Luigi Briguglio

Key Components – Overview

slide 1 This presentation is an overview of the work which was carried out during the 3-year CASPAR project and specifically concerns the CASPAR key components, which are the foundation and the framework for the preservation aspects in this context.

slide 2 The first part is dedicated to the key concepts taken from the OAIS reference model; then, we will present the digital preservation issues, the rationale which underlies the CASPAR solutions and the answers that the project has found to the digital preservation issue.

slide 3 One of the main questions that CASPAR is trying to address is “how can digitally encoded information still be understood and used in the future when the software, systems and the everyday knowledge will have changed?”

slide 4 The CASPAR project is mainly based on standard ISO:14721:2003 OAIS.
In this perspective, its architecture is defined to manage the key concepts of the OAIS reference model and support the main functionalities and responsibilities identified in it.
Moreover, the CASPAR project aims to define and implement interfaces and functionally independent components.

slide 5 In order to address the digital preservation issue, a best practice has been identified.
According to the OAIS information model, it is important to highlight that what has to be managed by means of the developed components consists of: the Information Package, the Content Information, the Preservation Description Information (PDI), the Data Object and the Representation Information (RepInfo) which is related to the Designated Community and to its Knowledge Base.

slide 6 The OAIS functional model identifies six macro components: Ingest, Access, Preservation Planning, Data Management, Archival Storage and Administration.
They are all described in the reference model in order to identify what their responsibilities and functionalities are.

slide 7 The mandatory responsibility of an OAIS archive is to provide the following:

- the negotiation and acceptance of submitted information;
- the long term preservation of information;

- to determine the Designate Community;
- to render the information preserved independently understandable to the Designated Community without having information from the providers;
- to follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies;
- to make the preserved information available to the Designated Community.

slide 8 Having presented the key concepts, it is now possible to tackle the digital preservation issues.

slide 9 Several questions can be addressed, and they concern the way to guarantee:

- the access to the digital information and its comprehensibility in the future;
- a proper information package management within the OAIS archive;
- a long term preservation maintenance of any information package and the retrieval of the Archival Information;

slide 10

- the intelligibility of the resources and of the related information within an heterogeneous Designated Community;
- that the preservation actors are informed about changing events;
- an adequate security access with the proper rights to the resources;
- the integrity and the identity of the any Archival Information.

slide 11 By using these issues, we can see the rationale behind the CASPAR solution.

slide 12 In this context the big question is “what has to be done to preserve the digital content object and guarantee that it can be used and understood in the future?”

slide 13 There are three main steps.
To preserve the digital content object, it has to be prepared and packed in an appropriate way to be ingested into the digital archive that will manage and maintain it for a long time.
In this preliminary phase, a package containing digital content objects and all necessary information is created and then stored.

slide 14 This is the Ingestion phase in which one has to:

- select the content information;
- create and register the RepInfo which allows one to understand the semantics and the structure of the information;
- create and register the Descriptive Information which is used for the retrieval aspect;
- create and store the Archival Information Package (AIP).

slide 15 According to the reference model, the main activities of the Ingestion are:

- the generation of the AIP;
- the receiving of the submission;
- the quality assurance;
- the coordination of the updates;
- the generation of the Descriptive Information.

All these activities work together with the Data Management, the Archival Storage, the Producer and the Administration.

slide 16 CAPSAR is trying to provide a best practice; it has created a sort of facade layer by introducing the Information Package Management where the main activities are:

- the ingestion of the Content Information;
- the creation of the Information Package by creating the RepInfo, the Descriptive Information and the PDI;
- the check of the Information Package and finally its storage on a long term basis.

These actions are part of the Ingestion Data Management and Archival Storage.

The main components and functionality which are promoted and supported by CASPAR are:

- the Packaging, i.e. the component which creates the package;
- the Data Store;
- the Virtualization;
- the RepInfo Toolkit;
- the Rep Info Registry;
- the Finding.

slide 17 After the Ingestion, the Access is the second step which is needed to make the digital content information available, usable and comprehensible.
In this case, the internal package of the digital archive system has to provide useable and understandable information to the consumer who requires it.

slide 18 The digital content object has to be found within the digital archive, through its descriptive information and checked for any restricting access right policy. After that, the processed information is prepared and packaged in such a way as to be accessed from the digital archive system and made available and intelligible to the consumer, taking into account the consumer's knowledge profile.

slide 19 The following actions are part of the Access functionality:

- the retrieval of the AIP by using the Descriptive Information;
- the check of the Access Right;
- the processing of the RepInfo by the Designated Community profile of the

requester;
- the creation and submission of the Dissemination Information Package (DIP);
- the request for further information.

The latter action is needed because, when a consumer browses the information maybe he or she could have a different Designated Community profile and, as a consequence, this person may not understand what the real contents that is being browsed are and could ask for further details.

slide 20 In the access, as is shown in the OAIS reference model, there are the activities of:

- coordination of access activity;
- generation of DIP;
- delivery response.

slide 21 The CASPAR best practice identifies an Information Access, i.e. a macro component which allows the consumer to search content information and obtain the Information Packages and relative contents and descriptions.
The Data Management and the Archival Storage are part of the access.
The main component is the Finding Manager which, together with the Knowledge Management System, the Packaging and the Data Store, is able to support this functionality.

slide 22 Several aspects concerning the Designated Community should also be analyzed.
The Designated Community is extremely important during the access to the information by the consumer.
Any change can have an impact on the Designate Community which can itself change in time, implying the need to deal with the change of its profile and its Knowledge Base.
The system should be able to identify the gap between two different Designated Communities. In this perspective the Knowledge Management System has been developed; it is part of the Data Management, of the Access and of the Preservation Planning, and, together with the RepInfo Registry and the Orchestration, could be able to deal with these issues.

slide 23 The most important aspect for a preservation system is the preservation planning.
The Ingestion and the Access are the two main activities which are usually involved in digital archives, but not in digital preservation archives.

slide 24 The third step concerns the digital content object within the digital archive: it has to be maintained in order to be accessed, used and understood for any changes during its long-term lifecycle. It implies that any impacting change in

the real world needs to be identified and notified, so that corrective actions can be carried out according to the preservation plan.

slide 25 The Preservation step can be simplified into three main tasks:

- the notification of changes;
- the performance of corrective action according to preservation plan;
- the update of AIP and PDI.

slide 26 According to the OAIS reference model, the main activities of Preservation in an OAIS archive are:

- to develop preservation strategies and standards;
- to develop packaging design and migration plans,
- to monitor the Designated Community;
- to monitor the technology.

slide 27 The CASPAR preservation best practice is mainly a communication management system which allows the notification and alert for change events impacting on long term preservation, and to trigger the preservation process in order to address the potential impact.
This is part of the Preservation Planning and Administration from the OAIS reference model point of view.
The main components are:

- the Orchestration, which notifies system changes, events and so on;
- the Knowledge Manager;
- the Authenticity;
- the Representation Information Registry.

Altogether, they are able to carry out the preservation functionality.

slide 28 Many important aspects need to be covered for the security reasons; these are: the need to deal with user accounts, roles, profiles, with the content access and the permission during the access and the ingestion; but also the need to deal with the digital rights and to guarantee the authenticity.
For these reasons a macro component has been identified, the so called Security Management.
It covers the aspect of preservation planning data management and administration.
Its main components are:

- the Data Access Manager & Security (DAMS);
- the Digital Rights Management (DRM);
- the Authenticity.

They all cooperate to provide this functionality.

slide 29 In general, there is a Data Producer who is able to use the Information Package Management and the Data Consumer who can ‘consume’ the

information through the information access, and together with the Communication Management, the Designated Community, the Knowledge Management and the Security Management, represent the main best practice that can be provided by CAPSAR in order to address the digital preservation issues.
This facade layer is built onto the CASPAR Foundation which is mainly composed of 11 key components, a framework and a platform.

slide 30 The system is really complex.
It has a platform which is mainly based on different operating systems (Linux, UNIX, Windows, etc.).
By using the Java platform, all the framework has been built, which is mainly a web service framework, by using the novelty solution of JAX-WS (which is an evolution of JAX-RPC); the Google web toolkit has been used as a development framework for the user interface, and Hudson and JTrac have been used for the development management.
The user can deploy its components in different application servers (Tomcat, Glassfish, WASCE-WebSphere Application Server Community Edition).
CASPAR has identified 11 components and made them available.

slide 31 The CASPAR project has tried to deal with several crucial issues and has come up with some answers.

slide 32 To guarantee that digital information may be accessed and understood in the future, an adequate OAIS RepInfo is needed.
CASPAR proposes the RepInfo ToolBox, the Virtualization and the Registry.

slide 33 In order to guarantee an information package management within an OAIS archive, an adequate OAIS Information Package must be created.
The Packaging is the solution proposed by CASPAR.

slide 34 To guarantee long-time preservation maintenance of any information package, the implementation of OAIS Archival Storage is necessary.
The CASPAR Preservation DataStores is a solution.

slide 35 An OAIS Finding Aids is necessary in order to guarantee the retrieval of Archival Information.
CASPAR offers the Finding as a solution.

slide 36 The Knowledge Management System is the CASPAR solution to guarantee the intelligibility within a heterogeneous Designated Community and its digital information, which requires the capacity to manage the Designated Community profiles and their Knowledge Base.

slide 37 In order to guarantee that preservation actors are informed about change

events, an adequate management of message exchange can be provided in CASPAR by the Orchestration Manager.

slide 38 The Data Access Manager & Security (DAMS) and the Digital Rights Manager (DRM) are the two solutions provided by CASPAR in order to guarantee adequate security access with the proper rights to any resource and functionality within an OAIS archive.

slide 39 The Authenticity Management System is the tool proposed by CAPSAR in order to guarantee an adequate integrity and identity for any Archival Information.

slide 40 The Data Producer, the Data Consumer and the Data Curator are the actors who are involved in the preservation process by using the key components in the different ways that have been explained previously.

slide 41 Further details, including all specifications and the source code are available on the developers and on the SourceForge community's websites.