

Humanities, Human Sciences and Digitalization

Jacques P. Dubucs

Ministère de l'Enseignement Supérieur et de la Recherche - France

Changing Publication Cultures in the Humanities

Hungarian Academy of Sciences
European Science Foundation
Budapest, 27-28 November 2009

Outline

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Doing research
in Humanities
at the digital
age

1 Digitalization of the paper-based heritage

2 Doing research in Humanities at the digital age



1. Digitalization of the paper-based heritage

1. Some data

- All the books ever written = 50 billions of bytes

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

1. Some data

- All the books ever written = 50 billions of bytes
- Production of information in 2006 = 150 quintillions (150×10^{18}) of bytes = 3 millions times the information content of all the books ever written

1. Digitalization of the paper-based heritage

1. Some data

- All the books ever written = 50 billions of bytes
- Production of information in 2006 = 150 quintillions (150×10^{18}) of bytes = 3 millions times the information content of all the books ever written
- Internet world traffic in May 2009 = 500 quintillions of bytes

1. Digitalization of the paper-based heritage

1. Some data

- All the books ever written = 50 billions of bytes
- Production of information in 2006 = 150 quintillions (150×10^{18}) of bytes = 3 millions times the information content of all the books ever written
- Internet world traffic in May 2009 = 500 quintillions of bytes
- Our paper-based heritage is becoming a little and decreasing parcel of the material that the human race is continuously producing

1. Digitalization of the paper-based heritage

1. Some data

- All the books ever written = 50 billions of bytes
- Production of information in 2006 = 150 quintillions (150×10^{18}) of bytes = 3 millions times the information content of all the books ever written
- Internet world traffic in May 2009 = 500 quintillions of bytes
- Our paper-based heritage is becoming a little and decreasing parcel of the material that the human race is continuously producing
- Digitalization is the process of integrating this parcel in the ocean of digital-born information

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital
Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce Odyssey by chance?)

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce Odyssey by chance?)
 - Doubts about the possibility of achieving humanistic ideals in computerized environments

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce Odyssey by chance?)
 - Doubts about the possibility of achieving humanistic ideals in computerized environments
- Technical : how to digitalize in accordance with scholarship requirements?

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce Odyssey by chance?)
 - Doubts about the possibility of achieving humanistic ideals in computerized environments
- Technical : how to digitalize in accordance with scholarship requirements?
- Scientific : what is the impact of digitalization on human sciences?

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce *Odyssey* by chance?)
 - Doubts about the possibility of achieving humanistic ideals in computerized environments
- Technical : how to digitalize in accordance with scholarship requirements?
- Scientific : what is the impact of digitalization on human sciences?
- Philosophical : what is the whole meaning of the digitalization enterprise

1. Digitalization of the paper-based heritage

2. Some questions raised by these data for Humanities

- Affective (anxiety of paper-born scholars)
 - Love or nostalgia of books as objects
 - Fear from commensurability between valuable and insignificant (is anything really reducible to bytes, as in Borel's story of typist monkeys which eventually produce Odyssey by chance?)
 - Doubts about the possibility of achieving humanistic ideals in computerized environments
- Technical : how to digitalize in accordance with scholarship requirements?
- Scientific : what is the impact of digitalization on human sciences?
- Philosophical : what is the whole meaning of the digitalization enterprise
- Managerial : how to conceive work environments suitable to the new situation?

1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital
Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes
 - Nothing specific : the situation was worse with books firing in Alexandria, where *content* has been lost



1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes
 - Nothing specific : the situation was worse with books firing in Alexandria, where *content* has been lost
 - Therapy : to change the culture of paper age (safeguard)

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital
Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes
 - Nothing specific : the situation was worse with books firing in Alexandria, where *content* has been lost
 - Therapy : to change the culture of paper age (safeguard)
- Fragility of hardware (but the same happens for paper (a fortiori, papyrus) : problems with ferro-gallic inks, etc. Moreover, solvable by re-embodiment of software into new hardware)

1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes
 - Nothing specific : the situation was worse with books firing in Alexandria, where *content* has been lost
 - Therapy : to change the culture of paper age (safeguard)
- Fragility of hardware (but the same happens for paper (a fortiori, papyrus) : problems with ferro-gallic inks, etc. Moreover, solvable by re-embodiment of software into new hardware)
- Obsolescence of computer software. Real problem : html websites will not survive their creators ; money is continuously lost to organize the migration of databases towards new software, etc. Solution : new culture, open sources.

1. Digitalization of the paper-based heritage

3. Fragility of digital data

- Experience of computer crashes
 - Nothing specific : the situation was worse with books firing in Alexandria, where *content* has been lost
 - Therapy : to change the culture of paper age (safeguard)
- Fragility of hardware (but the same happens for paper (a fortiori, papyrus) : problems with ferro-gallic inks, etc. Moreover, solvable by re-embodiment of software into new hardware)
- Obsolescence of computer software. Real problem : html websites will not survive their creators ; money is continuously lost to organize the migration of databases towards new software, etc. Solution : new culture, open sources.
- Reminder : Digitalization is (normally) not destruction, but protection, of the digitalized source !



1. Digitalization of the paper-based heritage

4. Philology revamped

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

- XXth Century : shift from philology (= textual scholarship : text is text) to hermeneutics (= Social and Literary Studies : text is context)

1. Digitalization of the paper-based heritage

4. Philology revamped

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

- XXth Century : shift from philology (= textual scholarship : text is text) to hermeneutics (= Social and Literary Studies : text is context)
- XXIth Century : back to text as *document*, i.e. as the text together with the evidence of its own production and reception

1. Digitalization of the paper-based heritage

4. Philology revamped

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

- XXth Century : shift from philology (= textual scholarship : text is text) to hermeneutics (= Social and Literary Studies : text is context)
- XXIth Century : back to text as *document*, i.e. as the text together with the evidence of its own production and reception
- Context can be viewed in such a frame as well.



1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

Document = textual object viewed from all these relevant perspectives

1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)

Document = textual object viewed from all these relevant perspectives



1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)
- a piece of German or French language

Document = textual object viewed from all these relevant perspectives

1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)
- a piece of German or French language
- a proof sheet (corrected or no)

Document = textual object viewed from all these relevant perspectives

1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)
- a piece of German or French language
- a proof sheet (corrected or no)
- a translation

Document = textual object viewed from all these relevant perspectives

1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)
- a piece of German or French language
- a proof sheet (corrected or no)
- a translation
- something Frege wrote about

Document = textual object viewed from all these relevant perspectives

1. Digitalization of the paper-based heritage

5. The very idea of document

The same object can be viewed as

- a piece of philosophical work
- a set of printed pages (can be even subcategorized, according to whether we speak about the book as type or about the book as token : in the last case, an item of Husserl's personal library, or something where Fermat has written (think to Fermat's copy of the French translation by Bachet of Diophantes' *Algebras*), etc)
- a piece of German or French language
- a proof sheet (corrected or no)
- a translation
- something Frege wrote about
- etc (more later on "etc")

Document = textual object viewed from all these relevant perspectives



1. Digitalization of the paper-based heritage

6. Textual object as text

We want firstly the text (= a correct transcript of the words)

- Not an easy work

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

6. Textual object as text

We want firstly the text (= a correct transcript of the words)

- Not an easy work
- Counter-example : bad Optical Character Recognition softwares, or correct OCR softwares applied to poor source (e.g. old micro-films, or poorly scanned books), do not do the job

1. Digitalization of the paper-based heritage

6. Textual object as text

We want firstly the text (= a correct transcript of the words)

- Not an easy work
- Counter-example : bad Optical Character Recognition softwares, or correct OCR softwares applied to poor source (e.g. old micro-films, or poorly scanned books), do not do the job
- Cheap scans spare money at the start, but they are quite expensive in the end (many examples, including in France)

1. Digitalization of the paper-based heritage

6. Textual object as text

We want firstly the text (= a correct transcript of the words)

- Not an easy work
- Counter-example : bad Optical Character Recognition softwares, or correct OCR softwares applied to poor source (e.g. old micro-films, or poorly scanned books), do not do the job
- Cheap scans spare money at the start, but they are quite expensive in the end (many examples, including in France)
- Lesson : scanning and, more generally, digitalization, should be supervised by scholars rather than to be left to technicians alone, specially when they work for companies or institutions that have no interest on quality and usability by scholars. Scholars should be associated to every stage of the digitalization process

1. Digitalization of the paper-based heritage

6. Textual object as document

General philosophy : textual objects are the only enduring products of human intellectual activity

- Textual objects are the only enduring products of human intellectual activity
- Writing is an activity, which has products of two types : enduring products (texts as objects) and transitory products (the thoughts that are created by reading them in the minds of the readers)
- The alleged eternal mental outputs are only the effect of re-reading textual objects at different times

Kazimierz Twardowski, *Actions and Products. Comments on the Border Area of Psychology, Grammar and Logic* (1912)

1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

First stage : Sleeping (and wet) papers in the cellars of Poznan University

[This is a part of a the project of digitalization of the Lvov-Warsaw Philosophical School archives, which is jointly performed in Paris and Warsaw (60 000 pages digitalized today)]



1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Second stage : high-definition scan of manuscripts in Warsaw

Humanities,
Human
Sciences and
Digitalization

Digitalization of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
in the digital
age

71

Fonctions et produits

Quelques remarques touchant ^{des} les confins
de la morphologie, de la grammaire et de la logique

§1. En comparant deux expressions ~~homonymes~~
~~apparentées~~ ^{apparentées} jointes de en français (comme
celles-ci :
marcher - marche; venir - voyage
sauter - saut, crier - cri, chanter - chant
parler - parole penser - pensée, errer - erreur,
piger - pignoler, trouver - trouvaille, ~~est~~
ou voit, que la première signifie une
^{action} fonction; quelle est la signification de
la seconde en rapport à la signification de
la première - c'est la question, dont
s'occupent les vers de lectures suivantes.

§2. A première vue on pourrait ^{supposer} ~~supposer~~
que la diversité ^{des} des deux ~~expressions~~ ^{expressions} n'est
que grammaticale et non logique, c'est à
dire, quelle ~~touchent~~ ^{touchent} seulement leur forme
et non leur signification. ^{Or} les mots
course, sont ~~de~~ ^{de} désignés de ~~comme~~ ^{comme} une
~~fonction~~ ^{fonction} ~~de~~ ^{de} désignant les mots venir,

1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts Second stage : high-definition scan of manuscripts in Warsaw

Humanities,
Human
Sciences and
Digitalization

Digitalization of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
in the digital
age

71

Fonctions et produits

Quelques remarques touchant ^{des} les confins
de la morphologie, de la grammaire et de la logique

§1. En comparant deux expressions ~~homonymes~~
~~apparentées~~ ^{apparentées} jointes de en français (comme
celles-ci :
marcher - marche; venir - voyage
sauter - saut, crier - cri, chanter - chant
parler - parole penser - pensée, errer - erreur,
piger - pignoler, trouver - trouvaille, ~~est~~
ou voit, que la première signifie une
^{action} fonction; quelle est la signification de
la seconde en rapport à la signification de
la première - c'est la question, dont
s'occupent les vers de lectures suivantes.

§2. A première vue on pourrait ^{supposer} ~~supposer~~
que la diversité ^{des} des deux ~~expressions~~ ^{expressions} n'est
que grammaticale et non logique, c'est à
dire, quelle ~~touchent~~ ^{touchent} seulement leur forme
et non leur signification. ^{Or} les mots
course, sont ~~de~~ ^{de} désignés de ~~comme~~ ^{comme} une
~~fonction~~ ^{fonction} ~~de~~ ^{de} désignant les mots venir,



1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasures, annotations etc)

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age



1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasures, annotations etc)
- Normalized "semantic" representation of the manuscript by means of the elements of an "ontology"

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age



1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasures, annotations etc)
- Normalized "semantic" representation of the manuscript by means of the elements of an "ontology"
 - This feature is the main difference with, e.g. Latex formats, that only keep the *formal structure* of the text (paragraphes, sections, chapters, footnotes, a.s.o.)

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasures, annotations etc)
- Normalized "semantic" representation of the manuscript by means of the elements of an "ontology"
 - This feature is the main difference with, e.g. Latex formats, that only keep the *formal structure* of the text (paragraphes, sections, chapters, footnotes, a.s.o.)
 - That is the part that achieves suitable immersion of the paper-based item into the digital ocean, by making it *commensurable* to every other piece of it (thus, the commensurability is not only by bytes!)

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Some data
about paper and
digital

Questions raised
for Humanities

Fragility of
digital

Philology
revamped

Doing research
in Humanities
at the digital
age

1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasures, annotations etc)
- Normalized "semantic" representation of the manuscript by means of the elements of an "ontology"
 - This feature is the main difference with, e.g. Latex formats, that only keep the *formal structure* of the text (paragraphes, sections, chapters, footnotes, a.s.o.)
 - That is the part that achieves suitable immersion of the paper-based item into the digital ocean, by making it *commensurable* to every other piece of it (thus, the commensurability is not only by bytes!)
 - The network of links created in this way provides the basis of hermeneutics in a new guise

1. Digitalization of the paper-based heritage

7. An example : the digitalization of Twardowski's manuscripts

Following stages

- Normalized "syntactic" representation of the manuscript as it stands (with erasions, annotations etc)
- Normalized "semantic" representation of the manuscript by means of the elements of an "ontology"
 - This feature is the main difference with, e.g. Latex formats, that only keep the *formal structure* of the text (paragraphes, sections, chapters, footnotes, a.s.o.)
 - That is the part that achieves suitable immersion of the paper-based item into the digital ocean, by making it *commensurable* to every other piece of it (thus, the commensurability is not only by bytes!)
 - The network of links created in this way provides the basis of hermeneutics in a new guise
 - Ontology is not fixed for ever in Aristotelian mood, but defined and continuously redefined by scholars community

2. Doing research in Humanities at the digital age

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Doing research
in Humanities
at the digital
age

- The idea of work in progress (books discussed and revised before publication)

2. Doing research in Humanities at the digital age

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Doing research
in Humanities
at the digital
age

- The idea of work in progress (books discussed and revised before publication)
- Collaborative work by means of wikis

3. New perspectives

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Doing research
in Humanities
at the digital
age

- Use of large databases (scientometry, quali-quantitative analyses)

3. New perspectives

Humanities,
Human
Sciences and
Digitalization

Digitalization
of the
paper-based
heritage

Doing research
in Humanities
at the digital
age

- Use of large databases (scientometry, quali-quantitative analyses)
- New way of working (link between scholars and engineers)