knowledge and skills development portfolio is envisaged that would enable participants to complete individual **CASPAR** training modules on an ad-hoc basis while others would achieve **CASPAR** Stars and, when they had collected a certain amount of training and abilities in **CASPAR** methods and technologies, they would be granted the status of **CASPAR** Certified Engineers.

A specific task will develop a tool for supporting the traditional face-to-face training. Therefore the **CASPAR** Training infrastructure will deploy a technological platform in order to support the training activities with value added services. This web-based platform will help trainers and trainees to use multimedia content and training aids; moreover, it will be a useful tool for post-project dissemination. In particular, it develops the following services:

The Lesson storage and Content management infrastructure will provide a web-based interface to access the overall services provided by the platform. The **CASPAR** digital preservation methodology will of course be applied.

The post-lesson support tool will provide an all-inclusive working space to afford interactivity between trainers in their working groups, and between trainers and trainees. This kind of communication service will allow the conducting of one-to-one (real-time or not) communications and real-time meetings, facilitating small or large group collaboration.

6.4 Management of the Consortium activities

The following sections list the main management activities. More detail is available in Section 7. A preliminary Work Package list covering the full duration of the project is shown in Table 2

Management of project strategy and direction

The overall responsibility for making sure that the project remains on course as intended by the Commission and the Consortium lies with the **IP Coordinator** (IPC). The IPC is responsible for maintaining consistency between the views of the Commission and the Consortium. Project management is further detailed in Section 7 and the Consortium Agreement. The **IP Coordinator** will be supported by the **IP Stream Directors** and the **IP Secretariat**, who together form the **Executive Project Management** (EPM).

Project Control and Co-ordination

The control of finances is managed by the IP Coordinator based on the **Consortium Agreement**, which also includes details of IPR management.

To monitor project progress, a concise **monthly report** from all parties involved in the work will be supplied to the IP Secretariat. Projections of possible delays are required in the latter in order to take pre-emptive measures. Work Package leaders and the IP Stream Directors will be contacted directly outside normal reporting in urgent cases. A more detailed **Quarterly Report** details work performed and progress towards all milestones and deliverables. In addition, a project report is compiled by the IP Stream Directors and approved by the IP Coordinator.

The **Executive Project Management** is responsible for operative project management at project level, Work Package Leaders at work package level. They operate using guidelines agreed by the IP Coordinator.

Scientific Relevance of Work

The project will appoint a **External Review Committee** (ERC) to ensure that the project addresses challenging issues well beyond the state of the art. Their recommendations shall be made known to the Commission, as shall the responses of the project.

Management of project impact

The EPM will decide on conferences, such as those of IEEE[26], PV[27], ECDL[28] etc or other events, as well as publications, as part of the dissemination work. For external awareness and dissemination, a project web site will provide information and demonstrations. After approval by the EPM, information will be made available via the web site. A set of tools for Demonstration purposes, and much of the Training material, will also be available on this site.

To facilitate the maximum exploitation of results, Business units of the Parties will be informed about the project and its status and the project will appoint an **Exploitation Advisory Committee** (EAC). The EAC will be recruited from the business units of Parties and advise the project on how to exploit the results of the project. Examples are exploitation by one of more Parties in new products or services, but also the generation of new business, e.g. as spin-offs from companies and universities, and joint ventures.

Table 2: Preliminary Workpackage lists (full duration 42 months)

WP	WP Name	Lead	mm	Start	End
1100	Review State of the Art	INA	47	1	4
1200	**CASPAR** methodology and models	UG	183	1	38
1300	Overall architecture	ENG	59	3	12
2100	OAIS-based components	CCLRC	144	1	30
2200	OAIS-based Storage	IBM	41	1	30
2300	OAIS-based access	CNRS	130	3	30
2400	Next generation components	CNR	62	25	38

3100	Framework Architecture	ACS	33	3	16
3200	Integration of existing component	ACS	86	9	27
3300	Framework technical testing	ESA	36	16	33
3400	Multi-industries long-term research perspectives	MW	36	25	38
4100	Generic testbed elements	ASMX	117	1	38
4200	Cultural data preservation testbed	UNESCO	72	7	33
4300	Contemporary art testbed	IRCAM	118	7	33
4400	Scientific data preservation testbed	ESA	86	7	30
5100	Exploitation and take-up	ENG	80	1	38
5200	Dissemination activities	MW	47.5	1	38
5300	Training	UU	56.5	6	38
5400	Demonstration activities	CCLRC	14	30	38
6100	Project management and coordination	CCLRC	44	1	38
6200	Project performance monitoring	CCLRC	28	1	38

6.B Plans

6.5 Plan for using and disseminating knowledge

The overall objective of the plan is to ensure the continued and wide use of the **CASPAR** results.

Users have been identified in the main following categories (bearing in mind there is some overlap in membership of these groups):

1) Data creators of all kinds

CASPAR Consortium members CCLRC and ESA are major data producers and, along with INA, IRCAM, CIANT and UNESCO, have close links with many other data producers. These will form the core of a natural clientele, which should grow as success is demonstrated.

We believe that, supported by the OECD declaration[29] and guidelines on access to research data from public funding, research data producers and others are increasingly concerned about the preservation of their data in a cost-effective way. **CASPAR** will work, initially with our natural clientele, emphasising the project's focus on international standards, the open framework which facilitates sharing of effort, tools which should be relatively easy to integrate into existing systems, together with the results of the research on cost models.

2) Professionals dealing with funding decisions about digital preservation

Funders need to have some reassurance that they are making the best decisions – in particular the choice of preservation service providers. The work on Certification and

Cost Models will be the areas **CASPAR** will focus during contacts with this group – using personal contacts, publications, conferences and specific focus groups.

3) Operational people managing preservation activities

These implementers of preservation systems will be most interested initially in the virtualisation of storage and other hardware and software, which will facilitate the migration between such systems. A specific focus group will bring such individuals together, but in addition there are specific workshops and conferences which will be targeted.

4) Data owners, data holders and data curators

Contacts with this group of users have shown that there is a demand for a "Roadmap for preservation". **CASPAR** will work with this group showing them a coherent, validated, broadly applicable framework for preservation. However this is not to assume that this will be an "easy sell". We anticipate that we will have to work over an extended period with this group, involving them throughout the development of the methodology, framework and tools.

Digital Rights Management is likely to be a topic which will be of interest to those commercial data owners such as artistic performers.

5) Preservation service providers

Certification services and components such as the Certification Assistant will be of initial interest to this group. However use of the shared infrastructure to which **CASPAR** would naturally lead will be applicable even within a single distributed organisation.

6) End-users of digital information.

Here we include the general public, discipline specific groups and those who produce end user applications. **CASPAR** should also consider information and dissemination targeted at the general public insofar as, for example, music is more and more elaborated and performed by non-professional and independent artists accessing new technologies. **CASPAR** will propose web-based structures to reach people outside the usual networks and institutions in order to provide them with solutions, methods and services drawn from the main **CASPAR** results.

The project plans two different phases for these activities:

- 1st phase (first 18 months): defining of dissemination and communication strategies (dissemination planning and implementation), aggregation of a critical mass;
- 2nd phase (second 24 months): continuous process of critical mass aggregation, progressive involvement of the **CASPAR** Preservation User Community in the validation activities.

The plan of using and disseminating the knowledge will be supported by a specific dissemination approach that will be personalised according to the target users, and for this different dissemination tools will be produced.

International organisations and projects

We plan active collaboration and teaming with relevant digital preservation initiatives outside the EU, and with National and International projects such as Chronopolis, INTERPARES and NARA to bring a global dimension to our work.

The decision to use OAIS as a basis for **CASPAR** and to involve other communities, such as the GRID community, to develop specific strategies and methodologies is in itself an open door for professional communities which, until now, have not been specifically concerned with digital preservation. The consequences of this effort will directly involve the educational sector in many areas including programmes dedicated to the digital preservation in the custodial communities (museums, archives and libraries), and also technological and scientific environments.

In addition we will be taking part in high profile events in the science, cultural heritage and performing arts domains, reflecting the natural constituencies of the **CASPAR** consortium members.

A number of International organisations and projects signed letters of support, and intend to work with **CASPAR**. These were:

- NARA, ERA project (Head of Research)
- NASA/Science Office of Standards and Technology (NOST)/National Space Science Data Center
- US National Virtual Observatory/CalTech
- StorageTEK
- UK Digital Curation Centre
- Library of Chinese Academy of Sciences
- InterPARES project
- San Diego SuperComputer Center/Chronopolis Project
- ECLAIR Laboratoire
- Centre Virtuel de la connaissance sur l'Europe
- Luigi Sturzo Institute
- Mediateca toscana - Pisa section
- European Business Register
- U.S. Geological Survey
- BRICKS Project
- Australian Partnership for Sustainable Repositories

Exploitation of results

The objectives of the exploitation activities are threefold:

- Aggregate and analyse the **CASPAR** target community (e.g. digital repositories and any other participants involved in digital preservation)
- Define and develop an appropriate cost/benefits model complete with specific metrics for costs of digital preservation
- Define a business model and a strategic positioning to produce an effective sustainability plan, including a socio-economic impact analysis.

The general exploitation approach is based on three general assumptions and guidelines:

- to be **user oriented**
- to be the **reference model** for digital preservation in Europe
- to use an **open and replicable approach**.

All of these guidelines will be used to define the user scenario requirements, the market analysis and the standard overview, after which the results will be included in the success indicators of the project and in the definition of the first sustainability model.

To be user oriented

Only the continuing and effective participation of users throughout the entire production cycle will guarantee that the system meets all the requirements for which it is designed and that it complies with the needs of each and every type of user – including proxies for those not yet born.

In line with this assumption, all three different **CASPAR** scenarios are defined with a real user involvement. The three scenarios have been chosen in order to offer a range of different user types and to be good examples for future scenarios, bearing in mind the business plan.

To be the reference model

The ambition of the **CASPAR** project is to define **the** methodology and infrastructure for digital preservation in Europe, and all exploitation activities will be conducted in accordance with this main goal.

To use an open and replicable approach

CASPAR will not be a closed system; the aim will be to interoperate with as many different systems as possible and to be operated or re-implemented by others outside the **CASPAR** project.

Functionally open**,** since the project allows for an evolution of the system with the passing of time related to the evolution of the regulations, standards and organisation of the digital preservation activities. It will be able to observe its connection with outside systems, if any, sharing information stored in both central and local structures.

Technologically open insofar as the development languages and environments are concerned. The latter will be identified in line with advantages and opportunities and may also be categorised by areas in order to guarantee both development and maintenance efficiency and full consistency with the architectural reference models.

Branding and consensus building

Exploitation activities include the task of defining and managing a proper communication strategy to validate **CASPAR** results. The main objective of the Communications activity is to create and manage the **CASPAR** "Brand" which will be fundamental to its gaining acceptance.

Brand management requires a strong commitment to communicating a core message to all targets over the long term. In the case of **CASPAR**, this means raising public awareness, attracting new signatories to User Groups, obtaining support from public and private entities, and recognition of **CASPAR** as the sector reference.

Exploitation activities schema

In order to reach the exploitation goal and to produce the results planned in this part of the project, the following activities schema will be implemented.

Business Concepts are evaluated internally and also by the Exploitation Advisory Committee. In developing these concepts we will use, and in turn revise, the Preservation Cost Model and the Exploitation Plan. The latter in turn will interact with the Communication Plan and our Socio-Economic analysis.

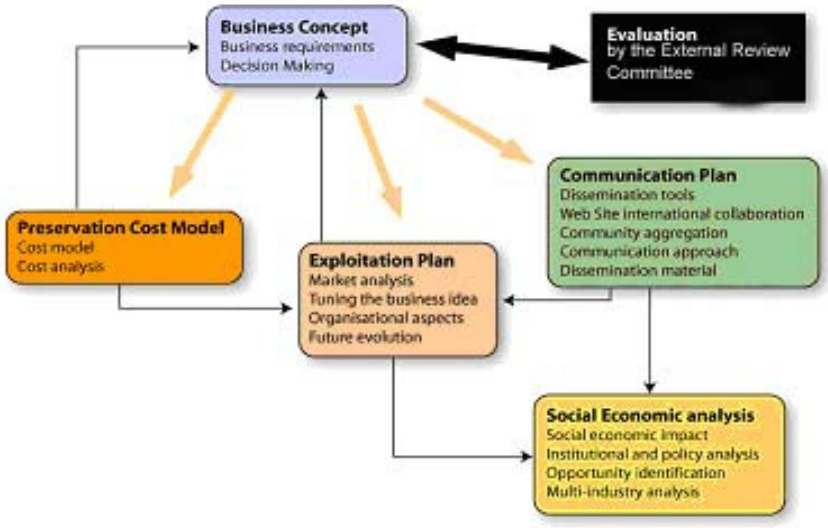


Figure 7 Exploitation schema

6.6 Gender Action Plan

Throughout **CASPAR**, gender issues will be seen as part of the more general issue of equal opportunities which also covers race, ethnicity, national origin, sexual orientation, religion, age, veteran status and disabling condition. All results coming from **CASPAR** will be multi-cultural, multi-lingual and gender non-specific and we will ensure that the interests and needs of both genders are taken into account, and the dissemination is made to both groups.

Employment on and participation in the project will aim to involve both genders without bias. This means using equal opportunities in the selection of staff; while we cannot mandate practices in partner organisations, the project will act as an advocate for equal opportunities within the partners.

A project gender action plan will be included in the Project Plan. The action plan will include measures purposed to encourage the involvement of women at all stages of research and development such as:

- disseminating best practise across the projects in the area of gender equality, and distributing the results of the evaluation of the action plans to the consortium;
- making links with networks of female scientists in the field of the project via organisations such as:
    - the British Computer Society Women's Group,
    - the Scottish Hopper Colloquium for Women in Computer Science,
    - the Femconsult Database (see http://www.cews.uni-bonn.de),

- linking with European researchers in Gender Equality via the Research in Gender Equality in Higher Education Conferences and the D.G. Research for Women and Science of the European Commission.

Once the plan is finalised, the IP Coordinator will monitor any project activities and ensure that the interests and needs of both genders are taken into account in accordance with the plan defined. Gender issues will be discussed in the General Assembly meetings, with a report by the IP Coordinator on the state of the application of the plan. The General Assembly will promote new actions, will suggest revisions of the policy, and will evaluate reported discrimination facts and related issues.

A gender action report will be delivered at the end of each phase of the project.

6.7 Raising public participation and awareness

Even more so than other areas, **CASPAR** has to be a team player, co-ordinated with a number of other projects if we are to seriously raise public awareness.

At the European level we anticipate working closely with Co-ordination Actions and other EU projects to present a steady number of press releases at European level and within individual states. **CASPAR** can provide raw material for these communications, across a wide range of disciplines and interests. Music and performing arts and cultural heritage should be particularly useful sources of interesting examples.

At the National level, the **CASPAR** consortium has organisations from 7 countries, and several members are significant memory institutions. We are therefore in a good position to emphasise more local concerns and local interest, using press releases and participation in relevant local events. Cross fertilisation of ideas should also be productive. For example in the UK the Digital Preservation Coalition (DPC) has been particularly effective in its Public Relations efforts, with high profile launches, preservation awards and press releases. **CASPAR** will advocate these techniques more widely.

UNESCO has an even broader reach, and an international profile. The links to globally recognised sites such as Machu Picchu, Venice and the Bamiyan Buddhas, should provide **CASPAR** with material which can catch the public's attention.

In addition to these broad public interest issues, as noted in Section 5.3, public awareness of how these issues are of direct importance to each individual needs to be raised. For example personal and family data such as digital photographs and other digital recordings, have very different preservation demands to, for example, the paper copies of photographs which can be used after sitting on a shelf untouched for decades.

6.C Milestones

6.8 Major Milestones over full project duration

Milestone no.	Milestone date (month)	Milestone title and goal	Key WorkPackages
M#1	m03	**Kick Off of CASPAR** • Consortium agreement completed	WP6100, WP6200
M#2	End m06	**CASPAR preservation requirement review** • State of the art completed and reviewed • Testbed Use Cases completed and requirements derived • Management tools in place	WP1100, WP4100, WP6100
M#3	End m12	**Review of CASPAR methodology and Updated Management Plan** • **CASPAR** guideline and conceptual model completed • Draft Framework Architecture completed • Simple object virtualisation model completed • Testbed design completed • Validation/Evaluation methodology review based on ERC report • Mandatory reports	WP1200, WP4100, WP4200, WP4300, WP4400, WP6100
M#4	End m15	**Review of Architecture and initial Exploitation & Sustainability plan** • Architecture Review • Component model specification review	WP1300, WP5100
M#5	End m18	**Review of CASPAR designs** • Component Model and overall component architecture design completed • Framework Architecture completed • OASIS-based components, storage and access analysis started • Integration of existing component started • Testbed specific R&D activities started • Training and dissemination plan completed • Review of revised evaluation criteria	WP1300, WP2100, WP2200, WP2300, WP3100, WP3200, WP4200, WP4300, WP4400, WP5x00
M#6	End m24	**Review of prototype implementations** • Complex object virtualisation completed • Mandatory Reports	WP1200, WP3100, WP3200, WP6100

M#7	End m30	**Review of CASPAR Framework** • OAIS components completed • DRM implementation completed • Data Store work completed • Testbed implementation reports	WP2100, WP2300, WP2200, WP4100, WP4200 WP4300, WP4400
M#8	End m36	**Review of Validation** • Draft Testbed reports completed • Report on Training results • Demonstration system design and initial results • Mandatory reports	WP4100, WP5200, WP5300, WP6100
M#9	End m42	**Final reports**	WP6100

7. Project management

The management of complex constructs like an Integrated Project requires very efficient and well-structured project organisation. Of particular importance are the distribution of responsibilities and the flow of information, both for controlling and reporting. A clear conflict-management process is required to ensure fast and acceptable conflict resolution, while reducing the risks of escalations of disputes. A thorough assessment and analysis of potential risks is also important to prepare remedial actions if required.

This section provides an overview of the project management with emphasis on the relationship with the Commission and critical areas of responsibility, summarised in Figure 8; more details will be found in the Consortium Agreement.

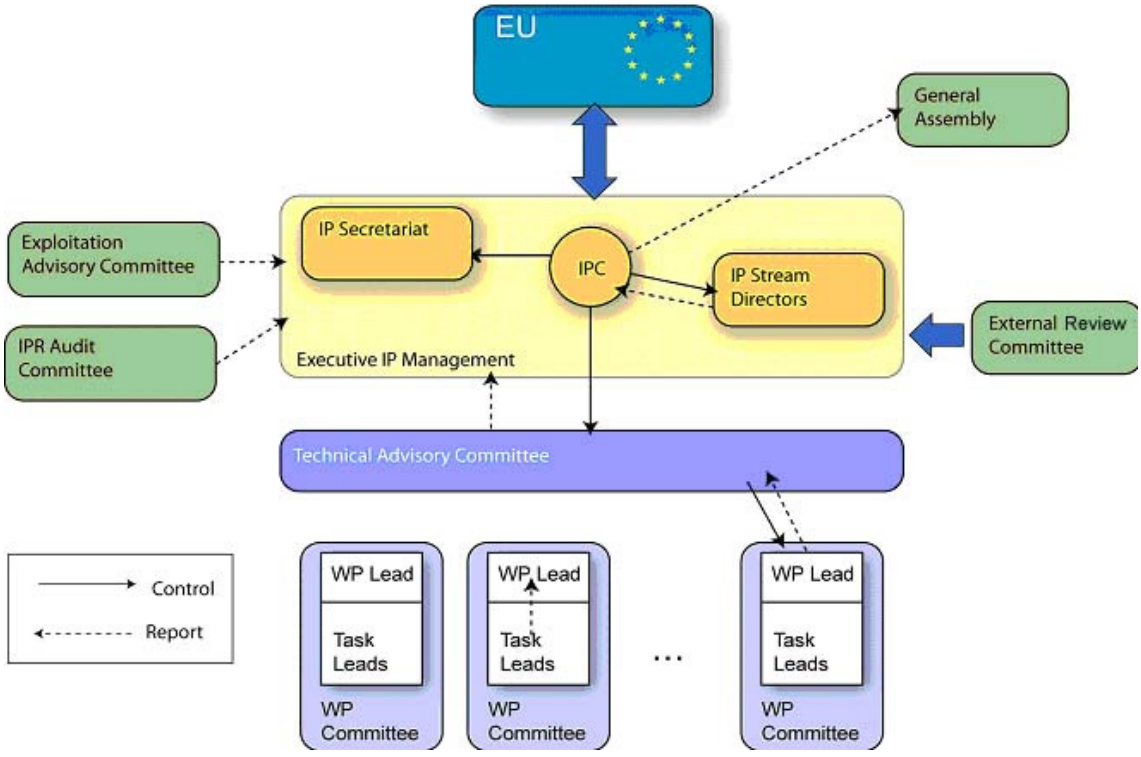


Figure 8: IP Management

Project Management Bodies

IP COORDINATOR (IPC)

The IPC is the intermediary between the Commission and the Consortium. The IPC is responsible for IP project control, for administrative and financial reporting, project time control, coordination issues and overall IP activities (e.g. training, dissemination) and is also responsible for the overall technical management and technical coordination within and between work packages. The IPC is the direct contact point to the Work Package Leaders (WPLs). The WPLs send all technical progress reports to the IPC. The

IPC is also responsible for the correct application of all EU rules, particularly concerning the handling of payments and maintenance of accounts.

The IPC responsibilities include:

- Overall responsibility for the success and smooth running of the project.
- Coordination of all activities and detection and correction of deviations from the plans. The EPM will be involved if necessary.
- Convening of the GA, preparation and follow-up of GA meetings.
- Drafting of a Quality Management Plan. The quality management plan will include procedures for
    1. Document issue and change control
    2. Software issue and change control
    3. Reporting and communication
    4. Corrective actions
    5. Tracking of action items
    6. Conflict resolution
- Monitoring the progress of the project and use of effort.
- Approves deliverables.
- Keeping partners informed about project progress.
- Reporting to the Commission and serving as the administrative liaison with the Commission.

The IPC will be David Giaretta (CCLRC).

IP SECRETARIAT

The IP Secretariat provides secretarial, administrative, financial and legal support to the IPC. The IP Secretariat will be a permanent contact person who also supports project participants, WPLs and EAC members. At least one person should be permanently available. The IP Secretariat will be based in the **CASPAR** Project Office at CCLRC. The IP Secretariat serves as project secretary and archive.

IP Stream Directors

The IP Stream Directors consist of a senior team initially named as Seamus Ross (University of Glasgow: Methodology), Luigi Fusco (ESA: Testbeds), Silvia Boi (MetaWare: Innovation), Carlo Meghini (CNR: Components level research), Ugo Di Giammatteo (ACS: Framework). The membership will be reviewed by the General Assembly annually.

The Stream Directors will have a supervisory brief over all the Work packages in his/her stream in order to ensure coherence and consistency between work packages and cross-fertilise ideas between streams, work packages and tasks.

EXECUTIVE IP MANAGEMENT (EPM)

The EPM is responsible for the day-by-day project management of the entire IP. It consists of the following entities:

- IP Coordinator (IPC)
- IP Secretariat
- IP Stream Directors
- others invited by the IPC for the efficient and balanced running of the project

The EPM will define and periodically update the technical project management plan, and will clearly maintain the WP interfaces and dependencies. The EPM will also be involved in resolving conflicts and attending all meetings and events where they are needed. The EPM should consult the GA on important issues and aim at achieving consensus, but the GA cannot overrule decisions of the EPM on day-to-day issues.

The EPM will organise all overall project reports and metrics, ensure smooth operation and timely financial transactions and organise audits by the Commission.

GENERAL ASSEMBLY (GA)

All partners of the Consortium are represented in the GA. The GA takes final decisions on the Consortium Agreement. The IP Coordinator will keep the GA informed about progress and achievements.

The GA shall solve conflicts within the Consortium that cannot be resolved by the EPM or the IPC.

The GA meets regularly twice a year, and if more than 30% of the GA members require it. It can take decisions by correspondence. The IPC will chair the GA.

EXTERNAL REVIEW COMMITTEE (ERC)

The **External Review Committee** (ERC) advises the EPM on its scientific direction. It reviews the progress made on a yearly basis and gives advice on the scientific aspects of the IP (e.g. new academic or technological achievements **CASPAR** should consider, new important trends, new societal developments the project should take into account, etc.). The ERC is also invited to propose ideas on generating new business and exploiting project results further.

High-profile industry and/or academia representatives of related areas constitute the ERC. The IP Coordinator, with the agreement of the Commission, selects and nominates the members of the ERC. The IPC shall chair the ERC. The IP Stream Directors shall be represented at the meetings, and the Commission IP Officer may also be represented. Suggestions from the ERC may be forwarded to the Commission.

In addition to the annual meetings, members of the ERC will

- perform the independent external audit of the effectiveness **CASPAR** in terms of digital preservation, including
- evaluation of the methods of validation of the **CASPAR** methodology
- evaluation of the effectiveness of the Testbeds.

It is proposed that specific funding should be set aside for these activities to help to ensure a high level of involvement of the ERC. The reports from the ERC will be made available to the Commission.

EXPLOITATION ADVISORY COMMITTEE (EAC)

The **Exploitation Advisory Committee** (EAC) shall be composed of high-profile business managers recruited from the Consortium partners. It is expected that about three experts should come from large corporations and SMEs. The EAC shall develop concrete proposals on how new business may be generated and how exploitation should be organised from the project results. Any GA member can propose EAC members. The EPM decides on the composition of the EAC. The IPC will chair the committee.

Management of Intellectual Property

Primarily, the project aims to enrich and expand the state of the art and state of practice of reliable software design and deployment methods in Europe. In order to carry out the work, the partners will develop and share know-how and technologies in many forms, including, but not limited to, algorithms, tools, experiences and methodologies. The know-how exchanged between the partners may include, in certain cases, pre-existing knowledge.

The Consortium Agreement will specify how IPR will be managed. The IPR Audit Committee (IAC) will monitor the management of IPR; it will meet at least once a year or upon request by the GA, to ensure that IPR issues are addressed. The IPR Audit Committee gives recommendations to the IPC on the handling of the assessed IPR issues. The IPR Audit Committee shall consist of not more than three experts with IPR background recruited from the Consortium partners.

Information Flow

The communication strategy shall guarantee that at any given time within the lifetime of the project the necessary information is provided in a timely manner and adequate for the different internal and external target groups. The communication objectives are the following:

- Information is provided at the right time and form to the right target.
- The people involved are aware of deadlines and understand their tasks/duties.

To implement the above, the IPC will use a set of communication tools, including:

- **Meetings:** Specific meetings for different purposes will take place. These include physical meetings, web meetings and audio-conferences.
- **Distribution lists:** Several distribution lists will be set up in order to liase with all the relevant individuals working for the project. They will be accessible from the intranet.
- **Project Internal Web Portal:** The IP Secretariat will set up a secure Web Portal for exchange of sensitive information within the project as well as with the Commission. This will include a Wiki for the rapid publication of news items, agenda, minutes and internal reports. Progress monitoring reports and reports to the Commission will be placed here unless they should be publicly available, in which case they will be on the Pubic Web Portal.
- **Project Public Web Portal:** The public portal (for which the name **http://www.casparpreserves.info** has been obtained) will be a showcase for **CASPAR** and its achievements, including all web-based applications that are produced. It will give access to the project's public documents. There will also be a Wiki that will allow individuals outside **CASPAR** to contribute ideas and suggestions.

Financial Management

In the Consortium Agreement the conditions on payment procedures and resource re-allocation, in case it becomes necessary, are laid out. The underlying principle will be balance on the one hand fairness and the ability to undertake capacity planning by

partners with, on the other hand, the ability to manage the risk to the Consortium and the Commission that work may not be completed satisfactorily. Payments will only be continued if the work committed to has been performed. If a partner is not able to deliver the work, the IP Coordinator may re-allocate the assigned tasks to other partners who offer to complete the work. This procedure ensures a high degree of security and flexibility, and will help to ensure that work will be delivered according to plan even if a partner may fail to perform the assigned work.

Measuring Project Performance

In **CASPAR** will measure performance relative to expectations, measure performance over time, and use those measures to provide feedback and take action using a Project Performance Assessment (**PPA**) process.

The PPA process will continuously monitor a set of Key Performance Indicators (KPIs) during the project development. The identification and organisation of KPIs will be carried on in WP 6, Task 6201, while Task 6203 will perform the actual PPA. Figure 9 depicts the preliminary organisation of KPI. A ***Balanced Scorecard***[30] performance methodology will be adopted to gather and evaluate KPIs measurement in different perspectives:

- innovation: asset value of R&D processes and technology
- operational: assessment of value creation
- strategic: assessment of market potentials and integration with business within multi-industry perspective

A specific **Project Performance Assessment tool** will support the continuous monitoring of the project performance. The tool will be accessed through the reserved area of the **CASPAR** Web-site, and will be able to collect KPIs measures from different actors of the overall **CASPAR** community (R&D, system integrators, users and potential customers etc.) through online questionnaires.

The PPA tool will support the testbed Testing activities planned in WP 4100 and will provide the basic material for the Validation reports.

We will clearly identify which KPIs are generally applicable to any digital preservation project, and which are specific to **CASPAR**, in order to facilitate the use of this tool by other projects.

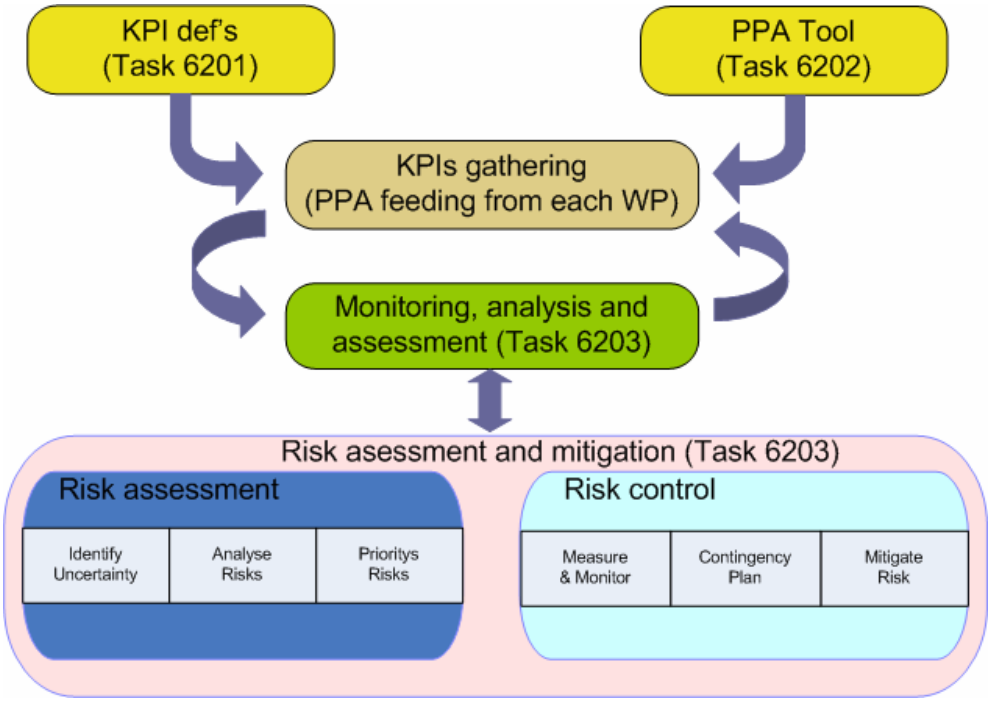


Figure 9 CASPAR overall PPA process

Table 3 Provisional Key Performance Indicators

Provisional organisation of Key Performance Indicators		
Level	Type	Indicators (examples)
INNOVATION (asset value of technology and R&D processes)	Scientific	Consistency with OAIS Capability of OAIS extensions to adaptively respond to new preservation requirements Level of virtualisation of digital information Capability for single components to feature innovative functions at different levels: Storage and Information Virtualisation, preservation orchestration, certification, authenticity, rights management…. Innovation in access mechanisms Contributions to standardisation bodies Adaptability to new domains and applications;

Provisional organisation of Key Performance Indicators		
	Technical	Coherence (e.g., knowledge capture & reconstruction, DRM etc.) Number and types of digital information virtualised Performance of services (e.g., Object Storing, Complex Object Mngnt...) Technical quality of Services and Applications Robustness of Services and Applications Achievement of infrastructure requirements Level of interoperability Methodology and quality of technical documentation
	Visibility	Papers, awards, citations Outreach through public bodies and networks; Impact of studies.
OPERATIONAL (assessment of value creation)	External	Consensus on standard, methodology and/or ontologies Availability/ acceptance of regulatory frameworks for preservation
	Validation	Ability to set sizeable user-groups Ability to set quality of user-group Number of trained researchers and users; Coverage of user's required function Usability Accessibility Quality & quantity of information sources Performance with and without semantic technology Ease of use Learning curve
	Demonstration	Level of use of Demonstration Web site List of potential data providers that have been contacted and signed SLA for demonstration. SLA signed with organisations to extend commercial partnership beyond demonstration completion

Provisional organisation of Key Performance Indicators		
STRATEGIC (portfolio assessment and integration with business)	Financial perspective	Forecast of return on investment Forecast of economic value added growth in net result and various cost ratios (preservation costs etc.)
	Market Perspective	Customer loyalty , acquisition, satisfaction , profitability Market share forecast Competitiveness level
	Internal Perspective	Effectiveness and efficacy of internal processes / organisations supporting **CASPAR**. Efficacy of new services. Expected Time to market of intermediate and final product Level of training of employees Level of reuse of technologies produced

Risk Management

As a technologically-challenging project, controlling risk within the **CASPAR** project is recognised by the project as an important task and strong controls will be applied. During the quarterly meetings, the EPM will hold a dedicated session to identify, evaluate and track project risks. For this purpose, the EPM will adopt the strategy described in the following.

Risks, generically, will be independently analysed across 3 dimensions – **costs**, **time**, and **quality of results**. The partners, through the responsible WP leaders, will strive to increase transparency in regard to dependencies and exposures (activities, time, resources, and cost). A central risk register will be created and managed by the IP Co-ordinator, and regularly reviewed by the EPM. Contingency plans and/or plans for corrective actions will be developed and implemented for all identified risks.

Two processes, the **Risk Assessment** and **Risk Control** are integrated parts of the overall **CASPAR** PPA as anticipated in Figure 9.

- Risk Assessment will take place at any time during the project and will allow project managers:
    a. To explore the entire project plans and look for areas of uncertainty.
    b. To specify how those areas of uncertainty can impact the performance of the project, either in duration, cost or meeting the users' requirements.
    c. To establish which risks should be eliminated completely, because of potential extreme impact, which should have regular management attention, and which are sufficiently minor to avoid detailed management attention (risk prioritising)
- The Risk Control has three tasks, as follows:

a. Take whatever actions are possible in advance to reduce the effect of risk. It is better to spend money on mitigation than to include contingency in the plan..

b. For all those risks which are deemed to be significant, have a contingency plan in place before it happens.

c. Track the effects of the risks identified and manage them to a successful conclusion.

As part of the PPA methodology, two reports will be filled online through the PPA tool:

- A **Risk Assessment Report**, to manage a proper risk evaluation and identification

Risk assessment report				
		Probability of risk		
Risk number	Risk description	Event probability	Risk Impact	Risk Factors
		A (1-3)	B (1-3)	AxB (1-9)

A **Risk Action Report**, to manage the proper risk mitigation action plan

Risk action report					
		Risk coverage			
Risk number	Risk description	Technical risk	Management risk	Financial risk	**Action plan**

Task 6203 embeds the risk management activities. Monitored Risks in **CASPAR** belong to the following categories:

- Management Risk
- Technological Risk
- Functional & Usability Risk
- Market Risk

The main risks identified at this stage in the project, for the whole duration of the project, are shown below; this list will be reviewed, and updated if necessary, by the EPM at the quarterly meetings, and will also be developed in parallel with the Quality Management Plan. The shaded ones are those which might be expected to be most important in the first 12 months.

Management Risks

Risk description	Evaluation	Resolution
Consortium is too numerous to be easily co-ordinated.	**Impact High, Prob. High** Should this problem occur its impact on the project would be significant. However the probability of occurrence is low due to the background of	The designated project manager has great experience in co-ordinating distributed research projects. Nevertheless the project management organisation is defined to be as representative as possible, maintaining

	experience in such projects by almost all the partners. In particular CCLRC, ESA, Engineering and MW have extensive experience in past research project co-ordination.	the correct efficiency and operability (see relevant chapter)
Consortium has no harmony.	**Impact High, Prob. Low** There are many reasons to believe that harmony will be the core of consortium, ranging from personal friendships to company alliances and recent experiences between partners.	Previous experience within single partners has been very positive and should be maintained at consortium level. Whole project meetings and workshops will be held to ensure that good communications are established between partners.
Too many diverging objectives between technical objectives (horizontal) and testbed oriented objectives (vertical).	**Impact High, Prob. Medium**. The **CASPAR** management structure has been specifically designed to minimise this risk and to ensure the proper collaboration between user and SW developers.	One of the key aims of the Innovation stream will be to avoid and manage this risk. This will stress the importance of the testbeds and potential users.

Technological Risks

Risk description	Evaluation	Resolution
Lack of Consensus on methodology, standards & technological approach	**Impact High, Probability Low** Lack of confidence is the danger against which many project efforts are focused, and it can become a serious threat if the demonstrator fails to prove its capacity to attract a large number of participants toward the core formed by the project partners.	The success of the **CASPAR** business model will be proportional to the dimension of the **CASPAR** community in terms of number of actors and acceptance. For this reason the process to organise and put in place the community will be started immediately and many efforts will be dedicated to the dissemination of results and to the aggregation of members. In matter of facts even before the start of the projects, more of 20 members have declared the interest in participating to **CASPAR** Community.
Technical problems arise during component development	**Impact Medium, Prob. Low**. Components could be developed independently by the various partners, without a detailed discussion about their functionality according to the specification document.	The design phase of the project will be led by Engineering , who have significant experience in system design. The framework architecture is designed in the light of maximum flexibility to simplify the integration of all technologies available.
The complexity of integration compromises system performance.	**Impact High, Prob. Low**. The computational complexity that eventually arises from the architecture may be a problem. The implementation of the core technology needs to be effective to give real time response to users.	The experience of CCLRC, ESA, CNR, Engineering, MW and all the academic partners is essential to solve this problem.

Market Risk

Risk description	Evaluation	Resolution
Lack of consensus on standard and methodologies.	**Impact High, Probability Low** While the potential of standards (e.g.,OAIS) is considerable, in practice the degree to which it will be able to drive a homogeneous preservation framework will depend on European stakeholders agreeing upon these standards. The main problem we	The issue is how **CASPAR** should contribute to prevent a proliferation of preservation standards. This will require specific, support to the (technical, regulatory) adoption of **CASPAR** adopted methodologies and standards.

	foresee here is that each 'standard' schema is unique and particular to a reporting application, and is 'owned' by a specific user community.	

Functional & Usability Risks

Risk description	Evaluation	Resolution
User requirements are not captured adequately	**Impact High, Probability Medium** Feedback by users from forums and prototypes could show mismatch to user requirements	Involvement of users in all stages of research and development, combined with workshops and iterative development of software should allow mismatches to be detected and corrected early.
User's functions do not respond to usability requirements	**Impact High, Probability Low** The Validation processes indicate that usability parameter do not respond satisfactorily to planned values,	A loop-back on the development activity has to be issued and responsibility for lack of usability has to be identified. Possible budget reallocations have to be considered.
Components' functionalities do not fulfil preservation application requirements	**Impact High, Probability Low** The testbed specifications were not delivered with adequate details or the component developments were not issued with a suitable methodology	A loop-back on the development activity has to be issued and responsibility for the missing functionality has to be identified. Possible budget reallocations have to be considered

8. Detailed implementation plan – first 18 months

8.1 Introduction - general description and milestones

The work-plan activities are organised hierarchically in three levels, namely:

- ✓ Stream (First digit)
- ✓ Work Package (Second digit)
- ✓ Tasks (Third and Fourth digits)

In the following diagram (Figure 21) the task level is not considered.

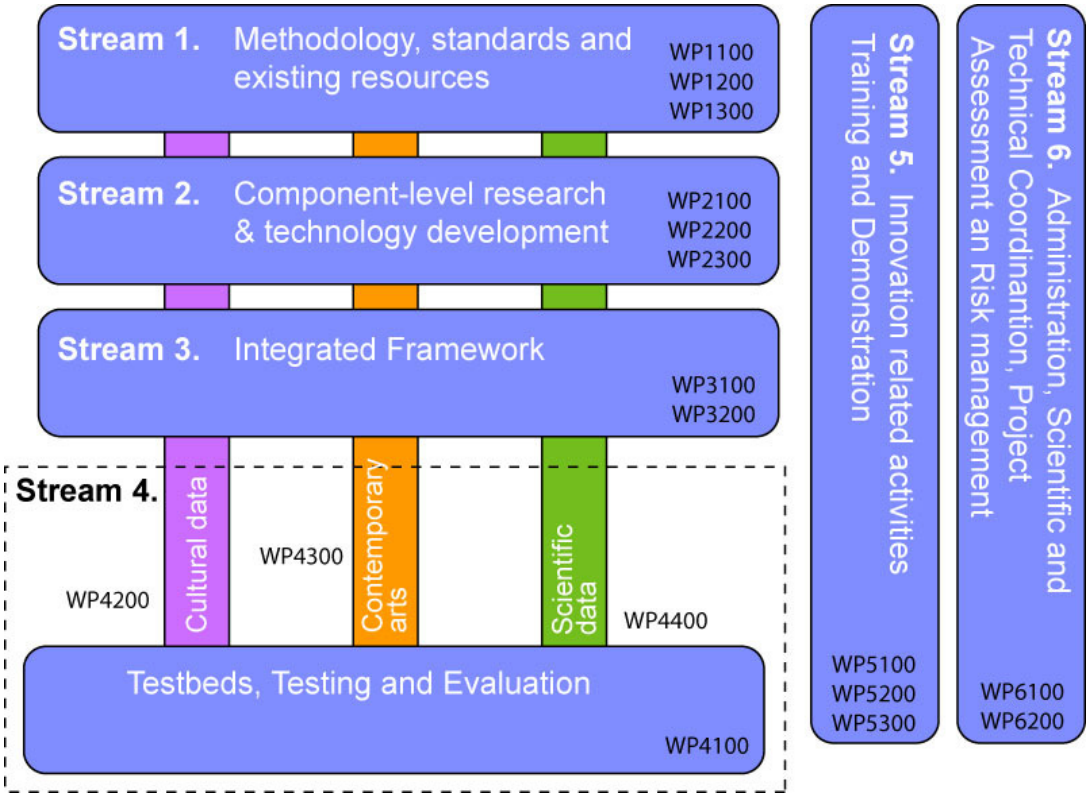


Figure 10 CASPAR Implementation plan structure

Stream 1: Methodology, Standards and Existing Resources

This stream consists of a group of workpackages that lay the foundations for the rest of the project. The guiding principle of **CASPAR** is the application of the OAIS Reference Model which is the only ISO standard addressing the long-term preservation of information. It is an abstract standard but one that contains some fundamental concepts and approaches to information preservation. The sequence of tasks is (1) review, (2) develop a specific **CASPAR** methodology and models, after which (3) an overall architecture is produced.

WP1100: Review of State of the Art

The main objective of this workpackage is to inform and motivate the early choices in the project by means of a critical and reasoned review of literature and current practices. The **CASPAR** consortium, during the preparation of this proposal, has chosen a number of standards or norms in order to propose an effective and operational solution to preservation problems, namely, OAIS, OSD, Semantic Web languages such as RDF and OWL, etc. The consortium has confidence that these main choices will be supported; however, the possibility of new developments and new ideas must be accepted.

Technological changes and research carry on apace and in this workpackage the Consortium will ensure that the project remains fully informed about the current state of the art.

WP 1200: CASPAR methodology and models

This workpackage lays down the roadmap for the **CASPAR** project. It is made up of two related parts: (1) methodology and (2) models.

The first of these starts with an extended, in-depth analysis of the implications of the OAIS Reference Model. Some preliminary work[31] on this has been done by members of the consortium under the aegis of the UK Digital Curation Centre, and that work was used in the development of this proposal. Even this brief analysis has indicated the many implications of OAIS; these range from the use of registries of Representation Information to the significance of a critical analysis of the many Persistent Identifier schemes. The components identified in the proposal will be further elucidated (and we must allow the possibility of component changes) in this phase. Underlying all of this is the idea of use and re-use of information as the touchstone of proof of preservation.

The in-depth analysis is expected to last for some time, not least because especially here the results must be proven, and improved, through the work on testbed validation. Unfortunately, there is insufficient time to do everything sequentially and therefore use will be made of many intermediate results and interactions between users, the workpackage leaders, task leaders, IP Stream Directors and IP Coordinator. The **CASPAR** Consortium recognises the criticality of the Methodology and will continually check and cross-check this vital keystone of the whole **CASPAR** structure.

One of the related intermediate results will result from a sister task on **CASPAR** guidelines. These guidelines will address some of the practical considerations for the following Streams, covering in particular those areas critical for interoperability with the other projects. This underpins the aim of facilitating the use of the **CASPAR** paradigm in as many projects as possible by minimising the cost of “buy-in” (this refers to effort that would be needed by the other project to use the **CASPAR** framework rather than payment).

The Guidelines and the first draft Methodology are required before work on the preservation related aspects of the **CASPAR** architecture can begin.

Carrying forward the subject of proof of preservation, this workpackage also covers the **CASPAR** work on certification. The intention is to build on the draft Audit and Certification document from the RLG/NARA task force and play a full role in its development into a full international standard with an associated accreditation and certification organisation. It is worth noting here that the IP Coordinator is very familiar with this process, being one of the authors of that document, and also instrumental in its origins – flowing from the OAIS Reference Model.

The second part of this workpackage delves into modelling of information objects – virtualising the information in the sense that a number of standard classifications of Information Objects will be developed to facilitate description and processing of the

underlying digital data. The modelling deals with some of the simpler syntactical and semantic aspects, developing in overlapping sequence virtualisation models for: a simple object, a complex object and objects that are produced on-demand.

The use cases and user requirements from the early stages of the testbed stream will inform this modelling work. In other words, the modelling must be based on practical requirements – in the case of **CASPAR** this is a very wide range of disciplines and users, in fact a range that goes beyond even the remit of the three testbeds. The **CASPAR** Consortium members have many links to government, academia and commerce, beside those to science, contemporary arts and cultural archives. The user base forms the real strength of **CASPAR** and the means by which the results will be tested and improved.

WP 1300: Overall Architecture

The overall architecture is determined by the conceptual model, the (draft) Methodology, and the object modelling. The architecture provides the specification of the interface of each component, the overall view of how components fit together, how they interoperate with themselves and other systems, and most importantly how this can work over time.

This last requirement is what singles **CASPAR** out from most other projects. There are many constraints on **CASPAR**. One is that unnecessary development must be avoided (and cannot be funded), and therefore contemporary state-of-the-art tools must be used, in order to avoid inventing things unnecessarily. Simultaneously, there is an absolute requirement that **CASPAR** builds for the future.

It is expected that the architecture will leverage the many recent developments of Web Services and the GRID. Yet, even as one thinks about that, the difficulty of the challenge becomes clear; what would this project have been planning to use even five years ago – and how would that have survived the rise of Web Services?

Stream 2: Component Level Research and Technology Development

In this Stream the core of the **CASPAR** project is implemented. It takes the guidelines, models and architecture from Stream 1 and produces workable components on which the subsequent streams can build.

WP 2100: OAIS-based components

If Stream 2 is the core, then this workpackage is its very centre. This work package creates essentially all the preservation components (apart from Preservation Storage, which deserved a workpackage of its own). The Access-related components, namely DRM, Authenticity, Identification and the Information Access Manager, are also dealt with in a dedicated workpackage.

The analyses that have been carried out in preparation for this proposal led us to have confidence in being able to deliver the required components, bearing in mind that a great deal of the fundamental design work will have been completed in Stream 1.

An important point to remember is that the **CASPAR** Consortium is aware that **CASPAR** will not solve all problems. The Work Package manager and Task Leaders will be keenly aware of functional “bloat” and “requirements creep”. The touchstone will be the **CASPAR** guidelines, Methodology and models, and the focus on validating preservation in an iterative manner. Once again, reliance will be placed on the breadth

of subject matter covered and the number of users brought along by the data holders within **CASPAR**.

This workpackage will focus on Representation Information, Preservation Description Information and Archival Information Packaging – all key OAIS concepts. It is worth stressing the importance being placed on the much fuller treatment of Semantics and its survival over the long term. Although the knowledge management work is emphasised in this workpackage, nevertheless its influence will be felt throughout the whole of **CASPAR** through re-use of tools and techniques in the project as a feature of its integration. To date, the most promising techniques to deal with knowledge come from the Semantic Web initiatives. This initiative uses languages like RDF (Resource Description Framework) or OWL (Ontology Web Language), which enables an explicit and declarative representation of knowledge that has its own precisely defined semantics. These languages provide an adequate framework to model the different levels of knowledge or metadata and to define the inference mechanisms needed to manage archives. However the same methodology must be applied to languages such as these as to any other information – in other words it itself will need, for example, Representation Information which will preserve its meaning and its use, and allow appropriate migrations and transformations over time.

WP2200: Storage

Storage is an essential component in any preservation system. To address the preservation-specific needs from the storage layer, we will define and implement a unified Preservation Object Datastore interface based on the object storage model.

Object storage is an emerging new storage paradigm. It raises the level of abstraction presented by today's storage devices. It enables the creation of self-managed, shared and secure storage. It moves lower-level functionalities such as metadata management and authorisation to the storage device itself, accessing the device through a standard object interface. The OSD standard interface is defined in the SNIA OSD working group[32] .

The OSD (Object-based Storage Device) Standard: Object storage, the future building block for storage systems, is a new and emerging interface for (networked) storage. It replaces today's block storage interface and will enhance the state of the art in storage technology. It is developed in SNIA OSD (Object Storage Device) working group , and involves a number of industrial companies including HP, IBM, Panasas and Seagate. The OSD standard is an evolving standard. The first version of an OSD protocol, embodied over SCSI, has been standardised as a T10 protocol. Currently, version 2 of the protocol is under development.

Preservation Object DataStores: extends the OSD work by adding the concepts required to support preservation. It becomes in effect an interface to OAIS Archival Storage, and underpins the Archival Information Package. It will thereby provide a powerful storage infrastructure for preservation environments. It will maintain the object metadata, or a reference to it, as an integral part of the object at the storage level. The object's metadata is managed, stored persistently and is recoverable with the object's data via built-in mechanisms. It facilitates the extension of object attributes in a flexible manner. This allows the incorporation of OAIS-related information into the object in a standard fashion. It allows secure and shared access to storage via the OSD security protocol, and can be extended to include support for collections of objects, multi-object operations, and advanced copy services on object such as snapshots and cloning.

The powerful preservation interface will insulate the preservation system from details of the storage hardware, whether it be, for example, disk or tape, local or distributed, and in particular will not depend on whether the underlying system is itself actually a specialised OSD. The important point is the **virtualisation** provided by the interface and implementations, built on a powerful, developing, industry standard. It will of course support the **CASPAR** authenticity and digital rights management components.

WP 2300: OAIS-based access

OAIS is a very powerful framework because it is organisational and not technical. This is the reason why it can be used as a common basis to integrate different levels of information and knowledge. However, in order to turn OAIS recommendations and structures into an operational system, the framework should be refined by distinguishing what are the relevant knowledge and information components, by designing them in terms of structures and tools and integrating them into the OAIS organisation. This is why this WP is entitled "OAIS-based access" insofar as the objective is to provide new, powerful information access tools, needed for long-term preservation, compatible with the organisation and structure recommended by OAIS.

CASPAR deals with complex contents coming from various contexts and communities. Complexity appears at several levels:

- Content objects may be virtual and dynamic insofar as they are dynamically built and presented on demand according to the user needs; the problem is to be able to preserve the structure and access to the required resources. This will build on the On-Demand Virtualisation tools from WP2200.
- Objects may vary in content according to the necessary migration or transformation. The paradox of preservation is that it is necessary to transform content and alter it in order to make it accessible in the future. As a consequence, it is necessary to preserve the authenticity of content and be able to keep some track of the way preserved contents have been transformed from their original or first forms. There arise very complex and deep questions about object integrity and invariance: to which extent a transformation is acceptable and how to make the transformation perceivable? An answer comes from the notion of content abstraction: for example, an abstracted form like a music score constitutes an invariant that should remain unchanged through the various transformations or interpretations. This idea should be generalised to propose several levels of abstraction where content can be transformed and preserved, following the evolving constraints of culture and technique, without losing its authenticity and integrity, and is closely connected with the other work on Virtualisation.
- **CASPAR** Objects are science, cultural and artistic contents associated with rights. Any preservation system should provide a way to represent juridical metadata and to manage their evolution. This implies the capability to recognise which materials are covered by which rights, who owns them, and to whom these rights have been transferred, under which conditions. The common players in such a process usually include an Author or Rights Holder, a Content Provider (i.e. the publisher of a book) and final Customer acquiring rights to use content under a specific license.

  In order to make this process work, a strong trust relationship must emerge between the rights holder and the repository, related both to the preservation of a material's integrity and to the enforcement of the conditions under which the rights holder allows access to the material. Such strong trust can only be achieved by adopting

technological measures and protocols, which are, of necessity, only briefly touched on in the current OAIS reference model.

This adds the need for an extension to the conventional DRM procedures to include the rights of copying and in some cases transforming the format or representation media of the protected material to fulfil the task of preserving it. Common DRM-aware distribution procedures can then be used for making the material accessible to the Designated Community, which may include collecting fees or transforming the materials in a suitable format, by applying copy protection techniques such as CSS or using tracking mechanisms, for example digital watermarks.

The long-term goal of preservation adds an additional layer of complexity to Digital Rights Management: copyright laws allow restriction of rights for a relatively short (usually 70 years) period of time and in a long-term scenario this has to be reflected in digital rights enforcement procedures. Moreover, the rights holder, content provider or legislative framework could change and evolve significantly in time: again these changes must be reflected in new policy descriptions and consequently in new access policies or procedures.

This also applies for new distribution tools, channels and strategies that may be developed in the future, for example new devices or media could alter the way we perceive the right to "print" or "broadcast" a material, thus requiring a dynamic adaptation of DRM policies. The display of scanned images from the fifties on the World Wide Web, the phenomenon of e-books, or the usage of small segments of music in mobile phone ring tones can give a basic idea of the future challenges we will be facing in the fields of rights protection.

- Digital Rights Management (DRM) deals with the definition and enforcement of rights governing the use and distribution of digital content. The main goal of DRM is to give authors the ability to manage the policies under which their material will be made available and to have them enforced at the time of distribution.

  This implies the capability to recognize which materials are covered by which rights, who owns them, and who these rights have been transferred to, under which conditions. The common players of such process usually include an Author or Rights Holder, a Content Provider (i.e. the publisher of a book) and a final Customer acquiring rights to use content under a specific license.

  In the preservation scenario, the institution providing an OAIS (Open Archival Information System) becomes part of the process as an additional actor. The Author or the Content Provider may want to rely on the institution's ability to preserve their material, also delegating the role of content distributor when external parties belonging to a Designated Community require access.

  In order to make this process work, a strong trust relationship must emerge between the rights holder and the OAIS, both related to the preservation of a material's integrity and to the enforcement of the conditions under which the rights holder allows access to the material. Such strong trust can only be reached by adopting technological measures and protocols, which are just partly specified by the current OAIS reference model.

  This adds the need of an extension to the conventional DRM procedures to include the rights of copying and in some cases transforming the format or representation media of the protected material to fulfil the task of preserving it. Common DRM-aware distribution procedures can then be used for making the material accessible to

the Designated Community, which may include collecting fees or transforming the materials in a suitable format, by applying copy protection techniques (i.e. CSS) or usage tracking mechanisms (i.e. digital watermarks).

The long term goal of preservation adds an additional layer of complexity to Digital Rights Management: copyright laws allow restriction of rights for a relatively short (usually 70 years) amount of time and in a long term scenario this has to be reflected in digital rights enforcement procedures. Moreover the rights-holder, content provider or legislative framework could change and evolve significantly in time: again these changes must be reflected in new policy descriptions and consequently in new access policies or procedures.

This also applies for new distribution tools, channels and strategies that may be developed in the future, for example new devices or media could alter the way we perceive the right to "print" or "broadcast" a material, thus requiring a dynamic adaptation of DRM policies. The display of scanned images from the fifties on the World Wide Web, the phenomenon of e-books, or the usage of small segments of music in mobile phone ring tones can give a basic idea of the future challenges we will be facing in the fields of rights protection.

The solution to overcome such issues is to treat IPR information with the same approach suggested to preserve context information, formats and knowledge in general. A semantic approach on modelling DRM policies, leveraging emerging standards on Semantic description and inference may result in a winning strategy to produce a dynamic DRM framework suitable for present and future challenges in intellectual property preservation.

A DRM ontology will be modelled after current state-of-the-art DRM policy declaration standards such as MPEG21-REL and ORDL, which correctly identify main entities in intellectual property management domain, but lack the dynamic nature necessary to face the mentioned emerging challenges in rights management oriented to long term preservation.

The ontology will include the necessary concepts for handling the evolving legal international framework related to copyright laws, including the much debated concept of Fair Use, but also the transition of materials into the so called Public Domain. The evolution of such ontology can be seen as the key tool in the consistency of the preserved materials rights information, both in the short and in the long term.

Such model will be then applied to the description of digital rights with a specific focus over preservation. A specific set of tools will be integrated as a component in the **CASPAR** framework for the declaration of rights and policies for long-term storage, along with all the necessary features needed for assembling distribution policies into Dissemination Packages.

These features will include a framework for managing plug-ins targeted to adapt content in formats suitable for distribution, such as digital image watermarking, copy protection and license tracking, along with suitable semantic descriptions of algorithms used. Such requirement has to be introduced in order to preserve the ability to recover the embedded information in the future.

The global approach in WP2300 is to manage this complexity by representing knowledge with semantic-based languages and formalisms having rich expressivity and

declarativity. Declarativity provides the possibility to migrate easily from one syntax to another. Expressivity provides the possibility to adequately represent knowledge, and the focus on semantics allows one to ignore ad-hoc and implementation details. Since long-term preservation has to manage change but cannot avoid it, technical choices should rely on formalisms that provide explicit meaning and enable smooth evolution between organisations and over time.

Stream 3: CASPAR Framework

Stream 3 brings the components together within a preservation framework, which provides the discipline within which the components work together to ensure preservation.

WP 3100: Framework architecture

The overall **CASPAR** system will rely on a software framework that will be the technological basis for the implementation and testing of the different testbeds. This framework is intended to be used throughout the project and should constitute the foundation for a European infrastructure to be used even beyond the project timeframe.

The definition of the technical framework architecture will encompass many interacting tasks.

First of all, we will define of the basic Middleware that will support the development, test and exploitation of the different software modules.

A component-based approach will be followed for the definition of this architecture.

It is expected that the architecture will leverage the many recent developments of Web Services and the GRID. Yet, even as one thinks about that, the difficulty of the challenge becomes clear; what would this project have been planning to use even five years ago – and how would that have survived the rise of Web Services?

A Service Oriented Architecture (SOA) will thus be created, taking advantage of its flexibility and its ability to include existing products and services, even based on different low-level technologies. This means that a series of basic rules will be set for the services to be plugged into the **CASPAR** framework. Software engineering methods and procedures will also be designed to facilitate the development of specific software components within the project.

The management of software configuration will be decided and applied, with the objective of maintaining the full control of modules developed or integrated during the project lifetime.

The preservation process flow will be analysed in detail in order to understand the logic relationships, the time constraints and the different maturity levels of the system component. On the basis of this analysis, the optimal process flow will be designed and described by means of specific instruments. The management of process flow will imply the monitoring and control of the various preservation tasks and the updating of preservation procedures as long as new application exigencies emerge, and as technology evolution allows the implementation of more effective services.

WP 3200: Integration of existing components

In the context of the **CASPAR** Framework implementation, special care must be devoted to integration of those components and resources already existing in project members' existing tools and procedures.

This workpackage aims to address all aspects of such integration activity:

- **Adaptation (and/or customisation) of existing components to be integrated**:

  Most of the actors and entities in the project come with their own infrastructure regarding information acquisition, management and storage. In other words, they come with their own processes and, sometimes, software components, each of them focused on the specific application context and working activity of the entity involved. In order to maximise effort, time-effectiveness and efficiency, the existing information management heritage shall be integrated as far as possible into the **CASPAR** system. This general principle implies, at a practical level, a customisation and adaptation of the existing infrastructure parts of each actor in order to make them fully usable and transparently accessible within the **CASPAR** software framework.

- **Analysis of the different platforms involved and coordination of component distribution among project actors**

  The physical locations hosting the components are, according to the general philosophy of the **CASPAR** Framework, potentially distributed all over the world and, in fact, they involve a number of different platforms (both hardware and software). This heterogeneous dissemination has to be planned, tracked and managed throughout the lifetime of the preservation activity.

- **Integration and testing of the integrated components**

  The actual component integration in the system framework has to be carried out and verified. Verification has to focus on the effectiveness of integrated components during full exploitation of the preservation chain. The status of the integration has to be clearly stated at any time during the integration phase. This requires the creation and regular updating of a detailed integration report.

Stream 4: testbeds, testing and validation

The objectives of this Stream will be to validate the solutions developed throughout the **CASPAR** project – **CASPAR** framework, generic models and tools – by applying these to specific corpuses (science, culture and contemporary arts).

The first workpackage addresses those tasks which are common to all three testbeds, while the others develop specific models and tools in order to cope with the specificity of each of the application fields.

There are two testbeds about very different types of cultural data, namely cultural heritage sites and contemporary arts; each of these spans an enormous range. The third testbed covers a huge variety of science data.

WP4100: Generic testbed elements

The objectives of this workpackage in the first 18 months are:

- to gather use cases and requirements for all the testbeds at the start of the project;
- to create a design for those elements and services that are common to all the testbeds;

- to create an agreed testbed evaluation methodology;

The reports on the R&D work of each of the testbeds will be integrated in this workpackage in order to ensure the maximum synergy between the testbeds, and to further identify common elements.

a. Use cases and Requirements

At the start of **CASPAR**, user requirements will be gathered for all the testbeds, covering all the testbed users. The organisations ESA, CCLRC, INA, CIANT and UNESCO will play leading roles in this exercise via their large user base. Users in this context include not just end-users but also their funders, suppliers, data producers and others. These use cases and requirements will inform not just the further testbed designs, but also the **CASPAR** Methodology and Architecture.

b. Generic testbed design and specifications

The three testbeds will share a number of common elements and so, to avoid duplication of effort, this work is carried out in a single task. It will build on the Requirements analysis. The process for performing cross-disciplinary validation will also be addressed here.

All services will be capable of working in a completely distributed environment and with a reliable solution that will provide services even after a partial failure of the system or after migration and reorganisation of resources used in testbeds.

c. Evaluation method definition

The general approach proposed by the **CASPAR** Consortium for the testbed evaluations was described earlier. This approach will be reviewed and developed into quantitative and qualitative evaluation criteria as a **Testbed Evaluation Plan** for each testbed, agreed by the External Review Committee and the Commission.

.

WP4200: Cultural data preservation testbed

This workpackage 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, and they are distributed and not all well documented..

The objectives of this workpackage will be to test the solutions developed throughout the **CASPAR** project – **CASPAR** framework, generic models and tools.

Specific R&D activities include:

- R&D of the large list of World Heritage sites that UNESCO has in order to identify pilot test-cases that will then be provided to **CASPAR** for testing. We foresee the selection of some sites in Europe where the complexity of existing information is overwhelming, as well as some sites in developing countries. For example, for the old city of Venice there is an enormous amount of information scattered all over the world.
- R&D for the selected sites with respect to the most representative types of data and information. For the selected sites, an analysis will be performed of the most

representative types of data and information relevant to the site and on which media this information is stored.

- R&D for any associated tool that is necessary to visualise the data. For the various types of information, mainly those stored in electronic form, specific software packages have to be used in order to make use of the data. An in-depth analysis is required of the most significant software packages used for the selected sites.
- Liasing with UNESCO partners in order to identify how access to data residing outside UNESCO will be provided to the **CASPAR** project for testing.

There are opportunities for cross-disciplinary work using science (Earth Observation) data to integrate with heritage site data. Similarly the contemporary arts data can be used in a cross-disciplinary way.

WP4300: Contemporary art testbed

This testbed is focused on the implementation of a testbed for contemporary arts: Contemporary music, performing arts, and other forms of technology-enhanced arts like video games

Contemporary music, as well as performing arts and video games, is an interesting case that produces very complex objects including specific hardware (such as MIDI instruments), instructions and specific equipments. The model can be described as an IFLA-FRBR based model with specific extensions such as interaction devices (specific sensors, specific instruments), and specific sets of instructions (including score, specific textual production instructions...), and can be summarised as follows:

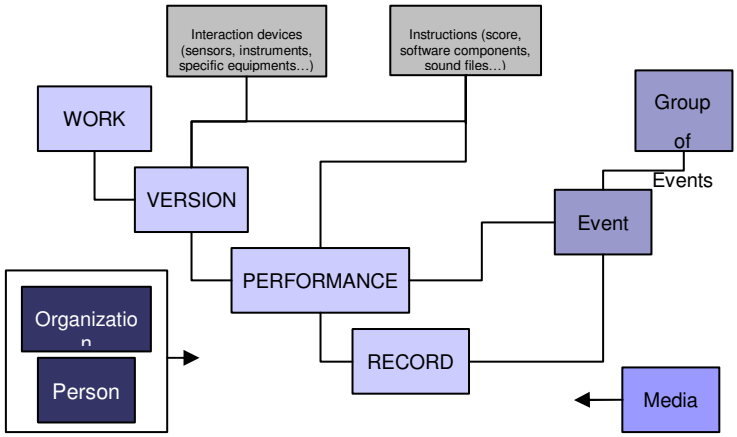


Figure 11 IFLA-FRBR based Contemporary art model

The model is extensible to all kind of contemporary arts with technology enhancement, where the same classes of extensions exist (scenarios instead of score, video files instead of audio files…). In addition, in the case of performing arts, the performer (or the performance) must be virtualised in order to become part of the preservation process. To this end, models of expression, gesture, and interactions with the performer must be elaborated and virtualised, and rendering devices must be developed in order to preserve intelligibility.

The objectives of the workpackage in the first 18 months will be to study these specific extensions, their relationship with the generic models, procedures and services, and to start the analysis and set-up of a generic and specific testbed infrastructure.

The following tasks are envisioned in order to reach these objectives:

Modelling

- Examine the process of production of complex musical works (as produced at IRCAM and GRM), for pieces of performing arts, and for video games
- Develop a description model to help gather required production elements
- Develop a description model for describing complex relations between production elements
- Develop classification schemes for specific elements (instructions, instruments, equipment)
- Develop classification schemes and models of interaction with performers (motion gesture analysis, events modelling...)
- Participate in standardisation efforts by defining formal models to be proposed to standardisation bodies.

Infrastructure and set-up

- User's interface design and implementation
- Develop specific agents in order to help gather required production elements and describe relations between these (includes automatic production of documentation from specific production tools)
- Analysis tools: Develop specific agents in order to help analysis and enrichment of content (intra-document browsing, synchronising, annotating motion gesture analysis and segmentation...)
- Repurposing: Develop specific agents in order to help repurposing.

WP4400: Science data preservation testbed

As for the other two testbed workpackages, the high-level objective of this workpackage is to provide a proof-of-the-**CASPAR**-concept for science data preservation by setting up and implementing a science data preservation testbed, taking into account the project findings, and to provide a source of feedback to other **CASPAR** Work Packages. In summary, this science data preservation testbed will assess, the Earth Science community requirements, develop the necessary specific services and prototype an Earth Observation science data preservation environment.

In particular, the science data preservation testbed workpackage will carry out R&D activities specific to the testbed, will consider the infrastructure, set up the testbed and implement it. Initially it will focus on information related to ESA Earth Observation missions/instruments, such as the SAR (Synthetic Aperture Radar) on board the ESA ERS-1 (European Remote Sensing) satellite mission, which was operated between 1991 and 2002. Together with the fact that the total amount of information generated over the indicated time-span is in the order of some hundred Terabytes, and the information is distributed and has a high level of complexity, it makes it a good case for prototyping in the **CASPAR** project. Initially we will restrict the testbed to a representative subset of data, complemented by all necessary related information.

ERS-1 SAR related information

As demonstrated in Figure 12, Earth scientists access not only the data from the satellite instrument to conduct their research. Often, they need access to complementary data from other instruments, including satellite sensors, air sensors, and ground sensors as well as to various models, algorithms, software and reports, publications and other

documents that have information explaining how to produce useful research products, science results and so on.

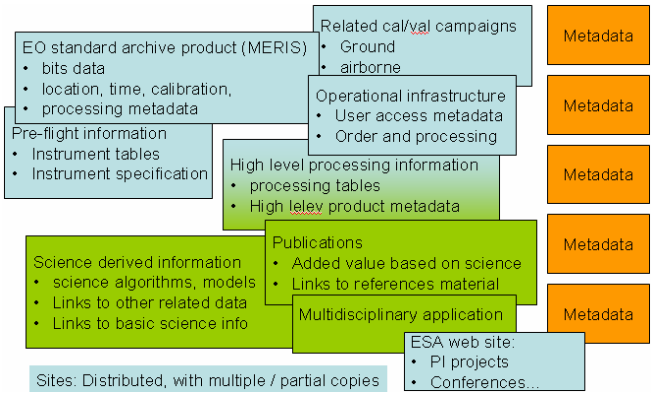


Figure 12 : Earth Observation activities

As such, the intended preservation infrastructure will deal with a mixture of selected simple/complex/on-demand objects based on SAR data, SAR data products, metadata available from catalogue databases (inventory and quick look images), processing strings, calibration campaigns (with other instruments), datasets in other locations, large volumes of documentation regarding the instrument, the processing, algorithms, tools as well as with information about dedicated workshops, conferences and ESA-supported application projects, operational services deployed, and some special datasets (e.g. Etna (Italy) mapping pre/post eruptions).

The key issue in these specific examples is the availability and reconstruction of detailed metadata associated with the very different types of objects to be managed.

For reference, the total amount of products (simple objects) related to ERS-1 can be estimated as few hundred thousand including: products in the catalogue database (inventory and quick look images), various processing strings, several calibration campaigns - with other instruments and datasets in other locations, large volumes of documentation about the instrument, processing, algorithms, tools, plus dozens of dedicated workshops, conferences, and over 500 ESA supported application projects, with over 100 operational services deployed, and a few special datasets (e.g., ETNA mapping over the life time of ERS). The testbed shall ensure access to above mentioned information, algorithms, tools etc.

Testbed specific R&D activities

The following table lists R&D activities that are considered relevant for the science data preservation testbed. Some will be considered within the first 18 months::

- The preservation life-cycle of scientific data: preparation of specific tools which will allow visualisation and navigation of the complex metadata associated with science data complex objects.
- Implementation of the OAIS Reference Model for Science data preservation.
- Compatibility with OGC efforts for geospatial data management.
- Evolution of the proposed SAFE (Standard Archive Format for Europe) archive format standards for Earth observation missions to include preservation requirements.
- Digital library integration.
- Metadata, Ontologies and semantics of science information objects: analysis of aggregation of satellite images and on-demand processed images from the Grid with corresponding text descriptions from news and from internal text archives as well as with other data available for the same period and same geographic area.

- The use of the Grid and on-demand (re)generation of information objects; some information objects are not stored as such but are generated on the fly. Preservation of such on-demand objects, that is, of the conditions under which they were generated and how to regenerate them is a specific research topic.
- Distributed partial copies of Earth science information objects; there are different problems related to availability of the 'same' data on different locations, e.g., authentication, authenticity and provenance. There is however also a problem related to partial copies. For example when satellite data are transferred from the satellite to the ground-station, it may happen that the reception is interrupted and received partially by different ground-stations. Research is needed to understand how to deal with these cases.

Stream 5: Innovation-related activities, Training & Demonstration

WP 5100: Exploitation and take-up

The consortium members will define how the exploitable aspects of the projects would be used and integrated within their business. They will develop an exploitation plan explaining in detail how the output of the project would be introduced into a common and individual strategy. Particularly, the consortium will have to identify, in agreement with the partners responsible for the development of the three scenarios, the services and/or the products that could be derived from the project results, as well as their positioning within the digital preservation environment, based on a costs/benefits analysis and an evaluation of the marketing strategy for digital preservation. Each partners' plan will be consolidated into a single exploitation plan for the **CASPAR** project.

In order to meet the objectives of the feasibility and market validation phase of **CASPAR**, the first steps are to carry out a comprehensive costs/benefits study and metrics in Digital Preservation; in addition, a market survey will be conducted. Information will also be collected on possible competitors and their products/services.

A detailed analysis of the technological framework will also be conducted to analyse all the technological aspects to be taken into consideration in a business model, such as future standards evolution, technological trends, legal aspects, risks, etc.

The next task will analyse the business benefits of the **CASPAR** sustainability model, focusing on the following targets:

- to define success indicators, in order to monitor the project activities and validate the project results;
- to optimise interaction among the actors of the Digital Preservation;
- to validate the possibility of cost reduction in using **CASPAR** technologies for digital preservation.

A preliminary sustainability plan will then be prepared, addressing market entry and technical deployment strategies, choice of criteria for best cooperation partnerships, organisational structure of operations, pricing, and exploitation strategies. The business plan defines the measurable parameters for the **CASPAR** solutions to attract potential customers interested in the Digital Preservation. Therefore, the **CASPAR** preliminary business plan will result in the definition of the success indicators for the **CASPAR**

project. Preliminary assessment will be carried out on main operational costs and revenue sources.

The goal for this work package is also the definition of a preliminary Communication strategy in order to understand the framework in which **CASPAR** will move in the future. The approach for the sustainability model is the understanding of the basic needs and requirements of the preservation user community, analysis of the market of reference, and definition of success indicators and guidelines for a preliminary sustainability model. Following this approach the activities will immediately start to aggregate **CASPAR** potential users through the creation of a **CASPAR** Preservation User Community; indeed preparation of the bid showed the health interest there is in this already, in Europe and World-wide. Following the Dissemination plan, the community activities will start during the period of this work package.

Engineering will coordinate all the activities in this workpackage.

WP 5200 Dissemination activities

MW will lead this WP, offering its wide experience gained in the field of dissemination and communication.

This workpackage has two aims:

Dissemination planning and implementation

The objective of this task is to identify and organise the activities to be performed in order to promote and disseminate the **CASPAR** project outcomes.

This workpackage will coordinate the activities correlated to the following subjects:

- Promotional activities for the dissemination of the results;
- Participation in focused workshops, local and international events and conferences, coordination with external organisation, special interest groups;
- Contribution to the building up of a **CASPAR** Preservation User Community aimed at involving a huge critical mass around the activities of the project and its initiatives.

The main purpose of the Dissemination activities is therefore to ensure the following actions are performed:

- To present the project results to the widest audience and enlarge the **CASPAR** Preservation Community.
- To target specific audiences that may benefit from the results: artistic digital data creators of all kind, professionals and decision makers dealing with digital preservation, data owners, data holders and data curators, end-users of digital information, etc.
- To use a variety of techniques and media as appropriate for the content to be delivered to the target audience.

The entire consortium will be involved in the participation in events, in the publication of papers regarding the project results and working progress, as well as in providing specific user requirements and suggestions for the potential evolution of the network.

Production of Dissemination material

This task will focus on the preparation of varied dissemination material (leaflets, brochures, flyers, a web site) that will be designed and analysed according to different communication needs, various event typologies and, of course, following the evolution of the project. In particular, the following deliverables will be realised:

- A **CASPAR** logo and a coordinated design will define a "**CASPAR** brand" with respect also to exploitation purposes. The Brand is not just a name or a logo; it is a whole anthology of attributes that describes the "personality", presence and performance of the **CASPAR** results, products and services. The **CASPAR** Brand will be fundamental to its gaining acceptance. Brand building can raise awareness, catapult a product or a service into the marketplace and increase interest with stakeholders. In the business world, the Brand creates "accountability";

- Realisation of a simple info-package, in a first instance to explain the **CASPAR** project objectives and expected results to a general public;

- Realisation of leaflets, brochures (suitable for both technicians and general public), booklets (detailed report, in easily understandable language, about the status of the project activities and results, released regularly), posters and gizmos (bookmarks, stickers, handouts and so on);

- Press releases, paper publications, logo and target articles in specialised media, press kits;

- Participation in the most important forums, international events and conferences organised around Europe and beyond related to preservation, and also to raise the interest of other similar projects interested in building common synergies and to maximise the impact of the project results;

- A project web site, with public and restricted sessions and interactive services and information. This will be an effective web-based working space where information and knowledge about preservation produced by **CASPAR** could be shared between experts;

- Electronic resources such as semester newsletters and mailing list, in order to permit and promote collaborative actions with other research communities dedicated to preservation.

WP5300 CASPAR Training activities

A significant amount of fundamental research is being undertaken as part of **CASPAR** and it is vital that there is a rich interaction between the research components of the project and the practical use of that work across the EU member states. There will be a focus on improving the knowledge and skills of professionals in the cultural heritage and scientific communities through the development of training materials and their delivery in different professional development environments. In the first eighteen months the main activities relating to **CASPAR** training will be:

CASPAR Training Needs Definition: The **CASPAR** team will begin by assessing the relationship between current community capabilities and the skills that would need to be developed if the community is to be prepared to take full advantage of the methods and technologies that **CASPAR** is delivering. The results of this activity will be knowledge about the target communities that will ensure that the training programme is appropriately structured and targeted to meet their various learning needs and that the effectiveness of the training can be assessed and its impact measured.

CASPAR Training Plan: The objectives and learning outcomes of the training programme as a whole will be defined, as will those of the individual courses. The teaching and learning methodologies that will best enable participants to achieve these objectives and outcomes will be laid out. The programme will ensure that it uses a range of teaching and learning techniques. **CASPAR** will use a variety of teaching scenarios (e.g. large and small group teaching, self-directed learning, and remote learning) and different approaches to teaching (e.g. lectures, worked examples, case-based learning, and practical exercises). The plan will provide the definition of the courses that **CASPAR** will run, how they will inter-relate and how they will achieve their learning outcomes. This will incorporate a schedule of course delivery locations, dates and timings.

CASPAR Training Development activity will cover the design of the **CASPAR** training courses including agreeing the **CASPAR** Training Methodology, which will indicate the learning and teaching methods to be used, preparing the briefing documents, drafting introductions for trainers, establishing training schedules, producing detailed course programmes and identifying how we will measure the success of the individual course and the programme as a whole. As well as resulting in specific courses, it will include a model/template that can be used to develop other courses. This will include the creation of online materials to enable training participants to prepare for the training courses and to allow them to take advantage of training aftercare.

CASPAR Training Delivery and assessment will begin towards the end of the first eighteen months with the delivery of the first two training courses and the awarding of certificates to those who successfully complete the training. The process of measuring the success of the programme will begin by measuring learning and measuring effective use of what has been learned through impact assessment.

CASPAR will also deploy a technological training platform in order to support the training activities with value added services. The platform will be more than just posting documents and presentations: it is a tool for placing and managing course content online, so that the service will engage trainers in collaborative activities in order to support the **CASPAR** learning goals. The platform will help the **CASPAR** Education team in the preparation phase, during the courses and in the results dissemination of the training at the end of the project, ensuring durable consistency of the training activities and also providing a support for distance (throughout Internet) learners. The features and other integrated tools offered for smooth and accessible usage of the platform will be related to lesson storage and content management, and post-lesson support. In the first 18 months the training infrastructure specifications will be defined and a first draft of the prototype will be realised.

This first 18 months of activity will lay the foundation for a dynamic, replicable and extendable educational framework that will run beyond the timeframe of the project and have professional impact for those who take advantage of the programme, which lasts throughout the EU-funded **CASPAR** project.

WP 6100 Project management and coordination

This workpackage covers the whole duration of the 18-month implementation plan and focuses on the day-to-day coordination of **CASPAR** and interfaces with the Commission. More particularly, it oversees programme management, is responsible for delivering a quality plan and assessing the quality of the deliverables, addressing conflict management, overseeing the implementation of the agreed confidentiality, IPR, with advice from the IPR Advisory Committee, and exploitation management issues in cooperation with the Exploitation Advisory Committee and financial management, assisted by the IP Secretariat (see section 7 for more on the management structure and roles). The overall coordination activity is also responsible for collecting, analysing and communicating to the Commission the periodic progress reports of the project and for overseeing the management of the project resources. Communication between the partners, between the partners and the Stream Directors and between the partners and the Commission will be facilitated by the **CASPAR** internal web portal.

WP6200 Project performance monitoring

Project performance monitoring will rely heavily on the set of Key Performance Indicators, which will be identified in Task 6201, and will be kept under review. Task 6202 produced the Project Performance Assessment (PPA) tool, which will provide rapid feedback to the project and the Commission about the state of the project.

Continual assessment of the performance of the project is carried out under task 6203, where the KPIs will be regularly gathered and reviewed using the PPA.

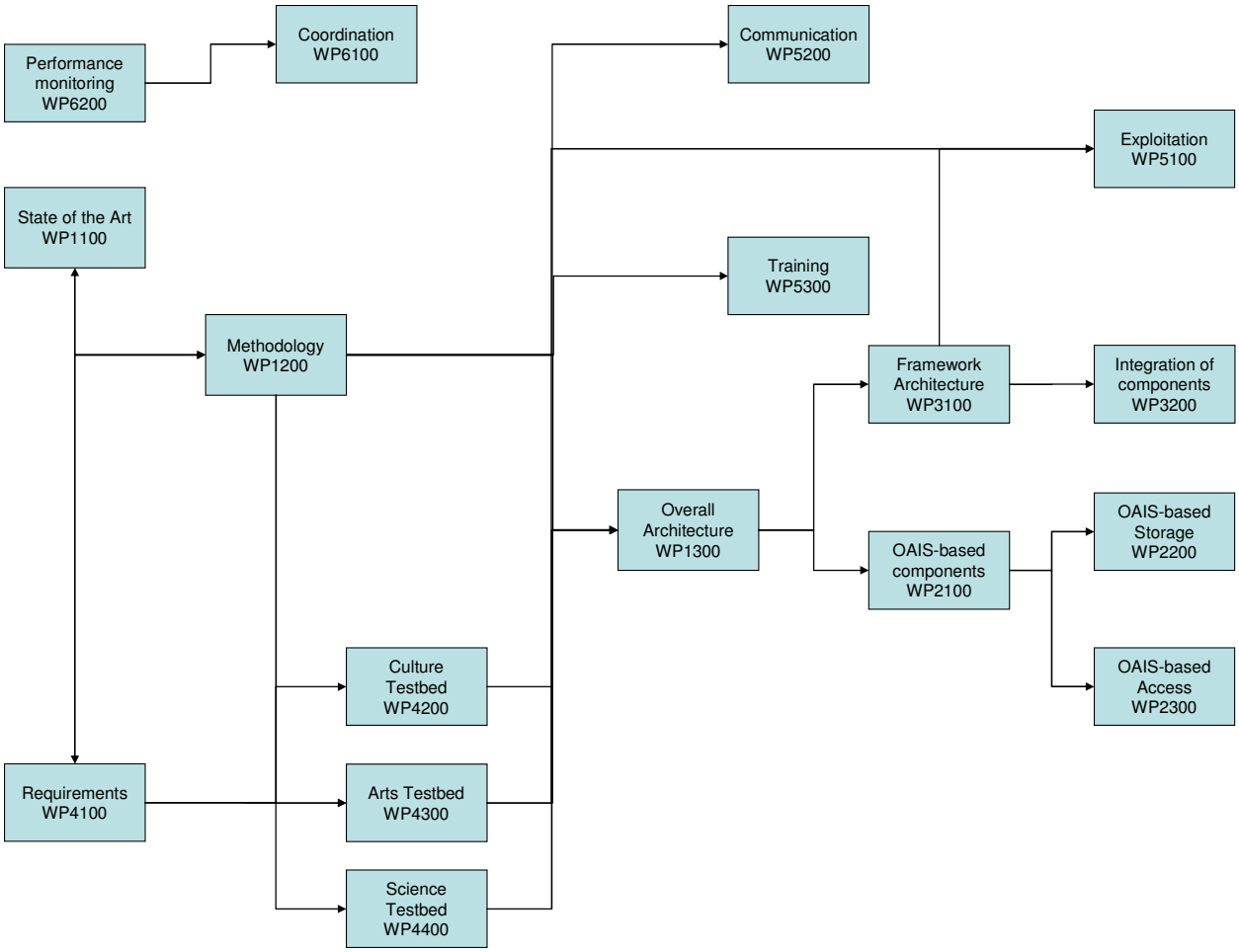
Milestone list (18 month plan)

Milestone no.	Milestone date (month)	Milestone title and goal	Key WorkPackages
M#1	m03	**Kick Off of CASPAR** • Consortium agreement completed	WP6100, WP6200
M#2	End m06	**CASPAR preservation requirement review** • State of the art completed and reviewed • Testbed Use Cases completed and requirements derived • Management tools in place	WP1100, WP4100, WP6100
M#3	End m12	**Review of CASPAR methodology and Updated Management Plan** • **CASPAR** guideline and conceptual model completed • Draft Framework Architecture completed • Simple object virtualisation model completed • Testbed design completed • Validation/Evaluation methodology review based on ERC report • Mandatory reports	WP1200, WP4100, WP4200, WP4300, WP4400, WP6100
M#4	End m15	**Review of Architecture and initial Exploitation & Sustainability plan** • Architecture Review • Component model specification review	WP1300, WP5100
M#5	End m18	**Review of CASPAR designs** • Component Model and overall component architecture design completed • Framework Architecture completed • OASIS-based components, storage and access analysis started • Integration of existing component started • Testbed specific R&D activities started • Training and dissemination plan completed • Review of revised evaluation criteria	WP1300, WP2100, WP2200, WP2300, WP3100, WP3200, WP4200, WP4300, WP4400, WP5x00

8.2 Planning and timetable

				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
	Methodology, standards and existing resources	1	38																																						
1100	Review State of the Art	1	4																																						
1200	**CASPAR** methodology and models	1	38																																						
1300	Overall architecture	3	12																																						
	Component level Research &Technology Development	1	38																																						
2100	OAIS-based components	1	30																																						
2200	OAIS-based Storage	1	30																																						
2300	OAIS-based access	3	30																																						
2400	Next generation components	25	38																																						
	CASPAR Framework	3	38																																						
3100	Framework Architecture	3	16																																						
3200	Integration of existing component	9	27																																						
3300	Framework technical testing	16	33																																						
3400	Multi-industries long-term research perspectives	25	38																																						
	Testbeds, testing and validation	1	38																																						
4100	Generic testbed elements	1	38																																						
4200	Cultural data preservation testbed	7	33																																						
4300	Contemporary art testbed	7	33																																						
4400	Scientific data preservation testbed	7	30																																						
	Innovation related activities, Training and Demonstration	1	38																																						
5100	Exploitation and take-up	1	38																																						
5200	Dissemination activities	1	38																																						
5300	Training	6	38																																						
5400	Demonstration activities	30	38																																						
	Administration, Scientific and Technical Project	1	38																																						
6100	Project management and coordination	1	38																																						
6200	Project performance monitoring	1	38																																						

8.3 Graphical presentation of work packages



8.4 Work package list /overview

Workpackage lists (18 month period month 1-18)

WP	Work Package Name	Lead contractor	Person Months in first 18 mth	Start month	End month	Deliv.No
1100	**Review State of the Art**	**INA**	**47**	**1**	**4**	**D1101**
1200	**CASPAR methodology and models**	**UG**	**146**	**1**	**18**	**D1201, D1202**
1300	**Overall architecture**	**ENG**	**59**	**3**	**12**	**D1301**
2100	**OAIS-based components**	**CCLRC**	**74**	**1**	**18**	**D2101**
2200	**OAIS-based Storage**	**IBM**	**30**	**1**	**18**	**D2201**
2300	**OAIS-based access**	**CNRS**	**70**	**3**	**18**	**D2301**
3100	**Framework Architecture**	**ACS**	**33**	**3**	**16**	
3200	**Integration of existing component**	**ACS**	**42**	**9**	**18**	
3300	**Framework technical testing**	**ESA**	**6**	**16**	**18**	**D3301**
4100	**Generic testbed elements**	**ASMX**	**68**	**1**	**18**	**D4101, D4102, D4103**
4200	**Cultural data preservation testbed**	**UNESCO**	**23**	**7**	**18**	see D4102 – integrated report
4300	**Contemporary art testbed**	**IRCAM**	**41**	**7**	**18**	see D4102 – integrated report
4400	**Scientific data preservation testbed**	**ESA**	**32**	**7**	**18**	see D4102 – integrated report
5100	**Exploitation and take-up**	**ENG**	**25**	**1**	**18**	**D5101**
5200	**Dissemination activities**	**MW**	**38**	**1**	**18**	**D5201, D5202**
5300	**Training**	**UU**	**23**	**6**	**18**	**D5301**
6100	**Project management and coordination**	**CCLRC**	**23**	**1**	**18**	**D6101, D6102, D6103, D6104, D6105, D6106**
6200	**Project performance monitoring**	**CCLRC**	**19**	**1**	**18**	**D6201, D6202**

Preliminary Workpackage lists (full duration 42 months)

WP	WP Name	Lead	mm	Start	End
1100	Review State of the Art	INA	47	1	4
1200	**CASPAR** methodology and models	UG	183	1	38
1300	Overall architecture	ENG	59	3	12
2100	OAIS-based components	CCLRC	144	1	30
2200	OAIS-based Storage	IBM	41	1	30
2300	OAIS-based access	CNRS	130	3	30
2400	Next generation components	CNR	62	25	38
3100	Framework Architecture	ACS	33	3	16
3200	Integration of existing component	ACS	86	9	27
3300	Framework technical testing	ESA	36	16	33
3400	Multi-industries long-term research perspectives	MW	36	25	38
4100	Generic testbed elements	ASMX	117	1	38
4200	Cultural data preservation testbed	UNESCO	72	7	33
4300	Contemporary art testbed	IRCAM	118	7	33
4400	Scientific data preservation testbed	ESA	86	7	30
5100	Exploitation and take-up	ENG	80	1	38
5200	Dissemination activities	MW	47.5	1	38
5300	Training	UU	56.5	6	38
5400	Demonstration activities	CCLRC	15.5	30	38
6100	Project management and coordination	CCLRC	44.5	1	38
6200	Project performance monitoring	CCLRC	28	1	38

8.5 Deliverables list (18 month period month 1-18)

Deliv No	Deliverable title	WP No.	Lead partic.	Estim. Person months	Nature	Deliv. date	Dissem. level
D5201	**CASPAR** Web site	WP5200	MW	3	O	M3	PU
D6101	Consortium and Project Management Procedures and Quality Plan	WP6100	CCLRC	2	R	M3	CO
D6102	**CASPAR** Management Web site	WP6100	CCLRC	3	O	M3	CO
D1101	Review of the State of the Art	WP1100	INA	46	R	M6	PU
D4101	User requirements and scenario specifications (will include use cases for various disciplines and analysis of common requirements)	WP4100	ASMX	65	R	M6	PU
D6201	**CASPAR** KPIs definition, organisation and metrics	WP6200	CCLRC	1	R	M6	PU
D6202	Balanced Scorecard-based PPA tool design and implementation	WP6200	MW	8	R, O	M6	PU
D1201	**CASPAR** Conceptual Model	WP1200	CCLRC	29	R	M12	PU
D1202	**CASPAR** guidelines	WP1200	GU	4	R	M12	PU
D1301	Overall component architecture design (and component model) (architecture provides views on the system so that one can understand how the individual components work together)	WP1300	ENG	56	R	M12	PU
D4103	ERC report on validation methodology	WP4100	ASMX	3	R	M12	PU
D5101	Dissemination and Use Plan	WP5100	ENG	10	R	M12	PP
D6103	Updated Management Plan	WP6100	CCLRC	2	R	M12	CO
D2101	Prototype OAIS-infrastructure and associated draft reports of knowledge management architecture and tools	WP2100	CCLRC	73	P, R	M18	PU
D2201	Prototype object data store	WP2200	IBM	30	P	M18	PU

	interface						
D2301	Report on OAIS-access model, including DRM, Authenticity and Persistent Identifiers	WP2300	CNRS	70	R	M18	PU
D3101	Report on Framework Architecture	WP3100	ACS	33	R	M18	PU
D4102	Integrated report of R&D activities on the Cultural, Performing Arts and Science data preservation testbeds (D4102)	WP4100	ASMX	162	R	M18	PU
D5301	**CASPAR** Training Plan	WP5300	UU	23	R	M18	PU

Mandatory Management Deliverables

D6104	Activity Report	WP6100	CCLRC	1	R	M12	CO
D6105	Management Report	WP6100	CCLRC	1	R	M12	CO
D6106	Updated Implementation Plan for months 19-42	WP6100	CCLRC	3	R	M18	PP

NOTES:

Software and associated documentation will be managed with a full configuration control system together with a bug and feature tracking system. Each prototype will be uniquely identified (module name, version, release date etc) and, where applicable, the various unit, integration and system tests will be available.

8.6 Work package descriptions :

Workpackage number	**1100**					Start month			1
Workpackage title	**Review State of the Art**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	7.0	4.0	3.0	0.0	2.0	0.0	0.0	3.0	2.0
Participant	ENG	CNR	Forth	CNRS	**INA**	IRCAM	UL	CIANT	
Person months	3.0	0.0	3.0	3.0	**6.0**	3.0	5.0	3.0	

Objectives

The objective of the WP is to review the state of the Art in the different fields covered by **CASPAR**. There has been increasing activity in these domains during recent years and proposals as well as initiatives are flourishing in many countries or international bodies. This WP aims at providing a clear cartography of different ongoing projects and at focusing on the main issues that remain critical for the development of long-term preservation solutions. Another crucial issue for this WP is to provide a common understanding of the state of the art within the **CASPAR** consortium in order to have a shared intellectual basis. Situated at the beginning of the project time line, this WP will be the work of only some partners but will disseminate its results to all partners.

Description of work

Task 1101 – Methodology review ***(Task leader CCLRC)***

The objective here is to document the different methodologies proposed in the domain of digital data curation and preservation. Preservation is not only a problem that technical answers should solve, but it essentially consists in managing a process and organising data and knowledge. This task will consider initiatives coming from different projects such as Chronopolis and InterPARES, but will focus also on the different **CASPAR** partners who are representative of their field: IRCAM and INA/GRM for music and contemporary arts, ESA for scientific data, UNESCO for cultural data.

The *work* will consist of the following issues:

- Review of existing projects and initiatives;
- Review of methodologies;
- Review of current practices by **CASPAR** content partners and users.

Task 1102 – Research and technological solutions ***(Task leader IRCAM)***

The objective is to document the different solutions coming from the technological solution providers. Because of the pervasive needs for data preservation, more and more industrial solutions are proposed on the market. These solutions often derive from research works. The task will cover these two aspects.

The *work* will consist of the following issues:

- Review of industrial solutions;
- Review of technological research solutions.

Task 1103 – Existing standards ***(Task leader CIANT)***

The objective is to review the status of a number of different standards, and draft standards. This is a crucial issue because standards are the main way to obtain interoperability and exchangeability of data. OAIS is very well known in the project. Besides this, standards from the documentation worlds such as METS, from the audiovisual world such as MPEG-7 and MXF, etc., will be considered in order to adequately situate **CASPAR** needs and solutions.

The work will consist of the following issues:

- Preservation standards

• Media specific standards
Deliverables D1101: Review of the state of the art (Month 6)
Milestones and expected results M#2: state of the art completed. Public. (Month 6) M#2: internal workshop to share results and build a common in-depth understanding of **CASPAR** issues. (Month 6)

Workpackage number	**1200**				Start month				1
Workpackage title	**CASPAR methodology and models**								
Participant	CCLRC	ESA	UU	UNESCO	**Glasgow**	ACS	IBM	ASMX	MW
Person months	15.7	12.2	9.0	8.0	**7.9**	3.0	20.1	4.0	2.4
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	3.0	6.6	4.3	19.2	15.5	11.8	0.0	3.0	

Objectives

The objectives of this WP are to produce the fundamental models and guidelines for the remainder of the **CASPAR** project. These models and guidelines must be produced in a timely way, to allow the other parts of **CASPAR** to proceed, and it must be possible to test these models against reality before finalising them.

In order to do this, a number of intermediate draft models will be produced. These will be used to produce prototype implementations, which can in their turn be used to validate the approach. The lessons learned will be fed back into the modelling process.

Closely related to this validation is the work carried on in this work package to take **CASPAR**'s research and advances into the relevant standards bodies.

Description of work

The tasks within this work package are as follows

Task - 1201 CASPAR overall methodology (Task leader CCLRC)

This task will critically review the details of the approach embodied in this proposal and revise and refine those details. In particular, the overall approach to preservation, solidly based on OAIS, will take into account the latest thinking in this field. **CASPAR** members have cross-membership with many of the key OAIS-related development groups.

Task - 1202 CASPAR guidelines (Task leader UG)

The **CASPAR** guidelines will provide the underlying consistency in later developments. This will cover everything from a checklist for Quality Assurance for future-proofing of **CASPAR** developments to software standards to be adhered to during implementation.

Task - 1203 Contributions to standards (Task leader UU)

CASPAR expects to make significant contributions to many important standards. This task has the responsibility of monitoring progress and advising on steps and resources needed to ensure adequate support for these inputs.

Task - 1204 CASPAR conceptual model and Requirements (Task leader CCLRC)

The **CASPAR** conceptual model begins with the OAIS Reference Model and begins the process of putting flesh on the bones. A good deal of preliminary work has been done under the auspices of the UK Digital Curation Centre. From that work it is clear that the conceptual model is very wide ranging, touching on the requirements for almost all of the subsequent work of **CASPAR**.

One essential input to this task are the use cases collected for the testbeds.

The following three tasks are closely related and build from one to the next, however each has its unique challenges.

Task - 1205 Simple digital objects virtualisation modelling (Task leader ACS)

The simplest of the three tasks, this one will begin by defining or collecting appropriate standard definitions of scientific, cultural and contemporary arts objects. The simple objects cover those features that are common to all. As an example, consider the concept of an image. This is quite fundamental, beginning with a very simple 2-dimensional array, but which then has specialisations ranging from "normal" TIFF images to highly specialised Earth Observation images.

Task - 1206 Complex digital objects virtualisation modelling (Task leader ACS)

Complexity is somewhat open ended and the specific areas developed here will be strongly influenced by the **CASPAR** Testbeds. Examples here include, continuing the image example from 1205, virtualising multi-band scientific images and moving images with sound.

Task - 1207 On-demand object modelling (Task leader ACS)

On-demand objects concern digital information which is created as required ("on-the-fly") from other, static, data. Examples include sub-sets of images, stereo images, motion analyses etc. The creation of such on-demand objects implies the need for data processing and workflow control, with a virtualised description of such steps. Because of this the task will have close connections with task 2101, which includes preservation orchestration.

Task - 1208 Certification (Task leader UG)

It is expected that the RLG/NARA audit and certification draft will develop into a full standard. Task 1203 will track that development. This task however lays the groundwork for the accreditation and certification processes which will flow from that standard. Work will include interactions with the RLG and NARA Task Force.

Deliverables

D1201: **CASPAR** Conceptual Model (Month 12)

D1202: **CASPAR** guidelines (Month 12)

Milestones and expected results

M#3: Review of **CASPAR** Methodology, Conceptual Model and Guidelines (Month 12)

Workpackage number	**1300**	Start month							3
Workpackage title	**Overall architecture**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	7.0	3.0	0.0	0.0	0.0	4.0	0.0	6.0	6.0
Participant	**ENG**	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	**14.0**	7.0	6.0	0.0	0.0	0.0	0.0	6.0	

Objectives

The objectives of this WP are:

- to collect, analyse and specify user and system requirements
- to specify system functionalities and architecture design
- to specify implementation features and any type of interface about the components

This work package influences the WP1100, WP1200, WP2100, WP2200, WP2300 and WP2400 and will be carefully co-ordinated with them.

Description of work

The goal of this workpackage is to develop the overall component architecture and specifications for the component model of **CASPAR**, ensuring the seamless integration of **CASPAR** components during the lifetime of the project

Task 1301 - Overall component architecture design (Task leader ENG)

The objectives of this task will be to collect and analyse the user requirements and design the overall component **CASPAR** architecture.

The process will involve analysts from project team who will elicit, analyse and transform these requirements into a formal specification. In addition, existing communities systems, processes, documents, and procedures will be reviewed and included as appropriate. From the available information the analyst will develop a list of the actors, use cases and system level requirements for the project.

Based on the gathered user and system requirements, the overall component architecture will be designed, as well as the functions and features that the new system will have, and how users would interact with the system will be described, based on the use case analysis.

Task 1302 - Component model (Task leader ENG)

Starting from the overall architecture design, the component model will be designed, including the class diagrams, and the interfaces and interactions specifications, hardware, system architecture will be described, including hardware, external interfaces and network requirements. Preliminary data elements, user interfaces, outputs and other interfaces will also be described.

Deliverables

D1301 Overall component architecture & Component Model (Month 12)

Milestones and expected results

M#4: Architecture Review (Month 15)

M#4: Component model specification reviewed (Month 15)

Workpackage number	**2100**	Start month							1
Workpackage title	**OAIS-based components**								
Participant	**CCLRC**	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	**15.7**	4.0	1.0	0.0	3.1	4.8	0.0	2.6	2.9
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	1.4	4.8	11.5	12.5	0.0	4.8	1.4	3.2	

Objectives

The objectives of this WP are to produce the fundamental implementations of the OAIS-unique components including those dealing with Representation Information.

The implementations will follow the **CASPAR** conceptual model interim draft, and the **CASPAR** guidelines.

Description of work

Task - 2101 OAIS-based structures and orchestration (Task leader CCLRC)

This task will produce implementation of the components such as the Representation Information Registry and the Key-Store.

Task 2102 Access tools to support virtualisation (Task leader ACS)

Here the Simple, Complex and on-demand virtualisation will be implemented

Task 2103 Knowledge manager (Task leader FORTH)

The objective is to develop strategies to deal with long-term preservation of knowledge associated with archives, in order to ensure their accessibility and intelligibility. Knowledge is expressed through various metadata coming from various communities. An issue is to harmonise these levels into a common framework allowing representation, access and evolution of these metadata. Our basic hypothesis is to rely on Semantic Web languages and formalisms: The SWeb (Semantic Web) technology (representation languages and query services) is very appropriate for this task due to its (1) expressiveness, namely ability to encapsulate existing metadata schemata and ontologies, as well as new ones; (2) formal well-foundedness, derived from Description Logics; (3) computational amenability; (4) worldwide scope and impact.

The work will consist in the following issues:

- Translation of the conceptual model for digital object preservation (PDI descriptions, knowledge of Designated Communities) that will be developed in Stream 1, into standard SWeb languages (i.e. RDF/S, OWL lite). Issues concerning the methodology for specialising and instantiating these ontologies will be elaborated.
- Supporting Ontology and Metadata Evolution. Preservation is a dynamic process. For this reason we will review and extend the existing methodologies and techniques for ontology and metadata evolution and we will devise techniques for supporting evolution in a space- and time-efficient manner.

Knowledge is one key factor for preservation according to the OAIS methodology. The Knowledge Management (KM) component will be responsible for the management of the metadata which are required by the OAIS model for the preservation of digital objects, and also the metadata required by other tools including the Certification Assistant and Orchestration Manager and in various other parts of the **CASPAR** Framework.

For instance, the OAIS Information Model comprises cyclic associations (allowing chains of unlimited depth at the object level), aggregation hierarchies, and generalization hierarchies (e.g. for the Software/Standards/Algorithms descriptions of Representation Information). Moreover complex digital objects are composed of multiple simple digital objects which are typically "multi-sourced" data coming in different formats and different versions (multi-formatted data). For this reason only a powerful knowledge

representation framework could address the needs of the current and future preservation environments and could increase the potential and impact of the **CASPAR** results.

Description of Semantic content is of course open-ended; this is the reason that OAIS introduced the Designated Community and associated Knowledge Base, in order to halt the recursion.

We potentially have the challenge of open-endedness in all knowledge-related areas. In some areas we will specify concepts analogous to the Designated Community to provide a natural "barrier", in other areas we will be limited by the resource available. We will ensure that we provide adequate Semantic capabilities throughout the **CASPAR** Components and **CASPAR** Framework to enable us to support our testbeds, and we recognise that this could limit the depth to which we can research in any one area.

Deliverables

D2101: Prototype OAIS-infrastructure and associated draft reports of knowledge management architecture and tools (Month 18)

Milestones and expected results

M#5: Review of OAIS infrastructure (Month 18)

Workpackage number	**2200**	Start month							1
Workpackage title	**OAIS-based Storage**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	**IBM**	ASMX	MW
Person months	0.0	0.0	0.0	0.0	0.0	1.5	**22.1**	0.0	4.0
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	0.0	0.0	2.7	0.0	0.0	0.0	0.0	0.0	

Objectives

This workpackage deals with the innovative concept of object-based storage and its applicability to preservation environments.

The objective of the workpackage is to raise the abstraction presented by today's data stores and create a common interface across all storage layers that support preservation-specific functions. Instead of traditional interfaces (such as file systems), it will enable the creation of self-managed, shared and secured data objects by moving lower-level functionalities such as metadata management and authorisation to the storage itself. It will support the virtualisation of complex information objects in a distributed environment. This interface will:

- Build on the new emerging object storage device (OSD) standard and technology.
- Extend the standard to support special functionalities and structures needed for a preservation environment, based on the OAIS model.
- Contribute preservation-specific enhancements to future versions of the OSD standard.
- Encapsulate existing data stores with the extended object-store layer.

Description of work

Task 2201 - Object store interface specification (Task leader IBM)

The OSD standard defines a basic object datastore API that is well suited for a distributed application. This API will be extended to support special needs for preservation environments. These include:

- Metadata extensions (in the form of attributes) to support an OAIS-based architecture
- Special functionalities, such as collections of objects and multi-object operations, to support virtualisation of complex objects

These extensions will be the basis for a specification for extending the OSD standard to address preservation needs.

Task 2202 - Encapsulate data stores (Task leader IBM)

The user data that is the basis for **CASPAR** testbeds, whether scientific, cultural or artistic, already resides in existing standard datastores such as file systems and tapes. The objective of this task is to develop a prototype of the extended Preservation Object Datastore interface to facilitate access to the existing data via the new API. This software will serve as an abstract layer on top of existing datastores. With this approach, there is no need to copy any data from where it currently resides; yet, it will allow object-based access to the data and metadata.

The prototype will provide a programming interface (in C) to be used by other software components, such as the virtualisation and possibly the registries.

Deliverables

D2201: Prototype object data store interface (Month 18)

Milestones and expected results

M#4: Review specification for a Preservation Object Datastore interface (Month 15)

M#5: Review prototype of an extended object data store interface, based on the specification, and integrated within a distributed protocol, will be released. This abstract layer will encapsulate an initial set of data at one of the scientific testbeds. (Month 18)

Workpackage number	**2300**	Start month							3
Workpackage title	**OAIS-based access**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	2.8	0.0	5.7	0.0	3.4	0.0	0.0	1.7	14.5
Participant	ENG	CNR	Forth	**CNRS**	INA	IRCAM	UL	CIANT	
Person months	0.0	19.4	8.5	**3.1**	0.0	8.6	1.0	1.6	

Objectives

The objectives of this WP are to specify and model knowledge components that will be based on the **CASPAR** Conceptual Model, and the **CASPAR** Framework architecture in order to implement a practical OAIS Access entity, using state of the art techniques and beyond, but always ensuring usability over the long term. Several difficulties should be overcome while designing tools for long-term preservation: technological changes, cultural shifts, rights and law evolution, etc. **CASPAR** should then be precise and rigorous without being dependent on a particular technique or ad-hoc solution. The approach adopted in WP2300 is to model knowledge using semantic-centric formalisms.

Several domains of knowledge should be addressed here, going from content to its more abstract representations and interpretations: identity of content (task 2301) persistent, unique and resolvable identification, authenticity management (task 2302), knowledge management (task 2303), DRM (task 2304) and knowledge interpretation (task 2305).

Description of work

Task 2304 - Unique identification, name resolution and resource discovery mechanism (Task leader @Semantics)

The objective is to develop a uniform and robust system for persistent, unique and actionable identification of content pieces. **CASPAR** contents are complex and composed of long-lived, heterogeneous pieces of information content. Capture and ingestion of content will be a progressive task with several problems where redundancy and conflict should avoided. Only a persistent and actionable identification system will provide a long-term insurance for a secure evolution of archives over time.

The work will consist of the following issues:

- Review of Persistent Identification systems
- Development of a persistent, unique, actionable identification model;
- Integrating the case of complex objects having several components with tight logical relations but more smooth physical dependencies.
- Extending the model to virtual objects that are reconstructed on demand, according to user needs.
- Implementing this model by the development of mechanisms which can be relied upon for persistence and resolvability.

Task 2303 - Authenticity management tools (Task leader UU)

The objective is to develop tools that ensure integrity and authenticity verification of contents and contextual information along the preservation process. Authenticity management monitors and manages protocols and procedures across the value chain to deliver the benefits of authenticity in production, processing and preserving systems. The most critical issues are the right attribution of authorship, the identification of provenance in the life cycle of digital resources, the insurance of content integrity of the digital components and their relevant contextual relationships and the provision of mechanisms and information to allow future users to verify the authenticity of the preserved objects or, at least, to provide the capability of evaluating their reliability in term of authenticity presumption.

The work will consist of the following issues:

- Authorship attribution mechanisms and provenance control

- Content and contextual relationships integrity control mechanisms
- Annotation process

Task 2302 - DRM (Task leader MW)

The objective is to model DRM policies, leveraging emerging standards on Semantic description and inference.

The work will consist in the following issues:

- A DRM ontology will be modelled after current state-of-the-art DRM policy declaration standards such as MPEG21-REL and ORDL, which correctly identify main entities in intellectual property management domain, but lack the dynamic nature necessary to face the mentioned emerging challenges in rights management oriented to long term preservation.
- Application to the description of digital rights with a specific focus over preservation. A specific set of tools will be integrated as a component in the **CASPAR** framework for the declaration of rights and policies for long-term storage, along with all the necessary features needed for assembling distribution policies into Dissemination Packages.

Task 2305 - Information access (Task leader CNR)

The objective is to implement a component supporting the access to the information stored in an archive endowed with long-term preservation information. Access means both generic querying and browsing functionality, defined on the basis of the OAIS information model, and scenario-specific functionality, taking into account the peculiarities of the scenarios data.

The work will address the following issues:

- Specification of a query language for querying an archive based on the content, the structure and the semantics of the stored information. The language will be built by integrating existing query languages defined on the different types of information.
- Design and implementation of a distributed query evaluation mechanism implementing the previously specified query language
- Design and implementation of a browsing mechanism for visually exploring the information space of the archive.

Deliverables

D2301: Report on OAIS-access model, including DRM, Authenticity and Persistent Identifiers (Month 18)

Milestones and expected results

M#5: Review of specifications (Month 18)

Workpackage number	**3100**	Start month							3
Workpackage title	**Framework Architecture**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	**ACS**	IBM	ASMX	MW
Person months	0.0	0.0	0.0	0.0	0.0	**13.0**	0.0	0.0	6.0
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	4.0	0.0	4.0	3.0	0.0	3.0	0.0	0.0	

Objectives

The objectives of this WP are:

- to define the technical environment for SW development
- to identify design and development standards and methods
- to design and implement the preservation process flow

The WP is linked with WP1300, WP3200, WP3300 and WP3400.

Description of work

The activities of this WP will aim at the definition of SW standards, procedures and tools to design, implement and verify the process flow.

Task 3101 - Basic Middleware (Task leader ACS)

This task will be focused on the definition of standards and methods that will be adopted in the SW development within the project. The SW infrastructure needed for the design, implementation and evolution of the system will be defined and justified. Different mechanisms for data exchange among the system's modules will be analysed, evaluated and selected.

A particular attention will be paid to the existing components to be re-used in the project, in order to maximise the effectiveness of their integration and to harmonise them with the SW infrastructure and with other components.

Task 3102 - Process flow design and management (Task leader ACS)

The system process flow will be analysed in detail, in order to identify the optimal phasing, the key actors and to limit possible weaknesses. The technical alternatives for the management of process flow will be taken into account. The baseline is to use BPEL to define the process flow, supported by a process flow management tool, open source if possible.

Deliverables

D3101: Report on Framework Architecture (Month 18)

.

Milestones and expected results

M#3: Draft Framework Architecture completed

M#5: Framework Architecture completed

Workpackage number	**3200**	Start month							9
Workpackage title	**Integration of existing component**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	**ACS**	IBM	ASMX	MW
Person months	5.5	2.9	0.0	0.0	0.0	**4.8**	6.2	1.3	7.7
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	4.3	2.9	2.8	2.1	0.0	0.0	0.0	1.1	

Objectives

The objectives of this WP are:

- to adapt and customise existing components for integration
- to coordinate the distribution of components among the project actors and on the different platforms
- to integrate the components in the framework

Interactions with WP3100, WP 3300 and WP3400 will be properly managed.

Description of work

This WP will be organised in different activities that will allow the integration and growth of **CASPAR**'s component based architecture.

Task 3201 - Acquisition and customisation of existing components (Task leader ACS)

Existing components will be analysed and evaluated with the aim of integration in the system process flow. Existing components will be adapted and customised to match with the SW framework of the **CASPAR** system. Customisation will be oriented toward the full exploitation of each component's features keeping in mind the possible evolution of the system's functionality.

Task 3202 - Distribution coordination (Task leader ACS)

Components will be distributed over different physical location to achieve the needed level of redundancy and to optimise the preservation process.

This activity will be designed, implemented and tracked by means of a rigorous method and using specific management tools.

Task 3203 - Component Integration (Task leader ACS)

The actual component integration will be carried out by inserting the components in the system framework and by verifying the functionality of integrated components. The integration will permit the full exploitation, in the preservation chain, of each acquired component and will be documented by means of a detailed report, constantly updated during the integration phase.

Deliverables

Integrated into D3101.

Milestones and expected results

None in this period

Workpackage number	**3300**	Start month							16
Workpackage title	**Framework technical testing**								
Participant	CCLRC	**ESA**	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	0.0	**0.3**	0.0	0.3	0.0	0.0	0.0	0.0	1.3
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	0.7	1.0	1.3	0.0	0.5	0.5	0.0	0.0	

Objectives

The objectives of this WP are to test the **CASPAR** framework according to

- Functional requirements and
- Performance requirements

Interactions with WP3100, WP 3200 and WP3400 will be properly managed.

Description of work

Task 3301 - Functional Testing (Task leader MW)

Existing components will be tested according to the functional requirements which were captured in WP1100 and WP4100.

Task 3302 - Performance Testing (Task leader ENG)

The performance of the components and framework will be tested to ensure that it is adequate for the testbed requirements, for example that it is fast enough, scalable enough etc.

These activities will be designed, implemented and tracked by means of a rigorous method and using specific management tools.

Deliverables

No deliverables because workpackage started in month 16.

Milestones and expected results

None in this period

Workpackage number	**4100**	Start month							1
Workpackage title	**Generic testbed elements**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	**ASMX**	MW
Person months	3.8	8.4	3.0	3.0	6.0	6.0	0.0	**10.0**	0.0
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	2.8	0.0	3.0	6.0	3.0	5.0	3.0	5.0	

Objectives

- prepare a *common set of services* for identification, resolution and on-demand aggregation of data make a *cross-testbed experiment* with emphasis on persistence of the solutions developed in **CASPAR**;
- validate system-wide solutions common for all test-beds;

Description of work

Task 4101 Use cases and requirements definition (Task leader UG)

This task is dedicated to the development of a set of use cases and scenarios from each of the testbed areas. It will involve detailed interactions with a large number and variety of users. Special care will be taken to include the requirements of the "unborn users".

Intercomparison of these use cases will be used as input to the identification of generic services (Task 4102), and also input for each of the separate testbeds

Task 4102 Generic testbed design and specifications (Task leader ACS)

The use cases and requirements from Task 4101 will be used to identify and develop the common services required.

Task 4103 Evaluation method definition(Task leader CNRS)

This task will define the evaluation criteria for the evaluation of the testbeds.

It will also prepare the integrated report of the R&D activities from each individual testbed, in order to further identify commonalities.

Deliverables

D4101: User requirements and scenario specifications (Month 6)

D4102: Integrated report of R&D activities on the Cultural, Performing Arts and Science data preservation testbeds (Month 18)

D4103: ERC report on validation methodology (Month 12)

Milestones and expected results

M#2: Review of User Requirements (Month 6)

M#3: Review of common service elements (Month 12)

M#5 Review of revised evaluation criteria (Month 18)

Workpackage number	**4200**	Start month						7
Workpackage title	**Cultural data preservation testbed**							
Participant	CCLRC	ESA	UU	**UNESCO**	Glasgow	ACS	IBM	ASMX	MW
person months	9.6	0.0	0.0	**0.2**	0.0	2.0	0.0	8.6	0.8
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	0.0	0.8	0.8	0.0	0.0	0.0	0.0	0.0	

Objectives

The objectives of the Work Package will be to do R&D identifying pilot test cases out of the UNESCO World Heritage List as well as identifying the most representative data and information for each site.

This will be done by managing and coordinating a network of partners that jointly with UNESCO will be deeply involved in testing all the software solutions on specific cultural and natural heritage sites.

This general objective is to be reached through specific objectives:

1. Definition of a number of pilot test cases on natural and cultural heritage sites.
2. Identification and description of the most representative data types as well as media on which the data is stored.
3. Identification, description and documentation about the most common software packages being used to visualize the selected data and information.
4. Establishment of negotiations and agreements with UNESCO partners with respect to data access and data sharing to set up the process for the **CASPAR** testbed.

Description of work

Management, liaison and coordination within the UNESCO network of partners. Liasing with UNESCO partners in order to identify how access to data residing outside UNESCO will be provided to the **CASPAR** project for testing.

In particular UNESCO will act as the main contact point of the project with the data holding sites, and will ensure that the **CASPAR** development team can be prepared for the different types of data and information that will be used during the testbed.

The tasks envisioned to reach these objectives are :

Task 4201 : Testbed specific R&D activities (Task leader UNESCO)

- R&D on the World Heritage List, selection of site pilot test cases. The intention is that sites will be selected which are well distributed geographically, but giving emphasis to European sites for which there is an enormous amount of widely distributed data.

- R&D with respect to the most representative types of data and information as well as the associated media for storage.

- Examine the process of production of complex world heritage site information

- Develop interfaces to these data and tools to enhance the cultural experience of users over the long-term.

Task 4202 : Testbed infrastructure and set-up (Task leader @Semantics)

- R&D with respect to computer packages required to visualize the selected data.

- User's interface design and implementation

- Develop specific agents in order to help in the gathering of the required elements of production and to describe relations between these (including automatic production of documentation from specific production tools).

- Analysis tools: Develop specific agents in order to help in the analysis and enrichment of content, (intra-document browsing, synchronizing, annotating,...)

Deliverables see D4102 – integrated report
Milestones and expected results

Workpackage number	**4300**					Start month			7
Workpackage title	**Contemporary art testbed**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
Participant	ENG	CNR	Forth	CNRS	INA	**IRCAM**	UL	CIANT	
Person months	1.0	0.8	0.8	5.0	5.8	**13.6**	8.4	4.8	

Objectives

The objectives of the WorkPackage will be to validate the solutions developed throughout the **CASPAR** project, and to develop specific models and solutions to cope with specific cases.

This general objective is to be reached through specific objectives:

1 - Design of a generic data model for electronic and digital works, implementation of a digital dissemination prototype platform and validation on an existing corpus.

2 - Work porting and repurposing, by identifying a set of generic functions implemented when porting existing works to updated technology.

3 - Work abstracting, by identifying and formalizing a set of abstract operations in a given application context

4 - Work analysis mediation, by identification and implementation of a set of generic analysis structures and computer-aided analysis tools

5 - Ensure generality of proposed solution by considering for applicability and extension toward different creative contents.

Description of work

Task 4301 : Testbed specific R&D activities : Modelling(Task leader UL)

- Examine the process of production of complex artistic works

- Develop a model of description to help gathering required elements of production

- Develop a model of description for describing complex relations between elements of production

- Develop classification schemes for specific elements (instructions, instruments, equipments)

- Develop classification schemes and models of interaction with performers (motion gesture analysis, events modelling...)

- Participate in standardization efforts by defining formal models to be proposed to standardization bodies

Participants : IRCAM, INA, CIANT, Leeds

Task 4302 : Testbed infrastructure and set-up : Development (Task leader CIANT)

- User's interface design and implementation

- Develop specific agents in order to help gathering required elements of production and describe relations between these (includes automatic production of documentation from specific production tools).

- Analysis tools: Develop specific agents in order to help analysis and enrichment of content, (intra-document browsing, synchronizing, annotating, motion gesture analysis and segmentation...)

- Repurposing : Develop specific agents in order to help repurposing

Participants : All

Deliverables

see D4102 – integrated report

Milestones and expected results

Workpackage number	**4400**					Start month			7
Workpackage title	**Scientific data preservation testbed**								
Participant	CCLRC	**ESA**	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	8.4	**5.6**	0.0	0.0	0.0	12.0	2.4	0.0	1.6
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	0.0	1.2	0.8	0.0	0.0	0.0	0.0	0.0	

Objectives

The high-level objective of this WP is to provide a proof-of-the-**CASPAR**-concept for science data preservation by setting up, and implementing a science data preservation testbed taking into account the project findings and provide a source for feedback to other **CASPAR** work packages. In particular, this science data preservation testbed will assess, develop and prototype an ESA European Remote Sensing Satellite (ERS) – 1 Synthetic Aperture Radar (SAR) science data preservation environment. The lower level objectives include:

1. The definition and set-up of an infrastructure for Earth science preservation testbed
2. The implementation of a science data preservation environment that takes into account
    a. simple and complex information objects
    b. on-demand information objects
3. Analysis of certain research areas of interest for the Earth science preservation community, in particular:
    - The preservation life-cycle of scientific data
    - Adoption of the OAIS Reference Model for Science data preservation
    - The SAFE archive format standards for Earth observation missions.
    - Compatibility with OGC efforts for geospatial data
    - Digital library integration
    - Metadata, Ontologies and semantics of science information objects
    - The use of the Grid and on-demand (re)generation of information objects
    - Distributed partial copies of Earth science information objects

Description of work

The science data preservation testbed workpackage will carry out R&D activities specific to the testbed, it will consider the infrastructure, set-up the testbed and implement it. It will focus on information related to a specific satellite instrument, namely the SAR (Synthetic Aperture Radar) instrument on board of the ESA ERS-1 (European Remote Sensing) satellite mission that was operated between 1991 and 2002. Given ESA has the mandate to archive and maintain all ESA controlled mission associated data access and processing services for 10 years after the mission end, for the specific case of ERS-1 SAR it will have to do so until the end of 2012.

This WP will work closely together with WP4100 to generate requirements and use-cases that are the basis for the infrastructure and set-up and testbed implementation. In particular, it will ensure the test-bed set-up and implementation to meet requirements of two particular high-level use-cases knowing:

- Monitoring and analysis of floods in Central Europe in 1993
- Monitoring and analysis of Etna eruptions over the last 20 years

Task - 4401 Testbed specific R&D activities (Task leader CCLRC)

This task will be oriented towards testbed specific R&D activities as identified for the science data preservation testbed and reflected in the objectives of this testbed that are presented above.

Task 4402 Testbed infrastructure and set-up (Task leader ACS)

This task covers all actions needed for the science data preservation testbed infrastructure and set-up.

The testbed infrastructure will be defined on top of the framework as being defined and integrated in Stream 3 and WP4102 taking into account the requirements as defined within WP4101.

Deliverables

see D4102 – integrated report

Milestones and expected results

Workpackage number	**5100**					Start month			1
Workpackage title	**Exploitation and take-up**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	1.5	0.6	1.4	0.6	1.3	0.8	0.0	0.8	4.6
Participant	**ENG**	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	**10.4**	1.5	0.0	0.0	0.3	0.3	0.0	1.1	

Objectives

This workpackage deals with, preservation cost modelling, market analysis, business planning and commercial exploitation for **CASPAR** services and product. After a market analysis, a preliminary business plan will be developed, addressing market entry and technical deployment strategies, choice of criteria for best co-operation partnerships, organizational structure of operations, and exploitation strategies. Assessment of main operational costs and revenue sources will be performed. Following the demonstrator stage, a final plan will be prepared.

The goal for this workpackage is the definition of the preliminary Sustainability Plan in order to understand the framework in which **CASPAR** will move in the future. The approach for the sustainability model is to keep together the theoretical understanding and the concrete working activities. Following this approach the activities will immediately start to aggregate and analyse **CASPAR** customers i.e. the digital preservation community.

Description of work

Task 5101 Preservation Cost Modelling(Task leader CCLRC)

Within this activity cost model for Digital Preservation will be identified and analysed. Through interviews of application scenarios' customers and other potential actors, traditional cost model for digital preserving will be identified and analysed. In this task the main user requirements according to the cost model will be identified. All these analysis will be based on:

- relevant and secondary needs and business requirements
- cost model and cost analysis, including an assessment of the main operational costs (installation, gearing to local needs, training, upgrading, maintenance, ...)
- decision making processes and economic flow analysis
- cost of real technology solutions adopted

Task 5102 CASPAR Preservation exploitation & business model (Task leader ENG)

This task deals with understanding the baseline framework on which business exploitation must be based. A market analysis and a general technological framework will be produced with input from the users of different application scenarios. In addition a deep analysis of all possible business model, including value chain actors will be conducted. These task will include:

- analysis of the overall market with main evolution trends foreseen and main competitors
- analysis and definition of the possible Business Model addressed by **CASPAR**
- definition of the future evolution with respect to the preservation technical standards

Task 5103 Multi-industry Preservation Socio economic study(Task leader ENG)

This task is devoted a complete socio economic analysis and study, including a multi-industry value chain analysis, the social impact of digital preservation and a macro-economic analysis of possible institutional solutions and approaches. In line with this approach the task will include:

- understanding of institutional approach to Digital Preservation
- identification of opportunities for social investments in digital preservation
- analysis of any kind of industries involved in the preservation context

- Analysis and identification of technological main trends with economic impact

Task 5105 – Communication and valorisation strategies (Task leader MetaWare)

This task foreseen the realisation of several activities, namely the aggregation and analysis of the target communities involved in the preservation area, the assumption of users requirements and the elaboration and adoption of a Sustainability plan according with a proper Market Analysis and Success Indicators. A preliminary business plan will be also developed, addressing market entry and technical deployment strategies, choice of criteria for best cooperation partnerships, organisational structure of operations, and implemented exploitation strategies. Each partners' plan will be consolidated into a single preliminary exploitation plan for the **CASPAR** project, covering in detail the following points:

- Definition of the Market analysis by presenting the current situation
- Preliminary overview of potential competitors and existing services.
- Analysis of risks and definition of Success Indicators.
- Description of the **CASPAR** services and their positioning
- Preliminary communication strategies
- Possible exploitation strategies

Deliverables

D5101: Dissemination and Use plan (Month 12)

Milestones and expected result

M#4: **CASPAR** preservation cost model review (Month 15)

M#3: Review of market analysis

M#5: Review initial Exploitation & Sustainability plan (Month 18)

M#5: Review socio – economic impact report and review (Month 18)

Workpackage number	**5200**					Start month			1
Workpackage title	**Dissemination activities**								
Participant	CCLRC	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	**MW**
Person months	2.2	1.2	0.2	2.2	4.2	0.2	0.0	1.2	**8.3**
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	4.4	2.2	4.2	2.2	1.2	1.2	1.2	1.2	

Objectives

The objectives of this WP are the definition and coordination of the dissemination activities. Those activities will focus on:

- Introducing and promoting goals, activities and expected results of **CASPAR** to national and international bodies;
- Creating and defining dissemination strategies and differentiated approaches according to specific target users, with respect to the developing process of the project and user communities' requirements.
- To ensure consensus on the **CASPAR** strategy, goals and services using a variety of communication and aggregation tools (**CASPAR** preservation user community).
- To establish a **CASPAR** network of special interest groups around the preservation topics, participating in concertation meeting with other projects, workshops and researches in preservation.

Description of work

Task 5201 Dissemination planning and implementation (Task leader Metaware)

This task foresees the planning and the implementation of the dissemination activities to be undertaken during the project life cycle, defining different type of actions:

- To show the project results to the preservation audience and enlarge the **CASPAR** Community;
- To target specific audiences that can benefit from the results and entities involved in preservation standards;
- To define a variety of techniques and media as appropriate for the content to be delivered to the target audience

The implementation of the dissemination activities implies going further than simple dissemination actions such as participating in events and distributing dissemination materials to targeted audiences. It focuses on attract potential users/customers by creating a worldwide community composed by all those organisations interested in preservation of digital resources.

The results of the Project will be disseminated via the publication of articles and technical papers in the specialised press, the organisation of focused workshops. Peer-reviewed and more general articles will be produced by many of the partners (large and small), covering the many different aspects of **CASPAR**.

Task 5202 Dissemination Material (Task leader Metaware)

A preliminary dissemination action will consist in preparing an info-pack explaining the **CASPAR** project objectives and expected results and in organising target oriented events addressed to wide audience and experts.

Different type of dissemination materials will be produced at each phase of the project, including:

- a web site that will be continuously populated and upgraded during the life of the project,
- paper based material such as brochures, leaflets, posters, bookmarks and other striking and impressive

equipment to be produced according to the promoting needs that will be come out during the achievement of on-going and intermediate results.
Deliverables D5201: **CASPAR** Web site (Month 3) D5202: Preliminary Dissemination plan (Month 12)
Milestones and expected results M#1: **CASPAR** Web site launch (Month 3) M#2: **CASPAR** Dissemination Material – 1st release (Information booklet) (Month 6) M#3: **CASPAR** Dissemination Material – 2nd release (Poster and brochure addressed to general public) (Month 12) M#5: **CASPAR** Dissemination Material – 3rd release (technical customised brochures addressed to specific public) (Month 18)

Workpackage number	**5300**				Start month				6
Workpackage title	**Training Activities**								
Participant	CCLRC	ESA	**UU**	UNESCO	Glasgow	ACS	IBM	ASMX	MW
person months	0.9	1.0	**4.3**	1.0	1.4	0.0	0.5	0.0	5.4
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	1.0	1.5	0.5	0.5	0.5	0.5	0.5	3.4	

Objectives

CASPAR Training will:

- provide a conduit between the innovations of the project and the exploitation and user communities to ensure that the **CASPAR** methods, products and services are as widely taken up;
- offer training opportunities to staff in cultural, artistic and scientific heritage institutions such as archives, libraries, data archives and universities to promote the take-up of **CASPAR** methods, practices, and technologies;
- provide capability building training for consultancies, SME technology developers, and vendors;
- provide certification for those trained to validate their use of planning tools, templates, construction of new testbeds, and content characterization.
- define the training infrastructure specifications. Based on the key audiences identified for the training, the user and system requirements will be collected in order to provide a services platform tailored to the "user needs"

This work will be co-ordinated with other EU projects.

Description of work

Task 5301: Training Planning (Task leader UG)

The **CASPAR** team will begin by assessing the relationship between current community capabilities and the skills that would need to be developed if the community is to be prepared to take full advantage of the methods and technologies that **CASPAR** is delivering.

Definition of the courses that **CASPAR** will run and how they will inter-relate. This will incorporate a statement of the overall training programme objectives and the detailed objectives for each course and the different learning and teaching methodologies which will be adopted to ensure the course achieve their intended learning outcomes, a schedule of course delivery locations, dates, and timings.

Task 5302 Training infrastructure (Task leader Metaware)

The goal of this task is the realization of a technological infrastructure supporting the training activities and providing a number of integrated tools to the involved actors. These services could be accessible through a web based portal. Two types of service will be deployed in this framework: contents management and post lessons support.

The first one will be a collaborative space where teachers will manage and provide lesson documentations, calendar and so on, and where the trainers could access archived lessons.

The Post Lessons Support will be an interactive environment where the actors (trainers and trainees) will have access to a set of communication facilities in order to expand their acquired knowledge and to strengthen the relationships created between workgroups and teams.

In the first 18 moths, this task will produce the definition of the training infrastructure specifications.

Task 5303: Training material preparation (Task leader Metaware)

Design of the **CASPAR** training modules including agreeing the **CASPAR** Training Methodology, preparation of briefing documents, drafting introductions for trainers, establishing training schedules, and detailed course programmes. This will include the creation of online materials to enable training

participants to prepare for the training courses and to allow them to take advantage of training aftercare. ***Task 5304: Training Delivery and Evaluation Plan*** Delivery of the first training courses and the awarding of certificates to those who successfully complete the training.
Deliverables D5301: **CASPAR** Training Plan (Month 18)
Milestones and expected results M#5: Review of preliminary specifications of the platform prototype (Month 18)

Workpackage number	**6100**					Start month			1
Workpackage title	**Project management and coordination**								
Participant	**CCLRC**	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
person months	**19.2**	0.9	0.0	0.0	0.0	0.9	0.5	0.5	0.9
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	

Objectives

- To provide overall project management and co-ordination.
- To ensure the quality management and assurance.
- To maintain the information flow between partners.
- To provide administrative and financial control according to work plan.
- To co-ordinate the dissemination and promotion activities and to present the project towards the European Commission.

Description of work

Task- 6101: Administrative and Financial co-ordination (Task leader CCLRC)

Establish and maintain financial records. Co-ordinate cost statements submission by all project partners, follow-up of EC payments, distribute partner shares according to consortium agreement agreed rules.

Maintain contractual documents (Consortium Agreement)

Coordinate internal and contractual periodic reporting

Organise and manage audits requested by the Commission. Of particular importance is the mandatory mid-term and end-of-term review, and technical or financial audits upon request by the Commission

Organise periodic project meetings and committees (IP Executive management committees, ERC etc) for project progress review, decision-making and conflict resolution.

Task 6102: Scientific and technical co-ordination (Task leader CCLRC)

Coordinate the technical activities; monitor the work being carried out, the results and the necessary changes to the work plan.

Organise Quality control, establish and benchmark project milestones. Co-ordinate timely production of deliverables, organisation of reviews, control of quality and consistency against technical and contractual aspects

Task 6103: Project management (EPM) (Task leader CCLRC)

Organise project launch: establish procedures, project management methods and tools (IP Management organisation and processes, quality plan, conflict resolution); prepare and organise project kick-off meeting.

Review and verification of business opportunities, exploitation, and use plans via the EAC

Handle IPR aspects and confidentiality

Co-ordination and controlling of each partner's results in order to allow for effective and efficient internal information delivery (e.g. deliverables).

Maintain and monitor the work plan, monitor project progress, identification and trouble shooting of

technical and organisational problems, technical co-ordination meetings

Task 6104: Web site, project management support (Task leader CCLRC)
Produce and maintain the **CASPAR** management web site

Deliverables

D6101: Consortium and Project Management Procedures and Quality Plan (Month 3)

D6102: Management Web site (Month 3)

D6103: Updated Management Plan (Month 12)

D6104: Activity Report (Month 12)

D6105: Management Report (Month 12)

D6106: Updated Implementation Plan for months 19-42 (Month 18)

Milestones and expected results

M#1: Consortium Agreement (Month 3)

M#5: Review of first phase Project Report, including the revised Project Implementation Plan for Months 19-42 phase (Month 18)

Workpackage number	**6200**					Start month			1
Workpackage title	**Project performance monitoring**								
Participant	**CCLRC**	ESA	UU	UNESCO	Glasgow	ACS	IBM	ASMX	MW
Person months	**2.7**	1.7	1.7	1.7	1.7	0.0	0.0	0.0	14.1
Participant	ENG	CNR	Forth	CNRS	INA	IRCAM	UL	CIANT	
Person months	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Objectives

To provide allow information on project performance to be gathered and organised in order to provide ensure that **CASPAR** is achieving its goals

Description of work

Task 6201: Key Performance Indicators (Task leader CCLRC)

Identify, review and organise the Key Performance Indicators (KPI) by which **CASPAR** performance will be measured. An initial set KPI organisation is set out in section 7.

Task 6202: Project Performance self Assessment tool (Task leader MetaWare)

Design and develop the Project Performance Assessment (PPA) tool which will record and analyse the KPI measures and provide rapid feedback to the project and the Commission. A Balanced Scorecard performance methodology will be used. The PPA tools will be accessed through the **CASPAR** internal web site which will be available to **CASPAR** project members and also to the Commission.

Task 6203: Project monitoring, self-assessment and risk management (Task leader CCLRC)

This is a continuous task aiming at reviewing KPI definitions and at collecting their measures by means of the PPA tool described in the previous task. This activity also has the task of monitoring and triggering critical situations and alerting the Project Management and Coordination in order to activate the appropriate mitigation and/or remedial actions. The PPA Tool, which will be accessible to the **CASPAR** constituency

Deliverables

D6201: **CASPAR** KPIs definition, organisation and metrics (Month 6)

D6202: Balanced Scorecard-based PPA tool design and implementation and the tools accessed by authorised users continuously feed with KPIs evaluations (Month 6)

Milestones and expected results

M#2: Review of KPI (Month 6)

9. Consortium

To carry out the ambitious **CASPAR** project, a consortium has been brought together which covers these important scientific and cultural data holdings, with the appropriate extensive scientific and cultural expertise, together with commercial partners with proven track records in the fields of knowledge engineering and modelling, including virtual reality software, and international leaders in the field of information preservation.

The **CASPAR** consortium has been established in order to conform to the following criteria:

- The excellence of partners at the International level in the key research areas addressed by the proposal, with complementary areas of competence to cover the demands of the project;
- Guarantee strong synergies with European and National Research programmes realised by each Member State, with partners being deeply involved in similar national and international projects;
- Guarantee an adequate level of manageability;
- Include the main value-chain actors in the digital preservation domain;
- Guarantee an adequate level of results exploitation and dissemination.

These criteria have been satisfied by selecting Organisations that have a long experience and a strong and recognised expertise in at least one of the following areas: preservation, knowledge management, access techniques, advanced storage technologies, software, exploitation, or content provision technologies. The distribution of expertise may be shown as follows:

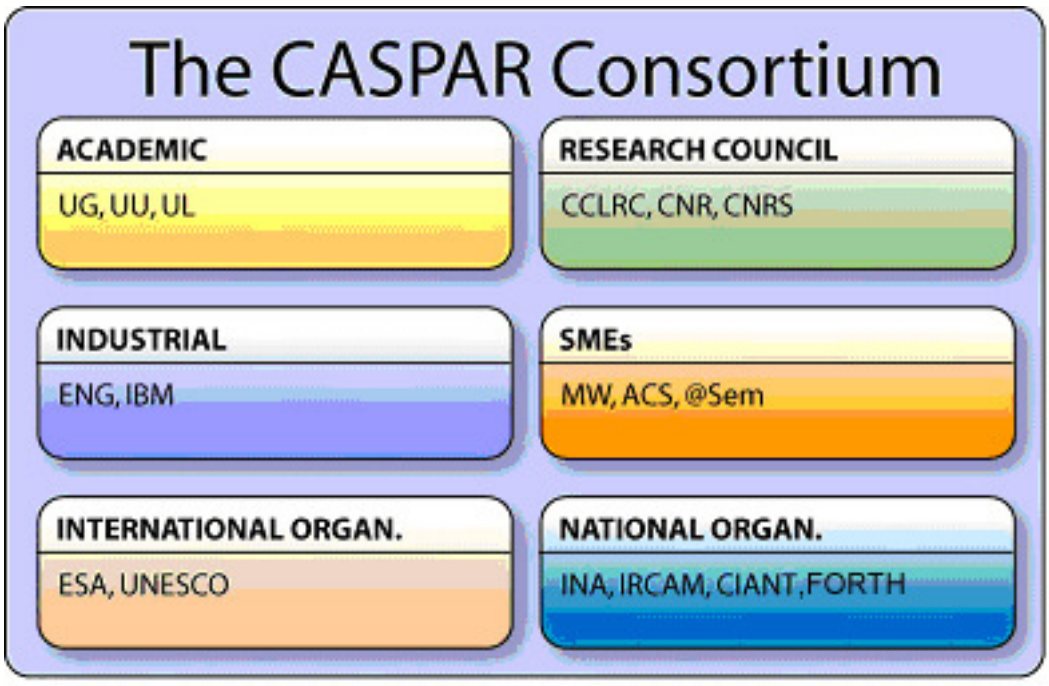


Figure 13 The CASPAR consortium structure

Many of the key personnel in these organisations are active members of international standardisation bodies (CCSDS, ISO, W3C, GGF, etc.). Moreover, the Consortium's members have long experience with digital preservation research projects also at National level, providing a solid "knowledge base" about features and drawbacks of existing solutions. The fulfilment of the second criterion is a direct consequence of meeting the first one.

9.1 CCLRC

The primary work areas for CCLRC will be providing the IP Coordinator, namely Dr David Giaretta, who has extensive credentials in digital preservation including Chair of the CCSDS Panel which originated the OAIS Reference Model, Associate Director (Development) of the UK Digital Curation Centre, member of the RLG/NARA Digital Certification Task Force, and leader of many research, archive and software projects.

The technical work of CCLRC will focus on *Conceptual Modelling and Standards* and the *Virtualisation Process* - areas where it has well-proven track records.

The CCLRC will also provide part of the Science test case for the testbed using current data environments. It hosts research environments for an extensive user community across the spectrum of science and engineering in the UK, and provides services to around twelve thousand visiting scientists each year.

Key people: **Dr David Giaretta MA MSc MBE** has had extensive experience in planning, developing and running scientific archives and providing and managing a variety of services to large numbers of users.

He chaired the panel which produced, and has made fundamental contributions to, the OAIS Reference Model which forms the basis of much digital preservation work far beyond repositories of scientific data. He is helping to developing the OAIS follow on standards; he is a member of the Digital Repository Certification Task Force. He co-organised a workshop entitled "Digital Curation: digital archives, libraries, and e-science" in 2001, bringing together, in many cases for the first time, a variety of communities with this common interest.

He has published a number of scientific papers in refereed journals and given presentations at many international conferences, scientific as well as technical. In addition he has broad experience in e-Science.

Dr Giaretta was awarded his degrees from Oxford University and in 2003 he was awarded an MBE for services to Space Science.

Dr Giaretta is Associate Director for Development in the DCC.

Responsibilities: IP Co-ordination, Methodology, Conceptual modelling

9.2 ESA

The European Space Agency (ESA) has the mandate to shape the development of Europe's space capability and ensure that investment in space continues to deliver benefits to the people of Europe. ESA's headquarters is situated in Paris and several technology centres are distributed across Europe and elsewhere.

ESRIN, the ESA establishment in Italy, operates as the European reference centre for Earth Observation (EO) payload data exploitation, acting as the central hub of a

network of globally distributed archiving and ground stations that receive, process and archive satellite EO data and distribute these to thousands of users worldwide. These activities focus on the development and operations of a trans-European and international EO payload data handling infrastructure, aimed at providing data and services to the European and global science, operational institutional and commercial user communities.

ESA is involved in several major initiatives towards a better exploitation of space EO data, such as GMES, a joint EU-ESA initiative, that aims to establish a European capacity for the provision and use of operational information for Global Monitoring of Environment and Security, to be operational by 2008. ESA has knowledge of a high number of ESA-funded technology based projects. Of particular interest are SpaceGRID, THE VOICE, and the ESA Grid on-demand portal. ESA also participates in various EC-funded technology oriented initiatives like DILIGENT.

Key people: Luigi Fusco, Senior Advisor for EO Applications, EO Science and Applications Department

Responsibilities: In CASPAR, ESA plays the role of both user and infrastructure provider for one of the proposed testbeds. ESA will make available the existing Grid and e-collaboration infrastructure at ESRIN, including the Grid on-Demand web portal via which access to selected ESA catalogue data, archived data, services and products will be given.

9.3 University of Glasgow (UG)

Humanities Advanced Technology and Information Institute (HATII) University of Glasgow. HATII is a world-leading Institute researching the application of advanced technologies to the cultural and scientific heritage and in the area of digital curation and preservation. It is home to ERPANET.

Key people: **Professor Seamus Ross** is Director of the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, Associate Director of the UK's Digital Curation Centre, and leader of the Digital Preservation Cluster of the DELOS Network of Excellence. Since 2001 he has also been Principal Director ERPANET (Electronic Resource Preservation and Network) (IST-2001-32706) and has been a lead partner in The Digital Culture Forum (DigiCULT Forum, IST-2001-34898), which monitors and assesses the research and technological developments in and for the cultural heritage sector in Europe. He has more than twenty years experience as a practitioner and researcher in information communication and technology and has for fifteen years been investigating problems associated with digital preservation and curation. He co-Chaired with Margaret Hedstrom of the University of Michigan the EU(DELOS)/NSF Workgroup on Digital Preservation and Archiving. Its report is Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation (2003) proposed an innovative research agenda in the area of digital preservation. His research focus on digital preservation including work on metadata, repository design, digital library design and services, ingest, and semantic metadata extraction. He earned his doctorate from Oxford University. Responsibilities:

Responsibilities: The University of Glasgow has the responsibility for the CASPAR Methodology and models, and also joint leadership of the Training workpackages. In

addition, Prof. Ross will play a pivotal role in linking Cultural data centres, and CASPAR's links to ERPANET and DELOS.

9.4 University of Urbino (UU)

The Institute, as part of the University of Urbino, organises special courses at master level and manages national and international research programmes in the area of archival science, electronic record-keeping systems, librarianship, information studies and legal studies for cultural heritage and digital preservation. The Institute has been involved in many national and international projects specifically related to the preservation environment: partner in ERPANET; partner in DELOS2 specifically in the Digital preservation cluster; as Italian coordinator in the InterPARES projects 1 and 2 ; in the NPACI environment with reference to the NARA-DOCT project for the preservation of Persistent Archival Objects; in the project funded by the Italian Ministry for University and scientific research in a industrial research project (3D Informatica, CNR, Consorzio Roma Ricerche) to run as testbed in evolving an Information Retrieval application; for the preparation of the dossiers on the policies for preservation existing at European level and on digital preservation disasters and emergencies for the International Conference of Florence on Digital Preservation /16-17 October 2003). The institute also runs many projects at application level.

Key people: **Mariella Guercio**, ISTBAL, Università degli Studi di Urbino, is a full professor in archival science and electronic record management at the University of Urbino where she entered in 1998.

For twenty years (1978-1998) she worked as State archivist for the Ministry of Cultural Heritage where she cooperated with the Authority for information technology in the public administration to define the Italian legislation related to the electronic recordkeeping systems. She chaired the ICA Committee on current records (1990-1992) and was part of the Committee on program management of the ICA itself. She chaired the Italian team for the international project InterPARES 1 (1999-2001) and she is part of the European team of InterPARES 2 (2002-2006). She has been a co-director of the European project ERPANET (2001-2004), a network for digital preservation. She runs many Italian projects funded by the Ministry for research and by Cultural Foundations in the area of digital preservation. She takes part of the DELOS project (2004-2007) with specific reference to the preservation workpackage. Since January 2006 she is part of the national committee created by the Ministry for Innovation to define the rules and the requirements for the preservation of the digital resources in the e-government environment. Since 2002, she is the director of the journal Archivi & Computer.

Responsibilities: The University of Urbino provides the leadership of the work on Authenticity and also shares responsibility for Training.

9.5 UNESCO

UNESCO is the United Nations Educational, Scientific and Cultural Organization.

Among a large range of cultural activities, UNESCO has always worked for the promotion of the conservation of the cultural and natural heritage of the World.

These two main items are protected under the World Heritage Convention.

UNESCO hosts the secretariat for the World Heritage Convention, the World Heritage Center (WHC). The World Heritage Convention calls for the protection of all types of natural and cultural heritage. In particular, the Convention inscribes sites in the World Heritage List, being sites that have a universally recognised value. The World Heritage Convention is working to ensure that future generations can inherit the treasures of the past.

The WHC handles all the documentation that is provided to support a nomination for a site to the World Heritage List, including additional documentation and updates required from the sites already inscribed in the World Heritage List.

The type of data required to completely document a cultural site (both cultural and/or natural) is complex. A minimum amount of this data has to be sent to UNESCO when a country wishes to propose a "site" for eventual inscription in the World Heritage List. However, the bulk of this data should theoretically reside at country level, even better at "site level", in order to be constantly improved and updated. This complex amount of data then becomes the main support tool for site conservation, and will provide the basis for one of the testbeds in CASPAR.

Responsibilities: UNESCO provides the Cultural data testbed and brings with it a large community of end-users for that testbed.

9.6 ACS

ACS Advanced Computer Systems SpA is a private Italian company developing software and technologically advanced solutions in the fields of Earth Observation, Large Digital Archives and Virtual Reality. In the field of VR, ACS developed 3D models of archaeological sites and end-to-end solutions for the exploitation of heterogeneous, multi-source, multi-dimensional data (EO images, geographical information, socio-economical data, meteorological data). ACS solutions in the VR field involve integration with advanced visualisation environments (Virtual Reality Theatres, featuring high-resolution, stereographic display). ACS is also very active in the research and development of interactive image mining tools that allow the intuitive retrieval in large image archives. For further information http://www.acsys.it/.

Key people: Ugo Di Giammatteo

Responsibilities: In the CASPAR project ACS will act as SW designer and developer and as system integrator. ACS will thus contribute to the implementation of the overall SW system and to the development and integration of the Virtual Reality tools in the system.

9.7 ASemantics (ASMX)

Asemantics is a European hi-tech SME software company whose mission is the development of Semantic Web technologies and applications for commercial and government markets. Their major customers are currently world leaders in Internet services, the Dutch and Italian governments and the European Space Agency. Company founders and partners are actively involved in open standards and open source development in the World Wide Web consortium, the Internet Engineering Task Force and Apache Software Foundation.

Key people: Zavisa Bjelogrlic, Libby Miller, Dan Brinkley

Zavisa Bjelogrlic has long experience in software/networking technologies with more than twenty years of a technical and management career in university/R&D, international organisations, industry and media/newspaper groups. Zac was developing web applications since the first days, mainly for the European Space Agency and for telecomunications industry. Dr Libby Miller has been part of the Semantic Web Community from the begining, working on a wide range of topics such as such as calendaring, query, vocabulary development and community building. Alberto Reggiori has more than 7 years of Web technologies experience. He has been working on metadata and RDF since 1999 and, before that, with XML and GIS technologies. He is the prime developer of RDFStore, a fast and native Semantic Web storage engine.

Dr Libby Miller has worked in the Semantic Web community for six years, `for five of these at the Institute for Learning and Research Technology at the University of Bristol, UK, and for the last year as Chief Scientific and Research Officer at Asemantics S.r.l.

She has been an active member of the Semantic Web Community from the beginning, working on a wide range of topics such as such as calendaring, RDF query, vocabulary development and community building. Dan Brickley and Libby developed the Friend of a Friend (FOAF) RDF ontology and community.

At ILRT she co-designed the SWAD-Europe European project and project managed it for a period. Since joining Asemantics she has been helping to design and plan commercial Semantic Web projects for data integration and data personalisation.

She is a member of the W3C Semantic Web Best Practices and Data Access working groups and an active member of the Semantic Web Interest Group. Libby earned a BA at Oxford University and a M.Sc and Ph.D from University of Bristol.

Dan Brinkley is a researcher, advocate and developer of Semantic Web technologies. He currently combines commercial work (through Semantic Web Vapourware Ltd.) with ongoing involvement in Web standards as a Visiting Fellow at the University of Bristol. Dan served for six years as a member of the W3C Technical Staff, and now brings over twelve years' experience of knowledge sharing technologies to his work on practical Semantic Web deployment.

Dan founded the Bristol University ILRT Semantic Web group in 1997, exploring RDF-based knowledge engineering issues from a digital library, Social Science and e-learning perspective. Dan was a PI on the HARMONY multimedia metadata project, and played a leading role in IST-funded research at ILRT including early work on RDF quality labelling, thesaurus, query and rules in the DESIRE project. Dan was Technical Director of the MedCertain (EU Safer Internet) project, applying W3C RDF technology to problems of quality and trust in consumer-facing health/medical Web content.

Responsibilities: Asemantics leads the workpackage capturing the generic testbed elements and will play a significant role in developing the preservation and access components.

9.8 IBM (Haifa Research Laboratory)

IBM Haifa Research Laboratory[33] is the largest IBM research laboratory outside the United States, with 500 researchers.

The Reliable Systems department at HRL is at the forefront of storage technology worldwide. This department is responsible, among other IBM technologies, for the development of advanced data recovery and business continuance functions for IBM's flagship disk storage subsystems. The department has recently invented the iSCSI protocol, one of the most active areas in the storage industry today, and is active today in developing solutions in the area of security for storage.

Key People: The technical contributions of IBM to CASPAR will be supervised by Dr Dalit Naor, head of the IBM Object Storage team in Haifa, by Dr Michael Factor, Senior technical staff member, Julian Satran, an IBM Distinguished Engineer, and by Dr Yaron Wolfsthal, Senior Manager for Reliable Systems Technologies.

Dr. **Dalit Naor** joined IBM in1996 and since then has been working on security technologies for the Internet, content protection and storage systems. She is leading the Object Storage implementation activity at HRL. She takes an active role in standardization efforts in the area of security for storage, including the T10 OSD standard and the IEEE P1619 encryption for storage. Dr. Naor holds a BsC from the Technion - Israel Institute of Technology, and MsC and PhD from the University of California at Davis, all in Computer Science.

Dr. **Michael Factor** is one of the architects of advanced copy functions for IBM's DS family of storage subsystems. He is taking a leading role in storage related research in IBM's lab including topics such as distributed storage systems, SAN-NAS convergence and advanced storage networking technologies, storage management and object based storage. He chairs the security subgroup of the SNIA OSD standization effort. Dr. Factor holds a B.S. from Union College, Schenectady, NY and M. S., M. Phil. and Ph. D. in Computer Science from Yale University. Dr. Factor has worked at the IBM Israel Haifa Research Lab since graduating from Yale University.

Julian Satran is with the IBM Research Laboratory at Haifa since 1987. Currently he is Distinguished Engineer and his areas of interest span system and subsystem architecture, networking, development and operating environments. He lead several pioneering research projects at the lab in clustering, file system structure (and object storage), I/O and networking convergence (iSCSI) , has driven an industry-wide effort to standardize iSCSI and is now driving an effort to standardize Object Storage. Mr. Satran graduated (MSc EE) in 1962 from the Polytechnic Institute Bucharest, Romania.

Dr. **Yaron Wolfsthal** is a senior manager for Reliable System Technologies in the IBM Haifa Labs. The department headed by Dr. Wolfsthal provides advanced tools and services for assuring system quality and reliability, including (1) solutions for elimination of software and hardware defects, and (2) advanced technologies in the area of data storage. Dr. Wolfsthal has 16 years of experience in various R&D and management assignments, and holds a BsC, MsC and PhD degrees in computer science, all from the Technion - Israel Institute of Technology

Responsibilities: IBM's HRL contributions to CASPAR will centre around the use of Object Storage for preservation systems. HRL brings its extensive experience with storage systems, and Object Storage in particular. The Object Storage team at HRL has been involved with this technology for over four years. It is an active participant in the OSD standard (Julian Satran from HRL is the co-chair of the standards group, and others are active members). This team developed an implementation of object

storage, including an open-source driver, and participate in a demonstration of this technology with other vendors.

9.9 CNR

The Institute of Information Science and Technologies (ISTI, http://www.isti.cnr.it/) is an institute of the Italian National Research Council (CNR). ISTI's domain of competence covers Information Science, related technologies and a wide range of applications. The activity of the Institute aims at increasing knowledge, developing and testing new ideas and widening the application areas. The "Networked multimedia information systems" laboratory is one of the research labs of ISTI, equipped with a staff of about 30 persons. Networked multimedia information systems indicate a specific area of research in information technology concerned with processing, management, distribution and realisation of composite information available through different media. Networked multimedia information systems play a major role in a number of applications, including educational applications, entertainment technology, ecommerce, medical, database, and library information systems. The NMIS Lab has participated, often with a leading role, to a number of European projects in the area of Cultural Heritage and Digital Libraries. In the Sixth Framework Programme, it is involved in the BRICKS Integrated Project and it is coordinating the Diligent Integrated Project as well as the DELOS Network of Excellence.

Key people: **Carlo Meghini** (http://pc-erato2.iei.pi.cnr.it/meghini/index.html) is a senior researcher at Istituto della Scienza e Tecnologie della Informazione (ISTI) of Consiglio Nazionale delle Ricerche (CNR) in Pisa. He is a member of the Networked Multimedia Information Systems Lab at ISTI (http://www.isti.cnr.it/ResearchUnits/Labs/nmis-lab/), and his main area of interest is Digital Libraries. He is currently part of the 6FP Integrated Project BRICKS (http://www.brickscommunity.org/), aiming at building a digital library system in the cultural heritage domain. Within BRICKS, he leads the work on semantic-based information access. He is also a member of the DELOS Network of Excellence in Digital Libraries, in which he conducts research on the Information Access and Personalization cluster. He has published more than 40 papers in international journals and conferences.

Responsibilities: CNR is Leader of WP2400 "Next Generation OAIS Components" and of task 2305 "Information Access"..

9.10 MetaWare

Metaware SpA (MW) was formed in the year 2000 and extended its partnership to Infocamere ScpA, the ICT company of the Italian Chambers of Commerce, with a focus on digital signature and information security.

Metaware core expertise lies in the fields of e-content management (digital content IPR protection), cultural heritage administration, information and transactional frameworks for public administration (one-stop-shop management), infrastructures (PKI) and software integration for digital signature.

MW is one of the key partners of the BRICKS project (*Building Resources for Integrated Cultural Knowledge Services* - IST) offering its expertise for technological implementation, exploitation dissemination and training.

Key people: Silvia Boi, Fiore Basile

Silvia Boi. Degree in Foreign Languages and Literature" (1989), Specialisation post-degree: "Diploma di Specializzazione Polivalente per minorati psicofisici, della vista e dell'udito", concerning the specialisation in teaching to psycho-physic impaired, hard of hearing, deaf, and blind students.

1991-96 She collaborated at the "Istituto di Neuro-Psichiatria dell'Età evolutiva" (INPE) Fondazione Stella Maris, within the psycho-pedagogical research "Aid in computer devices for disabled children ranging from nursery to elementary"

Since October 1996 has worked at Consorzio Pisa Ricerche offering support to project, activities. Experience has been broadened within the field of EU projects focusing in the cultural heritage sector.

Since January 2002 has worked at METAWARE SPA as responsible of European and National project on the cultural heritage sector.

Main Skills: Project quality assurance monitoring and management, Project management and organisation, EU programme expertise. Communication and dissemination activities, Web Publishing, Market planning. Expertise in EU project in the cultural sector.

Fiore Basile holds a degree in Computer Science from University of Pisa. He currently works as senior software engineer for Metaware S.p.A. He is responsible for management of software development activities in several European-founded and government related projects. His previous experience includes IT consulting, system administration, development of web-based enterprise applications, content management systems and PDA software. His current interests are related to semantic web technologies, mobile platforms, security, digital rights management and distributed service oriented architectures. He is also an expert on open-source and agile software development methodologies.

Responsibilities: MW offers its competencies both in the technological implementation of the CASPAR technical streams (especially the activities relating to the DRM and security and to the framework implementation) and in the dissemination and exploitation and management field. MW will also provide the support tool for PPA and remote training.

MW is in charge of WP 3400 Multi-industries long-term research perspectives, WP 5200 Communication and Dissemination, and is responsible for the following tasks: T2302 DRM, T5302 Training infrastructure, 5303 Training material.

9.11 INA

INA, a public-owned industrial and commercial company, is involved in several aspects of the audiovisual industry; audio and video archiving, restoration and commercialisation; legal deposit of French audiovisual radio and television broadcasting; professional training; technical and socio-economic research; production research.

INA is in charge of French Audiovisual Archives and a partner of the Radio and Television Broadcasters in France. About 950 persons work in the institution, in charge of Radio and TV Archives, professional training, research and experimental production, since 1975.

INA has been in charge of the Legal Deposit for national TV channels (public and private) and for public radio channels since 1995. In this context, INA has acquired experience in the analysis of user-based approaches and in the development of user-friendly interfaces for archive retrieval and consulting.

Key people: Daniel Teruggi, Philippe Poncin.

Daniel TERUGGI (born in 1952, La Plata, Argentine)

Daniel Teruggi studied composition and piano in Argentina. In 1977 he came to France and studied at the Paris Conservatory (Conservatoire National Supérieur de Musique de Paris) in the department of Electroacoustic Composition and Musical Research. In 1981, he became member of the INA-GRM where he first was in charge of the pedagogy of digital systems for composers, and then became Artistic Director of the group. In 1997 he become Director of the INA Groupe de Recherches Musicales.

Since October 2001 he directs the Research and Experimentation Department in INA.

PhD in Art and Technology in the Paris VIII University. He teaches Sound and Visual Arts, at the Paris I Sorbonne University. He is director of a Seminar on new technology applied to Music at the Paris IV University.

He has developed an important activity as composer and researcher, mainly on the relations between creation and technology and the concepts underlying interfaces.

Philippe PONCIN

Born in 1948, graduated from Ecole Nationale des Télécommunications in June 1973.

After joining the Archives Management of INA in 1977, he designed and set up the technical division, including all the storage, inventory and restoration services with dedicated buildings, equipment and technical staff trained for the purpose.

He was then in charge, as Deputy Director since 1984, of the evolution of the computerised information system (production and retrieval), and of the development of technical innovations for archival services (digital systems devoted to audio and video restoration, multimedia indexing and retrieving facilities)

He joined the Audio-visual Research Laboratory of INA in 1997, in charge of management of the European research projects (DiVAN and EuroDelphes for ESPRIT, Diceman for ACTS, Olive for Telematics, PrestoSpace for IST).

Responsibilities: multimedia indexing and archiving modelling and knowledge representation. INA will provide structured contents for experiments with computer-based music and performing arts in the context of innovative and long-term access scenarios.

9.12 University of Leeds (UNIVLEEDS)

University of Leeds (UNIVLEEDS, www.leeds.ac.uk) is one of the largest universities in the UK. The UNIVLEEDS Interdisciplinary Centre for Scientific Research in Music (ICSRiM, www.ICSRiM.org.uk) specialises in multi- and inter-

disciplinary research and involves members of staff from several Schools, including Music, Computing, Electronic Engineering, Psychology, Physics, and Sport and Exercise Sciences. ICSRiM provides a venue for research and development in a wide range of interdisciplinary research areas, including capturing, modelling and simulation of expressive performances; interactive gesture music, virtual and augmented instruments and others.

Key people: **Dr. Kia C. Ng** (www.kcng.org) is a senior lecturer at the University of Leeds where he is director and co-founder of ICSRiM. Kia's research specialism links together work in the School of Computing and the School of Music on computer music, computer vision, interactive multimedia system, cross-media and 3D gesture analysis and interface. Kia has an extremely broad ranging experience in multimedia imaging with world-leading research on pattern recognition, restoration, and translation of printed and handwritten music manuscripts. His Music via Motion (MvM) system, which provides interactive gesture control of musical sound, has been widely featured in the media, including the BBC, Sky, and amongst others in the Financial Times and the New York Times. His paper on 3D reconstruction of real environment won the prestigious U.V. Helava Award (Best Paper, ISPRS Journal, Elsevier Science). Membership of Professional and Scholarly Bodies include Fellow of the Royal Society for the Arts (FRSA); Chartered Scientist (CSci); Science Council UK; Chartered Engineer (CEng), Engineering Council UK; British Computer Society (MBCS); Institute of Directors (MIoD); etc.

Other related personnel include members of the Centre and the executive committee (www.ICSRiM.org.uk), the School of Music (www.leeds.ac.uk/music) and the School of Computing (www.comp.leeds.ac.uk), including Prof Roger Boyle, Prof David Cooper, Prof Clive Brown, Dr Luke Windsor, Dr Andrea Utley, Bee Ong, Garry Quested and Royce Neagle.

Responsibilities: Contemporary arts testbeds and data access methods

9.13 Engineering

Founded in 1980, Engineering Ingegneria Informatica SpA is the leader of the Engineering Group consisting of 12 firms specialising in different IT services, but sharing the same cultural identity and mission. Financed by private capital, the Engineering Group operates in more than 20 locations in Italy, and in 2002 achieved consolidated revenues of about 260M€ with almost 3000 employees, thus further consolidating the Group's position as one of the largest private enterprises in the system and business integration market sector in Italy.

The main areas of interest of the R&D Department are concerned with Service-Based Software Engineering, Multi-Agents System, Software Architecture, Knowledge Management, and Human-Computer Interaction. This policy - based on strict synergies between the operational and the R&D divisions - has proven essential to generate a wide number of opportunities and contracts with large organisations like Benetton, Telecom Italia, Monte dei Paschi di Siena, AIPA, Ministero dei Beni Culturali, and many others.

Since 1987, current and past research initiatives have been undertaken at European level, most recently within the Fifth EC Framework Programme - IST, TEN-TELECOM, eContent (namely Archware, Homey, P2P Architect, TNT, COMPONENT+, CBSEnet, Fin-tech, ECO-ADM, Infocitizen, BankSEC, IB,

Eusland, Chance, TradeX, Train.Me. CLARiFi, Ecolnet, I-MASS). Engineering R&D is now participating in the FP6 with several initiatives under different action lines. The long tradition in participating in research projects, combined with its intimate industrial nature, has enabled the Engineering R&D Department to be the coordinator of several of the abovementioned research projects both at the national and European level.

Key people: Francesco Nucci, Luigi Briguglio

Francesco Saverio Nucci - Born in 1964, he received a university degree in Computer Science in March 1990 at the University of Rome “La Sapienza”. After a grant for the Multimedia and Tutoring System at the Computer Science Dept of University of Rome he collaborated in the start-up of the Sespim Research Consortium in Naples in the field of Human Computer Interface and Multimedia Systems Application, following a project in collaboration with the Media Lab of the Massachusetts Institute of Technology, where he also attended some summer schools and executive courses.

In 1996 he was appointed as Technical Co-ordinator for the SESM Consortium and co-ordinate the negotiation for a ICT large Programme funded by the Italian Research Ministry. During these working activities he acquired skills in project management of research, development and technology based innovation, in the field of ICT and Multimedia Information Browsing. During years 1999 and 2000 he was involved in finance and change management programmes: he was involved in a post merger integration in Alenia Marconi Systems, an Anglo-Italian Joint Ventures and responsible of an internal change management project for the introduction of EVA (Economic Value Added) method in the Company.

In the April 2001 he joined the Engineering R&D Laboratory, becoming in 2003 Co-ordinator of European Research Projects, in this role he managed many European Consortia and applications projects. He was director of the BRICKS project, responsible for added value for the IP Diligent, member of the Advisory Board if the INTELCITIES project.

He is author of many international papers in Digital Libraries, Business Innovation and Multimedia Management. From 2002 he was appointed professor at the University of Rome "La Sapienza" in a Master for Business and Finance Management of Media.

Luigi Briguglio - received an Electronic Engineering degree at University of Padua - Italy in 1996, with a thesis concerning an "analysis and realisation of an audio amplifier with hi-fi characteristics". In 2001 he joined Engineering as Software Analyst at the R&D Laboratory in Rome. He covers several different operational roles within the company including, specially in IST European Projects: requirement and specification analysis, analysis and design of service-based and component-based architectures, analysis of development methodologies, thesis-stage tutor and training activity.

In 1997 he worked at TRAINET, a Telecom Italia Group company, as web application developer. In 1998, he was involved in research activity at INFM (National Institute for Physics of Matter) for design and realisation of a Snom microscope controller. Furthermore, a long experience in industrial system automation, specially in distributed acquisition systems and supervision software, has characterised his experience in analysis, specification and development of software

architectures within a quality controlled process, and preparation of technical proposals.

Responsibilities: Engineering will lead the work on the overall CASPAR architecture and also use their commercial experience to lead the exploitation of CASPAR results.

9.14 FORTH

Since its establishment in 1983, the Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH-ICS) has a relatively long history and an established tradition of internationally acknowledged excellence in conducting basic and applied research, developing applications and products, and providing services. FORTH-ICS represents Greece in the European Research Consortium for Informatics and Mathematics (ERCIM). FORTH-ICS is active partner of FP6 Networks of Excellences related to the current proposal, including DELOS II (Digital Libraries), and REWERSE (Reasoning on the Web with Rules and Semantics).

The FORTH-ICS Information Systems (ISL) Laboratory has many outstanding achievements in the Semantic Web area. ISL members have designed and implemented RDFSuite, a set of highly scalable tools used for managing RDF description bases and schemas. It includes a Validating RDF Parser (VRP), a RDF Schema Specific DataBase (RSSDB), an RDF Query Language (RQL) and an RDF View Language (RVL). ISL has also designed a novel Semantic Web Integration Middleware (SWIM) for mediating high-level queries to relational and/or XML sources using ontologies expressed in RDF/S. Furthermore, as part the basic research agenda of ISL, novel results on belief revision with application to ontology evolution have been recently produced and published. Specifically, a complete characterization of Description Logics with respect to the ability of defining consistent revision operators has been produced, paving the way for the much-needed development of tools that support a sound ontology evolution process.

Key people

Vassilis Christophides is an Associate Professor at the Department of Computer Science, University of Crete and a Researcher at the Information Systems Laboratory of the Institute of Computer Science, FORTH in Greece leading several EU projects related to the integration of heterogeneous and distributed information sources. His main research interests include Semantic Web and Peer-to-Peer information management systems, semistructured and XML/RDF data models and query languages as well as description and composition languages for e-services. He has published over 50 articles in international conferences and journals and has served on numerous conferences program committees, including ACM SIGMOD, VLDB, WWW, ICSW, ICWS. He co-chaired the Workshop on Databases and the Web (WebDB) at SIGMOD 2003, the Web Services Track of the IEEE International Conference on Services Computing (SCC 2004) and he was also Industrial and Application Chair for EDBT 2004.

Yannis Tzitzikas is currently Adjunct Professor in the Computer Science Dept. at University of Crete and Visiting Researcher in Information Systems Lab at FORTH-ICS. He was postdoctoral fellow at the University of Namur and ERCIM postdoctoral fellow at ISTI-CNR (Pisa) and at VTT Research Centre of Finland. He conducted his undergraduate and graduate studies in the Computer Science Department at University of Crete. His research interests fall in the intersection of Knowledge

Representation and Reasoning, Information Indexing and Retrieval, Conceptual Modeling, and Collaborative Distributed Applications. His current research revolves around faceted metadata and semantics, the P2P paradigm (focusing on query evaluation algorithms and automatic schema integration techniques), and flexible interaction schemes for information bases. The results of his research are published in more than 35 papers in refereed journals and conferences.

Grigoris Antoniou is Professor of Computer Science at the University of Crete and Head of the Information Systems Laboratory at FORTH. He has held professorial appointments at Griffith University, Australia, and the University of Bremen, Germany. He has a strong background in knowledge representation and reasoning, semantic web technologies, and their applications. He is (co-)author of a number of international books, the most recent being "A Semantic Web Primer", MIT Press 2004. He has published over 150 technical papers in scientific journals and conferences.

Anastasia Analyti earned a B.S. degree in Mathematics from University of Athens, Greece and a M.S. and Ph.D. degree in Computer Science from Michigan State University, USA. She worked as a visiting professor at the Department of Computer Science, University of Crete, and at the Department of Electronic and Computer Engineering, Technical University of Crete. Since 1995, she is a researcher at the Information Systems Laboratory of the Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH-ICS). Her current interests include the Semantic Web, conceptual modelling, faceted metadata and semantics, rules for the Semantic Web, bio-medical ontologies, contextual organization of information, contextual web-ontology languages, information integration and retrieval systems for the Web. She has participated in several projects and has published over 35 papers in refereed journals and conferences.

9.15 CNRS

The French National Centre for Scientific Research (CNRS) is a publicly funded research organisation that defines its mission as producing knowledge and making it available to society. CNRS has 26,000 employees (including 11,600 researchers and 14,400 engineers and technical and administrative staff). Its amounts to 2.214 billion euros for the year 2004.

The research planned in CASPAR will be carried out by the DOC group of the Heudiasyc laboratory. Heudiasyc was created in 1980 and since 1981 has been an associate laboratory of the CNRS, under whose auspices it now enjoys the status of a “joint” (CNRS-UTC) research laboratory.

Heudiasyc’s guiding principle is to bring together research in control, signal, image and computer science with an emphasis on human factors. The objective is to develop ways of representing, analysing and controlling technical systems that impact on and are constrained by the economic, social and han spheres.

The Doc group aims at studying knowledge engineering in the field of eLearning, document and content management and interactive systems. The main purpose is to make the most of knowledge representation and formal inference in real-life applications where close attention is paid to the users. Research consequently integrates user modelling, content semantics and knowledge theory. Heudiasyc/CNRS has led the MUSTICA project that dealt with contemporary music preservation, in

collaboration with the UCLA Center for Information as Evidence, in connection also with InterPARESII.

Key people: Bruno Bachimont, Stephane Crozat

Bruno BACHIMONT

Born in 1963

Bruno Bachimont is manager of the research direction of INA and also associate professor at the University of Compiègne (Université de Technologie de Compiègne) where he teaches computer science, logic and philosophy.

Bruno Bachimont is graduated of the French Ecole des Mines de Nancy (engineering school) and has received a PhD in computer Science from french university Paris 6 and also a PhD in Philosophy from the french school Ecole Polytechnique. He finally obtained a "Habilitation à diriger les recherches" in 2004.

Bruno Bachimont was previoulsy researcher in the field of artificial intelligence and medical information science at the Assistance Publique- Hôpitaux de Paris, which is the administration made up of the 50 hospitals situated in Paris.

Bruno Bachimont is author of many technical and philosophical papers and has written a book on artificial intelligence and knowledge-based systems.

Stephane CROZAT

Born in 1975

Stéphane Crozat is responsible for technical and scientific developments in the innovation unit "Content and knowledge Engineering" based at the university of technology of Compiegne (UTC).

Stéphane Crozat has received in 2002 a PhD degree from (UTC). His expertise domains are document engineering, XML based standards for document management and content modelling. He has contributed to the elaboration of the Scenari platform, which is an open-source platform proposing tools for document structuring and publishing in multi user and multi usage contexts.

Responsibilities: Heudiasyc/CNRS will lead the WP devoted to OAIS-based access to content. Deeply involved in knowledge technologies, Heudiasyc/CNRS participates in modelling tasks and in the artistic testbed.

9.16 IRCAM

IRCAM has, for 30 years, been a pioneer in developing real-time systems dedicated to audio and music information processing and has produced in particular the MAX® software, a graphical language for programming and executing real-time music processing algorithms, which has become a major standard for interactive music and multimedia artistic works. IRCAM also has competencies in database management and middleware, in audio synthesis and analysis, as well as in indexing and description of content. More recently, IRCAM has developed competencies in preservation (MUSTICA project) and tools for preserving intelligibility of content, such as OASIS for annotating content and Signed Listening for computer-aided hermeneutics. IRCAM is also engaged in content description and music notation standardisation, being an active participant in the MPEG 4 and MPEG 7 groups. Besides its activity as a research centre, IRCAM is also a contemporary music

production facility, and the IRCAM Production Department is very much involved in archiving, disseminating and adapting its electronic work corpus, composed of more users, and will help to define generic methods and models, and will also be able to validate the methodologies and tools developed throughout the CASPAR project.

Key people: Hugues Vinet, Scientific Director. Vincent Puig, Marketing & Communication Director. Jerome Barthelemy, Project manager at IRCAM

Hugues Vinet, Scientific Director. After a scientific and musical education, Hugues Vinet worked at the Musical Research Group of the National Institute of Audiovisual (Ina-GRM) as Engineer, and then Senior Engineer. Since 1994, he has been, as Ircam's Scientific Director, in charge of Ircam's R&D Department and of Ircam-CNRS STMS (Music and Sound Science and Technology) joint lab. He is Coordinator of the CUIDADO and SemanticHIFI European IST projects.

Jerome Barthelemy Project manager at IRCAM After a musical education and an experience of teaching in University and music schools, he gained experience in the development of Multimedia programs including a Musical dimension with Syrinx SA. He joined IRCAM in March 2001 for managing the WEDELMUSIC Project and developing WEDELMUSIC Local Distributor and Music Analysis tools. He is currently involved in the SemanticHIFI project at IRCAM.

Responsibilities: Leads the Contemporary Arts Testbed work package.

9.17 CIANT

CIANT has been active since 1998 at the international level in the field of culture, education and research related to new technologies.

CIANT is an independent non-profit organisation that initiates and supports partnerships with individuals as well as institutions having a background intersecting art, science and technology.

CIANT operates an art laboratory focusing on innovative experiments with and development of interactive media. They produce and co-produce artistic projects, co-ordinate residencies for artists and participate in the promotion of new media art and culture - they organise festivals, workshops and conferences.

CIANT stimulates networking and participation in public debates while promoting an interdisciplinary approach in the field of artistic and research creativity as well as technological innovation.

In the last couple of years they realised several international projects supported by the European Union, within programmes such as Information Society Technologies, CULTURE 2000 and MEDIA.

Key people: Pavel Smetana, Pavel Sedlak

Pavel Smetana is the Director of CIANT, focusing on institutional development of CIANT, involved in preparation, implementation and presentation of international projects in the field of art, culture, research and education, subsidised e.g. by European Union, Czech Ministry of Culture, the City of Prague. He is also a Professor and Head of Computer Arts and Virtual Reality Dept. Professor of 3D at the Academy of Fine Arts in Aix en Provence, France.

Pavel Sedlak is the Deputy Director of CIANT (International Centre for Art and Technologies) in charge of development and documentation. With background in philosophy of art and cultural management he managed several EU-supported cultural activities. He was the chief curator of the ENTERmultimediale 2 festival of digital art held in Prague (May 2005). He is the author of theoretical papers, magazine articles, interviews and reviews focusing on new media art. After his residencies at the ZKM - Center for Art and Media in Karlsruhe (2002, 2003, 2005) and the internship at New Media Dept. of the National Gallery in Prague (2002) he worked at the Centre for Science, Technology, Society Studies of the Institute of Philosophy of the Czech Academy of Sciences. Recently, he was a member of the Prague Mayor's committee on the city cultural strategy.

Glossary

Acronym	Definition
ACL	Access Control List
CCSDS	Consultative Committee for Space Data Systems
CIDOC	International Committee for Documentation
DRM	Digital Right Management
ebXML	Electronic Business using eXtensible Markup Language
IPR	Intellectual Property Right
ISO	International Standards Organisation
GGF	Global GRID Forum
GIS	Geographical Information System
ICOM	International Council of Museums
KPI	Key Performance Indicator
MoReq	Model Requirements for the Management of Electronic Record
MPEG	Moving Picture Experts Group
MPEG-REL	MPEG Rights Expression Language
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open GIS Consortium
ORDL	Object Relational Description Language
OSD	Object Storage Device
OSS	Open Source Software
PDI	Preservation Description Information
NARA	National Archive and Records Administration
SAFE	Standard Archive Format for Europe
SOA	Service Oriented Architecture
SRB	Storage Resource Broker
UDDI	Universal Description, Discovery and Integration
XML	Extensible Markup Language
XrML	eXtensible rights Markup Language

References

[1] http://public.ccsds.org/publications/archive/650x0b1.pdf

[2] http://www.ibeurope.com/Files/060lisbon.htm

[3] http://europa.eu.int/information_society/eeurope/2002/news_library/documents/eeurope2005/eeurope2005_en.pdf

[4] SPEECH/05/107 Brussels, 22 February 2005

[5] europa.eu.int/information_society/eeurope/2005/all_about/digital_rights_man/doc/directive_copyright_en.pdf

[6] Directive 2004/48/Ec of the European Parliament and of the Council of 29 April 2004

[7] Directive 96/9/ec of the European Parliament and of the Council of 11 march 1996

[8] Official Journal (L345/90), 31 December 2003.

[9] IST Advisory Group – see http://www.cordis.lu/ist/istag.htm

[10] http://www.rlg.org/en/page.php?Page_ID=580

[11] http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=298&categorie=6

[12] http://dl.kr.org/

[13] http://www.xrml.org/

[14] http://xml.coverpages.org/mpegRights.html

[15] http://www.eo.werum.com/en/mdm/eo/project/harm/index.jsp?detail=1

[16] http://www.loc.gov/standards/mets/ Metadata Encoding & Transmission Standard

[17] http://sindbad.gsfc.nasa.gov/xfdu/ being developed by the same group that developed OAIS

[18] The Open Geospatial Consortium (http://www.opengeospatial.org/)

[19] http://europa.eu.int/idabc/en/document/2303/5644

[20] http://europa.eu.int/information_society/eeurope/i2010/index_en.htm

[21] http://europa.eu.int/information_society/activities/digital_libraries/index_en.htm

[22] http://tfpa.kb.nl

[23] http://www.e-science.clrc.ac.uk/documents/staff/kerstin_kleese/156.pdf - *Applying the OAIS standard to CCLRC's British Atmospheric Data Centre and the Atlas Petabyte Storage Service*

[24] http://www.uml.org/

[25] http://www.service-architecture.com/

[26] See for example International Symposium on Emergence of Globally Distributed Data, June 19-24, 2005, Sardinia, Italy http://globalstor.org

[27] for example the PV2005 meeting - http://www.ukoln.ac.uk/events/pv-2005/

[28] see http://www.ecdl2005.org/

[29] http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html

[30] http://www.kpmg.nl/Docs/Information_Risk_Management

[31] http://dev.dcc.ac.uk/twiki/bin/view/Main/DCCApproachToCuration "Approach to Digital Curation"

[32] http://www.snia.org/tech_activities/workgroups/osd

[33] http://www.haifa.il.ibm.com/research.html and http://www.haifa.il.ibm.com/dept/storage/index.html