



*Empowering users: an active role  
for user communities*

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

**Parallel sessions II**  
***Sustainable policies for digital culture preservation***



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

## **Maria GUERCIO**

### **Introduction to the session: conceptual framework and chain of custody for sustaining the digital preservation\***

\* This introduction is partially based on the keynote contribution presented at the conference *Perspectives on Metadata*, held in Vienna, 12-13 November 2009, <https://fedora.phaidra.univie.ac.at/fedora/get/o:45908/bdef:Asset/view>.

#### **A premise**

This introduction will be dedicated to present a common perspective on digital preservation by assuming that basic requirements for its success have conceptual and organizational nature, as increasingly recognized by the literature and the research outputs in the field. The metadata for preservation, the early adoption of adequate formats, controlled methods and good technical standards for acquiring digital resources play their role for ensuring the sustainability of the function, but they need to be included within a comprehensive and convincing intellectual framework and well state responsibilities. If the specific applications and related tests are not included within a systematic and robust theoretical infrastructure, the fragmentation is not avoidable and the risks for failure increase. This is why we have to put the accent on the relevance of the main goal and principles of the entire system (the defense of its trustworthiness and credibility) and its roots (the conceptual framework) and on the correct identification of responsibilities and procedural rules (the custodial environment as a chain of custody and its certification), both required for developing new products and implementing the existing solutions.

This introduction will start from two assumptions:

1. first of all, the challenges still open, specifically for handling the creation and preservation of digital resources depends on the recognition of their dynamic nature and the related need for handling as part of continuing and ongoing processes: the digital world offers a rich series of tools for the identification and capture of metadata and information on the basis of their position and encoding: they can appear as attributes of the resource itself, i.e. in the face of the digital object, as logical and physical components of its form, they can play as external elements (i.e. in a database system), but they also can act as implicit information within the procedural, technological or juridical contexts and they have to be captured and, even more, understood and maintained;
2. a pragmatic effort is required but it must be strongly rooted on consistent theory and principles specifically if we want to play with advanced technologies): it must be able to combine the best models for interdisciplinary approach, to avoid a useless overloading of detailed but not always useful information and to take into account in the application the promising outputs of the most recent research projects (PLANETS, CASPAR, INTERPARES, PREMIS just to mention those already known for their successful achievements and presented and discussed in this conference.

InterPARES is here considered as a conceptual framework thanks to principles, policies and procedures tested in many case studies and based on a consistent dictionary. The OAIS standard is recognized as a reference model for information architecture but also – specifically in the CASPAR project - as an implementation system. The guiding principle of CASPAR has been the application of the OAIS Reference Model to research, develop and integrate advanced components to be used in a wide range of preservation activities and to create a specific framework as a software platform for preservation that enables the building of services and applications that can be adapted to multiple areas, specifically to cultural, scientific and performing arts domains (that is dynamic sectors which require very complex and really evolving solutions).

CASPAR and PLANETS conceptual models have included multiple relevant results achieved in the field of preservation in the course of the last decade research efforts: the principles of InterPARES itself, the OAIS general framework, the checklist for auditing digital repositories developed in the TRAC report (Trusted Repository Audit Checklist) and in the RAC recommendations (Repository Audit and Certification), the PREMIS schema developed as metadata for digital preservation, the ISO standard CIDOC (Conceptual Reference Model) for developing ontologies and mapping metadata schemas with semantic functionality. The motivation was the creation of digital repositories and the development of framework and services for preservation based on an integrated approach to be applied to differentiated and complex archival and information systems.

The contributions presented in this session have made constant reference to these results in the specific effort for developing concrete domain-centric solutions.

The definition (and the agreement on the role) of a conceptual framework for ensuring both the consistency and the efficiency of the digital repositories requirements and of the preservation action in terms of policy,

procedures and responsibilities is a key basic issue, a condition to transform into an interrelated approach the individual solutions based on metadata identification and extraction or on the development of persistent identifiers criteria as it will be illustrated and discussed further in the course of this session.

The solidity of this analysis and chiefly the consistency of its implementation need some general statements. Specifically we could/would agree at least on the fact that the handling of digital assets as reliable, accurate and authentic heritage implies the clarification of the principle of trustworthiness.

If we look at the applications developed at national level, in most cases we could see continuing and exacting attempt for integration of principles and tools as outcome of research projects and standards development. But the fragmentation is difficult to overpass and it is even more complex to build a organic scenario.

### **The conceptual framework and the principle of trustworthiness for digital preservation**

The information and record preservation is increasingly based on concept of trust, specifically if the environment becomes digital.

First of all, it is suitable to share the definition of this term and clarify the connection between the concept of trust and the nature and quality of the digital heritage to be preserved, because the questions related to the metadata collection but also those concerning the responsibilities and the technological and organizational contexts for preservation are involved in this analysis and cannot be used conveniently and efficiently without this clarification.

In the dictionary (Merriam-Webster, s.v.) trust is identified as “a charge or duty imposed in faith or confidence or as a condition of some relationship”, a sort of “glue which binds that relationship together”<sup>1</sup>, whose ingredients have to be identified and described for effectiveness of the custody.

The custody can play successfully its role if all the elements and activities involved in this function can imply or presume a trustful handling and accomplishment.

According to the recent CCSDS guidelines, still published as draft (Recommended practice: Requirements for bodies providing audit and certification of trusted digital repositories, <http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/ReqsForAuditors>) the trust is at the basis of the certification process and at the centre of the whole process for providing solidity and efficiency to the curation action in the digital world. It involves a large community:

“to give confidence to all parties that a management system fulfils specified requirements. The value of certification is the degree of public confidence and trust that is established by an impartial and competent assessment by a third-party. Parties that have an interest in certification include, but are not limited to

- the clients of the certification bodies,
- the customers of the organizations whose management systems are certified,
- governmental authorities,
- non-governmental organizations, and
- consumers and other members of the public”.

It requires the identification of reference principles able to inspire confidence. This kind of principles includes (according to the CCSDS report):

- “impartiality,
- competence,
- responsibility,
- openness,
- confidentiality, and
- responsiveness to complaints”.

Each single attribute should be evaluated and transformed into procedures, rules, tools and metadata collection in a way to provide frames and contents for the evaluation of requirements and the recognition of the quality of digital repositories and their management and preservation systems.

Specifically, a more detailed exam of the core definitions could be of help for investigating the efficient use of metadata finalized to

- foster the credibility of the repository as trustworthy custodian on the basis of its capacity of securing integrity and authenticity of their digital contents through a standardized accumulation of descriptive and management information,
- control the cost of descriptive function “by using a simple [and standardized] encoding scheme and by ingesting metadata on transfer from public sector institutions”,
- enlarge the range of interrelations by “exchanging finding aid metadata with metadata harvesters from all kinds of communities”.

<sup>1</sup> See Jennifer Borland, *Trusting Archivists*, in “Archivi & Computer”, 2009, 1, pp. 95-106.

We do not have here time for this analysis, but it is important to recognize, within this perspective, the risk of fragmentation in the collection of all these information elements<sup>2</sup> and the low capacity of the present schemas and standards to document comparatively processes and describe them with an holistic and dynamic approach, the only one capable of dealing with the continuing evolution of the technological complexity. Of course, this last aspect, the most crucial for preserving the digital resources, requires the design of the digital preservation work as a chain of custody based not only on content identification, description and protection but also and with an increasing emphasis on the requirements for certifying institutional dedicated repositories, common policies and well defined and documented responsibilities.

### **The chain of custody: requirements, policy, responsibilities**

“The enduring trustworthiness of our documentary heritage is becoming a central responsibility of its designated custodian”<sup>3</sup>, as neutral third party on the basis that “it has no reason to alter the records and no interest in allowing others to do so, and must have the knowledge necessary to implement procedures that ensure the integrity and accuracy of the records”<sup>4</sup>. This assumption is today at the centre of a common effort made by the professionals involved in digital documents and in digital forensics, all of them persuaded that the core concepts concern the creation of a multilayer approach able to verify the integrity and authenticity of the resources at various levels of analysis:

- on the basis of the elements on the face/in the form of the resource and its attributes and metadata,
- from the circumstances of its maintenance and preservation: “an unbroken chain of responsible and legitimate custody is considered an insurance of integrity until proof to the contrary”<sup>5</sup>,
- from the integrity of essential metadata related to the resources handling and preservation as a further requirement for attestation of integrity and authenticity (individuals/offices involved, indication of annotations, of technical changes, of presence or removal [and the related time] of digital signature and other digital seals, the time of transfer to a trusted custodian, the time of planned deletion, the existence and location of duplicates outside the system,
- as inference on the basis of the trustworthiness of the record/document/information system in which the records/documents/information exist.

As Luciana Duranti has recently clearly expressed, “the authenticity...is a removable responsibility, as it shifts from the creator’s trusted ...keeper, who needs to guarantee it for as long as the record is in its custody, to the trusted custodian, who guarantees it for as long as the record exists”<sup>6</sup>.

If the framework and some basic principles seem today accepted and constitute the basis for the future implementation, some relevant details stay undetermined.

### **What is still missing**

1. consistent and accepted terminology and definitions used across domains and requested to be well understood beyond the professional communities involved in digital curation environment with specific reference to the fact that:

- definitions related to the attributes of preservation are not clearly expressed and present dangerous ambiguities<sup>7</sup>,
- new terms or the revision of traditional expressions (i.e. significant properties<sup>8</sup>) can produce dangerous misunderstanding;
- OAISS glossary has still inconsistencies even if the standard is a fruitful framework for implementing digital curation/preservation environment and has the ambition and the capacity to define concepts for a

<sup>2</sup> See Kai Naumann, Christian Keitel, Rolf Lang, “One for Many: A Metadata Concept for Mixed Digital Content at a State Archive”, The International Journal of Digital Curation, 2009, 2, <http://www.ijdc.net/index.php/ijdc/article/viewFile/120/123>: “It is the diversity of these objects which represents the key challenge in devising a metadata concept to describe, preserve and distribute them. They all need to be located on the existing finding aid system, regardless of their media format”. See also PIKKA Heutonen, “Creating Recordkeeping Metadata”, *Atlanti*, 19 (2009), pp. 67-76.

<sup>3</sup> L. Duranti, From Digital Diplomats to Digital Records Forensics, in print.

<sup>4</sup> Ibidem, with specific reference to Bernard D. Reams Jr., L. J. Kutten, and Allen E. Strehler, *Electronic Contracting Law: EDI and Business Transactions*, 1996-97 Edition (New York: Clark, Boardman, Callaghan, 1997), p. 37.

<sup>5</sup> L. Duranti, From Digital Diplomats to Digital Records Forensics, cit.

<sup>6</sup> Ibidem.

<sup>7</sup> M. Day, Preservation metadata, <http://www.slideshare.net/michaelday/preservation-metadata>.

<sup>8</sup> The definition of significant properties is emblematic of the pointlessness of this new term: “the characteristics of digital objects which must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what the purport to record” (see Andrew Wilson, but also InSPECT - Investigating the Significant Properties of Electronic Content over Time). The term seems to concentrate what common sense normally does.



Empowering users: an active role  
for user communities

INTERNATIONAL CONFERENCE  
Florence 15<sup>th</sup> - 16<sup>th</sup> December 2009

general frame: the new version (under final approval) has not been able to solve all the uncertainties even if a serious improvement is easily recognizable.

2. development of interrelations and concrete and open cooperation among relevant projects and standardization process (like PREMIS, InterPARES, PLANETS, CASPAR, DRAMBORA, RAC, CIDOC) with the aim of building an interoperable framework and diminishing the present fragmentation for a better orientation of the users.

As a consequence:

3. integration of models, schemas and business solutions to be developed in the application scenarios for handling relevant tasks as:

- authenticity and its presumption,
- storage systems in independent environment,
- automated metadata extraction: on this last point, some efforts have been made recently, but the results are slow and not enough convincing. The time is not enough to enter into details. Two recent contributions to the field could be taken into account: Kim-Ross research on automated genre classification and the FinnONTO project developed in Finland<sup>9</sup>.

The complexity and the contradictions of the digital world could have two opposite consequences, as directly experienced by many e-government legal frameworks and preservation projects: frustration and inactivity on one side, free attitude for creating, testing and supporting innovation on the other side without avoiding or hiding difficulties. Of course the last possibility requires capacity, courage and most of all confidence on the professional accumulated knowledge. The session has offered the opportunity to share ideas and increase the quantity and the quality of this knowledge in one of the most complex and relevant task we have to face, rich of promises and contradictions. One more reason to thank the organizers for this event and all of contributors for their efforts.

---

<sup>9</sup> Y. Kim, S. Ross, "The Naming of cats. Automated Genre Classification", International Journal of Digital Curation, 2 (2007), 1, <http://www.ijdc.net>; Pikka Heutonen, "Creating Recordkeeping Metadata", Atlanti, 9 (2009), pp. 67-76. For the FinnONTO project see [www.seco.tkk.fi](http://www.seco.tkk.fi).