

## Reading digital documents: the need for a new paradigm

Bruno Bachimont

Université de Technologie de Compiègne

UMR CNRS 6599 Heudiasyc

BP 20259 60205 Compiègne cedex

Tél : 33 3 44 23 49 74 /

Mél : [bruno.bachimont@utc.fr](mailto:bruno.bachimont@utc.fr)

Website: [www.utc.fr/~bachimon.html](http://www.utc.fr/~bachimon.html)

Jean-François Blanchette

Department of Information Studies

University of California, Los Angeles

Los Angeles, CA 90005-1520

Tel: +1 310 267 5137; Fax: +1 310 206 4460

Mél : [blanchette@ucla.edu](mailto:blanchette@ucla.edu)

<http://polaris.gseis.ucla.edu/blanchette/>

Theme : Document Theory

Contribution : Powerpoint presentation + Full paper

Key words : Folder, philology, reading.

Paper and printed traditions have accustomed us with documents whose physical integrity ensures content identity and authenticity. A printed document is by itself a reference from which readings and interpretations can be drawn, those readings and interpretations being comparable because they *refer* to the *same* basis. As noted before by many, the possibility of having stable and objective documents opens the space for critics and hermeneutics, the space for intellectual work and the building of the theoretical reason that we encounter in science.

Digitization of documents has changed this situation insofar as what accounts for printed documents are now projected or reconstructed views, computed by rendition players or software from computable resources. Reader face to evanescent views on screen that exists only during the duration of the reading while what is stable and persistent are the computable resources. But direct access to those resources is impossible they cannot be viewed without the mediation of players. As a consequence, in the digital context, they are no longer anything equivalent to the stable printed document being the common reference for multiple readings.

In the digital context, the situation of the reader is comparable to an *inquiry* through which she tries to collect some indices, pieces of evidence coming from the multiple views she has of content, in order to reconstruct her own understanding and to objectivate what will be her reference. Because reference is missing, the reader builds one in order to make sense of the multiple proposed views. It should be noticed that there is a strong dissymmetry between authors and readers: most of the time, authors elaborate contents according to logical structures (for example XML structures like DTD or XML Schema) in order to obtain a structured document that constitutes their reference to consider the multiple views that can be produced thanks to the rendition devices and stylesheets. However, the reader cannot access the structured document so that she can be lost among the multiple views, lacking the common reference they derive from.

This situation is very close to those encountered by intellectual workers as physicians or jurists who have to build an interpretation and a synthesis from a collection of documents, facts, sheets, etc., that are the multiple facets of a single problem: a patient, a crime, etc. We call this kind of reading and interpretation the "folder paradigm". Drawing from this parallel, we can say that reading documents in a digital context amounts to lead an inquiry into a folder made up of computed temporary views on content.

Of course, there is not a strong discontinuity between printed reading and digital reading. For example most of documents that can be found on the web have their document characteristics and properties defined by reference to the printed tradition: the digitization and rendition of content are only a new way of transmitting content while its definition and integrity are defined elsewhere, according to the classical rules of the printed tradition. But, because the pervasive use of XML and Web technologies, that contributes to compute what is seen dynamically from what is stored, content is presented through views whose authenticity and integrity are attested by resources that are in such inaccessible. Reading digital contents in the perspective of the folder paradigm means that the reader adopts a hermeneutic approach where she selects what is relevant or not, qualifies it, annotates it, integrates it into new contents, etc. The reader browses contents in order to build her reading that can be materialized by a new document she writes or by a formal structure that memorizes paths and selections and enables to dynamically synthesize them into a computed view.

The folder paradigm raises several questions both on the theoretical and practical levels. On a practical level, issues are:

- What are the interpretative actions a reader can/should do to build her reading?
- How to relate those actions to the stored resources? How to relate them to the structured content playing the role of reference for the author?

On a more theoretical level, questions are:

- What is the real status of the resource in the digital context? Is it only resource? Is it content?
- What can be the role of a reference in a digital context where everything shown results from computations?

Finally, aside the folder paradigm, is there not the need for the development of a digital philology and hermeneutics, where we do not mean classical philology and hermeneutics exploiting digital tools, but where we claim for a new philology and hermeneutics dedicated to digital contents. Digital is not the solution it is the problem.

Our contribution will detail the folder paradigm and formulate some proposals for a digital philology and hermeneutics.