Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

Digitisation of Cultural Heritage Perspectives

Jacques Dubucs

Department of Human and Social Sciences

French Ministry of Higher Education and Research

Strategic Working Group - Societal and Cultural Innovation

ESFRI

Tallinn University

April 23-24, 2014

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
E L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

Outline

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

1 French Situation

2 Digital Cultural Heritage: Perspectives

French Situation

Early 2000's: *Errare humanum est*

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- 1997: Launching of *Gallica* by BNF
  Today:
  3,000,000 documents online
  500,000 books
  1,000,000 images
  50,000 manuscripts
  digitisation continues at the rate of 1500 docs a day

French Situation

Early 2000's: *Errare humanum est*

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- 1997: Launching of *Gallica* by BNF
  Today:
  3,000,000 documents online
  500,000 books
  1,000,000 images
  50,000 manuscripts
  digitisation continues at the rate of 1500 docs a day
- Scheme:
  Pure digitization (transformation of a material item into a digital item by means of scan)
  Texts as images (no navigability)
- Default: poor usability
- Diagnosis: silos of digitized cultural data (even with OCR patches) are not enough

French Situation

Since the mid 2000's: *Non perseverare HUMA-NUM est*

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Origins of change: collaboration Culture/Research (ministries and milieux)

French Situation

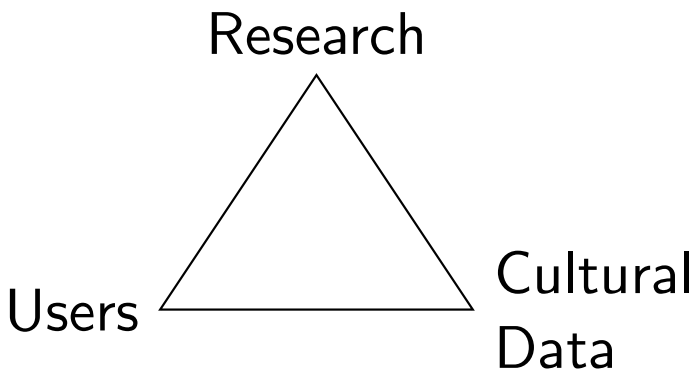
Since the mid 2000's: *Non perseverare HUMA-NUM est*

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Origins of change: collaboration Culture/Research (ministries and milieux)
- Objective: triangulation



French Situation

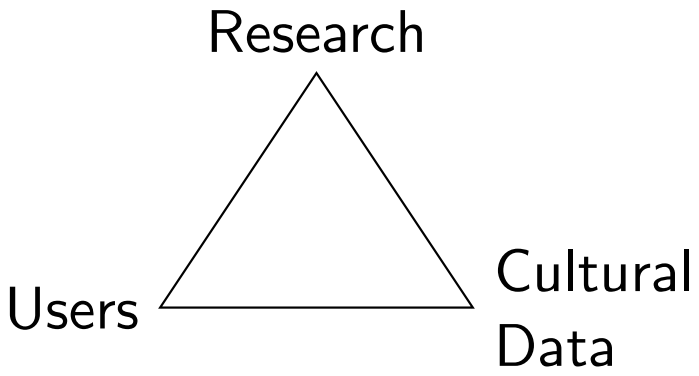
Since the mid 2000's: *Non perseverare HUMA-NUM est*

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Origins of change: collaboration Culture/Research (ministries and milieux)
- Objective: triangulation



- Result:

Huma-Num

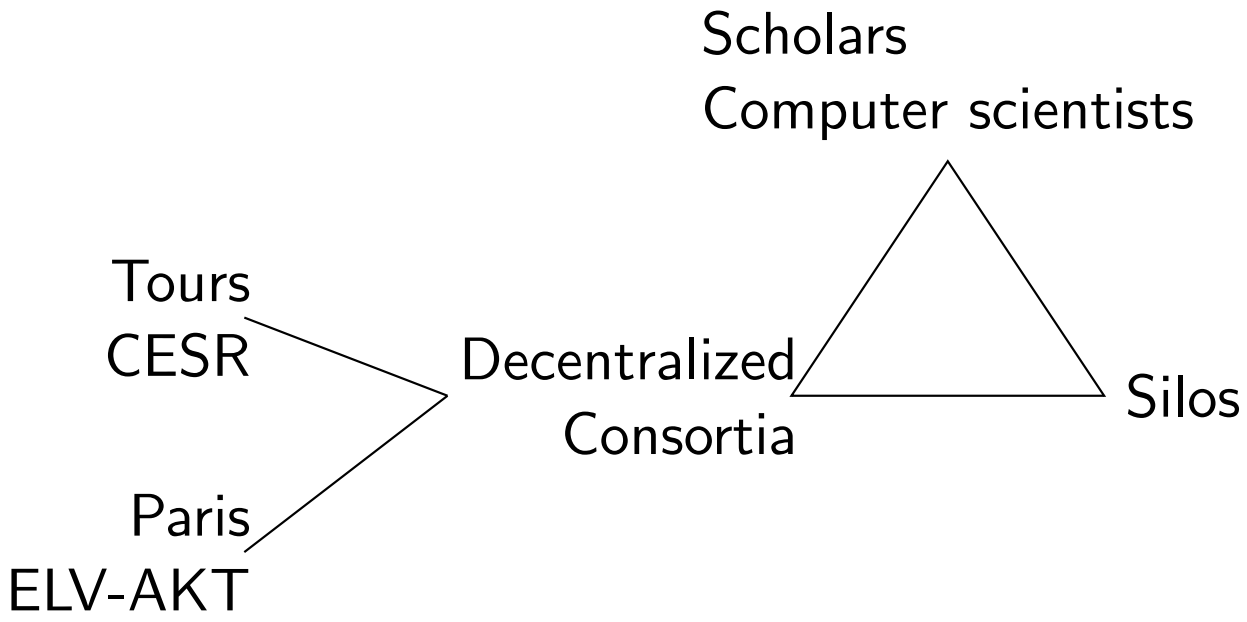
French Situation

HUMA-NUM sketchily described

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives



- Consortia
  - bring data (e.g. annotated corpora) to silos
  - use data (possibly from other consortia) from silos (access is **open** to them)
- A central team (**mix** of scholars and computer scientists) define the format of the data (e.g. XML/TEI) and the rules of maintenance/access of the silos

Digital Cultural Heritage: Perspectives

Tomorrow: what is going to be achieved across Europe - via DARIAH

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Multilingual Infrastructure
- Transnational Access
- European Cultural Common Knowledge (not the same as Unification)

Digital Cultural Heritage: Perspectives

After tomorrow: CH Big Data ?

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

DCH equivocal between

1. Digitized non digital-born CH
2. Digital-born CH

Digital Cultural Heritage: Perspectives

After tomorrow: CH Big Data ?

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

DCH equivocal between

1. Digitized non digital-born CH
2. Digital-born CH

- The problem: What does "Cultural" mean in DBCH ?
- Paper CH is self-authentifying: to exist as a book amounts to have been, at some stage, judged as worth to exist as such
- In digital realm, the material marks of *auctoritas* have just disappeared: situation of *Big Commensurability*: anything goes

Digital Cultural Heritage: Perspectives

After tomorrow: CH Big Data ?

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Some data
  1. All the books ever written = 50 billions of bytes
  2. Production of information in 2006 = 150 quintillions (150.1018) of bytes = 3 millions times the information content of all the books ever written
  3. Internet world traffic in May 2009 = 500 quintillions of bytes

Digital Cultural Heritage: Perspectives

After tomorrow: CH Big Data ?

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- Some data
  1. All the books ever written = 50 billions of bytes
  2. Production of information in 2006 = 150 quintillions (150.1018) of bytes = 3 millions times the information content of all the books ever written
  3. Internet world traffic in May 2009 = 500 quintillions of bytes
- Two possibilities
  1. To try to preserve all that *looks* "genuinely cultural" in that ocean, but what the standards for that ?
  2. To change the game: "cultural" just meaning "processed in such and such way", i.e. becoming all-encompassing

Digital Cultural Heritage: Perspectives

CH Big Data: Challenges

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- To do:
  1. Crossing Data (fiscality, health, ideology, a.s.o.)
  2. Connecting Infrastructures (CESSDA, SHARE, ESS, a.s.o.)
  3. Geo-localization of data
  4. Privacy issues (cryptography requested)

Digital Cultural Heritage: Perspectives

CH Big Data: Challenges

Digitisation of Cultural Heritage Perspectives

French Situation

Digital Cultural Heritage: Perspectives

- To do:
  1. Crossing Data (fiscality, health, ideology, a.s.o.)
  2. Connecting Infrastructures (CESSDA, SHARE, ESS, a.s.o.)
  3. Geo-localization of data
  4. Privacy issues (cryptography requested)
- The biggest challenge for SSH: Unknown Knowns (to find what is relevant in oceans of insignificance)