

Computer-Aided Hermeneutics: A Practical and Theoretical Approach to Digital Media Preservation

Bruno Bachimont* & Jean-François Blanchette†

Collective memory can only function insofar as the authenticity of its constitutive elements can be guaranteed. While digital media presents new opportunities for the constitution and sharing of collective memory, it comes at the price of considerable uncertainty regarding the possibility of long-term preservation of authentic digital media. "Digital preservation" has thus emerged as a pressing societal issue and a vibrant research area.

In the area of electronic records, several research projects have produced significant results regarding long-term preservation, authenticity, and appraisal (e.g., InterPARES). The problem of preserving relatively static electronic records pales however in comparison with that of preserving complex digital objects, such as digital media art, databases, multimedia, videogames. In particular, there has been little theoretical work in the archival field which could help address these issues.

The approach proposed here eschews the traditional archival distinction between preservation and access. Preservation is usually understood as the business of protecting the physical integrity of an artifact, while access is understood as the business of making that artifact available to some public, through some mechanisms (reading room, procedures, information system). In this paper, we argue that the business of preservation and access are intimately related, as two complementary ways of transmitting content overtime. That is, while preservation deals with the problem of transmitting *physical objects* through time, access deals with the problem of transmitting the *cultural competence* necessary to "read" the physical objects, so that they are intelligible.

Preserving invariants

Preservation is usually defined as a technical and physical process, with the objectives of preserving a physical document's formal properties of integrity and identity, so that it can be made manifest and perceivable. In this framework, preservation is independent of substance and meaning, even if those two elements are fundamental to the interpretation process.

In the framework we propose, preservation has three principal objectives: integrity, identity and *intelligibility*: (a) *integrity* consists in ensuring that archived artifacts, as

* Université Technologique de Compiègne, Centre de recherches, RH330, BP 20.529, 60205 Compiègne Cedex, France; Tel: +33 (0)3 44 23 4974; Fax: +33 (0)3 44 23 44 77; Email: bruno.bachimont@utc.fr; Web: <http://www.utc.fr/~bachimon/>.

† Department of Information Studies, University of California, Los Angeles, 218 GSEIS Bldg., Box 951520, Los Angeles, CA 90095-1520, USA; Tel: +1 310 267 5137; Fax: +1 310 206 4460; Email: blanchette@ucla.edu; Web: <http://polaris.gseis.ucla.edu/blanchette/>.

physical objects, remain the same despite the inevitable degradation brought about by time; (b) *identity* consists in being able to correctly recognize and identify archived items among others; (c) *intelligibility* consists in being able to access the archive item in a physical form enabling the reading process and interpretation of its content.

In archival practice, rooted in the printed tradition, these three characteristics of documents have a specific relationship to each other, insofar as the integrity and identity of the physical medium form the basis for the integrity, identity, and intelligibility of its content. However, the archival tradition in effect conflates what are distinct processes of documentary analysis. To see this, one can consider documents and archives as constituted of three distinct layers, that of *form*, *substance*, and *meaning*:

- *Form* is what is actually seen, heard, smelled or touched. It corresponds to the manifest, perceptible appearance of a document;
- *Substance* is what is considered to be the same between equivalent forms of a document, for example, the *content* of Hamlet, independently of its various published versions;
- *Meaning* corresponds to the some subset of the interpretative network associated with the document.

Only form corresponds to an effective physical object, while substance and meaning are *abstracted views* of that object. Form, substance and meaning establish a hierarchy of abstractions, according to *transformation rules* which postulate invariance. That is, substance is what is invariant with regard to different presentations of a content (e.g., different editions of the same book), following a set of transformation rules (e.g. typography changes are less serious than translations). In the same way, meaning is what is invariant with regard to different substances, following a set of transformation rules. The relationship of form, substance and meaning is expressed in the table below:

<i>Form</i>	What is on paper or screen
<i>Substance</i>	Invariant of presentations
<i>Meaning</i>	Invariant of content

In the case of textual documents, substance usually consists of the lexical encoding of the expression (that is, the choice of words) in a given format (that is, Unicode or a printed manual), and of the structuring of that encoding in sections, paragraphs, titles, etc. However, the choice of fonts and dimensions of margins are generally considered to be contingent with regard to substance and are left to the publisher's discretion. In general, the transformation rules that define substance as the invariance of presentations evolve according to cultural and historical imperatives.

Similarly, meaning is understood to be an invariant insofar as several different substances may express the same ideas or intentions. However, meaning is subject to *interpretative rules* that change according to context, history and culture. While conservation techniques (e.g., alkalization) ensure the physical integrity of documents over time, the *integrity of interpretation* is ensured through (1) philological and (2) hermeneutical tech-

niques:

- 1) **Philology** is the art of establishing the proper versions of texts and documents (f. ex., the reconstruction of the original texts of ancient authors (e.g., Aristotle), based on variant manuscript copies, provenance, dates). It is an **ortho-thetic** task whose purpose is to establish the true content of a spoken or written discourse, in a context of differing versions and physical manifestations of that discourse. *Ortho-thetic* means here that content is exactly (*ortho*) stated (*thetic*);
- 2) **Hermeneutics** is the art of interpreting the contents of documents. It proceeds through the characterization of the **meaning** of content, as precisely as possible, in a never-ending process of re-formulating and re-writing contents, so as to make ever more explicit the conveyed meaning.

Substance, as content abstracted from its contingent properties associated with its physical materialization, is the subject matter of philology. Meaning, as the interpretation of substance, corresponds to hermeneutics.

<i>Cultural counterpart</i>	<i>Document</i>	<i>Physical counterpart</i>
Philology Hermeneutics	Form Substance Meaning	Preservation

This stabilized hierarchy of “integrity rules”, from the physical to the interpretative, has no counterpart in the digital domain. For digital objects, no transformation rules have yet emerged that enable the abstraction of substance from form, and thus, of meaning from substance. That is, *we do not yet understand what rules of physical transformation will be considered to have no effect on a digital object’s identity, integrity and intelligibility*. And thus, digital preservation cannot (yet) be exclusively concerned with physical processes.

This implies that archiving can never consist in the mere physical preservation of content. That is, meaning can never be fixed once and for all, in some ahistorical time bubble. Rather, it must be constantly rearticulated and explicated so that it remains accessible through some hermeneutical network of texts, constantly at work minimizing the intelligibility gap between an archive and its users.

This hermeutical network of texts is enacted through the “never-ending process of re-reading, re-formulating and re-writing contents,” an intellectual activity, but not an abstract one. Interpretation is always performed through interactions with some physical medium. In the case of a book, for example, we will build our understanding of its content by making use of its various affordances – e.g., turning pages, locating chapters using the table of contents, comparing different paragraphs, underling certain passages, annotating margins, etc. These manipulations are, in effect, *cognitive operators*, insofar as they fashion our concepts and condition the relationships we can establish between them, that is, linguistic syntax reflects the operational syntax of manipulations (possible gestures and actions).

Thus, in the context of digital media, the development of interpretation rules implies the development of tools for “assisted readings,” tools which can perform effective manipulations on content and thus, enable the performance of cognitive operations necessary to re-formulate and re-write content. The following two research projects establish the foundations for the development of such tools, by delineating the appropriate “units of content.”

MUSTICA

The MUSTICA initiative, an international collaboration between two archival research centers (UCLA's Center for Information as Evidence, and UBC's InterPARES 2 Project) and two contemporary music institutions (the Institut de Recherche et Coordination Acoustique/Musique (IRCAM), and the Groupes Recherches Musicales (GRM)), performed research on the topic of contemporary music preservation between 2003 and 2004. We developed the **MUSTICA preservation tool** for the hundreds of electro-acoustic works the IRCAM has commissioned over the last 35 years. Such works typically involve combinations of traditional instruments, custom and off-the-shelf computer software and hardware, as well as various configurations of sensors and feedback mechanisms. The MUSTICA preservation tool provides the IRCAM with a mechanism to catalog all intellectual and material elements of a commissioned piece so that it may be recreated at a later time.¹

Record/Replay

Our follow-up project to MUSTICA, “Record/Replay”, aims to develop the theoretical and practical tools needed to ensure the long-term preservation, repurposing, and cultural appropriation of videogames. Such tools are becoming increasingly necessary as videogames constitute an important cultural form of our time, increasingly susceptible of transposition into other mediums (e.g., films, novels, machinima). Current solutions to the problem of preserving videogames (e.g., emulation) focus on user experience but neglect the needs of producers, yet the intellectual property value of videogames extends well beyond the end of their commercial production and creators and producers need to understand the history of their genre, in order to identify present and future trends.

Record/Replay approaches this problem from two separate angles: (a) *Knowledge Management*, i.e., developing tools which can help producers capture to the fullest extent possible the genesis of works, that is, the abstraction, description and organization of all the elements (units of content) that may be needed to recreate/repurpose the game at a later time (e.g., rules, maps, drawings); (b) *Curatorship*, that is, providing gaming communities with ways to access and interpret their cultural heritage, including tools to analyze, compare, quote, cite, sample, and otherwise manipulate games, thus fostering their cultural appropriation.

¹ A demonstration of the MUSTICA preservation tool is available on the IRCAM web site (<http://mustica.ircam.fr>).

