



Project no. 033572

CASPAR

Cultural, **A**rtistic and **S**cientific knowledge for **P**reservation, **A**ccess and **R**etrieval

Instrument: Information Society Technologies

Thematic Priority: 2.5.10 Access to and preservation of cultural and scientific resources

D1202: CASPAR GUIDELINES



Document identifier:	CASPAR-D1202-TN-0101-1_0
Submission Date:	15-05-2007
Due Date:	15-05-2007
Work package:	1200
Partners:	All Partners
WP Lead Partner:	CCLRC
Document status	FINAL

Abstract: This document provides the Roadmap which links together the various CASPAR resources.



Delivery Type Report
Author(s) CASPAR Consortium

Approval David Giaretta

Summary

Keyword List

Availability PUBLIC

Document Status Sheet

Issue	Date	Comment	Author
0_0	30 April 2007	Initial draft bringing together contributions	David Giaretta
1_0	15 May 2007	Final version	David Giaretta





Project information

Project acronym:	CASPAR
Project full title:	Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
Proposal/Contract no.:	IST-2006-033572

Project Officer: Carlos Oliveira

Address:	<p>INFSO-E3 Information Society and Media Directorate General Content - Learning and Cultural Heritage</p> <p>Postal mail: Bâtiment Jean Monnet (EUFO 1167) Rue Alcide De Gasperi / L-2920 Luxembourg</p> <p>Office address: EUROFORUM Building - EUFO 1167 10, rue Robert Stumper / L-2557 Gasperich / Luxembourg</p>
Phone:	+352 4301 33052
Fax:	+352 4301 33190
Mobile:	
E-mail:	Carlos.Oliveira@ec.europa.eu

Project Co-ordinator: David Giaretta

Address:	<p>STFC (formerly CCLRC), Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK</p>
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	d.i.giaretta@rl.ac.uk





CONTENT

1	INTRODUCTION	5
1.1	PURPOSE OF THIS DOCUMENT.....	5
1.2	CASPAR OBJECTIVES	5
1.3	GLOSSARY.....	7
2	CASPAR OVERVIEW	9
3	DELIVERABLES AND THEIR RELATIONSHIPS	10
3.1	SIMPLE VIEW	10
3.2	DETAILED ITERATIVE VIEW	13
4	PROJECT RESOURCES	14
4.1	CASPAR WEB SITE	14
4.2	CASPAR PRIVATE WIKI	14
4.3	CASPAR CENTRAL UML REPOSITORY	14
4.3.1	UML repository web snapshot	14
5	PROJECT TASK ORGANISATION	15
5.1	CASPAR WORKPACKAGES	15





1 INTRODUCTION

1.1 PURPOSE OF THIS DOCUMENT

This document provides a roadmap to guide the reader, whether an implementer, reviewer or simply an interested party, to understand the relationship between the various documents, formal releases, internal documents and other project resources within the CASPAR project.

For a complete understanding of the project the Description of Work¹ [A1] should also be read, as should the OAIS Reference Model.

1.2 CASPAR OBJECTIVES

From the CASPAR Description of Work, the objectives of the project can be stated as follows:

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

Goal 1: build a pioneering preservation environment, based on a full use of the OAIS Reference Model¹ 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 and use methods 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.

¹ http://www.casparpreserves.eu/Members/metaware/ReferenceDocuments/caspar-description-of-work/at_download/file





Validation of these objectives is to be via a number of Measurable Objectives, addressing preservation aspects including

- a sound theoretical basis and in particular alignment with the OAIS Reference Model,
- "accelerated lifetime" tests involving hardware, software and the knowledge base of the Designated Community,
- an increase in the trustworthiness of archives using CASPAR.





Applicable documents and reference documents

Applicable documents

- [A1] Description of Work, April 2006
http://www.casparpreserves.eu/Members/metaware/ReferenceDocuments/caspar-description-of-work/at_download/file
- [A2] Risk Form

Reference documents

- [R1] CASPAR proposal, Sept 2005
- [R2] D4101 USER REQUIREMENTS AND SCENARIO SPECIFICATIONS (CASPAR-D4101-SCEN-0101-1_0)

1.3 GLOSSARY

[Ax]	Applicable Document
[Rx]	Reference Document
CASPAR	Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
DoW	Description of Work
EC	European Commission
EPM	Executive Project Management
IPC	IP Coordinator
IST	Information Society Technologies
PACP	Partner Administrative Contact Point
PO	Project Officer
PPR	Project Progress Report
PQE	Project Quality Engineer
PTCP	Partner Technical Contact Point
R&D	Research and Development
SQE	Stream Quality Engineer
ST	Stream
TN	Technical Note
WP	Work Package
WPL	Work Package Leaders
Designated Community	An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. (OAIS definition)
Archival Information Package (AIP)	An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS. (OAIS definition)





Content Information	The set of information that is the original target of preservation. It is an Information Object comprised of its Content Data Object and its Representation Information. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc. (OAIS definition)
Knowledge Base	A set of information, incorporated by a person or system, that allows that person or system to understand received information. (OAIS definition)
Representation Information	The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol. (OAIS definition)





2 CASPAR OVERVIEW

The CASPAR deliverables up to month 12, and other resources, may be best understood from the point of view of OAIS, and in particular the Functional Model and the Information Model. OAIS provides a well accepted terminology and framework for discussion of digital preservation, but one which must continue to be tested by direct comparison with real case studies. However OAIS is not all encompassing, as is acknowledged within the document itself. Therefore we supplement OAIS terms in order to address issues of preservation over time of, for example, Digital Rights Management and Authenticity, which are not covered in depth by OAIS.

Figure 1. indicates the flow of information into and out of a repository over time, and the types of information (“metadata”) which must be collected about it. CASPAR will produce tools and techniques to support capture and use of this information. The details of this diagram are discussed at length in the DoW, the Conceptual Model and in other CASPAR documents.

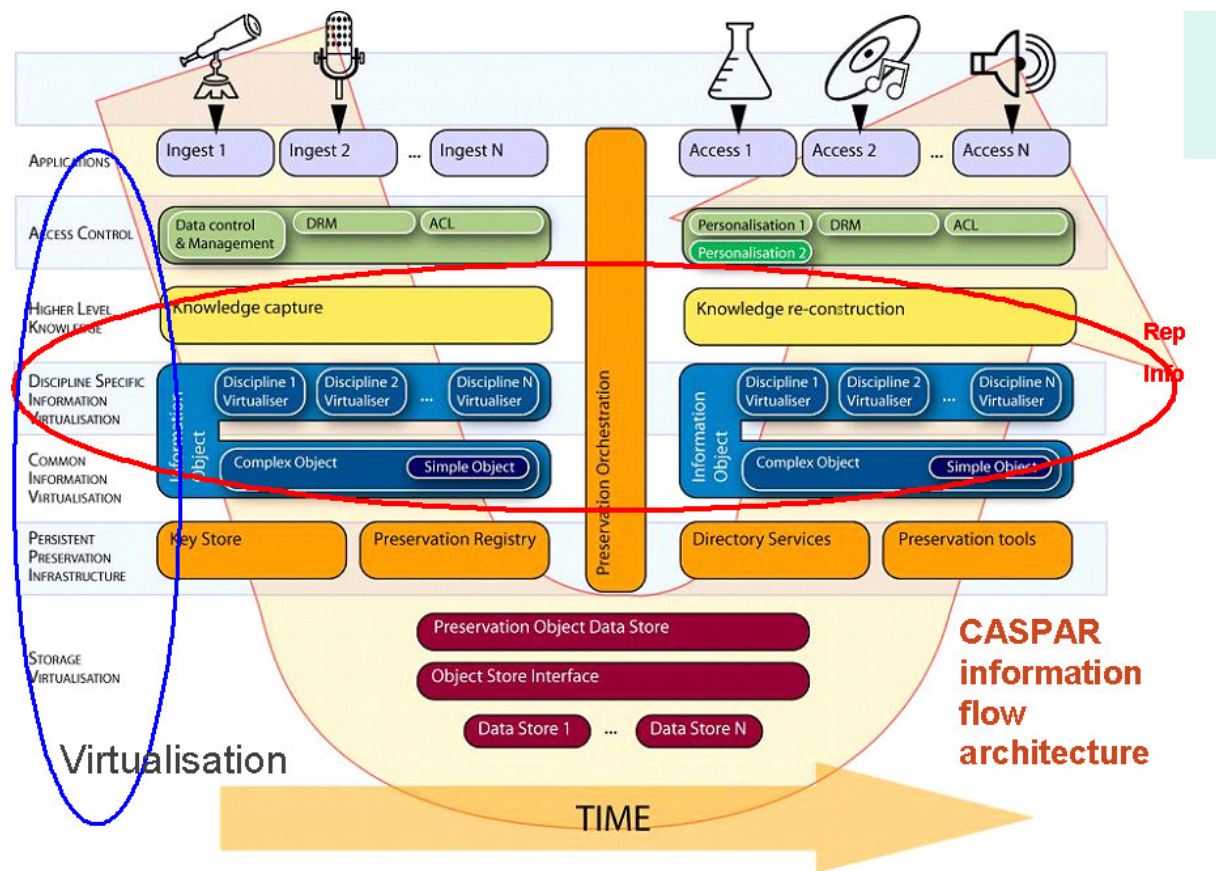


Figure 1 CASPAR information flow architecture





3 DELIVERABLES AND THEIR RELATIONSHIPS

The deliverables are given identifiers in the form of “*Dwwnn*” where *ww* indicates the Work Package under which the work was done and *nn* is a sequential number for deliverables from that Work Package. By month 12 the RTD deliverables are

- **D1101**: Review of the State of the Art
- **D1201**: Conceptual Model
- **D1301**: Architecture
- **D4101**: User Requirements and Scenario Specifications
- **D5101**: Dissemination and Use Plan

Suggested reading order:

1. *Conceptual Model (D1201), with selected reading of the State of the Art (D1101)*
2. *User Requirements and Scenario Specifications (D4101)*
3. *Dissemination & Use Plan (D5101)*
4. *Architecture (D1301)*

3.1 SIMPLE VIEW

At the simplest level the relationship between the documents may be described as (1) collect user requirements (**D4101**) and combine this with information about the current state of the art (**D1101**) in order to produce (2) a conceptual model (**D1201**) followed by (3) the architecture (**D1301**), as illustrated in

Figure 2. The remainder of the project would then involve implementing the architecture and testing, disseminating and training about the results of this work.

This may look like the classic, and now somewhat discredited, “waterfall” model of software development, but as will be discussed in section 3.2, the real situation is more iterative.



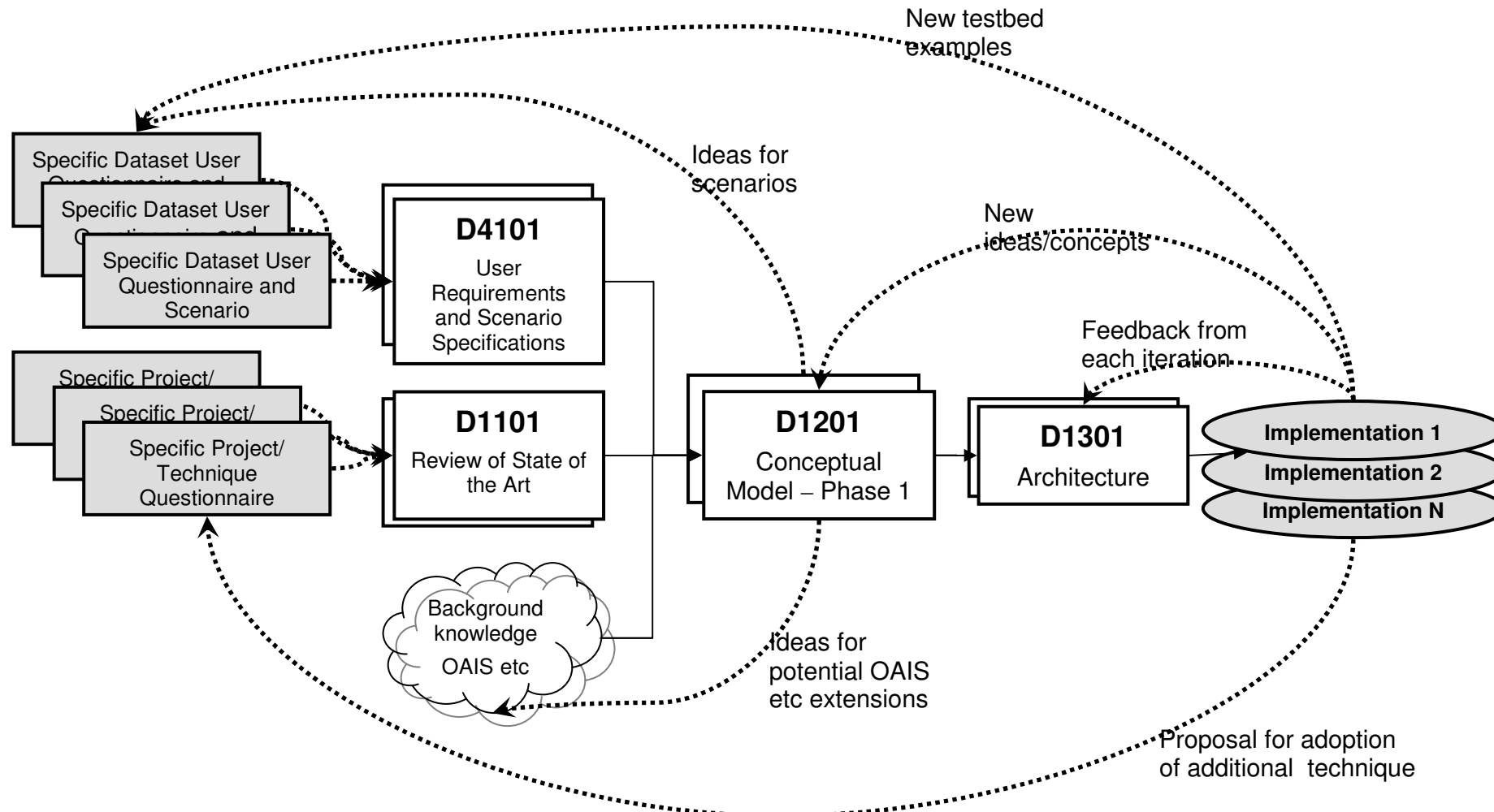


Figure 3 Iterative processes





3.2 DETAILED ITERATIVE VIEW

The process is iterative in that we expect the Conceptual Model and the Architecture to evolve as we start to produce implementations through cycles of increasing functionality. In addition we need to continue to expand the base of scenarios throughout the project. These iterative cycles are illustrated in Figure 3.

Feeding into the **D1101** (State of the Art) are documents about individual topics. Each document is a completed questionnaire, the headings of which follow the OAIS Functional and Information Models plus additional areas such as DRM. The questionnaire template is available at http://www.casparpreserves.eu/Members/cclrc/ReferenceDocuments/caspar-state-of-the-art-questionnaire/at_download/file. The headings may be grouped as follows

- OAIS Functional Model:
 - Ingest, Access, Storage and storage virtualisation,
- OAIS Information Model:
 - Representation Information, including the special cases of
 - Virtualisation and representation information
 - Higher-level knowledge
 - Annotation,
 - Preservation Description Information
 - Packaging
 - Description Information,
- Additional topics:
 - Access control, including DRM,
 - Preservation orchestration
 - Authenticity,

Similarly **D4101** (User Requirements and Scenario Specifications) is built on a number of questionnaire based case studies, each focussing on one archive or dataset. The questionnaire headings are again guided by OAIS ideas supplemented by topics such as DRM, and the template is available at http://www.casparpreserves.eu/Members/cclrc/ReferenceDocuments/caspar-test-case-questionnaire/at_download/file.

These individual questionnaires are available on the CASPAR Web site (<http://www.casparpreserves.eu>) apart from those which have not been approved for public release; these latter questionnaires are available on the CASPAR internal Wiki.

Throughout the remainder of the project new ideas will be generated, some of which will suggest new topics/datasets for questionnaires. These questionnaires will themselves influence new versions of the Conceptual Model. The implementation of the architecture will proceed in a number of distinct iterations, following modern software best practice. Each iteration will implement a certain subset of functionality and will generate feedback to the Architecture and Conceptual Model² as ideas are tested against real implementability.

We believe that we will also test OAIS concepts and, from the Conceptual Model, we may develop extension to existing OAIS concepts and also propose new concepts.

² Note that the revised versions of the deliverables, following month 12, are not themselves named as official deliverables to the EU, but will be made available on the public web site.





4 PROJECT RESOURCES

Information about the project is available from three main sources.

4.1 CASPAR WEB SITE

The CASPAR public web site (<http://www.casparpreserves.eu>) which contains:

- all public deliverable (see <http://www.casparpreserves.eu/caspar-project/deliverables>)
- reference documents, including the Description of work (see http://www.casparpreserves.eu/caspar-project/reference_documents)
- CASPAR Preservation User Community members
- publications and presentations about CASPAR
- press coverage
- links to training materials which will be produced in the course of the project

4.2 CASPAR PRIVATE WIKI

The CASPAR wiki (<http://wiki.casparpreserves.eu>) is password protected and contains all CASPAR working documents, including

- calendar of events
- discussion pages covering a number of specific topics
- details of individuals working on CASPAR
- details of meetings - planning and agendas for forthcoming meetings and minutes of meetings which have taken place
- drafts of papers
- a set of pages arranged according to the Work Package structure so that the information about a particular Work Package or Task can be located.

4.3 CASPAR CENTRAL UML REPOSITORY

This repository is accessed by the commercial tool Enterprise Architect and contains the software implementations, including

- Use Cases,
- a variety of other UML diagrams, plus
- requirements,
- source code

This provides a great deal more detailed information to support **D1301**.

4.3.1 UML repository web snapshot

Snapshots of the contents of the repository are made every week and are available via the CASPAR Wiki, and currently available at <http://church.eng.it:8080/model/>.





5 PROJECT TASK ORGANISATION

The CASPAR project is organised into Tasks which address a specific piece of work. Related tasks are grouped into Work Packages, each with a Work Package leader. Streams are made up of related Work Packages and each Stream has a Stream Director who in then a member of the Executive Project Management (EPM) team.

Naming of tasks is done hierarchically – a Task number “ijnm” tells us that

- it is in Stream “i”
- it is in Work Package “ij00”
 - this Work Package is the “j”th Workpackage in Stream “i”
- “nm” is the Task within the Work Package, with numbering starting from “01”

5.1 CASPAR WORKPACKAGES

The Work Package breakdown is available in the Description of Work, and their inter-relationship is shown in Figure 4

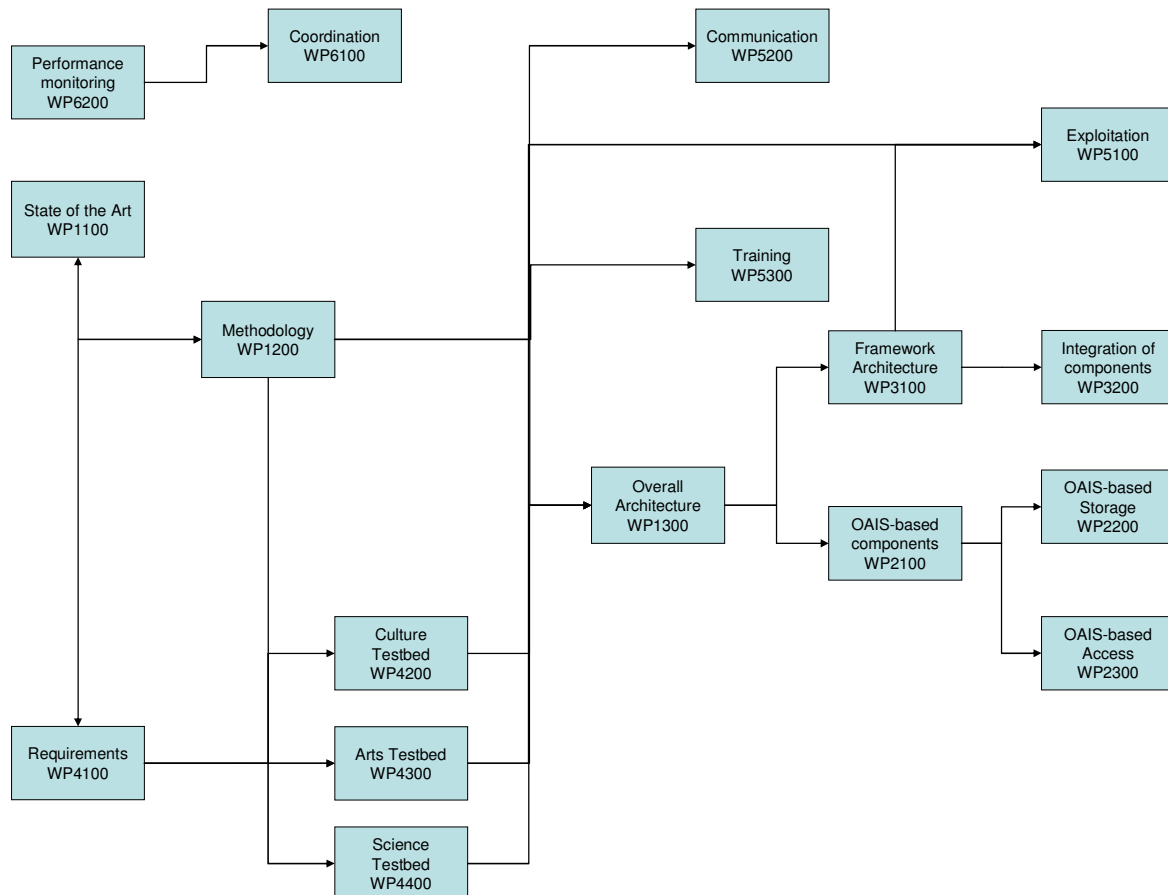


Figure 4 CASPAR Workpackage relationships

ⁱ <http://public.ccsds.org/publications/archive/650x0b1.pdf>

