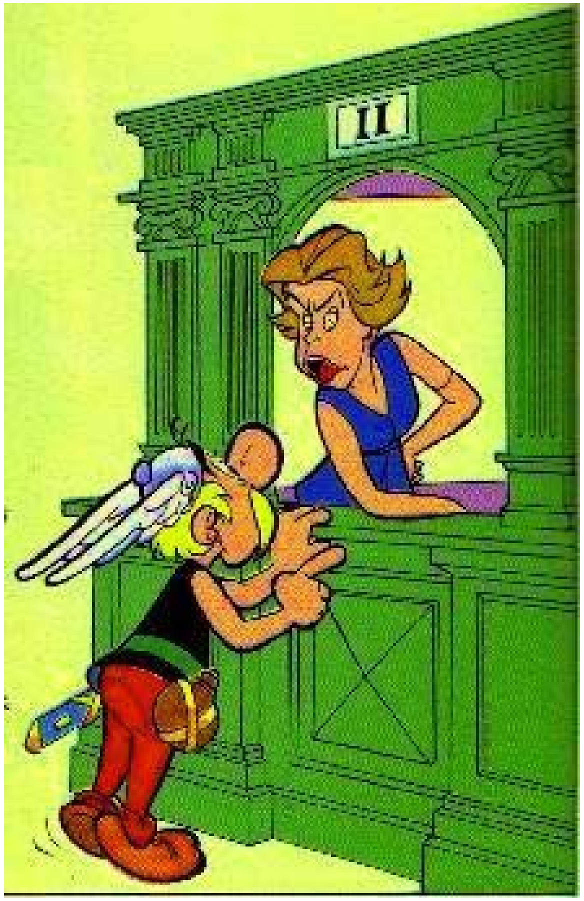


The role of standards and recommendations for **flexible** and **sustainable** ERMS

Mariella Guercio, University of Rome Sapienza, Digilab

maria.guercio@uniroma1.it

topics

- Basic principles for enhancing the digital record/document management
- Standards and functional requirements for ERMS
- ERMS principles and standards for digital continuity and preservation

basic principles for enhancing the digital record/document management

the challenges for creating, managing and preserving digital records

- The scale of the digital documentary evidence goes **beyond the capacity of single institutions or sectors**
- **Stakeholders** within and beyond the existing professional community need to be **mobilised for an active participation**
- More **coordination** is required and has to be organized within the corporate body (i.e. contact person for each unit)
- The solutions must be **sustainable and efficient**
- Tackling these challenges needs concerted and **orchestrated** efforts, as well as **synergies** of ideas, technology solutions and practices
- The **first condition** for success is related to the capacity of **understanding** the challenges and the risks involved and of identifying the **fundamental concepts** and methods to support convincing solutions
- The **second condition** concerns the **development of standards and recommendations** as technical, methodological and strategic framework to define efficient recorded information management systems

the principles: authenticity, reliability, integrity and usability

- **Reliability**: a record is reliable if it is capable of standing for the facts to which it attests
- **Integrity**: the capacity of a record of being complete and unaltered in all essential aspects
- **Usability**: a record is usable if it can be tracked, retrieved, rendered and interpreted
- **Authenticity**: a record is authentic when
  - it is what it purports to be and is free from tampering or corruption
  - its provenance and context are adequately documented
  - its creation is confirmed at a given and well defined time

the principles: authenticity is crucial and difficult to support …

- Most active systems that are presumed to contain records do not and cannot prove their **authenticity. Well formed** records need:
  - **fixed form and contextual relationships**
  - **stable content** and controlled specifications of the **variability of any property**.

- Inactive records often cannot be preserved because:
  - they were not created and/or maintained in static environments and in **preservable formats**
  - their formats are **obsolete**, their **evidence** is **lost**,
  - users are **unable** to **understand** or use them

… and to assess

- It is a complex task in digital environment due to the:
  - variety of the **new products**
  - need for **new concepts** and their related **requirements**
- The authenticity assessment implies **documentation** of the procedures and related decisions to be applied both by the creator and the preserver
- The **records lifecycle management systems** or records programs for business continuity are traditional requirements whose relevance is more crucial in the digital environment
- **Guidelines** are needed for the creator and preserver to support and accurately document decisions related to an increasing number of traditional and new requirements like:
  - file **formats**, wrappers and encoding **schemes** in the active and semi-active phases
  - quality and **granularity** of metadata/representation information required for reference, provenance, context, fixity and rights management (0AIS PDI-Preservation Description Information)
  - level and nature of the **records aggregations**
  - early decisions (and clear responsibilities) for the **storage** of inactive resources
  - efficient **timing** related to the records transfer into archival repositories

GARP – Generally accepted recordkeeping principles

- It is a **statement** created by ARMA international as a common **set of principles** for qualifying and making auditable a recordkeeping system, by identifying the distinctive characteristics of effective **information governance**
- It could be defined as a ***summa* of the disciplinary knowledge and professional expertise**, selected within international recommendations, projects and best practices
- The principles are **not detailed** and have to be supported by other specific tools whose definition implies multiple steps, chiefly the analysis of existing **standards**, national legislation, sectorial rules and internal procedures

which principles - 1

(http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles)

- **Principle of Accountability**
  A senior executive (or a person of comparable authority) shall oversee the information governance program and delegate responsibility for records and information management to appropriate individuals. The organization adopts **policies and procedures to guide personnel** and ensure that the program can be **audited**.
- **Principle of Integrity**
  An information governance program shall be constructed so the information generated by or managed for the organization has **a reasonable and suitable guarantee of authenticity and reliability**.
- **Principle of Protection**
  An information governance program shall be constructed to ensure a **reasonable level of protection** for records and information that are private, confidential, privileged, secret, classified, or essential to business continuity or that otherwise require protection.
- **Principle of Compliance**
  An information governance program shall be constructed to **comply with applicable laws** and other binding authorities, as well as with the **organization's policies**.

which principles - 2

(http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles)

- **Principle of Availability**
  An organization shall maintain records and information in a manner that ensures **timely, efficient, and accurate retrieval** of needed information.
- **Principle of Retention**
  An organization shall maintain its records and information for an appropriate time, **taking into account its legal, regulatory, fiscal, operational, and historical requirements**.
- **Principle of Disposition**
  An organization shall provide **secure and appropriate disposition** for records and information that are no longer required to be maintained by applicable laws and the organization's policies.
- **Principle of Transparency**
  An organization's business processes and activities, including its information governance program, shall be **documented in an open and verifiable manner**, and that **documentation shall be available** to all personnel and appropriate interested parties.

from principles to standards - 1

The relevance of standards depends upon broad participation in their development and, after they are developed, widespread recognition of their utility

- to limit the **fragmentation** induced by the present and future digital ubiquity
- to make effectively **available** the needed information/records
- to support **retention** and **disposal** and reduce **costs** and **risks** for storage
- to set controls and measures for conducting i**nternal audits**
- to ensure **interoperability, mainly based on automatic approaches**

from principles to standards - 2

- To promote the **automation** of the processes and make them **sustainable**
- To support the creator’s **accountability** and **efficiency** through the records adequate organization
- To limit the present drift which has transformed any **individual** into an **information/record manager (but without any professional competence)**

standards and functional requirements for ERMS

why standards on ERMS/EDMS are relevant

(http://www.arma.org/docs/standards/arma-intl-stndsdevprog-policies-procedures-v2012-01.pdf)

- they provide **guidance for the implementation of policies, systems and procedures** for the management o f recorded information throughout its life cycle
- they ensure **consistency** in the management of records and information throughout the enterprise and the RIM profession
- they ensure that **valuable information assets are protected** and remain **accessible** and **retrievable** throughout the information life cycle
- they ensure that **historical records are preserved** for future generations
- they establish uniform and readily **understandable terminology** for materials, supplies and procedures
- they establish **criteria for the selection** of products specific to a particular need
- they **eliminate** [mitigate] **misunderstanding and confusion** between suppliers and buyers relative to the specifications for equipment, materials and/or supplies on which standards are adopted
- they **advance the professionalism** of the records and information management discipline
- they **enhance interoperability** between systems
- they **promote efficiency** and **cost savings** by reducing wasted effort, ensuring consistency of procedures over time and reducing risk exposure

… and why not always exhaustive and efficient

- Too **many** but not necessarily those required.
  - ERMS standards developed by ISO, by ICA, by DLM Forum (MoReq) at international level or by national bodies:
    - ISO 15489 on record management
    - ISO 23081 on metadata for record management
    - ICA-REQ on principles and functional requirements for records in electronic office environments (2008)
    - MoReq2 (2008) and MoReq2010 (2011)
  - Standards for trusted digital repositories on the basis of ISO 14721 OAIS and *audit cheklist* developed by RLG-NARA in 2007 (ISO 16363: 2012 but also ISO 17068:2012 and DRAMBORA recommendations)

the main problems

- Too **complex** (MoReq2) or too **generic** (ISO RM 15489) in defining the basic requirements
- **Delayed** in comparison with the dynamic evolution of the innovation but also too much **controlled** by the market trends
- Too **rigid** (MoReq2) or too **flexible** (MoRe2010) but not enough smart and manageable (hundred of pages, thousands of rules) with respect to the aim of supporting the corporate business purposes and routines
- Too often **conflicting** (even within the same standard body): MoReq, but also ISO for trusted digital repositories
- More **technology-oriented** than required (i.e. security standards have been over-estimated)

what is missing

- **Mapping** among standards (with attention to the functional requirements and to their metrics and parameters)
- Assessment of their role and their **applicability**
- Critical analysis of the **industrial interests** behind the standard definition and approval; i.e. why MoReq2 and MoReq2010 in less than 3 years?
- **Concrete and independent system for certification and compliance** (at least at European level): an exacting effort is under development for ISO 16363 – Digital repositories audit and certification
- The first step should be the definition of **common basic requirements** (if any)

the devil always lurks in the details: the functional requirements - policies

- **Manuals** for defining documentary procedures in the electronic environment could be an obligation for public administrations (Canada, Italy) and a vital suggestion for private sector (not included in any standard, but increasingly relevant)
- Policies for protecting **privacy** (European directive and national legislation)
- Policies for handling special workflows of records type (i.e. **e-mails records**): MoReq, Ica-Req
- **Retention** policies including rules for transfers (national legislation or standard)
  - **Procedures** for transfer and e-archiving (UNI Sincro in Italy based on PREMIS dictionary)

the devil always lurks in the details: functional requirements – digital records transfer

data transfer: how and what to transfer with specific reference to the "range of technologies from which the transfers originate" and to the nature of data,

time of transfer: rules to ensure the promptness ("as soon as possible according to the digital continuity risk") of transfer according to sustainable negotiated procedures within the organization

internal organization of the data at the transfer time: definition of information package for submission

Transfer integrity: rules to document the transfer and the eventual changes required for records intelligibility

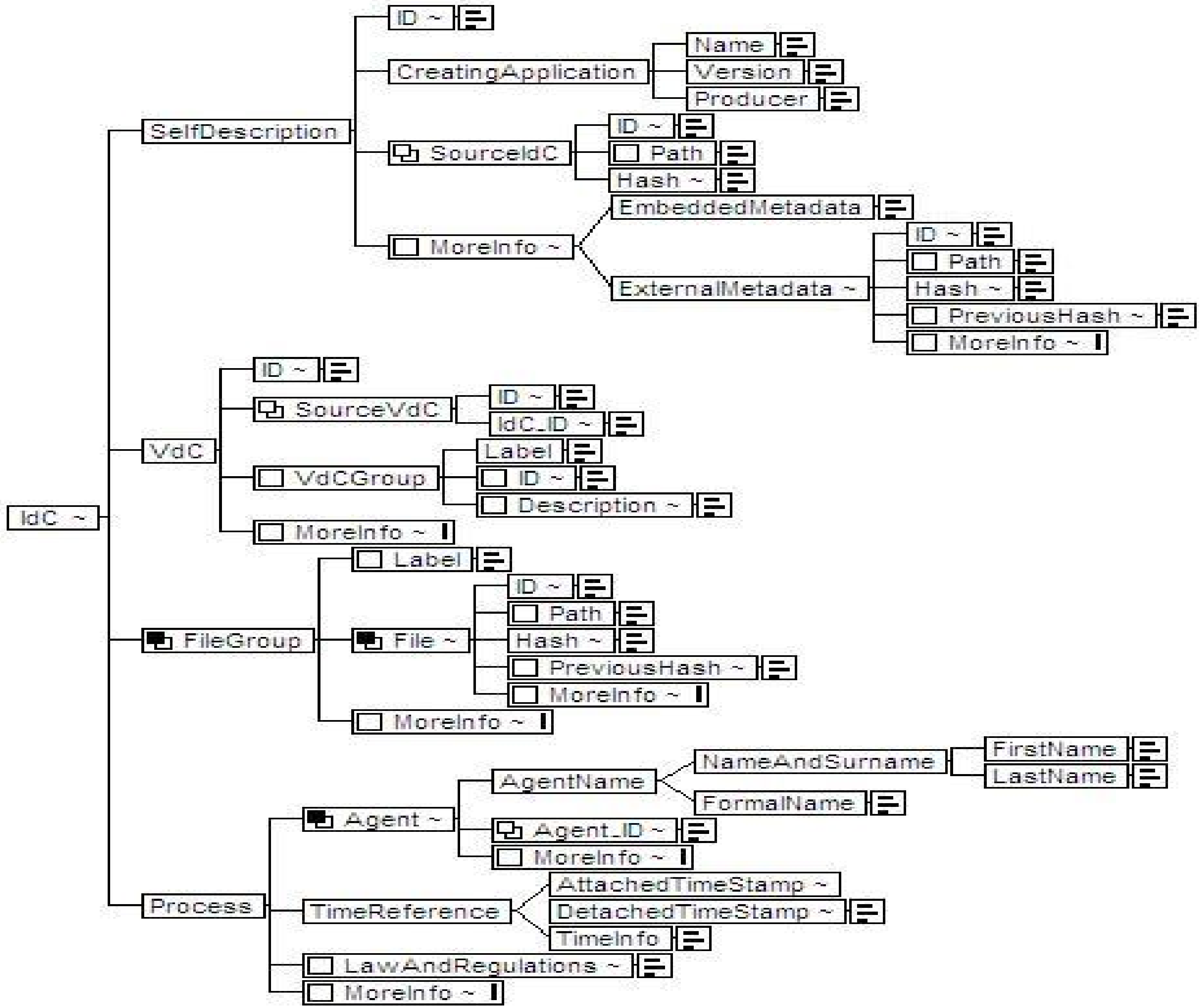
the UNISInCRO standard: the elements

1. Agent (chi interviene nel processo)
2. Agent_ID
3. AgentName
4. AttachedTimeStamp
5. CreatingApplication
6. Description (tipologia del volume)
7. DetachedTimeStamp (data e ora del volume)
8. EmbeddedMetadata
9. ExternalMetadata
10. File (indicazioni sul formato)
11. FileGroup (criteri logici di aggregazione)
12. FirstName
13. FormalName
14. Hash
15. ID (identificatore univoco)
16. IdC (indicazione dei contenuti in modo indipendente dal supporto)
17. IdC_ID
18. Label
19. LastName
20. LawAndRegulations (norme applicate)
21. MoreInfo
22. Name
23. NameAndSurname
24. Path (localizzazione)
25. PreviousHash (catena delle impronte)
26. Process
27. Producer
28. SelfDescription
29. SourceIdC
30. SourceVdC
31. TimeInfo (data di realizzazione dell'indice)
32. TimeReference
33. VdC
34. VdCGroup
35. Version

the UNISInCRO standard: the attributes

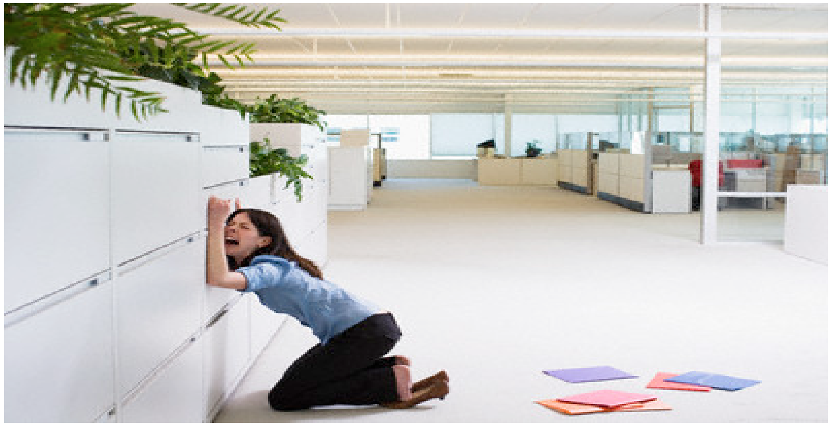
1. CanonicalXML
2. Encoding
3. Extension
4. Format
5. Function
6. Language
7. Normal
8. OtherRole
9. OtherScheme
10. RelatedIdC
11. RoleScheme
12. Scheme
13. Type
14. Url
15. Version
16. XMLScheme

the UNISInCRO standard: the XML scheme



the devil always lurks in the details:
the functional requirements – capture and classify - 1

- **Unique identification** based on persistent or at least verifiable date references for any digital records and their relevant components
- **Functional classification** interconnected with filing system to support the record function and their maintenance



the devil always lurks in the details: the functional requirements – capture and classify - 2

Unique identification of records according to ISO 15489

- **9.3 Records capture**

  The purpose of capturing records into records systems is:
  - to **establish a relationship between the record, the creator and the business context that originated it**,
  - to **place** the record and its relationship within a records system, and
  - to **link** it to other records.

Unique identification of records according to MoReq2010:

- it is defined as a **universal unique identifier** and has to be applied to any entity/component or to a record itself
- it is **not related to the concept of records capture**
- the certification function for a record or its components existence is based on **timestamp** (very complex and expensive technological system for dating digital entities)

the devil always lurks in the details: the functional requirements – capture and classify - 3

Classification according to MoReq2010

Classes and aggregations are separate entity types
Classes are held within the classification scheme
Aggregations are not specified as files, sub---files and volumes and cannot be defined on the basis of the classification
All records must be placed into an aggregation
All aggregations and records may have multiple classifications
Classification can be defined as a thesaurus

the devil always lurks in the details: the functional requirements - filing system

- Classification and **filing system** must be **integrated** for a functional organization of the records and their efficient retention and appraisal:

  the records within a file must
  share the same class code (MoReq1, very detailed rules)

- The relationships defined by the classification system and the filing plan have **a stable nature and this stability has to be identified, maintained and proved over time**

MoReq 2 – testing results for …

Test Module 3: Classification Scheme and File Organisation: 93 requirements, 61 accepted, 32 missed (65/6%)

Test Module 4: Controls and Security: 56 requirements, 40 accepted, 16 missed (71/4%)

Test Module 5: Retention and Disposition: 72 requirements, 48 accepted, 24 missed (66/6%)

Test Module 6: Capturing and Declaring Records: 92 requirements, 34 accepted, 58 missed (36/9%)

Test Module 7: Referencing: 14 requirements, 6 accepted, 8 missed (42/9%)

Test Module 8: Searching, Retrieval and Presentation: 54 requirements, 33 accepted, 21 missed (61/1%)

Test Module 9: Administrative Functions. 58 requirements, 35 accepted, 23 missed (60/3%)

MoReq2010: a low level approach for common recordkeeping requirements

- The revision has implied a **lower level of complexity for compliance** referred to recordkeeping requirements (to meet the needs of private sectors): the certification is very easy to obtain and can be required only for specific areas
- The technical requirements for **interoperability** and for **security** are very strict

open questions (beyond existing standards): policies and programs

- **What is the level of responsibilities** and **competencies** required for building and applying policies?
- What kind (if any) of **self-auditing tools** are required to verify the consistency and adequacy of policies?
- Are archivists and records managers **willing to engage with IT staff and developers** to:
  - **articulate recordkeeping principles**
  - incorporate them into current recordkeeping systems to form the basis of digital preservation programs (and **are they able** to do it at local level?)

open questions (beyond existing standards): technical aspects

- **Metadata are** expensive if not guided by semi-automating procedures starting from the active phase:
  - **when** to integrate the information required for preservation?
  - **who** will have this responsibility?
  - **how much** information/documentation have to be included for integration or excluded to avoid cumbersome redundancy?

- **storage**: outsourcing versus in-house; cloud computing risks (also for temporary records)

- **formats**:
  - the **timing** and the responsibility for converting records into an archival format
  - the preservation of the **original bitstream –** is it a problem?

crucial problems (beyond existing standards) for professionals

- Unsustainably **massive** quantities of **poorly managed data**
- **Lots of record-making systems**, too few record-keeping systems
- **Disconnect between business processes and recordkeeping** (e.g. reliance on printing-to-paper)
- Organisations often **don't know what information** they have got
- **Vital records can be hard to identify**
- the **ability to use digital records for as long as required** and through organisational, business & technological changes
  - is more than simply **storing** digital records
  - requires a range of **good records and information management**, including good metadata
  - needs good **recordkeeping to be an organic part of business**
  - includes infrastructure, tools, policies and processes.

how to prevent the corporate Alzheimer for digital records?

- Can a well organized ERM program **solve** the main challenges or at least **mitigate** the effect of obsolescence and human errors and **preserve the specific properties** of the digital records?

- Is it possible to plan the preservation as a **progressive and modular program?**

- Are there **professional profiles** that are:
  - experts in **digital recordkeeping** and **preservation**
  - able to manage **auditing** and **accreditation** processes?

- Will organizations be **motivated to (early) transfer** their digital records to more **controlled (and expensive) repositories** in absence of future space constraints?

it is time to run the preservation function in the form of digital continuity

- The **preservation service** (not the **accountability)** can be **delegated**
- The **trust** placed on custodians has to be based on qualified processes, **to document since the records are created**
- The solutions must be based on the systems and data **interoperability**
  - to mitigate the technological obsolescence
  - to improve the access
- **Digital preservation has to be interpreted in the form of digital continuity (as component of the business process for the active records), with the aim of ensuring ongoing access to essential evidence**

- Standard are relevant but the **flexibility** inherent to the technological innovation implies
  - a **proactive role** of the professionals and
  - the inclusion of ERMS/EDMS requirements in the **routine process functionality** of the business systems as part of the policy for information governance

References for standards

- Generally Accepted Recordkeeping Principles, http://www.arma.org/r1/garp
- ISO standards:
  - ISO 15489-1:20001 Information Documentation – Records Management – Part 1: General
  - ISO/TR 15489-2:2001 Information Documentation – Records Management – Part 2: Guidelines
  - ISO 23081-1: 2006 – Information and Documentation – Metadata for records – Part 1 – Principles
  - ISO 23081-2:2009 Information and documentation – Managing metadata for records – Part 2: Conceptual and implementation issues
  - ISO/TR 26122:2008 Information and documentation – Work process analysis for records
  - ISO 30300:2011 Information and Documentation – Management Systems for Records – Fundamentals and Vocabulary
  - ISO 30301:2011 Information and Documentation – Management Systems for Records – Requirements
  - ISO-16363:2012 e ISO-DIS 16919:2011: Audit and certification for digital repository

References for projects - 1

- **M. Factor, E. Henis, D. Naor, S. Rabinovici-Cohen, P. Reshef, S. Ronen, G. Michetti, M. Guercio**, *Authenticity and Provenance in Long Term Digital Preservation: Modeling and Implementation in Preservation Aware Storage, TaPP '09. First Workshop on the Ttheory and Practice of Provenance. San Francisco, 23 February 2009* http://static.usenix.org/event/tapp09/tech/full_papers/factor/factor.pdf
- **D. Giaretta, B. Matthews, J. Bicarregui, S. Lambert, M. Guercio, G. Michetti and D. Sawyer**, *Significant properties, authenticity, provenance, representation information and OAIS, iPRES 2009, The Sixth international conference on the preservation of digital objects: proceedings*, California Digital Library, 2009, pp. 67-73
- **M. Guercio**, *Modeling authenticity in CASPAR (2009),* http://www.casparpreserves.eu/training/advanced-digital-preservation-training-lectures/03.html
- **M. Guercio,** *The Italian case: legal framework and good practices for digital preservation*, in *CULTURAL HERITAGE on line – "Trusted digital repositories & trusted professionals. Firenze 11-12 December 2013*, Firenze, 2013, http://nbn.depositolegale.it/urn:nbn:it:frd-9406

References for projects - 2

- **Aparsen Project, D24.1 - Report on authenticity and plan for interoperable authenticity evaluation system,** 2012, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D24_1-01-2_3.pdf
  - Detailed analysis of the state of the art (projects and standard); proposal of a common view for capturing and evaluating authenticity evidence in a standardized way; development of a consistent methodology and of concrete guidelines to allow interoperability and support changes in data holders and processing workflows; analysis and discussion of secure logging mechanisms
- **Aparsen D24.2 - Implementation and testing of an authenticity protocol on a specific domain**, 2012, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D24_2-01-2_2.pdf
  - Test the methodology and the guidelines to check how they specialize on specific environments, case study analysis in different environments, to explore the current practices and to propose improvements , proposal and implementation of authenticity protocols (according to the CASPAR methodology)