

Authenticity in CASPAR

- slide 1** The four speeches of Maria Guercio, Giovanni Michetti, Luigi Briguglio and David Giaretta, are dedicated to authenticity in CASPAR.
- slide 2** This introduction speaks about the principles and methods, while Giovanni Michetti's statement presents the conceptual model that has been developed and used in the concrete testbed implementation.
The presentation made by Luigi Briguglio is dedicated to the relationship between Authenticity and Preservation Orchestration Manager, i.e. the kind of relationship between the concepts, the models, tools and the concrete development of components in CASPAR.
David Giaretta dedicates his speech to a prototype of a tool related to authenticity in CASPAR.
There is no part specifically concerning authenticity in the artistic domain, but in the IRCAM presentation we can find a great deal of material dedicated to how the authenticity tools and concepts have been used in this context by a specific testbed.
- slide 3** Many other materials on authenticity are available: in two different presentations, IBM explains how to use authenticity and provenance concepts in long term digital preservation, and what Preservation DataStores (PDS) is about and how it works.
- slide 4** Further publications dedicated to the specific aspects of the preservation process are available on the web.
- slide 5** Here, a general overview shows the effort made and the achievements reached in this specific area of the research project with reference to several specific concepts and their crucial role in the digital preservation environment; the need for specific tools for managing authenticity; the methodology which has been developed for using, implementing and updating these tools; the basic document for understanding the CASPAR perspective in this area.
- slide 6** This presentation is focused specifically on the significance of authenticity in the preservation process and on the tools for managing it.
- slide 7** All research projects related to digital preservation consider authenticity a central

issue.

slide 8 The CASPAR project has developed its own point of view but the workgroup has decided not to start from scratch and has chosen to use, as a base, what is considered the best in the most important recent projects in this environment, i.e. OAIS and InterPARES.

On the basis of the outputs of these two projects, it is possible to say that the fundamentals of authenticity concern the evaluation of the identity and the integrity of the digital objects.

The concepts of identity and integrity, which have been developed in the 8 years of the InterPARES project, are accepted today in the main ISO standards dedicated to metadata for record management and in descriptive standards.

Speaking about the preservation, we need to define the elements related to the accuracy, the reliability and the completeness of the information objects; these elements must be captured and maintained in the repositories to allow the user to evaluate the identity and integrity of these resources to be preserved.

Understanding these elements is not enough. They have to be organized into a framework to be useful in the effective life of the repositories and for the producer of digital materials. This framework is supplied by the OAIS reference model that is very important in the implementation of the preservation concepts and preservation activities.

The compliance to OAIS conceptual model also permits the description of the dynamicity of the authenticity profile and of the process that must be followed to ensure it.

The aim of this process is to gather, protect and evaluate information; this objective has been reached by identifying a set of attributes, mainly concerning integrity and identity, which allow the users to verify the degree of reliability of the object.

slide 9 Specific tools and methodologies for managing authenticity are required due to the complexity of the preservation function.

These tools should help the repositories to monitor and manage protocols, procedures and rules, across the custody chain, which is a chain of documented responsibilities and policies.

These protocols and procedure should be available, if possible, since the initial creation of the objects.

slide 10 An authenticity position paper has been created which is available on the web; it is just a basic document describing fundamentals that have been used to develop this project.

Its objective is to define how and on what basis authenticity has to be managed

in the digital preservation process to ensure the trustworthiness of digital resources.

Having recognized the general need for a set of shared definitions, as part of the basic efforts of the project a great deal of work has been spent in the definition of a glossary which has also been based on the main results obtained by the international research project InterPARES.

The main concepts used in the position paper have also been implemented in the US by the National Archives and Records Administration (NARA) as one of the most important projects in the field, and are also explicitly used in the recent regulations (under definition) by the Italian government for digital repositories and in the most important projects and implementation at an international level.

slide 11 Integrity and identity may seem two simple terms superficially, but really they contain a great deal of crucial issues which are complex to understand and even more so to implement as concrete tools, as specific, acceptable and reliable rules.

slide 12 The integrity of a resource refers to its wholeness; a resource has integrity when it is complete and uncorrupted in all its essential respects. So we must consider what is essential to preserve and if we have to preserve all the functionalities the digital object has in his creation environment, what the degree of quality and quantity of details is that we need to declare and prove that these resources are really complete and able to answer to the integrity.

The bit flow perhaps is not always preserved but, this is not always necessary.

What must be preserved is not the physical integrity of a resource, but the completeness of its 'intellectual form' and the ability to perform the original function for which the resource was created.

slide 13 Identity is another crucial point and it is not only a question of persistent identifiers in a digital repository.

The debate and the literature on the main projects up to now has attempted to define persistent identifiers which would be able to answer the request for the identity of the complex objects that have to be preserved, but this has, to a certain extent, been a failure.

Identity refers to the whole characteristics of a resource that uniquely identify and distinguish it from any other resource. These characteristics refer not only to the internal conceptual structure of the resource, but also to its general context (administrative, legal, documentary, technological and even 'social').

slide 14 As explained above, tools and methods are necessary to cope adequately with authenticity since it is a complex dynamic process which requires a great deal of

knowledge which, in turn, is not static. Also knowledge is dynamic and it has to be connected to the changes in the communities involved in the specific domains and for which a right level of granularity and details has to be chosen.

An enormous quantity of information must be acquired, captured and maintained, and it is necessary to deal with it as neutrally and as automatically as possible, but also in a feasible way for the repositories to disseminate this information and to make it understandable to its users in the future.

This is the reason why OAIS is an extremely important conceptual model; in no other models do we find the same capacity to build a system where this information is organised with different goals but at the same time connected to each other.

slide 15 Now, the presentation will tackle the critical issues that concern the tools for managing authenticity.

slide 16 Various complex requirements need to be accepted and understood before evaluating the results and outputs of CASPAR efforts.

Authenticity cannot be evaluated by means of a boolean flag telling us whether a document is authentic or not. The digital environment is never simple as it can provide us with an enormous quantity of information which has to be maintained. Moreover, users have to deal with all this information that originates from scientific, cultural and artistic environments.

One cannot say that the authenticity has degrees of reliability, but there are degrees in the capacity of presuming the authenticity of the digital resources; the certainty about authenticity is an objective.

We can be more or less certain of what we are going to leave in heritage to the future generations. Preservation requires changes to ensure accessibility, it is impossible to preserve things as they originally were, there will be a certain quantity of loss even if they are understandable and readable.

slide 17 The tools which have been built, help to identify the specific mechanism to ensure the maintenance and verification of the authenticity in terms of identity and integrity of the digital objects. These tools have to provide content and contextual information relevant to the authenticity depending on the kind of the resources available.

slide 18 Several issues should be considered in building the tools for managing authenticity: the right attribution of authorship, the identification of provenance in the life cycle of information objects (authorship and provenance are related to identity aspects); the assurance of content integrity of the whole relevant digital components and the relevant contextual relationships; the provision of

mechanisms to allow future users to verify the authenticity, to deal with all these details and to make them understandable enough to be disseminated at different levels.

The general idea is not to have the truth, but to have the capacity to presume the level of quality of the material which is preserved.

slide 19 These requirements imply the need to work on many things: mechanisms for authorship attribution and provenance control; content and contextual relationships; mechanisms for controlling the integrity and annotation processes. All these aspects have to be documented as much as possible and at every stage in the life cycle, to obtain a sort of 'authenticity card' for the digital objects in the repository.

Of course, it is necessary to deal with metadata and schemas for metadata that already exist, even if they would seem too constricting, because it would not be possible to do the same things without these tools and to be so concrete. A critical approach to these standards is needed.

Some examples of these are: PREMIS, which is a standard of metadata for technological preservation; descriptive standards like EAD and EAC; CIDOC-CRM, that is very useful in dealing with semantic knowledge for mapping and using what already exists.

All these materials developed by other communities have to be implemented in such a way that could be useful for our specific environment.

slide 20 The solutions that have been found are: identifying a set of attributes to catch relevant information for the authenticity as it can be collected along the life cycle of objects belonging to different domains; developing a conceptual model to describe the dynamic profile of authenticity, that is also the dynamic profile of preservation processes.

slide 21 What already exists, as PREMIS, ISAD and the other descriptive standards, should be taken into account, of course with reference to the OAIS model which is assumed as the base of the work done by CASPAR.

There is the awareness that, although this project has reached extremely important results, it is not complete; at present a valid framework has been built, but it is necessary to go on with the work.

slide 22 To establish an appropriate methodology implies several problems. One of these concerns the need to clearly identify the different levels of granularity of the information required to preserve the resources and ascertain their authenticity depending on the Designated Community and the changes in its Knowledge Base, in order to avoid the occurrence of both overload and lack of information

at the same moment.

A variety of domains have to be considered, taking into account that some concepts should be cross-domain but some others are strongly dependent on specific environments.

The attention to the peculiarity of the specific domain could also help solve the overlapping of concepts coming from different schemas (it may not be easy to decide whether an element has to be mapped onto this or that OAIS conceptual element). In any case, the authenticity team recognized that its aim is to find a set of information elements and assign them to an OAIS category: it is just a formal convention and so some uncertainties can be resolved in the course of the concrete analysis and testbeds.

Giovanni Michetti's presentation completes this one explaining the model and the procedures to preserve and evaluate authenticity.