

# **Sens et information : les enjeux de l'ingénierie documentaire**

Bruno Bachimont

Institut National de l'Audiovisuel

Université de Technologie de Compiègne

# Plan

- Numérique
- Document
- Ingénierie documentaire :
  - Indexation
  - Structuration
- Questions ouvertes

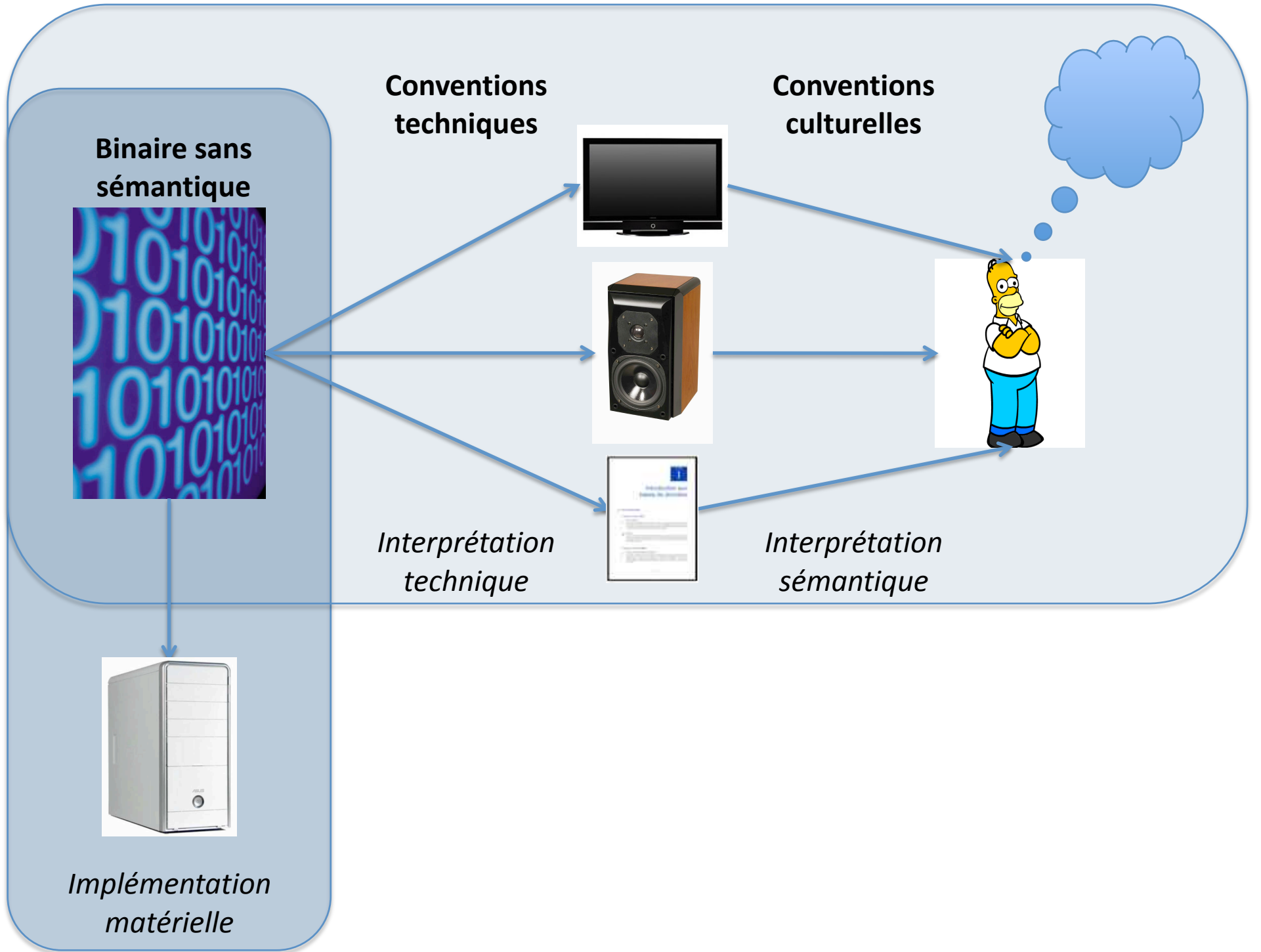
Numérique

# Le numérique...

- **Discrétisation:**
  - Se ramener à un ensemble fini d'entités primitives et élémentaires:
    - Vides de sens;
    - Indépendantes les unes des autres;
    - Distinguables de manière univoque et mécanique.
- **Manipulation:**
  - Appliquer des règles formelles (vides de sens) à ces entités:
    - Applicables par une machine.
    - Neutre par rapport à l'utilisation prévue.

# Caractéristiques du numérique

- Autothéticité:
  - Le numérique n'a pour seule effectivité que lui-même sans référence à d'autres ordres de réalité.
- Double coupure:
  - Coupure sémantique :
    - Le numérique n'a pas de sens ou d'interprétation propre;
  - Coupure matérielle:
    - Le numérique n'a pas d'ancrage matériel : il est neutre fonctionnellement vis-à-vis de la matière qui le réalise.



# Noème du numérique

## Photographie

(Barthes)

- Ça a été !

## Numérique

(BB, anonyme du XXIe siècle)

- Ça a été manipulé !

# Tendance du numérique

## *Fragmenter et recomposer*

- Le numérique fonctionne comme une bombe à fragmentation car il explose le contenu en unités arbitraires par rapport au sens ;
- La recombinaison obéit à des lois du calcul: les unités arbitraires sont recombinaisons et recombinaisons librement, sans contrainte liée au contenu.



# Grammatisation

- Décomposition et structuration des contenus du fait de leur matérialisation technique:
  - La technique de l'écriture et la constitution de la grammaire
  - La technique grammaticale (latine) et la linguistique
  - ...
- Processus associé à la globalisation des systèmes techniques et l'intégration des contenus dans le cadre numérique.

# Raison computationnelle

**Raison graphique  
(Goody)**

- Liste
- Tableau
- Formule

**Raison computationnelle  
(BB)**

- Programme
- Réseau
- Couche

Document

# Quelques repères

- Contenu:
  - Forme d'expression interprétable associée à un véhicule matériel
- Inscription:
  - Contenu fixé sur un support
- Document :
  - Inscription dans un contexte de production et de réception.

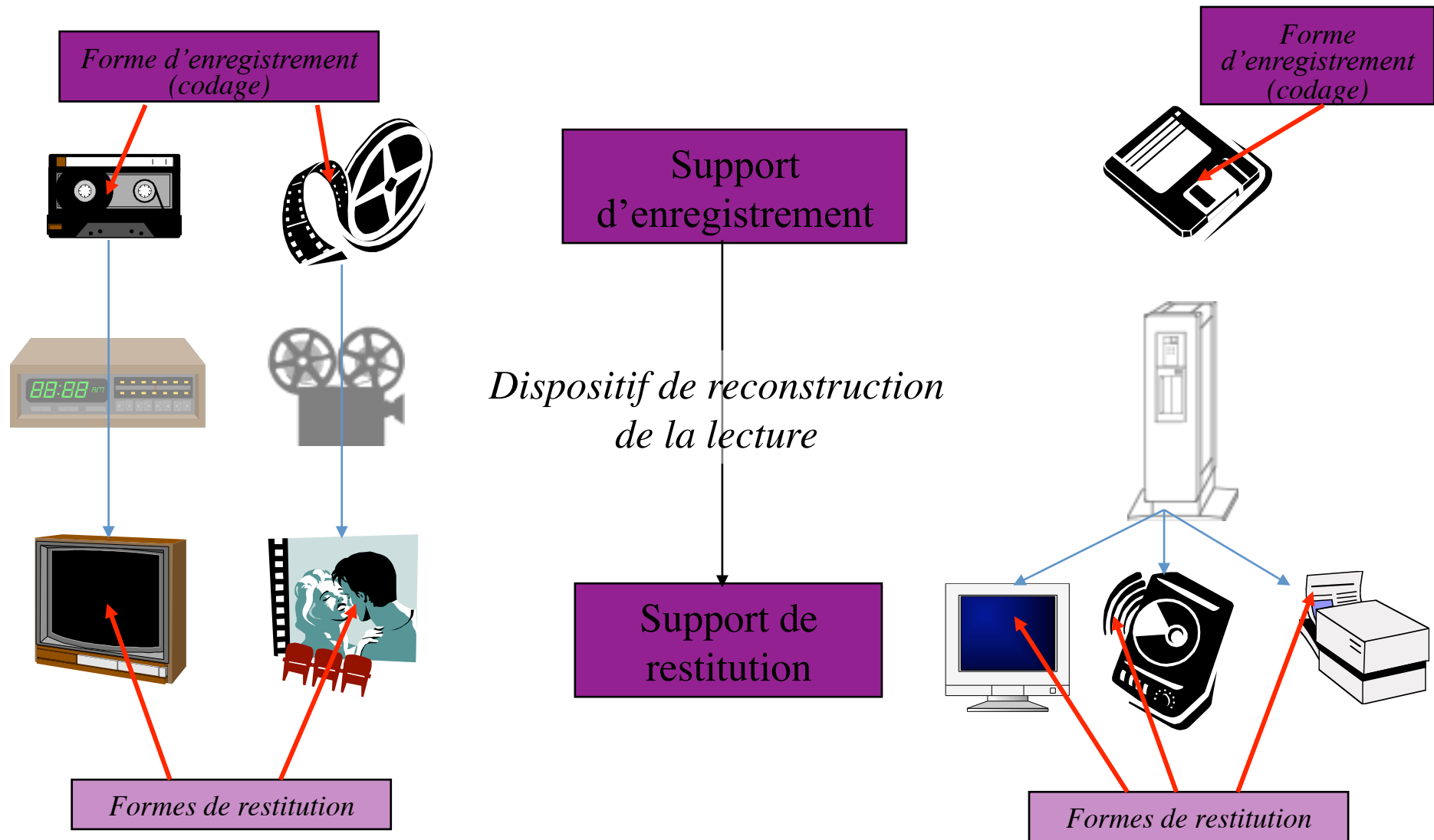
# Inscription documentaire et numérique

- Format d'inscription :
  - c'est le format physique dans lequel l'information est consignée sur le support : l'encre du papier, le code binaire, etc.
- Format de manipulation :
  - c'est le format sous lequel l'information consignée est manipulable : le code binaire, la feuille de papier, l'alphabet ascii, le pixel, etc.
- Format d'interprétation :
  - c'est le format sous lequel l'information s'interprète : les unités signifiantes. Elles sont dégagées par le parcours interprétatif, et amorcées par la segmentation perceptive du format d'inscription (le blanc entre les mots).

# Conséquences

- On pose que :
  - une inscription est numérique
    - quand les formats d'inscription et de manipulation sont confondus.
  - une inscription est documentaire
    - quand le format d'interprétation et le format de manipulation sont irréductibles l'un à l'autre et en un rapport arbitraire : manipuler le contenu à des conséquences arbitraires et non prévisibles sur l'interprétation.
  - Une inscription est formalisée
    - quand le format de manipulation devient une syntaxe permettant d'assumer et de contrôler une interprétation, en général définie par rapport à un modèle du monde.

# Les dimensions documentaires

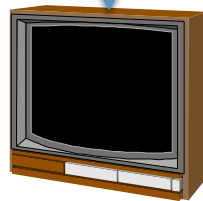


**Livre :**  
Consultation



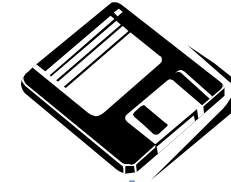
*Ce qui est consulté est identique à ce qui est conservé : archiver c'est garder intact.*

**Audiovisuel :**  
Reconstruction

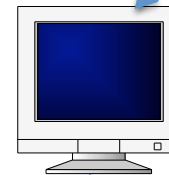


*Ce qui est consulté est reconstruit de manière dédiée à partir de la ressource: archiver c'est reconstruire à l'identique.*

**Numérique :**  
Ré-invention

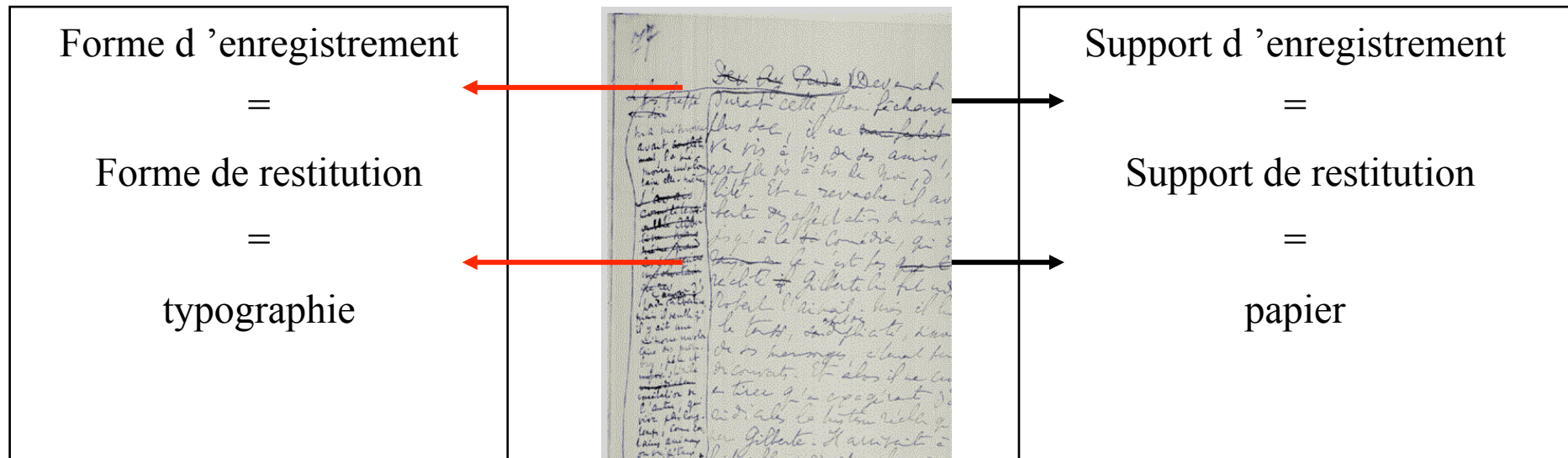


*Ce qui est consulté est reconstruit de manière plurielle à partir de la ressource: archiver c'est confronter des variantes à une ressource anonyme*





# Un cas particulier: le papier



*Dispositif de lecture : le lecteur*

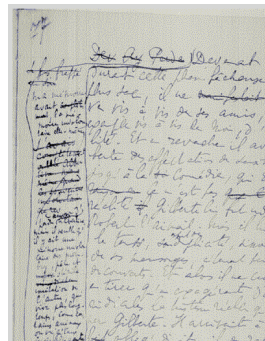
# Conséquences pour les documents papiers

- ce qui conservé est ce qui est relu.
- symétrie entre l'enregistrement et la lecture :
  - l'enregistrement est une écriture, i.e. expression en une forme sémiotique d'un contenu, c'est-à-dire une autre inscription puisque le contenu est inaccessible en tant que tel ; la lecture est l'inscription en une forme orale ou mentale du contenu, c'est-à-dire de la forme sémiotique du document.
- Il n'y a pas de dispositif d'enregistrement ou de restitution, mais seulement de lecture et d'écriture.

# Documents temporels

## Formes **statiques** / **spatiales**:

Les structures interprétables sont présentées simultanément sur le support de lecture ;  
L'ordre et le rythme de lecture sont déterminés par le lecteur.



## Formes **dynamiques/temporelles**:

Les structures interprétables sont présentées successivement sur le support de lecture ;  
L'ordre et le rythme de lecture sont imposés par le dispositif de lecture.



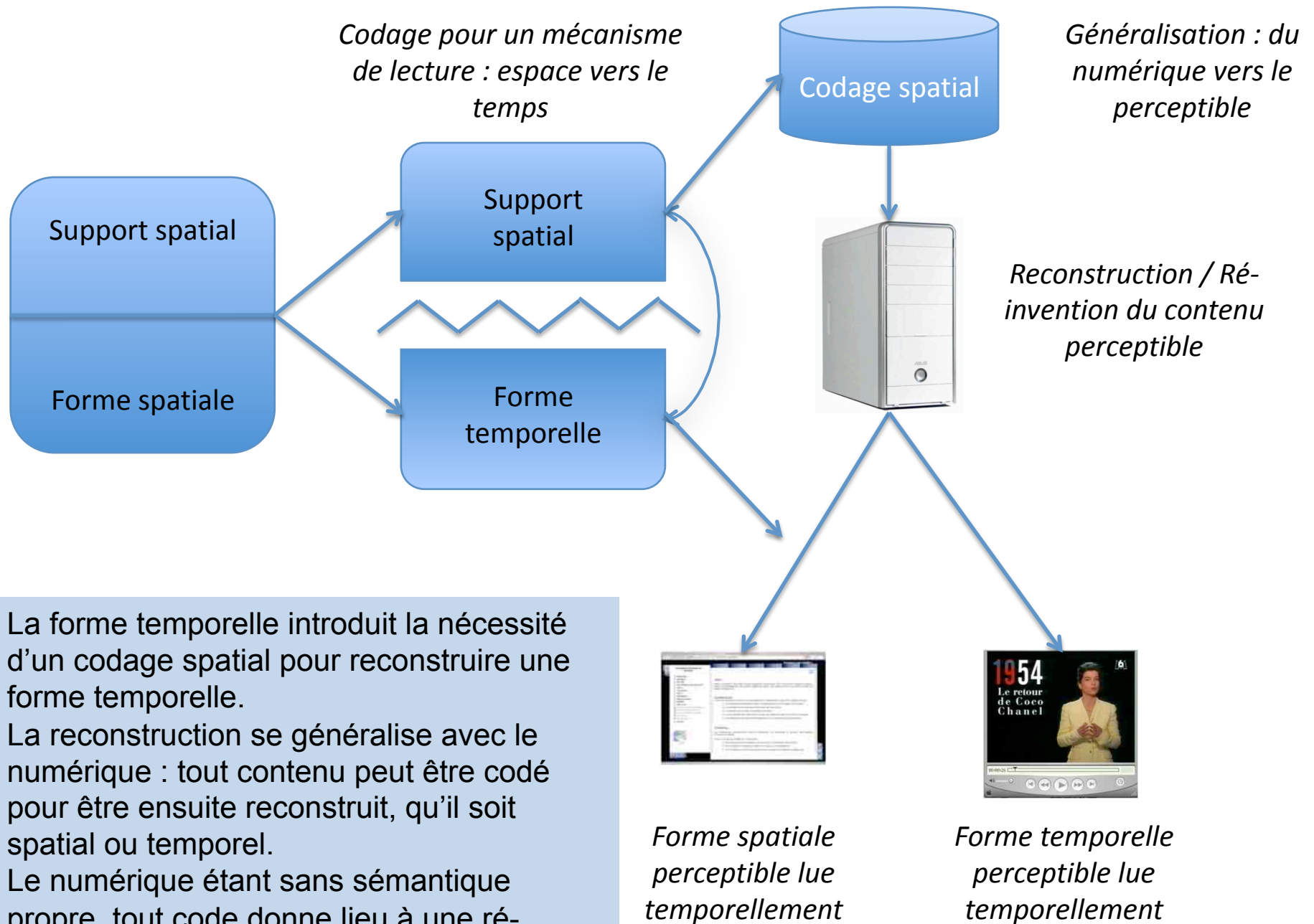
# Formes temporelles: la lecture reconstruite

- Une difficulté :
  - La lecture est un processus temporel imposé par la forme de restitution ;
  - L'enregistrement est spatial et statique ;
- Conséquences :
  - L'enregistrement ne peut être la restitution;
  - L'enregistrement est un code qui programme la reconstruction temporelle de la forme de restitution.
  - Le dispositif de lecture est nécessairement un procédé mécanique ou calculatoire.
  - le document archivé n'est plus le document, mais son enregistrement: La forme d'enregistrement assure la permanence du contenu sous une forme spatiale.

***Nécessité d'une reconstruction de la lecture***

# Document numérique

- Le document numérique n'est pas un document :
  - l'enregistrement numérique constitue une ressource à partir de laquelle un document peut être reconstruit sur un support d'appropriation ;
  - le document est une entité dynamique n'existant que lors de la projection ;
- Évolution avec le numérique :
  - Les documents habituels confondent formes et supports d'appropriation et d'enregistrement ;
  - le numérique distingue appropriation et enregistrement et rompt l'unité documentaire ;



1. La forme temporelle introduit la nécessité d'un codage spatial pour reconstruire une forme temporelle.
2. La reconstruction se généralise avec le numérique : tout contenu peut être codé pour être ensuite reconstruit, qu'il soit spatial ou temporel.
3. Le numérique étant sans sémantique propre, tout code donne lieu à une ré-invention en fonction du codage adopté

# Indexation

# Manipuler, interpréter

## Unités technique de manipulation (UTM)

- Ce que la technique saisit pour transformer le contenu matériel.
- Définies a priori.

## Unités sémiotiques d'interprétation (USI)

- Ce qui fait sens dans une interprétation.
- Définies a posteriori, comme résultant de l'interprétation et non comme la conditionnant.

L'enjeu est d'articuler UTM et USI : surmonter l'arbitraire de l'interprétation.



# Manipuler et interpréter

Problème - Cohérence entre les UTM's et les USIs :

- ce que l'on sait manipuler n'a pas toujours du sens ;
- ce qui a du sens n'est pas adressable par la technique.

Ceci est un texte qui ne dit pas grand chose car  
il n'est là que pour servir d'exemple.

Cependant, il doit se prêter aux besoins de  
l'illustration du propos et s'adapter à ses  
exigences.

Par exemple, à la queue leu leu est une  
expression lexicalisée où les blancs sont non  
signifiants.

# Document numérique

- Inscription numérique :
  - Les manipulations sont des calculs :
  - Les UTM sont calculables.
- Document numérique :
  - Inscription numérique pour laquelle les unités techniques de manipulations sont dans un rapport arbitraire à l'interprétation et aux unités sémiotiques d'interprétation.
- Document formalisé :
  - Le contenu est reformulé en une syntaxe formelle exprimant et contrôlant l'interprétation associée : les UTM et les USI sont en correspondances bi-univoques.

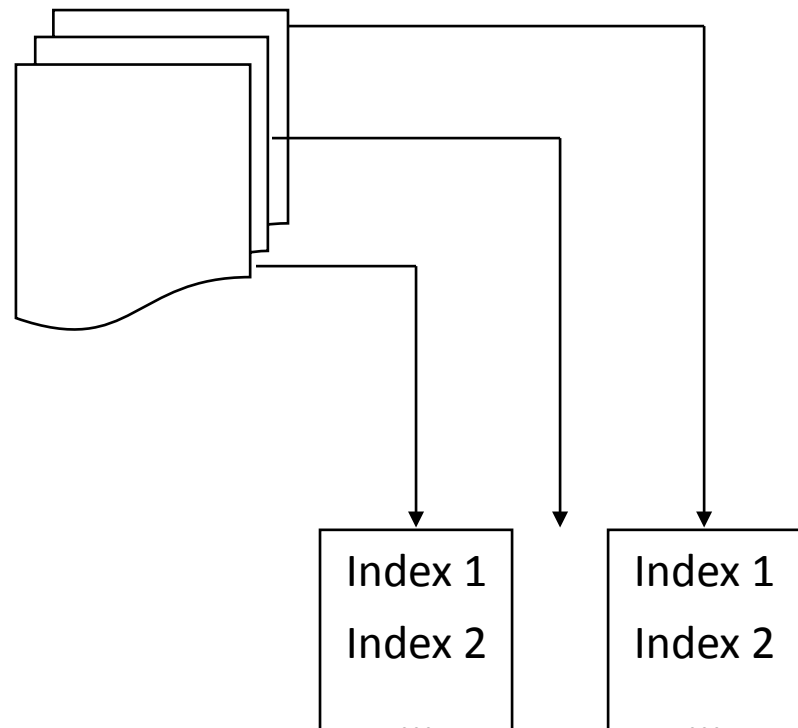
# Indexation documentaire

## Contexte : support non numérique

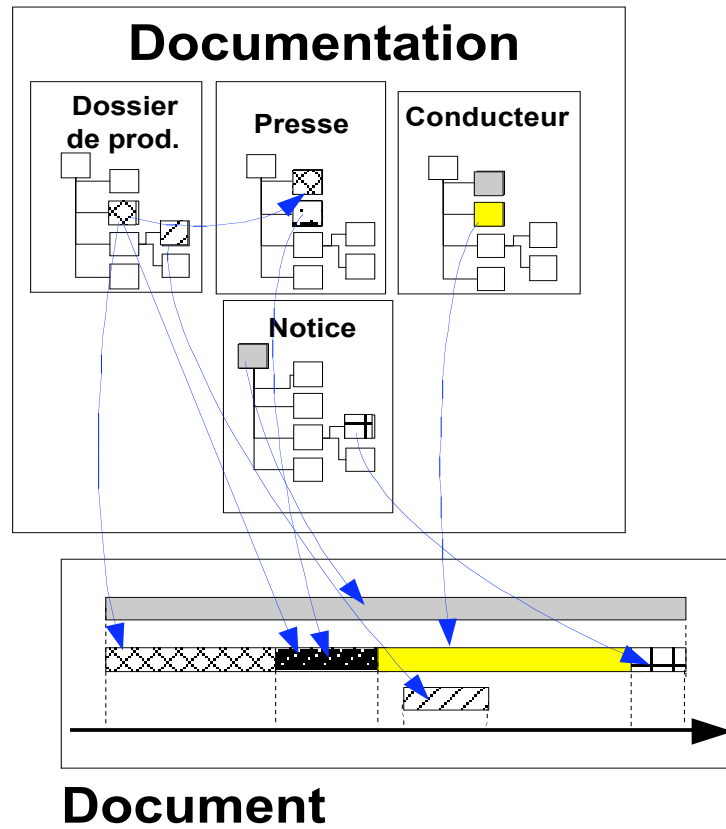
- Les UTM sont les unités documentaires physiques : livres, cassettes, documents.
- Inutile de dégager des USI plus fines, car elles ne sont pas manipulables.

UTM=USI=document

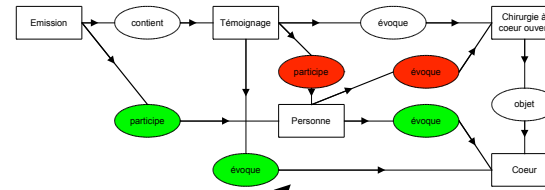
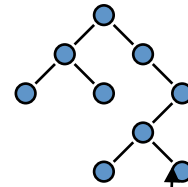
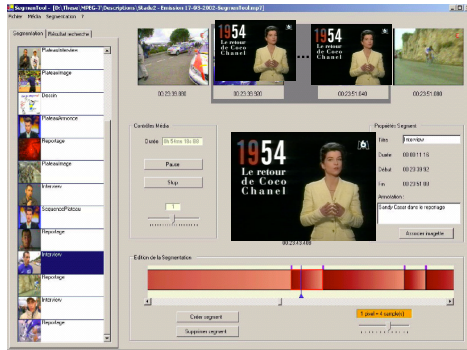
L'indexation reste globale



# Indexation fine du contenu

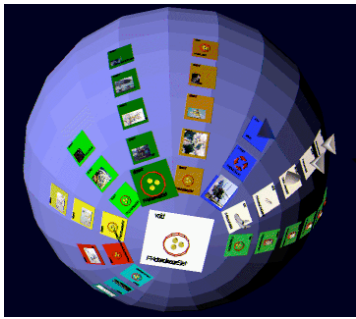


# Du signal au contenu culturel



*Raisonner  
Représenter  
Manipuler*

Bla bla bla Bla bla bla  
Bla blaBla bla bla Bla bla bla  
Bla bla  
Bla bla bla Bla bla bla  
Bla bla Bla bla bla Bla bla bla

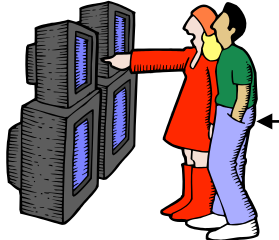
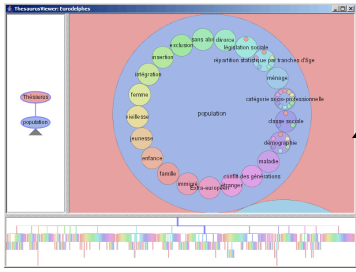


*Visualiser*



*Caractériser  
(extraire)*

*Comprendre*



# Structuration

# Paradigme documentaire-1

Jean Dupont  
3, avenue des Tilleuls  
99620 Nulle Part

Le dossier 32 doit être traité par M.  
Durand pour dans 15 jours.

# Paradigme documentaire-2

<p>

Jean Dupont

3, avenue des Tilleuls

99620 Nulle Part </p>

<p>

Le dossier 32 doit être traité par M.

Durand pour dans 15 jours.

</p>



## Paradigme documentaire-3

```
<p>  
<prenom>Jean</prenom> <nom>Dupont</nom>  
<numero>3</numero>, <typeVoie>avenue</  
typeVoie> des Tilleuls  
<cp>99620</cp> <ville>Nulle Part</ville>  
</p>  
<p>  
Le dossier <nbDossier>32</nbDossier> doit  
être traité par M. <nom>Durand</nom> pour  
dans <delais>15</delais>  
<uniteDuree>jours</uniteDuree>.  
</p>
```

# Paradigme « données »

```
<personne prenom="Jean"  
          nom="Dupont" />  
<adresse numero="3"  
          typeVoie="avenue"  
          nomVoie="rue des Tilleuls"  
          cp="99620"  
          ville="Nulle Part" />  
<tache  
  dossier="32"  
  responsable="Durand"  
  delais="15"  
  uniteDuree="jours" />
```

# Paradigme formel

## contenu

Jean Dupont  
3, avenue des  
Tilleuls  
99620 Nulle Part

Le dossier 32 doit  
être traité par M.  
Dupont pour dans 15  
jours.

## Reformulation logique pour le traitement du contenu

```
[Personne:#1]-  
  (a-pour-nom)->[Nom:Dupont];  
  (a-pour-prenom)->[Prenom:Jean];  
  (a-pour-adresse)->  
    [adresse:#2]-  
      (numero)->[integer:3];  
      (nomrue)->[string:"Nulle Part"];  
      (typeVoie)->[token:avenue];  
      (cp)->[CodePostal:99960].
```

# Deux tendances distinctes

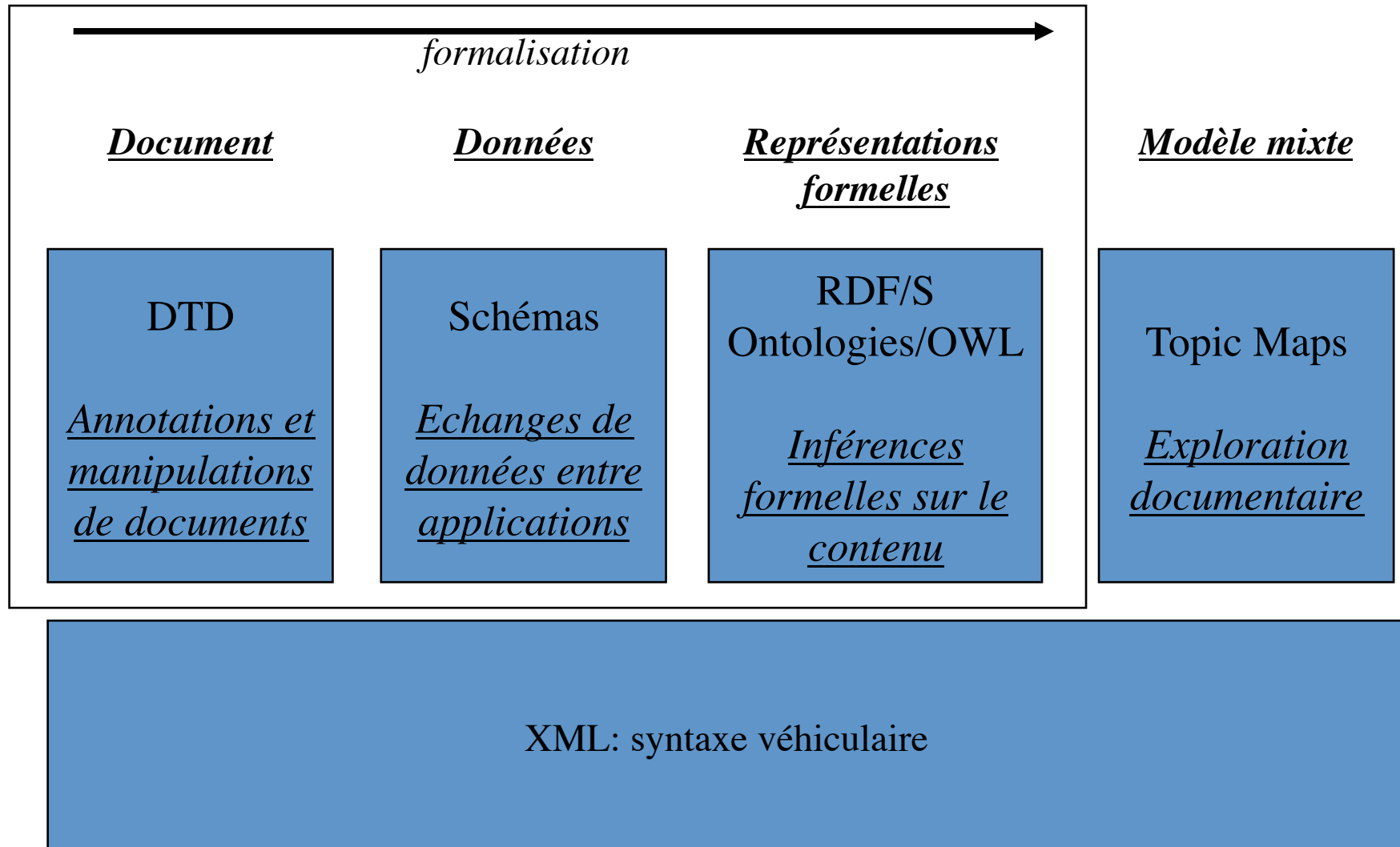
## *Paradigme documentaire*

- Enrichir le contenu par des balises ;
- Ces balises instrumentent le contenu pour son exploitation;
- Ces balises ne remplacent pas le contenu
  - Si on les enlève, on conserve le contenu.

## *Paradigme formel*

- Annoter le contenu par des représentations formelles ;
- Ces représentations remplacent le contenu pour son exploitation.
  - Si on les enlève, le contenu disparaît.

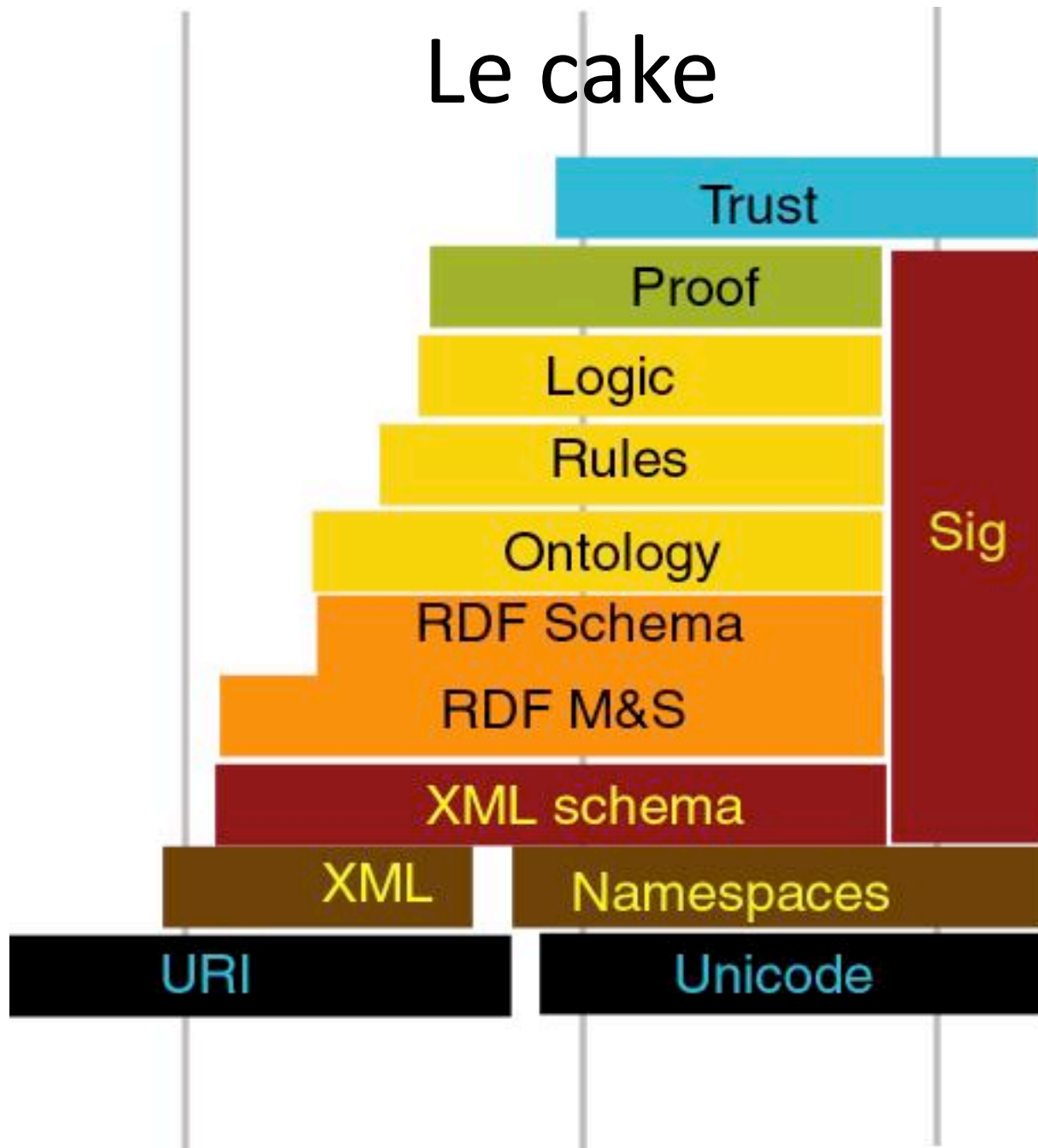
# Un même outil, différentes utilisations



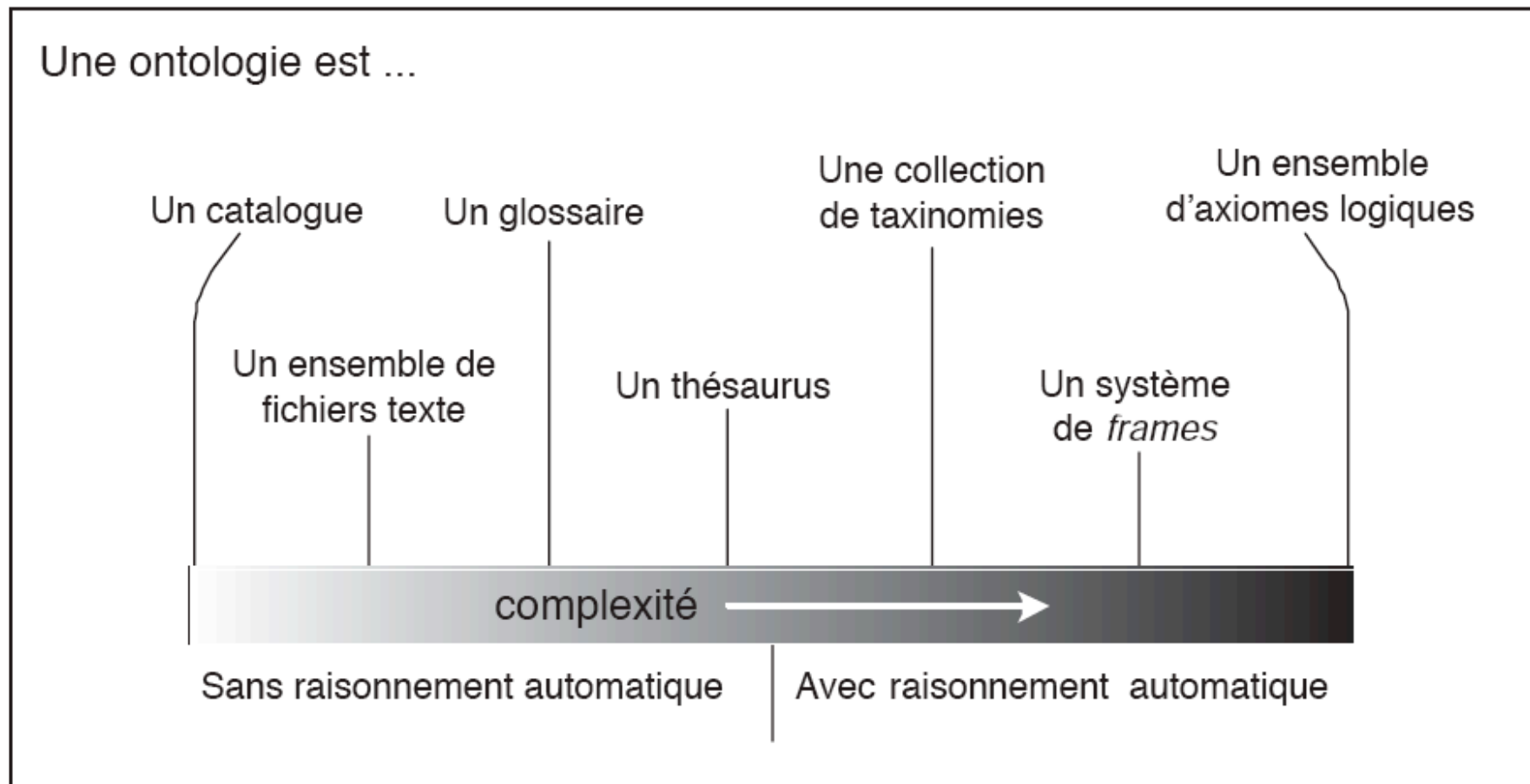
# Web sémantique: comment ?

- Représenter **formellement** la signification des contenus ;
  - Représentation logique des connaissances ;
- Reasonner **formellement** sur cette représentation ;
  - Inférence.

# Le cake



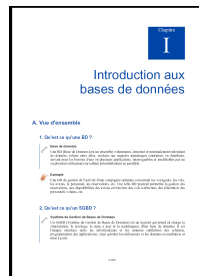
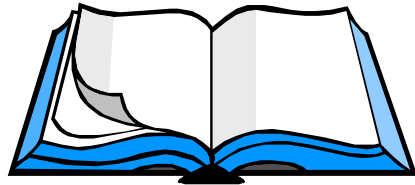
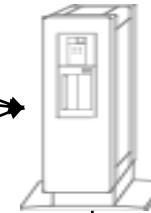
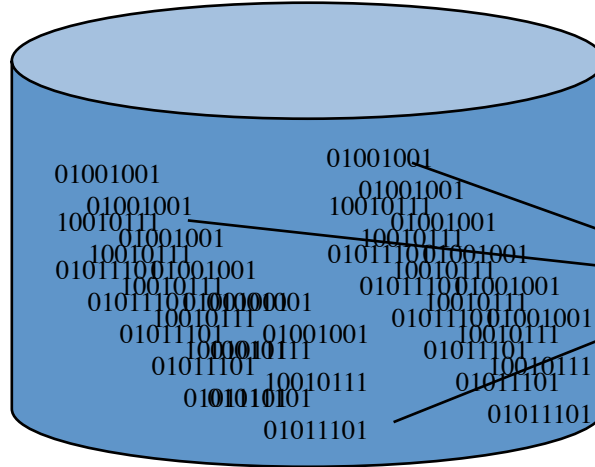
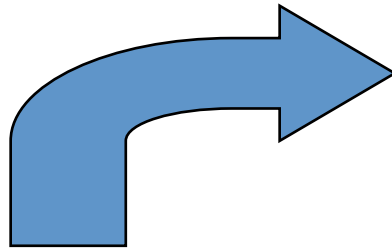
# Ontologie : des formes voisines





Tendances :  
ressources et variations

# ... aux bases numériques



Titel - v.0.1 - [lots résultats]

Historique des recherches - Répertoire - 3 - Total: 3

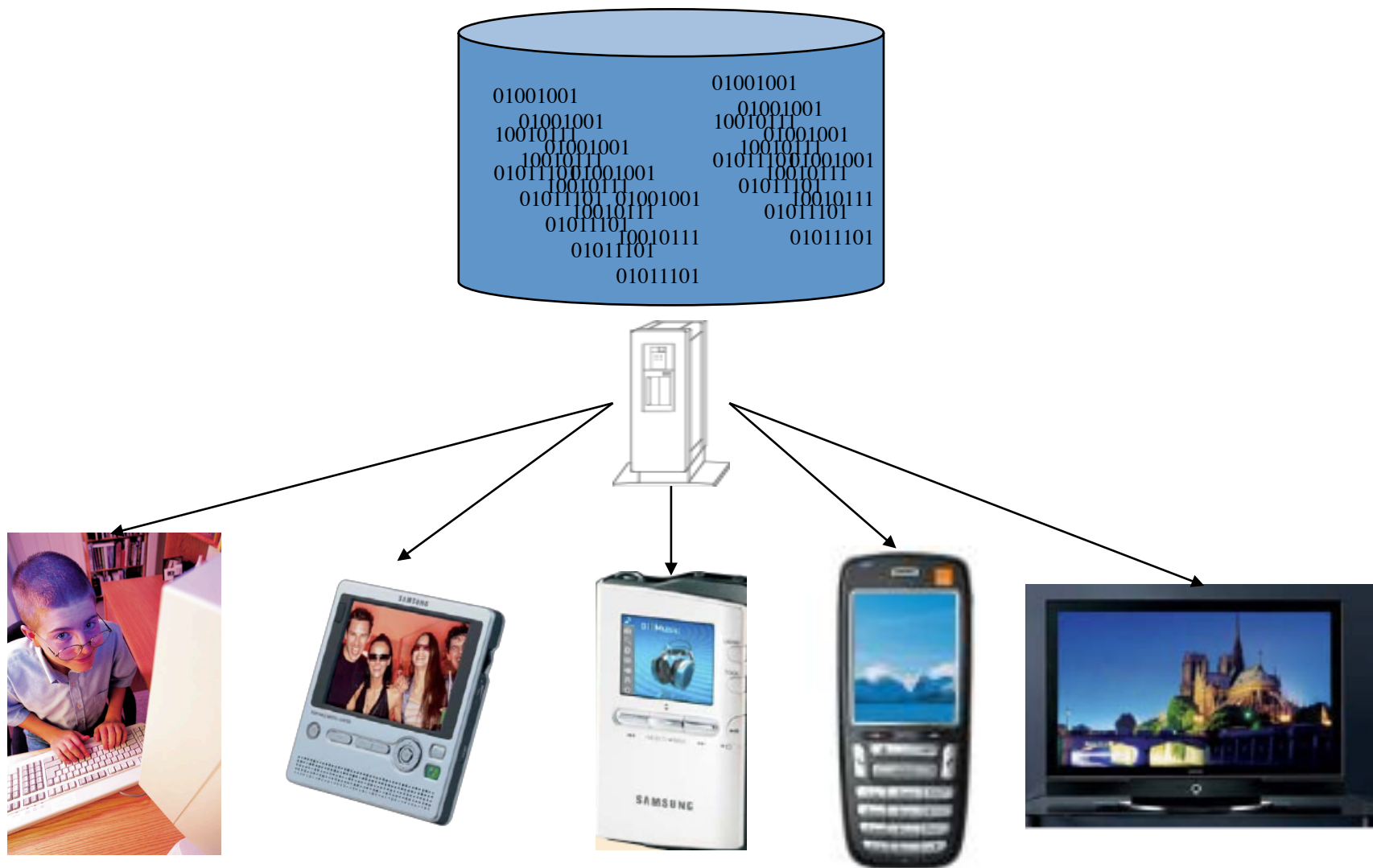
Recherche	Emissions = E...	(TS CHAMPS Cont. bea)	932	21/01/2004 10:20:14
ppoussel	Emissions Nat...	(Ts Col. + moie)	278	21/01/2004 09:20:35
	Evén	(From Cont. CONFERENCE DE PRESSE)	39	20/01/2004 16:34:26

Recherche en cours: [TS CHAMPS Cont. bea]

Nombre de page: 10 Documents ramenus: 932 Tri par: Asc. Desc.  
Enregistrement courant: 0 Date de diffusion

Document	Score	Date de diffusion
CAF97071858 JT 13h LES BEATLES	Inconnu	10/12/1963
EMISSSEM SEPT JOURS DU MONDE	Score: 79	10/01/1964
EMISSSEM LA BEATLEMAMIE	Score: 83	00:10:40
CAF97071869 JT 13h ARRIVEE DES BEATLES	Inconnu	15/01/1964
EMISSSEM ARRIVEE DES BEATLES	Score: 79	Inconnu
CPFB650005 AGE TENDRE ET TETE DE BOIS	Score: 81	00:52:31
EMISSSEM AGE TENDRE ET TETE DE BOIS : EMISSION DU 22 JANVIER 1964	Score: 81	00:52:31
90862 AGE TENDRE ET TETE DE BOIS	Inconnu	22/01/1964
EXYBAY parade des BEATLES par JOHNNY ET H SALVADOR	Score: 79	00:01:51
AFEB001067 LES ACTUALITES FRANCAISES	Score: 83	29/01/1964
EMISSSEM LA DERNIERE FOLIE : LA PERRUQUE BEATLES	Score: 83	00:00:59
CAF97071863 JT 13h BEATLES RETOUR DE LONDRES	Inconnu	06:02/1964
EMISSSEM BEATLES RETOUR DE LONDRES	Score: 77	Inconnu
CAF97071874 JT 13h ARRIVEE DES BEATLES A NEWYORK	Score: 81	11/02/1964
EMISSSEM ARRIVEE DES BEATLES A NEWYORK	Score: 81	Inconnu
CAF98046300 JT 20h LE RETOUR DES BEATLES	Score: 85	23/02/1964
EMISSSEM LE RETOUR DES BEATLES	Score: 85	Inconnu
AFEB001005 LES ACTUALITES FRANCAISES	Score: 81	04/03/1964
EMISSSEM L'ARRIVEE DES BEATLES : CA, C'EST DU SPORT!	Score: 81	00:01:16
CAF97071865 JT 13h LES BEATLES	Score: 79	04/03/1964
EMISSSEM LES BEATLES	Score: 79	00:00:41
CAF98046309 SEPT JOURS DU MONDE	Score: 81	20/02/1964
EMISSSEM SEPT JOURS DU MONDE	Score: 81	00:03:00

# Des supports diversifiés



# Des mutations:

- Du document à la ressource
- De la recherche à la composition
- De l'indexation à l'éditorialisation

# Du document à la ressource

- Ce qu'on appelle document tend à n'être plus que le « contenant » d'origine à partir duquel extraire des fragments;
- Ces fragments sont des ressources, dont le sens viendra des conditions d'exploitation et d'utilisation.
- La cohérence et cohésion documentaire d'origine n'est plus constitutive du contenu de la ressource.

# De la recherche à la composition

- La principale motivation applicative de l'indexation est la recherche d'information:
  - Trouver le document ou contenu qui exprime l'information recherchée.
- La principale motivation applicative des métadonnées est la sélection de ressources pour créer de nouvelles informations:
  - Ce n'est pas la ressource en tant que telle qui aura une valeur, mais le contexte dans lequel elle sera intégrée.

La documentation rendait compte de l'origine dans les termes de l'usage, elle tend à reconfigurer l'origine pour l'usage.

# De l'indexation à l'éditorialisation

- Objectif:
  - Trouver des ressources pour créer de nouveaux contenus
- Contrainte:
  - Le fragment est décontextualisé de son contexte d'origine ;
  - Nécessité de le recontextualiser dans son nouvel environnement:
    - Support spécifique;
    - Harmonisation avec des fragments hétérogènes.
- Créer de l'information et du contenu pour l'intégrer aux ressources sélectionnées

# Plusieurs visions

- Posture « généalogique »:
  - La ressource sélectionnée est enrichie pour être resituée dans son contexte d'origine
  - Travail éditorial qui publie le travail documentaire.
- Posture « amnésique »:
  - La ressource sélectionnée est enrichie dans un nouveau contexte oublieux de l'ancien.
  - Le travail éditorial est une création en coupure avec le travail documentaire effectué sur la ressource.



# Ina.fr

**ina** **RECHERCHER**  
100 000 émissions radio télé  
OK

**ma zone mémoire**  
Retrouvez vos listes de médias et votre parcours en un coup d'oeil  
Mon parcours (3) afficher

**mon espace personnel**  
Consulter mon compte  
panier: 0 | préférés: 1  
Identifiez-vous | Inscrivez-vous

**EXPLORER**  
thèmes  
personnalités  
époques  
programmes

**DECOUVRIR**  
rendez-vous  
parcours surprise  
journal de votre naissance

**PARTAGER**  
forums thématiques  
les préférés de ...

**APPRENDRE**  
après les cours  
aller plus loin  
recevoir la lettre d'info  
entrez votre e-mail OK

**BANDES ANNONCES A VISIONNER**  
le maître des mystères, Cézanne, apprendre sur ina.fr ...

**FICHE MEDIA**  
**Théâtre Jean Louis BARRAULT**  
Ouvert le dimanche  
FR3 - 23/05/1982 - 00h52m43s

Entouré de P. TESSON et de J. palette, J. I barrault parle de sa carrière au cinéma, au théâtre, de ses goûts littéraires, de ses maîtres, de sa philosophie de la vie. Indépendamment du théâtre et de la littérature, barrault est également passionné par la peinture et la musique. Après un reportage intitulé "les folies bourgeoises" consacré à l'architecture des maisons de la côte Normande, Jean Louis BARRAULT avoue son goût pour les vieilles pierres mais aussi pour l'architecture de théâtre : il insiste sur la nécessité de la communication entre les comédiens et la salle.



ina.fr

**REBONDIR**  
vers d'autres médias

- Arts
- Sciences humaines
- Spectacle
- Barrault, Jean Louis
- théâtre
- Tesson, Philippe
- Palette, Jim

Recherche

**EN SAVOIR PLUS**

- Jean-Louis BARRAULT : portraits
- Arts
- Sciences humaines
- Spectacle

envoyer à un ami  
ajouter à mes préférés

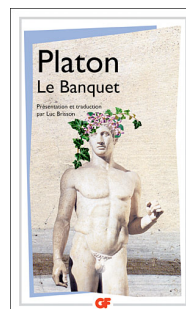
télécharger pour visionner en plein écran  
location 48h 1.50 €  
achat 4.00 €  
ajouter au panier

retour  
média précédent  
média suivant

49



Identité numérique



Un objet matériel unique

Identité matérielle



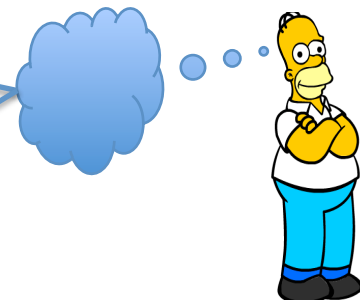
Une mise en forme matérielle particulière

Identité expressive

```
APOLLODORE
[ ]
Je crois que je suis assez bien préparé à vous faire le récit que vous me
demandez : car, tout dernièrement, comme je me rendais de ma maison de
Phalère à la ville, un homme de ma connaissance, qui venait derrière moi,
m'aperçut, et m'appela de loin :
GLAUCON
Homme de Phalère ! s'écria-t-il en badinant, Apollodore ! ne peux-tu
ralement le pas ?
APOLLODORE
Je m'arrêtai, et l'attendis.
GLAUCON
Apollodore, me dit-il, je te cherchais justement ; je voulais te demander
ce qui s'était passé chez Apollon, le jour où [ ] Socrate, Alcibiade et
plusieurs autres y soupèrent. On dit que toute la conversation roula sur
l'amour. J'en ai bien su quelque chose par un homme à qui Phénix, fils de
Philippe, avait raconté une partie de leurs diatribes, mais cet homme ne
put rien me dire de certain sur le détail de cet entretien ; il m'apprit
seulement que tu le savais. Conte-le-moi donc ; aussi bien est-ce un
devoir pour toi de faire connaître ce que dit ton ami ; mais avant tout,
dis-moi, étais-tu présent à cette conversation ?
APOLLODORE
Il paraît bien, lui répondis-je, que ton homme ne t'a rien [ ] dit de
certain, puisque tu parles de cette conversation comme d'une chose arrivée
depuis peu, et comme si j'avais pu y être présent.
```

Invariants propres à la forme d'expression

Identité sémantique

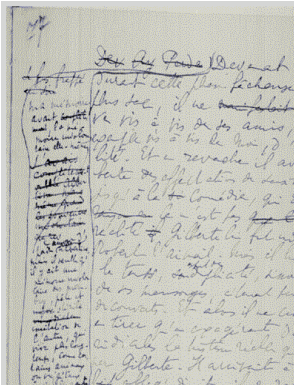


Commentaires et explicitations du contenu

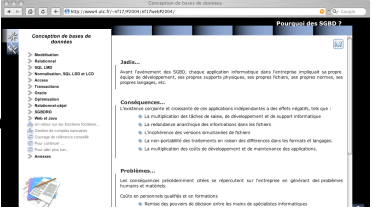
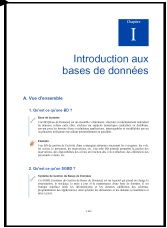
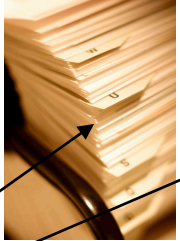
# Entre philologie et herméneutique

*Philologie:  
Etablir des versions de référence*

temps avant l'élaboration. C'est évident, comme l'écrit Ovidien dans le *Metamorphoses*. Il n'y a pas de grand nombre de mots. Mais l'usage de nos livres s'explique par le fait que nous sommes nés pour lire, non pour écrire. C'est ce qui nous rend capables de comprendre ce que nous lisons, et non de ce que nous écrivons. C'est pourquoi, si nous sommes capables de lire, nous sommes capables de comprendre ce que nous lisons, et non de ce que nous écrivons. C'est pourquoi, si nous sommes capables de lire, nous sommes capables de comprendre ce que nous lisons, et non de ce que nous écrivons.



Reziropeizop epozir fsdkf mlsdkfml  
skdfml ksmdkf mlsdkf mksdmf ksmkf  
mlkffrlmakk mlk amk amlk m km lkm  
lkm l kmk mlk mlk mlk mlk mlk mlk  
mkl mlk mlk mlk mlk mlk mlk mlk  
mlk mlk mlk ml mlk mlk mkm kml kmk  
mlk ml l k lmk mlk lmk lmk lmk l m  
kjldkds jkqsjdkqj lksqjd kjqslkdj  
pzeerskdsqmlkmlqskd kmlqskd lq  
Repzoirpoezi zi poezir qkdlmsqk  
mlsqkdlkdsmlq rezr ezrez erzr zer z  
zer zer ezr zer ze ze ze z z z z er ezrz  
ze ze zer zer ze zer ezr zer zer



*Herméneutique : construire des interprétations*

# Conclusion

- Le langage permet la réflexivité
  - Autotranscendance du langage
- L'écriture réifie l'exprimable
- Le numérique rend manipulable l'exprimé
- L'ingénierie documentaire explore les possibilités de la manipulation et du sens documentaire qu'elle peut avoir.





**Puisse Titivillus ne pas trop  
remplir sa besace !**

