



**HAL**  
open science

# Du document physique à la base de données : les chaînes de numérisation, procédés et problématiques.

Morgane Pica

## ► To cite this version:

Morgane Pica. Du document physique à la base de données : les chaînes de numérisation, procédés et problématiques.. Master. Du document physique à la base de données : les chaînes de numérisation, procédés et problématiques., Université de Caen, UFR HSS, laboratoire CRISCO, France. 2019. halshs-04785287

**HAL Id: halshs-04785287**

**<https://shs.hal.science/halshs-04785287v1>**

Submitted on 15 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Du document physique à la base de données :

les chaînes de numérisation,  
procédés et problématiques.



# Humanités numériques

- Discipline au croisement de la recherche en Sciences Humaines et Sociales et des techniques numériques.
- Formation en général hybride : formation théorique dans la recherche et formation pratique en techniques du numérique.

# Humanités numériques

« Les HN peuvent s'appuyer sur une histoire passée mais n'ont pour objet que **l'hyper-présent ou le futur à venir**. Elles servent à comprendre le monde de demain, à le faire advenir dans de meilleures conditions et à l'expliquer aux citoyens, aux dirigeants, etc. Le chercheur en Humanités est ici dans une toute autre situation par rapport à ce qu'il a l'habitude de faire dans sa pratique. Il doit **inventer, comprendre et analyser** en s'appuyant sur une solide culture classique scientifique qui le met en position de mieux **anticiper les changements dans la société**. Tel est, à mon sens, le rôle des HN. Bien loin des outils numériques utilisés pour notre recherche dans des secteurs disciplinaires donnés... »

Source : « Les Humanités Numériques : une nouvelle discipline universitaire ? », Suzanne Dumouchel, in *Digital Humanities à l'Institut historique allemand*, publié le 19/06/2015, consulté le 05/04/2019

# Humanités numériques

Concrètement :

- Édition numérique,
- Bibliothèques et expositions numériques,
- Traitement automatique des textes et images,
- Bases de données appliquées aux objets patrimoniaux,
- Archivage numérique,
- Programmation web pour institutions culturelles...

# Numérisation

- Production de la représentation numérique d'un objet physique.
- « Conversion » de données depuis un support physique vers une forme et un support numériques.
- Adaptation, voire remaniement des données brutes.

# Édition

- 1) « Action de faire paraître un texte et d'en assurer la diffusion auprès du public, directement ou par le truchement d'intermédiaires. »
- 2) « Ensemble des opérations intellectuelles et matérielles par lesquelles le texte d'une œuvre est établi. »

F. Duval  
*Les mots de l'édition de texte*  
Les manuels de l'École des chartes, 2015

# Édition numérique

- « Édition dont les données sont encodées et accessibles numériquement. »

F. Duval

*Les mots de l'édition de texte*

Les manuels de l'École des chartes, 2015

- Un texte sur un support numérique et dans un format numérique, natif ou non.

# Édition numérique

Pour nous :

- Travail scientifique, utilisant les moyens numériques, d'adaptation pour le public d'une œuvre rendue obscure à l'œil moderne par l'état de la langue ou de l'écriture dans lesquelles elle a été initialement conçue et/ou représentée.
- Travail d'enrichissement scientifique des données brutes d'une œuvre par des moyens numériques.

L'utilisation du numérique change la forme mais non le fond de la démarche. Elle change également les possibilités de l'éditeur.

# Édition numérique

Pourquoi une édition numérique ?

- Faciliter le partage des travaux scientifiques.
- Faciliter la recherche et l'analyse automatiques.
- Profiter des nombreuses possibilités structurelles et sémantiques des langages numériques.
- Repenser l'approche d'une œuvre en s'affranchissant des contraintes du format papier.

# Édition numérique

Pourquoi une édition numérique ?

- Faciliter le partage des travaux scientifiques.
- Faciliter la recherche et l'analyse automatiques.
- Profiter des nombreuses possibilités structurelles et sémantiques des langages numériques.
- Repenser l'approche d'une œuvre en s'affranchissant des contraintes du format papier.

# Chaîne de numérisation

L'ensemble des procédés, physiques et numériques, permettant de passer d'un support papier originel à sa représentation numérique.

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

# Les étapes de la numérisation

- 1) **Travail initial.**
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

# 1) Travail initial

- Travail codicologique et bibliographique.
- Définition des buts et cibles.
- Évaluation du travail à venir.

# Les étapes de la numérisation

- 1) Travail initial.
- 2) **Numérisation.**
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

## 2) Numérisation

Acquisition d'images de la source.

En fonction de :

- Accès à la source.
- État de la source.
- Taille de la source.
- Matériel.
- Buts scientifiques et techniques.

## 2) Numérisation

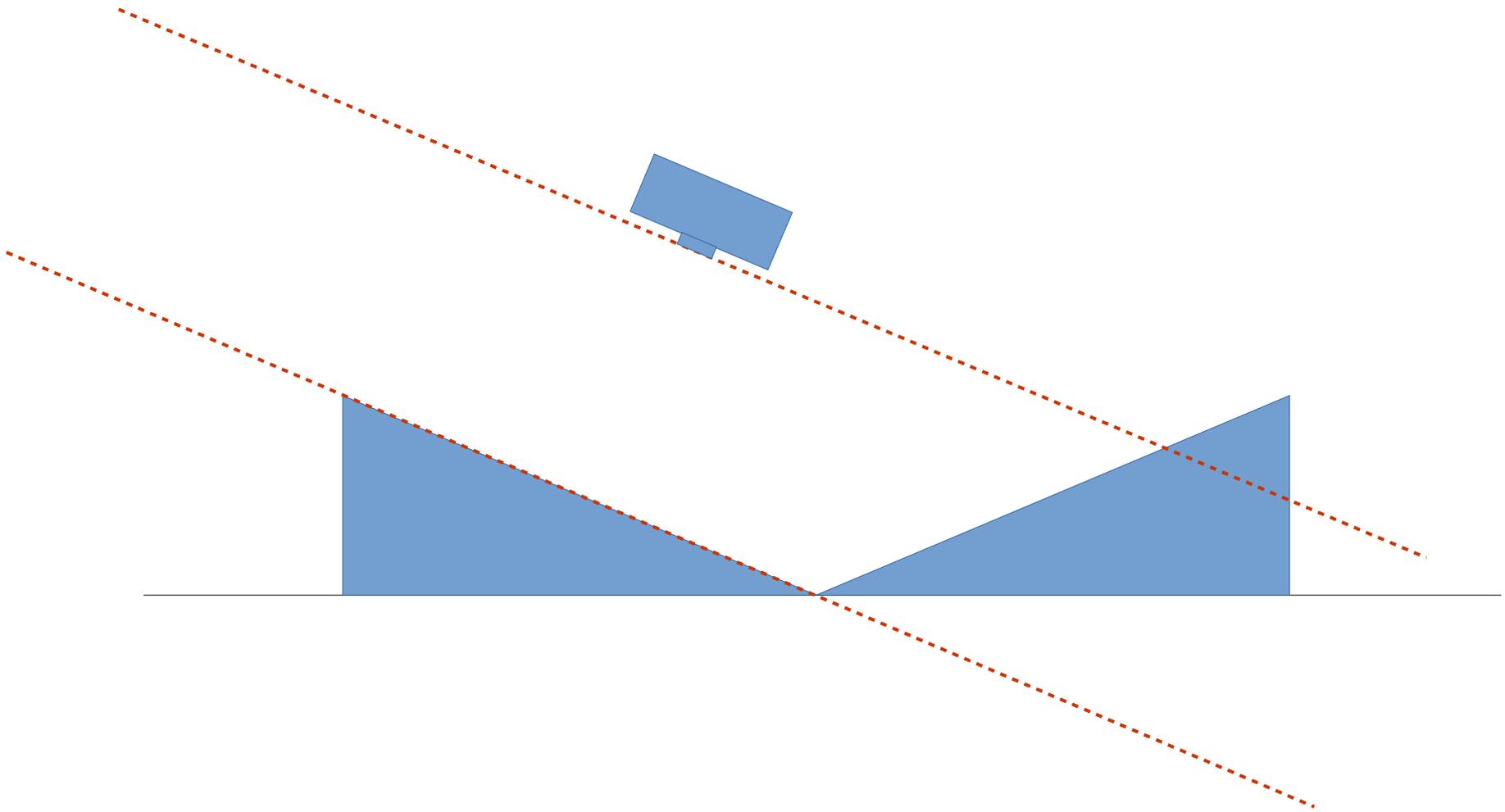


## 2) Numérisation

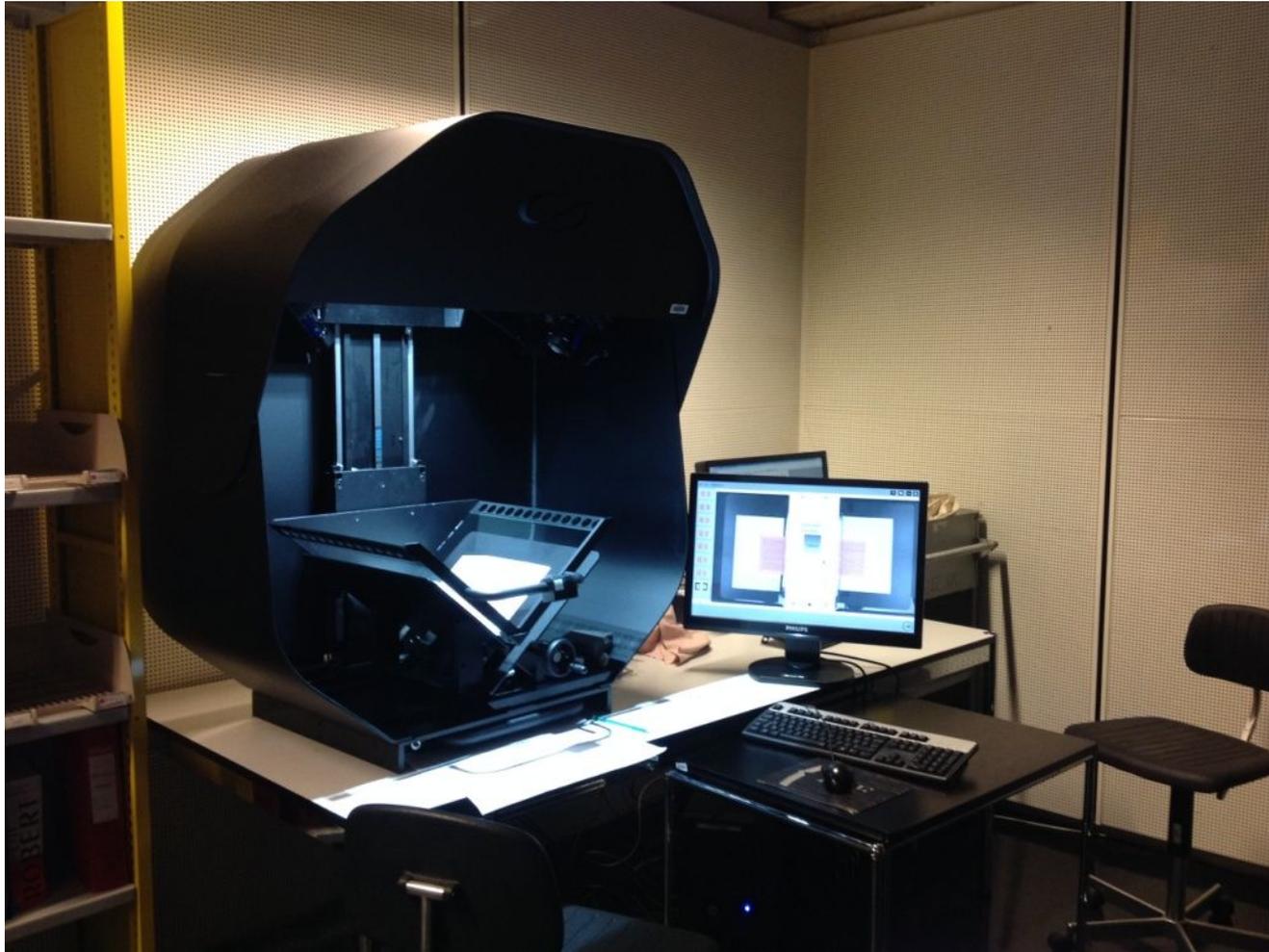


Source : Au delà des murs Créations - <https://www.audeladesmurscreations.com/product-page/ref-113h-housse-pour-113-113p>

## 2) Numérisation



## 2) Numérisation



Source : « Dans les coulisses de Gallica », <http://peccadille.net/2015/03/03/dans-les-coulisses-de-gallica/>

# 2) Numérisation

aucun rufus que vide theobont ut  
pouffis p' l'eu leur aage  
et l'arable de levitaci  
et p' l'arable de levitaci  
de marfon l'ou ou t'edou  
les esch' r' g' d' p' l'ou m' f' e  
t' f' r' e' f' r' a' c' h' r' e' t' ou p' l' e' m' f' e  
h' o' u' d' u' f' i' e' n' g' l' e' p' e' a' u' m' a' l' i' e' t'  
g' r' a' f' i' m' e' d' i' t' a' m' u' d' h' e' l' d' i' c' t'  
et r' e' f' t' o' r' e' t' o' u' r' l' e' p' t' e' m' e'  
ou d' l' a' y' d' e' f' t' r' e' p' r' i' m' e'  
p' o' u' r' e' t' q' u' e' l' a' l' a' f' y' p' p' u' e'  
p' m' a' l' a' y' a' u' d' e' f' t' o' u' d' e' n' e'

**S**ignificat h'acis domer and  
des d'ies et d'ies p' anons  
q' p' d' d' u' l' e' p' t' r' u' m' p' h' o' m' a' g' e'  
q' u' e' l' e' t' h' e' l' e' p' e' t' d' e' u' t' a' g' r'  
h' o' m' e' s' p' e' n' e' u' t' p' d' e' u' m' f' i'  
d' e' l' e' t' d' o' r' e' e' t' t' r' e' t' g' u' i' s' t'  
ou p' e' t' q' u' e' d' h' e' l' i' a' f' r' o' n'  
t' u' t' i' o' n' e' e' l' e' p' o' s' s' e' s' s' i' o' n'  
t' r' o' i' e' n' e' t' o' u' r' d' e' l' a' y' s' t' o' u'  
et l' e' u' t' a' g' r' a' l' l' e' t' h' e' l' i' a' f' r' o' n'  
r' u' g' t' e' a' u' r' h' e' l' i' a' f' r' o' n' e' t' d' i' t'  
et d' e' h' o' r' n' o' m' u' e' l' h' o' m' e' h' o' r' n'  
ou p' h' o' u' e' y' a' u' d' m' a' n' s' h' u' d' u'  
p' a' n' s' h' e' l' i' a' f' r' o' n' e' t' d' e' u' d' o' u'  
ou p' e' l' e' g' t' e' t' r' e' l' y' a' g' e'  
d' e' u' r' h' e' l' i' a' f' r' o' n' e' s' u' r' z' h' o' m' a' g' e'  
et p' e' t' p' o' p' u' l' e' p' e' m' e' p' l' l' e'  
h' e' l' i' a' f' r' o' n' e' t' h' o' m' a' g' e' o' u' p' l' l' e'  
q' u' e' h' e' l' i' a' f' r' o' n' e' n' u' l' p' e' p' o' p' u' l' e' t' r' a' n' s'  
p' a' n' s' h' o' m' a' g' e' p' n' o' n' d' u' g' t' a' r' e'  
m' o' d' d' e' f' i' c' i' p' o' d' m' p' r' o' u' e' n' d' u'  
q' u' e' h' e' l' i' a' f' r' o' n' e' n' o' d' e' m' u' r' p' u' n'  
p' e' t' q' u' e' l' e' f' r' a' n' c' h' i' s' e' p' e'  
et m' i' l' d' i' g' n' i' t' e' d' a' u' d' g' u' i' s' t'  
et t' o' u' r' p' o' u' q' d' d' o' u' r' e' t' a' u' r' e' u'  
h' o' m' a' g' e' a' u' d' h' e' l' i' a' f' r' o' n' e' p' a' u' e' d'  
t' a' u' s' e' t' a' u' r' e' f' r' o' n' e' p' o' p' u' l' e' d' u' b' i' t'  
h' e' l' i' a' f' r' o' n' e' t' d' e' u' d' o' u'  
t' o' u' r' h' e' l' i' a' f' r' o' n' e' o' u' d' i' g' n' i' t'  
p' u' n' e' t' f' i' c' i' p' e' d' e' g' a' u' d' e' p' u' n' e'  
q' u' e' d' e' u' r' p' a' u' e' p' o' l' i' t' e'  
h' e' l' i' a' f' r' o' n' e' p' e' t' p' o' u' r' a' u' d'

les d'ies qui sont cultiver  
pouffis p' l'eu leur aage  
et l'arable de levitaci  
que n'y p' d' u' m' e' p' r' a' c' t' a' u' s'  
et r' e' t' n' y' s' e' f' t' c' o' m' p' t' e'  
l' a' y' m' e' a' u' r' e' z' a' c' q' u' i' e' r' e'  
ou t' o' u' r' l' e' t' o' m' e' n' d' p' l' a' c'  
p' l' a' d' e' m' a' g' t' o' u' r' f' i' m' e'  
f' a' c' h' r' e' s' q' u' e' n' d' u' s' p' o' t' l'  
d' e' n' o' r' m' e' d' e' a' e' s' t' a' b' l' i' e' s'  
d' e' h' e' l' i' a' f' r' o' n' e' e' t' n' y' s' t' e'  
d' e' d' u' r' e' q' t' n' e' p' o' d' s' u' b' i' m' f' e'  
q' u' e' l' t' i' m' e' n' t' a' n' e' m' e' s'  
p' d' o' m' e' t' p' e' l' o' n' l' e' s' d' o' m' m' e' s'  
t' o' u' s' t' i' m' e' n' t' d' e' l' o' u' g' t' i' p' e' s' e' l' l' e'  
t' h' e' l' i' a' f' r' o' n' e' h' e' l' i' a' f' r' o' n' e'  
c' o' m' e' m' o' u' l' i' n' s' p' u' n' e' c' u' s' t' o' m'  
q' u' e' p' f' y' p' o' u' r' s' e' u' l' t' r' o' u' n' d'  
t' r' o' u' l' i' n' s' a' l' e' a' y' q' u' o' u' l' i' e' r' e' o' f'  
q' u' e' p' o' u' b' z' m' i' t' e' t' o' m' e' s' o' u' r'  
p' o' p' a' n' t' e' p' u' l' e' p' l' a' c' p' u' l' l' o' n'  
d' e' h' e' l' i' a' f' r' o' n' e' m' o' u' l' i' n' s' e' f' t' o' u' r'  
o' u' e' t' f' r' o' n' d' e' h' a' u' b' e' t' t' r' a' u' s'  
ou h' e' l' i' a' f' r' o' n' e' a' p' p' t' r' a' u' s'  
ou a' n' t' i' q' u' e' s' t' r' o' u' n' t' u' r' e'  
ou a' n' t' i' q' u' e' s' t' r' a' n' s' p' o' n' a' n' t'  
ou a' a' n' d' p' a' u' e' t' r' o' u' n' t' u' r'  
l' e' s' m' o' u' l' i' n' s' o' u' h' e' l' i' a' f' r' o' n'  
d' u' f' i' e' n' a' c' q' u' i' s' t' e' t' d' o' m' e' t'  
m' a' i' s' l' e' s' a' u' e' s' e' t' o' q' h' o' m' e' t'  
l' e' s' c' o' n' s' t' i' m' e' n' t' m' e' s' t' e' l' l' e'  
p' i' e' t' h' o' u' s' z' l' a' u' d' e' s' f' i' e' r' s'  
q' u' e' r' n' n' u' l' t' i' p' e' e' t' m' u' l' t' i' g' u' i' s' t'  
d' o' u' l' t' i' m' e' n' t' n' e' p' o' d' s' u' b' i' m' f' e'  
M' a' i' s' p' l' u' s' s' i' b' u' r' a' c' o' u' s' t' i' m' e' n' t'  
d' e' d' o' r' e' p' l' l' e' a' u' e' u' r' h' e' l' i' a' f' r' o' n'  
p' a' u' h' e' l' i' a' f' r' o' n' e' d' a' u' r' e' t' r' o' u' n' t' u' r'  
e' t' p' l' u' s' s' i' p' a' y' p' d' i' s' t' i' n' c' t'  
t' o' u' r' s' p' r' i' m' a' g' t' m' o' t' o' r' s' d' i' t' l'  
p' a' u' r' d' e' l' a' r' e' p' u' r' q' u' e' l' l'  
d' e' h' e' l' i' a' f' r' o' n' e' n' e' d' u' s' t' i' n' t'  
e' t' p' l' u' s' s' i' l' a' n' g' d' e' p' r' o' u' e' n' d' u'  
d' e' l' a' m' o' r' a' c' i' l' p' a' u' e' s' p' a' n'  
q' u' e' d' e' u' r' m' p' l' l' e' h' o' m' a' g' e'  
h' e' l' i' a' f' r' o' n' e' e' t' h' e' l' i' a' f' r' o' n' e' d' u'  
q' u' e' l' e' u' t' a' g' r' a' h' o' r' n'

aucun rufus que vide theobont ut  
pouffis p' l'eu leur aage  
et l'arable de levitaci  
et p' l'arable de levitaci  
de marfon l'ou ou t'edou  
les esch' r' g' d' p' l'ou m' f' e  
t' f' r' e' f' r' a' c' h' r' e' t' ou p' l' e' m' f' e  
h' o' u' d' u' f' i' e' n' g' l' e' p' e' a' u' m' a' l' i' e' t'  
g' r' a' f' i' m' e' d' i' t' a' m' u' d' h' e' l' d' i' c' t'  
et r' e' f' t' o' r' e' t' o' u' r' l' e' p' t' e' m' e'  
ou d' l' a' y' d' e' f' t' r' e' p' r' i' m' e'  
p' o' u' r' e' t' q' u' e' l' a' l' a' f' y' p' p' u' e'  
p' m' a' l' a' y' a' u' d' e' f' t' o' u' d' e' n' e'

**S**ignificat h'acis domer and  
des d'ies et d'ies p' anons  
q' p' d' d' u' l' e' p' t' r' u' m' p' h' o' m' a' g' e'  
q' u' e' l' e' t' h' e' l' e' p' e' t' d' e' u' t' a' g' r'  
h' o' m' e' s' p' e' n' e' u' t' p' d' e' u' m' f' i'  
d' e' l' e' t' d' o' r' e' e' t' t' r' e' t' g' u' i' s' t'  
ou p' e' t' q' u' e' d' h' e' l' i' a' f' r' o' n'  
t' u' t' i' o' n' e' e' l' e' p' o' s' s' e' s' s' i' o' n'  
t' r' o' i' e' n' e' t' o' u' r' d' e' l' a' y' s' t' o' u'  
et l' e' u' t' a' g' r' a' l' l' e' t' h' e' l' i' a' f' r' o' n'  
r' u' g' t' e' a' u' r' h' e' l' i' a' f' r' o' n' e' t' d' i' t'  
et d' e' h' o' r' n' o' m' u' e' l' h' o' m' e' h' o' r' n'  
ou p' h' o' u' e' y' a' u' d' m' a' n' s' h' u' d' u'  
p' a' n' s' h' e' l' i' a' f' r' o' n' e' t' d' e' u' d' o' u'  
ou p' e' l' e' g' t' e' t' r' e' l' y' a' g' e'  
d' e' u' r' h' e' l' i' a' f' r' o' n' e' s' u' r' z' h' o' m' a' g' e'  
et p' e' t' p' o' p' u' l' e' p' e' m' e' p' l' l' e'  
h' e' l' i' a' f' r' o' n' e' t' h' o' m' a' g' e' o' u' p' l' l' e'  
q' u' e' h' e' l' i' a' f' r' o' n' e' n' u' l' p' e' p' o' p' u' l' e' t' r' a' n' s'  
p' a' n' s' h' o' m' a' g' e' p' n' o' n' d' u' g' t' a' r' e'  
m' o' d' d' e' f' i' c' i' p' o' d' m' p' r' o' u' e' n' d' u'  
q' u' e' h' e' l' i' a' f' r' o' n' e' n' o' d' e' m' u' r' p' u' n'  
p' e' t' q' u' e' l' e' f' r' a' n' c' h' i' s' e' p' e'  
et m' i' l' d' i' g' n' i' t' e' d' a' u' d' g' u' i' s' t'  
et t' o' u' r' p' o' u' q' d' d' o' u' r' e' t' a' u' r' e' u'  
h' o' m' a' g' e' a' u' d' h' e' l' i' a' f' r' o' n' e' p' a' u' e' d'  
t' a' u' s' e' t' a' u' r' e' f' r' o' n' e' p' o' p' u' l' e' d' u' b' i' t'  
h' e' l' i' a' f' r' o' n' e' t' d' e' u' d' o' u'  
t' o' u' r' h' e' l' i' a' f' r' o' n' e' o' u' d' i' g' n' i' t'  
p' u' n' e' t' f' i' c' i' p' e' d' e' g' a' u' d' e' p' u' n' e'  
q' u' e' d' e' u' r' p' a' u' e' p' o' l' i' t' e'  
h' e' l' i' a' f' r' o' n' e' p' e' t' p' o' u' r' a' u' d'

les d'ies qui sont cultiver  
pouffis p' l'eu leur aage  
et l'arable de levitaci  
que n'y p' d' u' m' e' p' r' a' c' t' a' u' s'  
et r' e' t' n' y' s' e' f' t' c' o' m' p' t' e'  
l' a' y' m' e' a' u' r' e' z' a' c' q' u' i' e' r' e'  
ou t' o' u' r' l' e' t' o' m' e' n' d' p' l' a' c'  
p' l' a' d' e' m' a' g' t' o' u' r' f' i' m' e'  
f' a' c' h' r' e' s' q' u' e' n' d' u' s' p' o' t' l'  
d' e' n' o' r' m' e' d' e' a' e' s' t' a' b' l' i' e' s'  
d' e' h' e' l' i' a' f' r' o' n' e' e' t' n' y' s' t' e'  
d' e' d' u' r' e' q' t' n' e' p' o' d' s' u' b' i' m' f' e'  
q' u' e' l' t' i' m' e' n' t' a' n' e' m' e' s'  
p' d' o' m' e' t' p' e' l' o' n' l' e' s' d' o' m' m' e' s'  
t' o' u' s' t' i' m' e' n' t' d' e' l' o' u' g' t' i' p' e' s' e' l' l' e'  
t' h' e' l' i' a' f' r' o' n' e' h' e' l' i' a' f' r' o' n' e'  
c' o' m' e' m' o' u' l' i' n' s' p' u' n' e' c' u' s' t' o' m'  
q' u' e' p' f' y' p' o' u' r' s' e' u' l' t' r' o' u' n' d'  
t' r' o' u' l' i' n' s' a' l' e' a' y' q' u' o' u' l' i' e' r' e' o' f'  
q' u' e' p' o' u' b' z' m' i' t' e' t' o' m' e' s' o' u' r'  
p' o' p' a' n' t' e' p' u' l' e' p' l' a' c' p' u' l' l' o' n'  
d' e' h' e' l' i' a' f' r' o' n' e' m' o' u' l' i' n' s' e' f' t' o' u' r'  
o' u' e' t' f' r' o' n' d' e' h' a' u' b' e' t' t' r' a' u' s'  
ou h' e' l' i' a' f' r' o' n' e' a' p' p' t' r' a' u' s'  
ou a' n' t' i' q' u' e' s' t' r' o' u' n' t' u' r' e'  
ou a' n' t' i' q' u' e' s' t' r' a' n' s' p' o' n' a' n' t'  
ou a' a' n' d' p' a' u' e' t' r' o' u' n' t' u' r'  
l' e' s' m' o' u' l' i' n' s' o' u' h' e' l' i' a' f' r' o' n'  
d' u' f' i' e' n' a' c' q' u' i' s' t' e' t' d' o' m' e' t'  
m' a' i' s' l' e' s' a' u' e' s' e' t' o' q' h' o' m' e' t'  
l' e' s' c' o' n' s' t' i' m' e' n' t' m' e' s' t' e' l' l' e'  
p' i' e' t' h' o' u' s' z' l' a' u' d' e' s' f' i' e' r' s'  
q' u' e' r' n' n' u' l' t' i' p' e' e' t' m' u' l' t' i' g' u' i' s' t'  
d' o' u' l' t' i' m' e' n' t' n' e' p' o' d' s' u' b' i' m' f' e'  
M' a' i' s' p' l' u' s' s' i' b' u' r' a' c' o' u' s' t' i' m' e' n' t'  
d' e' d' o' r' e' p' l' l' e' a' u' e' u' r' h' e' l' i' a' f' r' o' n'  
p' a' u' h' e' l' i' a' f' r' o' n' e' d' a' u' r' e' t' r' o' u' n' t' u' r'  
e' t' p' l' u' s' s' i' p' a' y' p' d' i' s' t' i' n' c' t'  
t' o' u' r' s' p' r' i' m' a' g' t' m' o' t' o' r' s' d' i' t' l'  
p' a' u' r' d' e' l' a' r' e' p' u' r' q' u' e' l' l'  
d' e' h' e' l' i' a' f' r' o' n' e' n' e' d' u' s' t' i' n' t'  
e' t' p' l' u' s' s' i' l' a' n' g' d' e' p' r' o' u' e' n' d' u'  
d' e' l' a' m' o' r' a' c' i' l' p' a' u' e' s' p' a' n'  
q' u' e' d' e' u' r' m' p' l' l' e' h' o' m' a' g' e'  
h' e' l' i' a' f' r' o' n' e' e' t' h' e' l' i' a' f' r' o' n' e' d' u'  
q' u' e' l' e' u' t' a' g' r' a' h' o' r' n'

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) **Acquisition du texte.**
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

# 3) Acquisition du texte

- Acquisition manuelle (lecture et saisie humaines).
- Automatisation :
  - OCR (Optical Character Recognition) : caractères majoritairement prédéfinis, conçu pour les textes contemporains imprimés.
  - HTR (Handwritten Text Recognition) : apprentissage progressif du programme permettant de prendre en compte les variantes d'une même lettre dans une écriture manuscrite.
    - Sortie en texte brut.

# 3) Acquisition du texte

## Junicode

Projet de créer des caractères Unicode pour retranscrire les sources antiques et médiévales dans une police gratuite et standard.

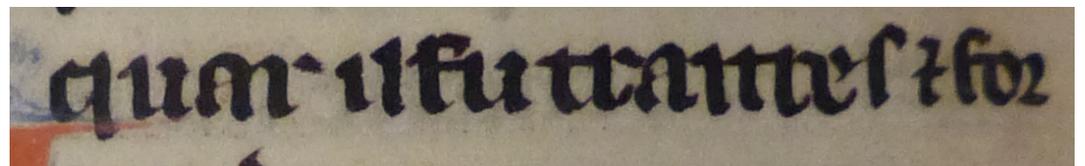
Œ œ þ ȝ Ɔ ʝ ʃ ft ā t̄ ð ʀ

Développée par Peter Baker à l'université de Virginie dans le cadre du projet MUF1.

### 3) Acquisition du texte

#### Entraînement d'un modèle d'HTR

- Transcription initiale manuelle d'un extrait représentatif.
- Synchronisation texte-image pour le modèle (caractère par caractère).



quar il fu traictes ] fo2

quar il fu traictes ] fo2

# 3) Acquisition du texte

## Entraînement d'un modèle d'HTR

- Transcription initiale manuelle d'un extrait représentatif.
- Synchronisation texte-image pour le modèle.
- Apprentissage initial suivant la synchronisation.
- Test initial sur de nouvelles pages.
- Correction du test.
- Reprise de l'apprentissage, et encore, et encore... jusqu'à satisfaction.

# 3) Acquisition du texte

## 4 DE JURISDICTION. ARTICLE PREMIER.

Le Bailli, ou son Lieutenant, connoît de tous crimes en premiere instance.

Bailli signifie la même chose que Gardien ; comme Baillie signifie Garde & Protection. D'où vient que Baillifre, dans les anciennes Ordonnances & quelques Coutumes, signifie Tuteur ou Curateur. Le Bailli donc étoit comme le conservateur du Peuple & des Loix : mais ayant été remarqué, que les Baillis s'attachoient plus volontiers aux exercices des Armes qu'aux fonctions de Judicature, le Roi Louis XII ordonna que leurs Lieutenans feroient licenciés en l'Etude du Droit Romain, & leur attribua toute la Jurisdiction ; de forte qu'il ne reste aux Baillis pour marque de leur ancien pouvoir, que la préférence, la voix honoraire non délibérative, & leur nom mis dans le titre des Sentences (1) : les Ordonnances d'Orléans, de Moulins & de Blois ayant disposé, que les Baillis fussent des Officiers Militaires & de Robe-courte. Quand donc cet Article attribue au Bailli la connoissance de tous crimes en premiere instance, cela s'entend de ses Lieutenans : mais cette compétence est limitée par plusieurs exceptions ; car il y a des Crimes prévôtaux & préfidiaux, dont la connoissance en premiere instance appartient aux Prévôts des Maréchaux de France, aux Vice-Baillis & aux Préfidiaux. Voyez l'Ordonnance criminelle, au Titre de la compétence des Juges, aux Articles XI, XII & XV. Le Vicomte de l'Eau, qui a la police sur la Riviere de Seine jusqu'à

Un Arrêt du Conseil du 10 Décembre 1744, fait Règlement entre les Avocats & Procureur du Roi des Juridictions ordinaires, Règlement utile pour affermir la paix entre ceux qui, par état, doivent veiller à la tranquillité publique : mais on s'écrite journellement de nouvelles difficultés sur son exécution. Voyez un Règlement de la Cour de 6 Juillet 1763, & le Règlement de Justice du 18 Juin 1769.

La Stance & les fonctions des Officiers d'un même Siège dépendent ordinairement des Concordats & de la possession qui n'est point abusive.

(1) Le Bailli étoit autrefois un Officier très-important, il réunissoit en lui les fonctions militaires & civiles ; il étoit établi pour veiller, dit l'ancien Coutumier, au maintien des Etats de nos Ducs, conserver les intérêts de la Patrie, & faire régner au-dedans la paix & la tranquillité, en réprimant les attentats, soit par la force, ou par la févérité des loix. Nous avons conservé la formule de son serment, qui nous retrace, avec le tableau de ses devoirs, les prérogatives de sa dignité. Comme le commandement des armes ne lui permettoit pas toujours de s'occuper de la distribution de la justice, il se faisoit représenter par des Lieutenans qu'il créoit & destituoit à son gré, il les multiplioit suivant la distance des lieux & l'étendue de son Bailliage : ces Officiers n'auroient pu sans doute l'empêcher de juger ; aussi dès que le Bailli cessoit d'être Guerrier, il redevenoit Magistrat, & il jugeoit lui-même ; rarement ses jugemens étoient attaqués, on s'en plaignoit quelquefois. Les grands pouvoirs font assez près de l'abus, la carrière séduisante des armes porta peu à peu le Bailli à négliger les fonctions de la justice ; dès qu'il commença de les mépriser, il en fit un honneux trafic ; il falut, pour rétablir l'ordre, multiplier les Ordonnances. Dans les derniers temps, on ne lui a enfin conservé, dans les tribunaux, que des honneurs, & les révolutions ont amené les troupes réglées. Ainsi nos Loix anciennes nous ont appris ce que le Bailli a été, nous sçavons par nos usages ce qu'il est.

## 4 DE JURISDICTION. ARTICLE PREMIER.

Le Bailli, ou son Lieutenant, connoît de tous crimes en premiere instance.

Bailli signifie la même chose que Gardien ; comme Baillie signifie Garde & Protection. D'où vient que Baillifre, dans les anciennes Ordonnances & quelques Coutumes, signifie Tuteur ou Curateur. Le Bailli donc étoit comme le conservateur du Peuple & des Loix : mais ayant été remarqué, que les Baillis s'attachoient plus volontiers aux exercices des Armes qu'aux fonctions de Judicature, le Roi Louis XII ordonna que leurs Lieutenans feroient licenciés en l'Etude du Droit Romain, & leur attribua toute la Jurisdiction ; de forte qu'il ne reste aux Baillis pour marque de leur ancien pouvoir, que la préférence, la voix honoraire non délibérative, & leur nom mis dans le titre des Sentences (1) : les Ordonnances d'Orléans, de Moulins & de Blois ayant disposé, que les Baillis fussent des Officiers Militaires & de Robe-courte.

Quand donc cet Article attribue au Bailli la connoissance de tous crimes en premiere instance, cela s'entend de ses Lieutenans : mais cette compétence est limitée par plusieurs exceptions ; car il y a des Crimes prévôtaux & préfidiaux, dont la connoissance en premiere instance appartient aux Prévôts des Maréchaux de France, aux Vice-Baillis & aux Préfidiaux. Voyez l'Ordonnance criminelle, au Titre de la compétence des Juges, aux Articles XI, XII & XV. Le Vicomte de l'Eau, qui a la police sur la Riviere de Seine jusqu'à Un Arrêt du Conseil du 10 Décembre 1744, fait Règlement entre les Avocats & nous ont appris ce que le Bailli a été, nous sçavons par nos usages ce qu'il est.

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) **Conception de la base de données.**
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

## 4) Conception de la base de données

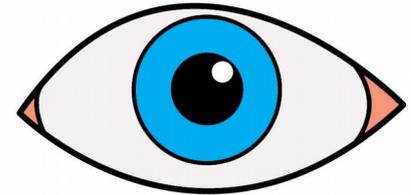
- Base de données textuelle, très souvent en XML-TEI (grandes et vastes possibilités et utilisation très large garante d'interopérabilité et pérennité).
- À nouveau, quelle finalité ?
- Quelle serait la structure logique du *corpus* ?
- Quel temps de travail a-t-on ?
- Quels outils sont prévus ?

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) **Encodage du texte.**
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) Maintenance.

# 5) Encodage du texte - XML

## Texte



Un **texte** est une série orale ou écrite de mots perçus comme constituant un ensemble cohérent, porteur de sens et utilisant les structures propres à une [langue](#) (conjugaisons, construction et association des phrases...)1. Un texte n'a pas de longueur déterminée sauf dans le cas de [poèmes](#) à forme fixe comme le [sonnet](#) ou le [haïku](#).

L'étude formelle des textes s'appuie sur la [linguistique](#), qui est l'approche scientifique du langage.

- Police plus grande
- Placé au-dessus
- Séparé du corps



= titre

# 5) Encodage du texte - XML



Texte

Chaîne de caractères,

- Police : « Liberation Sans »
- Taille : 22 px.
- Couleur : noire.
- Alignement : gauche.

# 5) Encodage du texte - XML



Chaîne de caractères

- “Texte”

Enrichissement structurel

- = titre

Enrichissement formel

- police
- couleur
- position...

# 5) Encodage du texte - XML



Enrichissement structurel

➤ = titre

XML

## 5) Encodage du texte - XML

- = *eXtensible Markup Language*
- Langage d'encodage ( $\neq$  programmation)
- Grammaire : langage à balises.

`<titre>Texte</titre>`

`<titre type="principal">Texte</titre>`

- Finesse de l'encodage jusqu'au caractère.
- « eXtensible » = mélange de vocabulaires XML

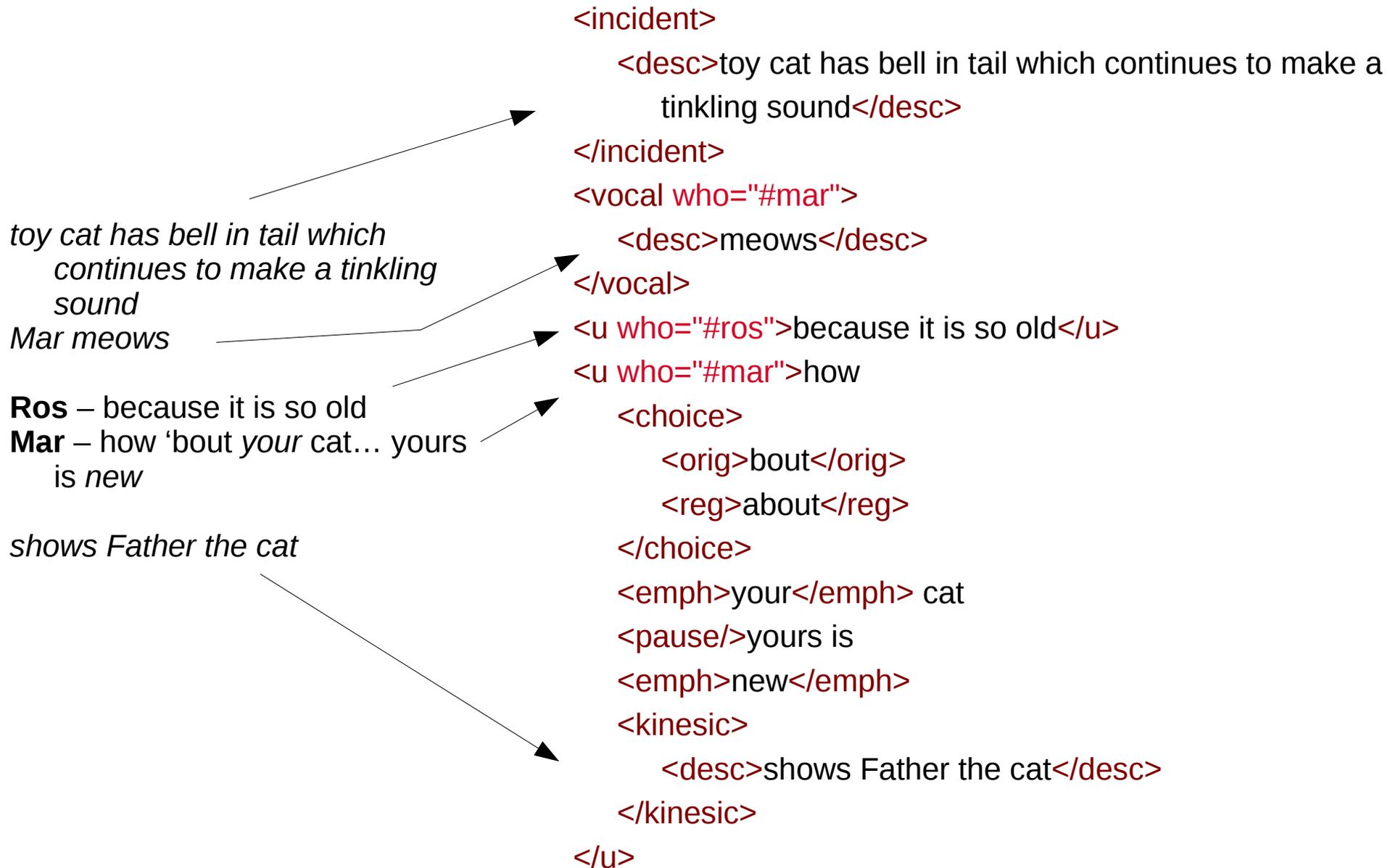
## 5) Encodage du texte - TEI

- = *Text Encoding Initiative*

<https://tei-c.org>

- Nom de domaine = vocabulaire spécifique  
= set de balises prédéfinies
- Encodage sémantique et structurel des textes oraux et écrits, quel que soit le support originel.
- Description fine et précise des supports originels et métadonnées.

# 5) Encodage du texte - TEI



## 5) Encodage du texte - TEI

Document daté un 10 avril (par le scribe)  
entre 851 et 867 (par l'éditeur).

<date>

<date when="--04-10">

**un 10 avril,**

</date>

<date notBefore="0851" notAfter="0867" cert="high">

**entre 851 et 867**

</date>

</date>

# 5) Encodage du texte - TEI

Sujet

Coutume

Commentaire français

Commentaire latin

De assise,

Angel. in tract. malefici, in verbo / dicitur / Citois se defendendo. 3af. in l. vt vim. col. ij. ff. de insti. ture. Alexan. consil. cix. incepti. vltis et vt potter consideratis. in pumo volumi. Nec debet insultatus expectare primam percussione. sed debet incipere vt delectant. Bal. Bar. et omnes. in l. c. vnde vi. leg. etiam 3af. in d. l. vt vim. Et ibidem q non est necesse quod imminet periculum mortis: sed factis est probare quod imminet periculum alicuius iniurie personalis / secundum Bar. in d. l. i. col. c. vnde vi. Sed an sit licitum pro defensione amici conlinguunt vel alterius occidere aggressorem? vt in glo. et per Bar. Bal. et 3af. ibidem in d. l. vt vim. Bal. in repetitio. d. l. ff. c. vnde vi. 30an. de anan. c. ij. de homicid. Et concludit q pro defensione confanguinito- rum sit licitum id quod pro suspensus: quia vt dicit 3afon in d. l. vt vim. quando fit mentio de se vel sua persona. talia verba non tantum comprehendunt propriam personam / sed etiam alios quos circa se necessarios habet per leg. ij. ff. sed quod bigamus. sibe peni. lega. et ibi textus notabilis. idem de comite in via dicit Bar. et Bal. in leg. vt vim. i. quarta questio. quia vt ibidem d. l. alias effertur veredicia. ideo potest contumeliam suam populare. Item apud laborem / tenetur. ff. de iniurijs. Co- rreus in consil. c. i. in p. pro defensione. in col. in pumo volu- imidem dicit Bar. in defensione extranei. in d. leg. vt vim. et Col. in dicto consil. Et ibidem q licitum est alium offendere pro defensione rerum. iuno licitum est occidere: vt tenet glo. in d. l. vt vim. Ratio secundum 3afon. ibidem est quia res sunt secundum sanguis hominis. glo. iuncto tex. in l. aduocat. C. de aduoca. Di- uer. iud. et ibi Bal. Item Bal. in l. i. no nictus. C. de ope. libert. dicit vt per eundem 3afon. in d. l. vt vim. equius vt per sfraciscum aretinum in consil. lxx. incipiens. sicut decet. penult. col. vbi dicit q licet repercuere et defendere res manus non emte auz glo- dio in cap. dilecto. circa p. i. p. q. de sententia excomm. lib. vi. nec pro sutione arma capere licet nec alium occidere nisi inta- dano armis vtatur / secundum Bar. et alios in d. l. i. c. vnde vi. 3ino. 30an. andr. et 3no. in cap. olim caulam. de restitu. spo- ra. cluere 3af. in d. l. vt vim. iij. col. post Cy. in d. l. i. c. vnde vi. et Angel. in q. ius emem gnam in hunc de iure na. q. et et in an. propter solas minas licitum sit alium occidere / et ibidem dicit q concludit etiam post Bal. in l. i. in fin. C. quando lice. sine iud. se vindic. Quod si aliquis immetor mortem et est in actu / ruc est licitum eum occidere / alias non: quia tunc potesta est in acta que idem operatus fuit in habita secundum 3p. voluit de mar. in l. i. in p. i. in q. na. col. ff. de feca. idem in l. penult. iij. col. ff. de quest. Bal. in l. multis. in fi. C. de lib. cau. Et circa istam mate- riam defensionis. est etiam notandum q licet debeat defendens vti moderamine vt in d. l. i. et d. l. vt vim. Si tamen excessu in mode- rati pena arbitrio iudicis / secundum eisdem doct. Paul. de cast. consil. c. xj. incipit in causa incertorum. col. ff. vbi dicit q licet ag- gressor: postea fugiens et insequutus ab aggresso fuit occisus / q videretur potius ad vindictam fuisse q ad tutelam. s. amen no de- bet morte condamnari propter dolorem qui eum mouit ad in- quendum et occidendum. l. strachus. C. de adul. idem dicit in quolibet alio actu defensionis secundum eisdem doct. in d. l. vt vim. vbi inter alia dicit Bal. q ictus non possunt dari cum men- sura: vt defendens habeat lateram et ponderet ictus / refert et se- quitur C. eps. in d. consil. c. x. incipiens. v. g. celli. penult. col. de leg. ca. ff. de p. motu. Et quia homo intus volens aut timens in- motus non sit in plenitudine intellectus / fm Bal. in cap. i. §. i. in fi. de pace iureu. fir. et quod di. id Bal. in d. l. vt vim. iij. col. Quod insultatus non potest bene deliberare et q sepe excidit oc- casio rationem. l. quequid calote. ff. de resu. iur. in cap. l. quis in l. i. quis non dicam. C. de episcop. et cleric. punitur ergo occi- dens de excessu / non de homicidio / et mitior pena q si non esset data occasio. ad hoc est tex. in l. si adulterum / §. si imperatores. verli. et magis. et ibi Ang. ff. de adul. quem tex. ad hoc reputat singulariter 3af. in d. l. vt vim. iij. col. vbi inter alia dicit q in du- bto excessus presumitur non dolo sed culpa processu alleg. Sa- lice. in d. l. strachus. An autem insultatus debeat fugere / dicit Bar. in d. l. vt vim. in fi. q non nisi sit talis qui verendum non timet. 3af. in d. l. i. c. vnde vi. et vt vim. iij. col. in re. per d. l. i. vt vim. in fi. et ca. ff. i. col. circa med. de confue. alleg. l. i. c. §. plagio. §. rabernarius. ff. ad leg. aquil. et ibidem Cy. et no. per 3archid. in cap. de his q. n. q. magis dicit. Bal. in ca. ij. col. de vita et honeste. per d. l. i. §. qui cum aliter. ad. l. aquil. et quia vt dicit ibidem / fuga est refugium miserorum. Et

ne peut être terminée  
fors en l'assise ou en les  
chiquier / ou par de  
uant le p'ice: par quoy  
la fin en puisse estre re-  
cordée.

De assise. L. chap. iv.  
Assise est vne  
court en la  
quelle ce q  
est fait doit  
auoir par

incp. queritur / ideo dicit Bal. in l. i. in. iij. colla. C. qui accusa. no poss. q. ius nature pmitit q quis defendat se per fas et nephas et q naturalis vltendi bulcedo sic iustis est / refert 3p. poli. de ma. ff. in d. l. i. c. vnde vi. c. ij. et c. significat de bo- micide. et competet secun- dum eum nebum homini- bus sed etia buntis. l. i. §. sed arites. ff. i. quibusd. paup. fecit. dca. et talis de- fensio a principi tolli non potest / refert ibi scrib. in §. Sed naturalis iustis de iu- re natur. gen. et cui. odral- dus de laude consil. p. iij.

Secunda additio.  
In textu ibi.

Ace cry zc.  
Additio. facit quod di. Bar. in l. vt vim ad fi. ff. de insti. c. iure vbi dicit q insultat? debet clamare succu- rite / vel occurrere / vt non an l. itaq. et ibi glo. ff. ad l. aquil. vt sit conuocatio hominum ad cuius auxilium debet vniuersis ve- nire vt dicit doct. in d. l. vt vim. Bar. in l. i. in fin. in l. i. in fin. in fi. ff. de iur. iud. de i. c. colla. de re iud. q si aliter fit inuisa guerra amici etiam non vocari possunt eum iunare alleg. Bar. in l. probibitum de iure fisci. libro. x. c. et possunt ca- pere aduertarios / arg. l. i. de malef. et 2d. ar. sed an extranei in auxilium alicuius aggressi conuocari possunt occidere aggressor? dicit Bar. Bal. et communiter omnes. in d. l. vt vim. q. ff. Ratio secundum Bal. in d. l. vt vim. in. vj. q. est quia aggressor facit iniuriam reipublice. ideo quilibet de populo potest resistere / in ca. dilecto. de sententia excommunicationis. lib. vi. et hoc maxime si sint conuocati in auxilium vt per eisdem doctores. Sed an inuiti debeat iurare aggressum glo. in l. culpa a carere. ff. de reg. iur. tenet q non in contrarium videtur esse tex. in ca. qua- re in l. i. iuncta glo. de sententia excommunicationis. et tex. in di. ca. dilecto. in cap. non inferenda et ibi glo. nota. et ca. ff. xij. q. iij. quod dicit vt poli. alios. in d. l. vt vim. vj. colla. Sullermus le rouille alij conuenio.

De assise. L. chapitre. iv.

Assise est vne court en la quelle ce q est fait doit auoir pardurable ferme / car se len nye ce qui a este fait es plets de viconte / on le peut ameder p vne simple desfene: mais en ce qui est fait en assise / ne recoit len au- cune desfene. zc. Sur ce chapitre on doit noter par ces mots: [se len nye ce qui a este fait es plets de la viconte on le peut ameder par vne desfene] cest a entendre que par voye de boeance len est receu a dire contre ce qui a este fait en viconte / et annuller ce qui y a este fait par affermant facit contraire ou desluctif de ce qui a este fait en viconte et que len contend annuller.

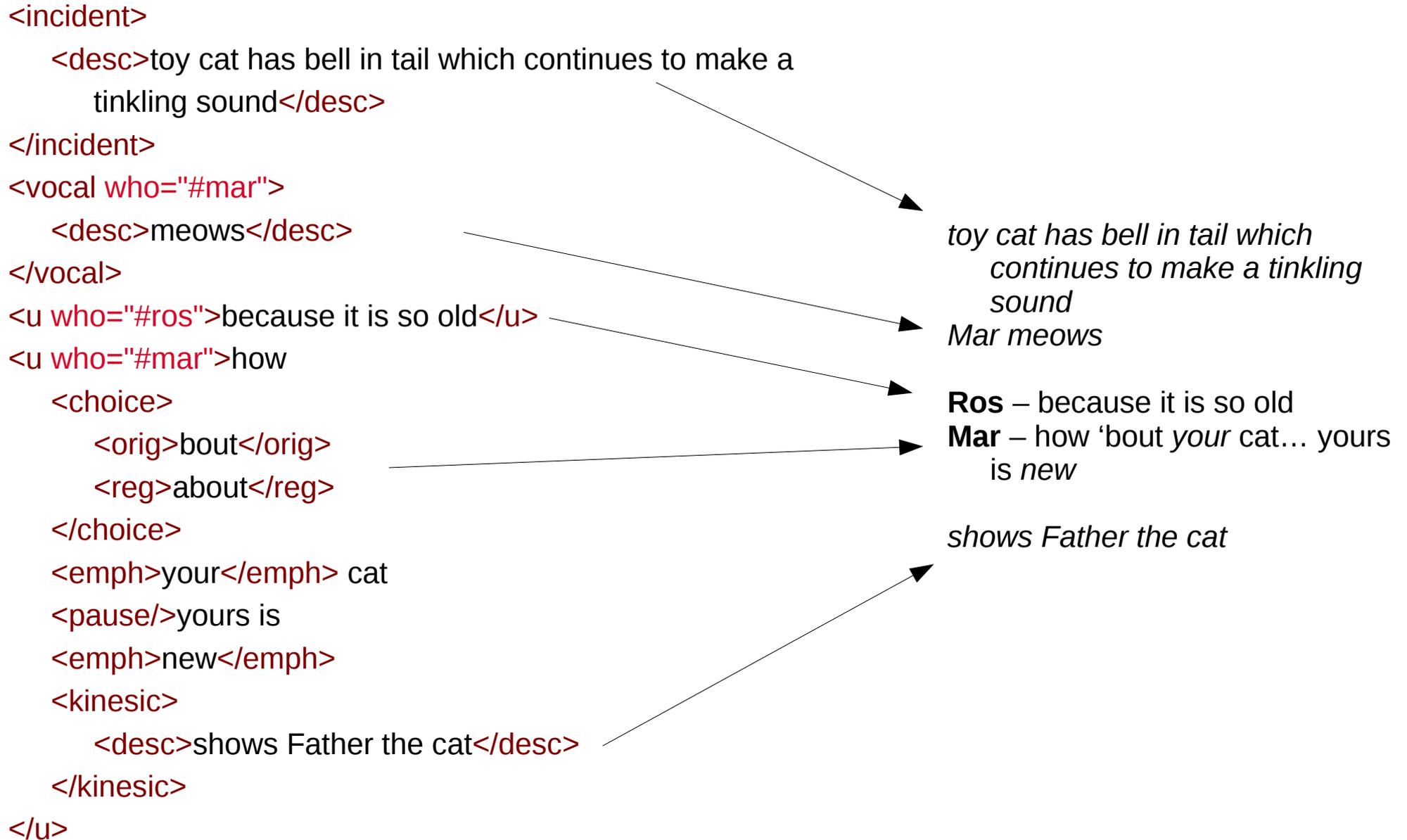
Item len doit noter par ce mots. Mais ce qui a este fait en assise zc. que quant on se met en record d'assise eu quel on ne se met point si no de lacoed des parties. Lombies quilz fe doivent mettre de vout quant le cas le requiert. Le qui y est recorde et confirme a tousiours doit auoir ferme / car on ne sen peut do- loir. Et la cause est pource quil est fait par record / qui est vne voye en quoy les parties se submettent de

Entête

Texte principal  
(gauche, droite)

Glose encadrante  
(gauche, droite)

# 5) Encodage du texte - TEI



# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) **Conception des outils de consultation.**
- 7) Mise à disposition du public.
- 8) Maintenance.

# 7) Outils de consultation

Site internet :

- Serveur,
- Base de données (= fichiers XML-TEI)
- Système d'interrogation (pour naviguer dans le XML-TEI)
- Transformations (pour l'affichage du XML),
- Structure HTML/etc.

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) **Mise à disposition du public.**
- 8) Maintenance.

# Les étapes de la numérisation

- 1) Travail initial.
- 2) Numérisation.
- 3) Acquisition du texte.
- 4) Conception de la base de données.
- 5) Encodage du texte.
- 6) Conception des outils de consultation.
- 7) Mise à disposition du public.
- 8) **Maintenance.**

## 8) Maintenance

/! TRÈS IMPORTANT /!

- Pérennité des outils et de la base de données.