Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# MODELING AUTHENTICITY

**Mariella Guercio – Giovanni Michetti**
University of Urbino

*CASPAR Training Day for the Cultural and Scientific Domains*
January 12, 2009
ICCU (Rome, Italy)

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# 1. INTRODUCTION

## Authenticity: A Key Component in the Preservation Process

- The preservation as developed by the main international projects in the sector (InterPARES and OAIS) requires that the elements related to the accuracy, the reliability and the completeness of the information objects are captured and maintained in the repositories to allow the users to evaluate their **identity** and their **integrity** (InterPARES project)
- These elements have to be organized according to a **conceptual model** (OAIS compliant) able to describe the dynamic profile of the authenticity as a **process** aimed at gathering, protecting and/or evaluating information/set of attributes mainly about identity and integrity



## The need for an Authenticity Management Tool (AMT)

- The complexity of the preservation function in the digital area requires the development of specific **tools** able to ensure that the main elements and procedures relevant for the quality of the preservation are maintained, and the authenticity of the preserved information objects can be presumed
- The CASPAR project has identified the need for an Authenticity Management Tool with the capacity of **monitoring and managing protocols and procedures across the custody chain** in order to deliver the benefits of authenticity into information systems, from the creation to the preservation phase



## The Authenticity Position Paper

- Goal: to define how and on what basis authenticity has to be managed in the digital preservation process to ensure the trustworthiness of digital resources
- The paper has also tried to define the conceptual basis of authenticity for the CASPAR project in terms of a common glossary
- The glossary and the analysis of the key components of authenticity are based on the main results of international community projects, specifically InterPARES, and focused on the interconnections between these results and the OAIS conceptual model



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# 2. CRITICAL ISSUES

## 2.1. Integrity and Identity

## Integrity

- The **integrity** of a resource refers to its wholeness. A resource has *integrity* when it is complete and uncorrupted **in all its essential respects**. The verification process should analyse and ascertain that they are consistent with the inevitable changes brought about by technological obsolescence
- While **the maintenance of the bit flow is not always necessary**, the completeness of the 'intellectual form' is required, especially with respect to the original ability to convey meaning e.g. maintenance of colours in a map, columns in a spreadsheet, etc. In other words, the physical integrity of a resource i.e. the original bit stream can be compromised, but the content structure and the essential components must remain the same



## Identity

- A crucial point is that *identity* must be intended in a very wide meaning: the identity of a resource refers not only to its unique designation and/or identification
- **Identity** refers to *the whole* of the characteristics of a resource that uniquely identify it and distinguish it from any other resource, i.e. it refers not only to its internal conceptual structure but also to its **general context (administrative, legal, documentary, technological, some could even add social)**



## Need to cope with authenticity

- Need to develop **tools** and **methods** that ensure authenticity of objects information along the preservation process
- The main issue is to **document** them as automatically and neutrally as possible on the basis of an adequate methodology OAIS compliant



Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# 2. CRITICAL ISSUES

### 2.2. Tools for Managing Authenticity

## Requirements

- **Authenticity cannot be evaluated by means of a boolean flag** telling us whether a document is authentic or not
- **There are degrees in the capacity of presuming the authenticity of the digital resources**: the certainty about authenticity is a goal
- We have to design all the mechanisms and tools keeping in mind that
  - **we could have alteration, corruption, lack of significant data etc.**
  - **we need changes to ensure accessibility**
  - **we need tools, mechanisms and *weights* to understand their relevance and their impact on authenticity**



## Requirements

- Authenticity Management Tools have to identify mechanisms for ensuring 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 authenticity**, i.e. to the identity and integrity profile, **all along the whole preservation process by capturing and making understandable over time all the required information**

## Requirements

- The main issues for the AMT are:
  - the right attribution of **authorship**
  - the identification of **provenance** in the life cycle of information objects
  - the insurance of content **integrity** of the whole relevant digital components and their relevant contextual relationships
  - the provision of mechanisms to allow future users to **verify** the authenticity of the preserved information objects or at least to provide the capability of evaluating their reliability in term of **authenticity presumption**



## Requirements

So these requirements imply working on:

- authorship attribution mechanisms and provenance control
- content and contextual relationships
- integrity control mechanisms
- annotation process

- **Every relevant aspect has to be described and documented at every stage in the life cycle** so to have, any time is needed, a sort of 'Authenticity Card' for any object in the repository



## Solutions

- **Identify a set of attributes** (someone call them metadata :-) in order to catch relevant information for the authenticity as it can be collected **along the life cycle** of objects belonging to different domains. This means analysing and evaluating the main and most promising metadata schemas and their basic components (i.e. the weakness and strength of metadata sets like PREMIS)

- **Develop a conceptual model** to describe the dynamic profile of authenticity i.e. **to describe it as process** aimed at gathering, protecting and/or evaluating information mainly about identity and integrity



## Methodology

- **Authenticity Team started taking into account PREMIS, and mapping ISAD and other descriptive standards onto OAIS** just to have a very general idea of some fundamental information elements which are to be preserved for 'authenticity purposes'

- This was assumed **as a starting point** to find some more elements by taking into account other resources (i.e. ISAAR, EAD, EAC, InterPARES, …)

- **CIDOC CRM** was assumed as a suitable means of expressing concepts and as a resource giving us clues about relevant aspects needed for consideration, especially about dynamic aspects (temporal entities)



## Methodology

Problems:

- **level of granularity**. Authenticity fundamental requirements must be clearly identified in order to avoid at the same time overload and lack of information (a relevant aspect for scientific but also cultural domain intended as dynamic environment with significant values in the current life of the creators and preservers like performing arts, digital music, protecting memory institutions)

- **variety of domains**. Authenticity methodology and concepts are cross-domain but their deployment is strongly dependent on specific environment. For example:
  - the Reference Information for a book could be ISBN, very specific and not suitable for other typologies
  - the authorship concept is quite 'easy' for a book but what about the author of a movie , or other cultural products in the performing arts?



## Methodology

Problems:

- **overlapping of concepts** coming from different schemas. It's not easy to decide whether an element has to be mapped onto either this or that OAIS conceptual element (e.g. whether the ISAD element "System of arrangement" belongs to either OAIS Provenance or OAIS Context). Anyway, the Authenticity Team recognizes that the its aim is to find a set of information elements and assign them to an OAIS category: it's just a formal convention and so some uncertainties can be resolved

Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# 3. AUTHENTICITY IN CASPAR

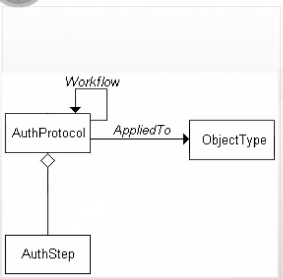
## 3.1. The Model

## Authenticity Protocol (AP)

- The protection and assessment of the authenticity of digital object is a **process**. To manage this process, we need to define the procedures to be followed
- We call one of these procedures an **Authenticity Protocol** (abbreviated as **AP**)
- We use UML notation to express the model



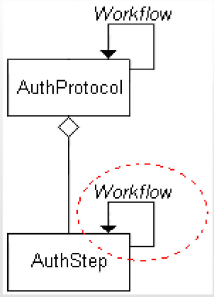
## Authenticity Protocol (AP)




- An AP is a set of interrelated steps, each called **Authenticity Step** (abbreviated as **AS**)
- An AP is applied to an **Object Type**, i.e. to a class of objects with uniform features for the application of an AP
- Any AP may be recursively used to design other APs, as expressed by the general ***Workflow*** relation



## Authenticity Step (AS)

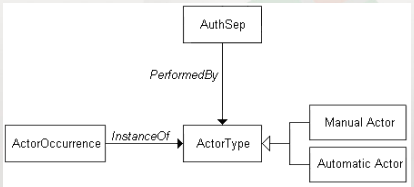



- Every AS models a part of an AP that can be executed independently as a whole, and constitutes a significant phase of the AP from the authenticity assessment point of view. The relationships among the steps of an AP establish the order in which the steps must be executed in the context of an execution of the protocol
- To model these relationships we can use any workflow model. We do not enter into the details of this modeling here, and simply denote as ***Workflow*** the set of required relationships



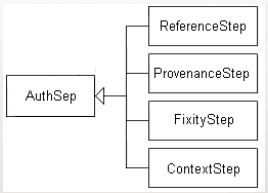
## Authenticity Step (AS)

- An AS is performed by an **Actor Type**, a class of either human or non-human agents instantiated through the **Actor Occurrence** class. The **Actor Type** is a generalization of both **Automatic Actor** and **Manual Actor**, the former performing tasks in an automatic way (hardware/software), the latter requiring human intervention





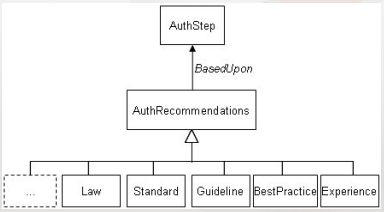

## Authenticity Step (AS)




- There can be several types of ASs. According to OAIS, we distinguish Steps based on the kind of PDI required to carry out the AS. Consequently, we have four types of steps:
  - **Reference Step**
  - **Provenance Step**
  - **Fixity Step**
  - **Context Step**

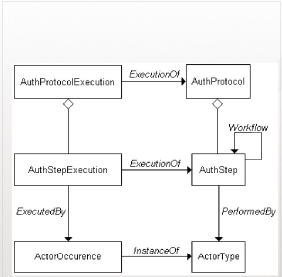
## Authenticity Recommendations

- Since an AS involves a decision, it is expected that it 'contains' at least information about:
  - good **practices**, **methodologies** and any kind of **regulations** that must be followed or can help in the analysis and evaluation
  - possibly the **criteria** that must be satisfied in the evaluation






## Authenticity Protocol/Step Execution (APE/ASE)

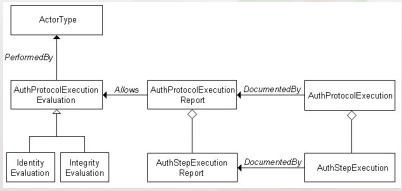



- APs are executed by an actor on objects belonging to a specific typology. The execution of an AP is modelled as an **Authenticity Protocol Execution** (**APE**)
- An APE is related to an AP via the ***ExecutionOf*** association and consists of a number of execution steps (**Authenticity Step Execution - ASE**)
- Every ASE, in turn, is related to the AS via an association analogous to the ***ExecutionOf*** association, and contains the information about the execution, including:
  - **the actor who did the execution**
  - **the information which was used**
  - **the time, place, and context of execution**
- Every ASE is executed by an **Actor Occurrence**, i.e. an instantiation of the **Actor Type**



## Authenticity Report

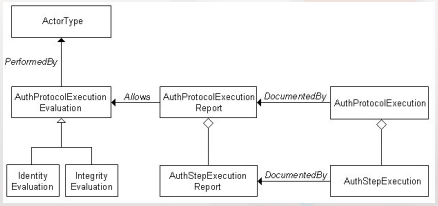
- Different types of ASEs will have different structures and the outcomes of the executions must by documented in order to gather information related to specific aspects of the object, e.g. title, extent, dates, and transformations
- An **Authenticity Step Execution Report** simply documents the step has been done – via the ***Documented By*** relation – and collects all the values associated with the data elements analysed in a specific **Authenticity Step Execution**






## Authenticity Evaluation

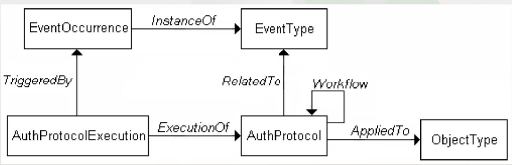
- The report provides a complete set of information upon which an entitled actor (manually, or automatically by means of a metric) can build a judgment, an **Authenticity Protocol Execution Evaluation** which states an evaluation about the authenticity of the resource referring to both the identity and the integrity profile





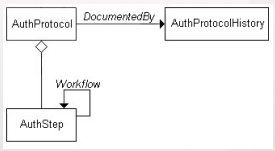

## (Authenticity) Event

- Authenticity should be monitored continuously so that any time a resource is somehow changed or a relationship is modified an **Authenticity Protocol** can be activated and executed in order to verify the permanence of the resource's relevant features that guarantee its authenticity
- Any event impacting on a resource – and specifically on a certain type of a resource – should trigger the execution of an adequate protocol: the **Authenticity Protocol Execution** is ***triggered by*** an **Event Occurrence**, i.e. the instantiation of an **Event Type** that identifies any act and/or fact related to a specific **Authenticity Protocol**



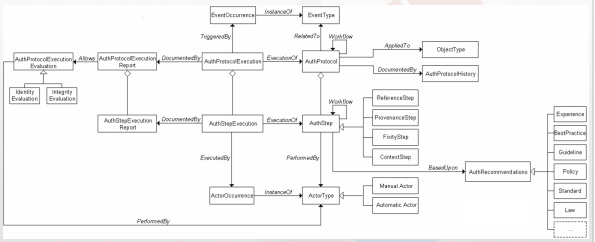



## Authenticity Protocol History




- The authenticity of a resource is strongly related to the criteria and procedures adopted to analyse and evaluate it: the evolution of the **Authenticity Protocols** over time should be documented – via the ***Documented By*** relation – in an **Authenticity Protocol History**
- The evolution of an AP may concern the addition, removal or modification of any step making up the AP, and the change of the sequence defining the *Workflow*. In any case both the old and the new step and/or sequence must be retained for documentation purposes
- When an AS of an AP is changed, all the executions of the AP that include an ASE related to the changed step, must be revised, and possibly a new execution is required for the new (modified) step

# Overall Authenticity Model






# Authenticity Model: An Example

| Prot | Obj | Descr | Steps |
|---|---|---|---|
| Prot1 | PDF | … | S1, S2, S3 |
| Prot2 | MP3 | … | |
| Prot3 | JPEG | … | |

| Steps | | Actor | Def | Rep |
|---|---|---|---|---|
| S1 | R1 | Boss of CASPAR | •Do this | Date, Place, Actor |
| S2 | P1 | Anyone | •Do that | … |
| S3 | F1 | Internal Operator | •Do | … |

| ProtExecution | Obj | Descr | StepExecution |
|---|---|---|---|
| Prot1-070726 | Myfile.pdf | … | exeS1, exeS2, exeS3 |

| StepExecution | Actor | Def | Rep |
|---|---|---|---|
| exeS1 | R1 | David | • Do this | 07-07-26, Rome, David |
| exeS2 | P1 | Monica | • Do that | … |
| exeS3 | F1 | Carlo | • Do other | … |



# Authenticity Model: Data Dictionary

**Authenticity Protocol**

- A process designed to assess the authenticity of a resource.
  Any Authenticity Protocol:
  - is composed by (at least one) **Authenticity Steps**
  - is applied to an **Object Type**
  - is documented by an **Authenticity Protocol History**
  - is related to an **Event Type**, i.e. it refers to a specific type of fact and/or action having an impact on the resource
  - is instantiated through an **Authenticity Protocol Execution**
  - may be recursively used to design other **Authenticity Protocols**

**Authenticity Step**

- A single phase of the Authenticity Protocol, aimed at analysing specific aspects of the resource. Accordingly, there are four different kind of Authenticity Steps: Reference Step, Provenance Step, Fixity Step, Context Step.
  Any Authenticity Step:
  - is performed by an **Actor Type**, i.e. a class of human or automatic agent
  - is based upon **Authenticity recommendations**, i.e. good practices, methodologies, rules, criteria, and any sort of control over the resource
  - is instantiated through an **Authenticity Step Execution**
  - is linked to other **Authenticity Steps**, in a set of relationships, according to an order established to design the specific **Authenticity Protocol**



# Authenticity Model: Data Dictionary

**Authenticity Protocol Execution**

- The execution of the process as modeled by the Authenticity Protocol
  The Authenticity Procotol Execution:
  - is composed by (at least one) **Authenticity Step Execution**
  - is triggered by an **Event Occurrence**
  - is documented by the **Authenticity Protocol Execution Report**

**Authenticity Protocol Execution Report**

- The report providing evidence of the execution of an Authenticity Protocol
- It is a composed by joining together each report (Authenticity Step Execution Report) resulting from the execution of the Authenticity Steps defining the Authenticity Protocol
- It allows a human or automatic agent (Manual Actor and Automatic Actor) to make the final evaluation about the authenticity of the resource



# Authenticity Model: Data Dictionary

**Authenticity Step Execution**

- The execution of the single phase of the Authenticity Protocol, as modeled by the Authenticity Step.
  The *Authenticity Step Execution*:
  - is a phase of the **Authenticity Protocol Execution**
  - is documented by the **Authenticity Step Execution Report**
  - is executed by an **Actor Occurrence**, i.e. an instantiation of a human or automatic agent (**Manual Actor** and **Automatic Actor**)

**Authenticity Step Execution Report**

- The report providing evidence of the execution of an Authenticity Step. It is a part of the Authenticity Protocol Execution Report

**Authenticity Protocol Execution Evaluation**

- The final assessment about the authenticity of the resource, performed by an Actor Type on the basis of the Authenticity Protocol Execution Report. The evaluation may refer to either the identity of the resource (Identity Evaluation) or its integrity (Integrity Evaluation), or both



# Authenticity Model: Data Dictionary

**Identity Evaluation**

- The evaluation of the authenticity of the resource with special regard to its identity, performed by an Actor Type on the basis of the Authenticity Protocol Execution Report

**Integrity Evaluation**

- The evaluation of the authenticity of the resource with special regard to its integrity, performed by an *Actor Type* on the basis of the *Authenticity Protocol Execution Report*

**Event Type**

- Any act and/or fact related to a specific Authenticity Protocol, i.e. that needs for attention with regard to the authenticity of the resource.
- It is instantiated through an Event Occurrence

**Event Occurrence**

- Any instantiation of an Event Type. It triggers the execution of an Authenticity Protocol

**Object Type**

- A class of objects having uniform features with regard to the application of an Authenticity Protocol

## Authenticity Model: Data Dictionary

**Reference Step**
- An Authenticity Step devoted to gather information about the identification of the resource

**Provenance Step**
- An Authenticity Step devoted to gather information about the history of the resource

**Fixity Step**
- An Authenticity Step devoted to gather information about the bit integrity of the resource

**Context Step**
- An Authenticity Step devoted to gather information about the relationships of the resource to its environment



## Authenticity Model: Data Dictionary

**Authenticity Protocol History**
- A report providing evidence of any changes of the *Authenticity Protocols*

**Actor Type**
**Actor Occurrence**
**Manual Actor**
**Automatic Actor**
**Authenticity Recommendations**
**Experience**
**Best Practice**
....

**Applied To**
- Association representing application

**Based Upon**
- Association representing control

**Documented By**
- Association representing documentation

....



## Authenticity as a Process

- According to the above, the Authenticity Management component deals with Authenticity Protocols which are **processes** defined for **specific types of objects** in order to guarantee their identity and integrity
- In this perspective, **Create/Manage Authenticity Protocol** and **Execute Authenticity Protocol** are the main features provided by the Authenticity Management component. And for that reason a part of the Authenticity Management component is a Process Editor

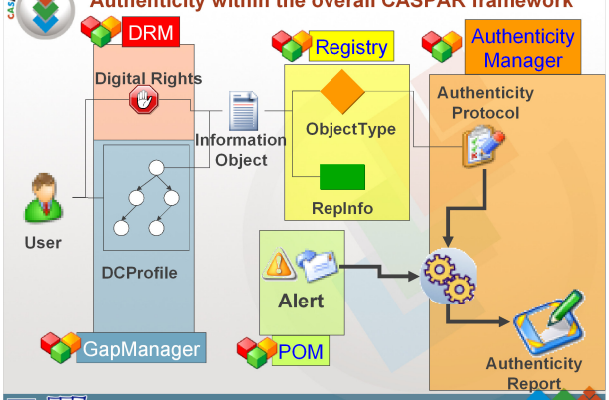


Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval

# 3. AUTHENTICITY IN CASPAR

### 3.2 Authenticity within the Overall Caspar Framework

## Authenticity within the overall CASPAR framework






## The role of the testbed partners

- IBM, IRCAM, UNESCO and ESA have been involved for the **validation of the conceptual model** and for **testing/verifying the Authenticity Model and consequently refining it**


Cultural, Artistic and Scientific knowledge
for Preservation, Access and Retrieval
Thank you for your attention
m.guercio@mclink.it
giovanni.michetti@uniroma1.it