The Long-term preservation of digital heritage

Luciana Duranti
Director, InterPARES Project

Firenze, 15 December 2009

Digital Preservation

The whole of the principles, policies, rules and strategies aimed at prolonging the existence of a digital object by maintaining it in a condition suitable for use, either in its original format or in a more persistent format, while protecting the object's identity and integrity, that is, its authenticity.

Why Authenticity?

- Records are trusted sources of evidence. Their value to their creator and their users is linked to their authenticity
- What about other types of digital heritage?
- As the concept of original disappears in the digital world, we are looking for certainty that what we observe is what it claims to be

How Can We Do That?

- Repository institutions must first and foremost ensure the integrity of their own systems and processes.
- They must also create mechanisms that allow for the determination of trustworthiness of digital entities based on their source and on the method of transmission, and then adopt the necessary strategies to manage and preserve them over time.

A Sustainable Approach

The preservation methodology must

- allow the digital entities preserved to continue to be readable and useable regardless of any technological changes to the underlying hardware or software environments, and
- account for these changes, allowing the entities and related metadata to continue to be migrated to newer technology as needed to avoid technological obsolescence.

Case Study: cIRcle

- A digital repository for the management, dissemination and preservation of the intellectual output of a university and its community members
- It includes both *records and publications*: e.g., preprints and post-prints of articles, books, theses and dissertations, raw research data files, working papers, class notes, exams, handouts, committee meeting minutes, unpublished conference presentations, etc.

cIRcle

- As of November 6, 2009 there were 14,073 items in cIRcle totaling 130GB
- It uses DSpace open source software; it assembles communities and collections; it accepts all formats
- It needs:
  Acquisitions policy related to material and sources
  File format guidelines for types of material
  Metadata guidelines for records and publications, for duplicates and “originals”

Most Importantly

It needs a

- Preservation policy (no UIR in NA has one)
- Copyright and intellectual rights policy (70.8% do not have one)
- Policies implementation procedures
- Criteria for regular assessment and evaluation of policies and procedures

Why So Many Policies?

- Ongoing copying and transformative migration are required for reasons of security (which is based on redundancy) and of continuing access
- The authenticity of digital material is dependent upon the maintenance through time of its identity and of its integrity

Why Authenticity, Again?

- Because of intellectual rights, which are attached to the authentic version of the digital object and, specifically, to its form, to the mode of expression
- Intellectual rights comprise economic rights and moral rights.
- Economic rights (e.g. copyright) enable the owner of it on a work to make commercial gain from the exploitation of that work. They can be waved or signed out.
- Moral rights are rights to reputation, attribution and association: they cannot be waved

Moral Rights

- **Right of reputation** is over the integrity of a work—such that no one, even the copyright owner, is allowed to distort, mutilate or otherwise modify the work in a way that is prejudicial to the author's honour or reputation;
- **Right of attribution** is the right to be associated with the work as its author by name or under a pseudonym, and the right to remain anonymous
- **Right of association** is the right to refuse to allow the work to be used in association with a product, service, cause or institution in a way that is prejudicial to the author's honour or reputation

Protecting Authenticity

- Not just maintaining identity (metadata)
- The main challenge: integrity
- Ghirardini and Faggioli: any form of conversion modifies the nature of the material, thus the native format must be kept or the process/system must be proven reliable beyond reasonable doubt.
- Digital Forensics in general considers open source the best tool for preservation as it satisfies the requirements of objectivity, transparency, verifiability and repeatability

Proving System Integrity

- As it regards migration, open source would allow a practical demonstration that the software could not simultaneously manipulate the content of the files while copying them, and that nothing could be altered, lost, planted, or destroyed in the process
- “Repeatability,” which is one of the fundamental precepts of digital forensics, is also supported by the accurate documentation of each and every action carried out on the material

What Other Integrity Shall We Prove?

- Data integrity: data are not modified accidentally or intentionally (value or content)
- Duplication integrity: the process of creating a duplicate of the data does not modify form and composition (either intentionally or accidentally): the duplicate is an exact copy of the original data set (form)
- The integrity of the duplicate should be linked to a time stamp as the inference of integrity is linked to the passage of time: when did the system have integrity?

Preservation Principles

- Non-interference: the method used does not change the entities (emulation)
- Identifiable interference: the method used does alter the entities but the changes are documented and identifiable (migration)
- What principle applies depends on the materials and on which of their characteristics we are aiming to protect
- In a way, it is simpler with records: protecting evidence is easier than protecting all the rights that might be embedded in a research product or in a work of art

To Sum Up

To define the most appropriate preservation strategy for IR it is necessary:

- to identify in which way recommended digital preservation strategies may infringe existing intellectual rights legislation as it applies to published and unpublished material;
- to establish what long-term preservation measures would be possible in the context of the existing legislation and to test them on IRs in course of development to assess their impact on the continuing authenticity and accessibility of the digital material; and
- to determine what changes to the law are required to ensure that the proper long-term digital preservation strategies can be applied so that the research output of universities can remain attributable and accessible in its authentic form for as long as needed.

STAY TUNED…

InterPARES 3
www.interpares.org

University Institutional Repositories Copyright and Long-Term Preservation
http://uir-preservation.org/