



HAL
open science

**Master Pro “ Patrimoine écrit et édition numérique,
option HN ” Initiation à l’encodage des textes
patrimoniaux Formation XML/TEI**

Nicole Dufournaud

► **To cite this version:**

Nicole Dufournaud. Master Pro “ Patrimoine écrit et édition numérique, option HN ” Initiation à l’encodage des textes patrimoniaux Formation XML/TEI. Master. France. 2015. hal-03778050

HAL Id: hal-03778050

<https://hal.science/hal-03778050>

Submitted on 15 Sep 2022

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.

Master Pro « Patrimoine écrit et édition numérique,
option HN »

Initiation à l'encodage des textes patrimoniaux
Formation XML/TEI

Introduction à XML/TEI

CESR Tours – 12 janvier 2015



Nicole Dufournaud

Chercheuse en histoire – Membre associée LaDéHis, EHESS

Biographie

- De 1982 à 1993 : administration de la recherche au Centre mondial informatique et ressource humaine ; participation à l'informatisation dans le milieu financier
- De 1996 à 2013 : participation dans les projets TEI
 - En 1996-1997 : projet Emma au CNRS, IRISA, EMN
 - De 1998 à 2007 : travaux universitaires en histoire
 - 2006/2007 : projet Millefeuille avec AN, ENC, Paris I
 - Mai 2007-mars 2010 : projet BVH au CESR Tours
 - Avril 2010 au 31 août 2012 : projet ENBaCH à l'EHESS
 - Octobre 2012 à juillet 2013 : projet Millefeuille à l'ENS Lyon pour les AN Paris ; projet Ampère (Centre Koyré) ; projet Archives e-LV à l'IHPST
- Depuis 2004 : cours XML/TEI Master Pro, CESR
- Depuis septembre 2013 : projet KinSources à l'EHESS

Objectifs

- Rappel des notions de base :
 - L'édition de textes
 - Le vocabulaire informatique
 - Codage et encodage
 - Langage, format
- Structuration de données : du document au texte
 - Un outil d'édition, de diffusion et d'analyse : TEI (Text Encoding Initiative ou Text Encoding for Interchange !)
 - Les principes de balisages
 - Méthodologie : introduction à tous les types de structures de texte
- Finalités du travail de structuration de données

Les guides

- Les diaporamas en format pdf
- Des références choisies :
 - Une bibliographie : imprimée et en ligne
 - Des sites modèles et des références
- Un extrait des Guidelines TEI :
 - Avec des exemples à compléter
- Les Guidelines TEI en ligne sur <http://www.tei-c.org/>
- Manuels d'encodage XML/TEI
- Les listes de discussions : TEI Liste, TEI FR, DH

L'édition des textes : l' « ecdotique »

- Science de l'édition de textes non numérique -

- **Les données textuelles sont compliquées :**
 - Mots composés de caractères
 - Concepts
- **Éditions critique ou savante**
 - Edition critique ou apparat critique : rétablir la source originale (philologie)
 - avec des notes philologiques
 - Edition savante : enrichir les textes édités
 - avec des notes érudites

L'édition des textes

- Une définition pour « édition savante » ou « apparat savant » :
 - C'est l'enrichissement de qualité « érudite » qui accompagne un texte.
- Qu'entend-on par « accompagner un texte » ?
 - Trouver les citations et les vérifier
 - Trouver les références et les sources de l'auteur
 - Annoter
 - Introduire les textes
 - Traduire en français les parties en langues étrangères
 - Commenter : historique, philologique, littéraire, philosophique...
 - Construire des index
 - Établir un glossaire, etc.

L'édition des textes à l'ère du numérique : les besoins

- Que peut faire un utilisateur, un internaute, pour :
 - consulter et rechercher en plein texte ?
 - enrichir, publier et diffuser des documents ?
 - analyser, indexer et visualiser des informations ?
 - des textes :
 - manuscrits ou imprimés ?
 - actuels ou anciens ?
 - qui ont été saisis par l'utilisateur ou par d'autres ?

La réponse est :

STRUCTURER LES DONNEES !

Structurer des données

- C'est convertir des données (textuelles) dans un format exploitable par un ordinateur
- Dans le contexte des documents numériques, c'est :
 - Utiliser un langage formel de marquage pour décrire un texte sous la forme d'une suite de phénomènes textuels et paratextuels qualifiant et **structurant** le texte.
- Le document numérique est alors exploitable par l'humain ET par la machine
- Les données doivent être réexploitables

Exemple de format structuré : quel est ce format ?



Codage et encodage numériques

- **Codage informatique :**
 - Représentation numérique (binaire) de données
 - Exemple du code ISO 639 pour la représentation des noms de langues
http://www.iso.org/iso/fr/home/standards/language_codes.htm
et <http://www-01.sil.org/iso639-3/codes.asp>
- **Encodage (textuel) :**
 - Transformation dans un code qui structure les données textuelles
 - Dans quel but ?
 - Structurer pour structurer
 - Éditer des textes numériquement
 - Analyser des textes pour mieux les exploiter

Langages, éditeurs, normes et formats

Langages de programmation

Premier vers 1840 : Ada Lovelace

1936 : Turing

1954 : Fortran

1959 : Lisp

1968 : Logo

1971 : Pascal

1972 : Smalltalk, C, Prolog

1983 : Ada, C++

1991 : Python

1995 : Java, PHP

```
#include <stdio.h>
#include <stdlib.h>
#include <sys/types.h>
#include <arpa/inet.h>

void serveur1(portServ ports)
{
    int sockServ1, sockServ2, sockClient;
    struct sockaddr_in monAddr, addrClient, addrServ2;
    socklen_t lenAddrClient;

    if ((sockServ1 = socket(AF_INET, SOCK_STREAM, 0)) == -1) {
        perror("Erreur socket");
        exit(1);
    }
    if ((sockServ2 = socket(AF_INET, SOCK_STREAM, 0)) == -1) {
        perror("Erreur socket");
        exit(1);
    }

    bzero(&monAddr, sizeof(monAddr));
    monAddr.sin_family = AF_INET;
    monAddr.sin_port = htons(ports.port1);
    monAddr.sin_addr.s_addr = INADDR_ANY;
    bzero(&addrServ2, sizeof(addrServ2));
```

Langage C

Langages, éditeurs, normes et formats

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Commentaire -->
<ex:collection
  xml:lang="fr"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://www.w3.org/1999/xhtml"
  xmlns:ex="http://exemple.org"
>
  <élément>Texte</élément>
  <dc:title>Asterix le Gaulois</dc:title>
  <ex:livre attribut="valeur" type="BD">
    <dc:title>Asterix chez les Belges</dc:title>
    <!-- élément répété -->
    <dc:creator>René Goscinny</dc:creator>
    <dc:creator>Albert Uderzo</dc:creator>
    <dc:description>
      <b>Asterix chez les Belges</b> est un album de
      <a href="http://fr.wikipedia.org/wiki/Bande_dessinée">bande dessinée</a>
      de la série Asterix le Gaulois créée par René Goscinny et Albert Uderzo.
      <br/><!-- élément vide -->
      Cet album publié en 1979 est le dernier de la série écrit par René Goscinny.
    </dc:description>
  </ex:livre>
</ex:collection>
```

Langage XML

Langage de balisage

1986 : SGML

1990 : HTML

1998 : XML

Langages, éditeurs, normes et formats

- Les normes ISO :
 - Langues et caractères
 - Langue ISO 639-3 : frm (pour la Renaissance)
 - Codes et langages
 - Date et heure : 2012-04-02
- XML est une norme ISO qui sert à décrire des documents structurés

Un langage informatique : XML, en vrac !

- XML représente :
 - les données structurées hiérarchiquement
 - une arborescence composée de parents et d'enfants
- XML est « indépendant » des outils :
 - Emacs
 - Oxygen
 - Eclipse, etc...
- XML est un langage pérenne et relativement accessible
- XML a une syntaxe qui doit être respectée

Affichage sous forme d'arbres

- Arborescence XML

```
<parent>
```

```
  <enfant1>Claude</enfant1>
```

```
  <enfant2>
```

```
    Dominique
```

```
    <petit-enfant1>Bébé</petit-enfant1>
```

```
  </enfant2>
```

```
</parent>
```


Les différents formats d'encodage

- Avant de structurer, il faut choisir !
 - Les formats diffèrent selon l'encodage voulu
- Plusieurs encodages possibles selon les besoins :
 - HTML : pour afficher sur le Web
 - LATEX : pour la publication papier
 - EAD : pour le catalogage en bibliothèque
 - EAC : pour le catalogage des archives
 - Dublin Core, etc...

- Word

Le texte est centré
et en *italique*

- LaTeX

```
\begin{center}  
  Le texte est centré\\ et en \textit{italique}  
\end{center}
```

- HTML

```
<p style="text-align: center;">  
  Le texte est centré<br>et en <em>italique</em>  
</p>
```

- TEI

```
<p rend="center">  
  Le texte est centré<lb/>et en <hi rend="it">italique</hi>  
</p>
```

Les formats et leur extension

- Word : .doc
- OpenOffice : .odt
- Excel : .xls
- HTML : .htm
- XML : .xml

Exemples : rapport-aeres.docx

Quel est la différence entre .doc et .docx ?

Editeurs de textes graphiques et structuration des données

- Différences entre un « éditeur », un « traitement » et un « formateur de textes »
 - Aspect graphique : **Word**
 - Anciennement « traitement de textes »
 - « Éditeur de textes graphiques »
 - « WYSIWYG »
 - Aspect structurel hiérarchique : **Latex**
 - Anciennement « formateur de textes »
 - « Éditeur de documents structurés »
 - « WYSIWYM »

Le format RDF

Ressource Description Framework

- Le langage du « Web sémantique »
- RDF est compatible avec la TEI
- RDF utilise des triplets : un « triplet » est une association entre un sujet, un prédicat et un objet
- RDF est organisé en « graphe » constitué de nœuds et de liens
- RDFa :
 - Répondre aux besoins importants de présentation, de mise en forme sur des documents peu structurés avec les attributs dans RDF

Clarifier le vocabulaire

- « Edition numérique » ou « Edition électronique » ?
 - L'électronique renvoie au « matériel » alors que le numérique renvoie à la dématérialisation, à l'informatique
 - Ne pas utiliser le mot « digital » en français
- « Informatique »
 - Domaine d'activité scientifique qui traite l'information automatiquement grâce à un ordinateur
- Exemple de la « typologie »
 - C'est une démarche pour définir un certain nombre de types classés afin d'en faciliter l'analyse (Guidelines de la TEI).

Accepter le vocabulaire et les traditions des autres disciplines !

Editeur graphique et structuration des données

- Pour diffuser un texte en ligne, pourquoi NE PAS utiliser HTML fait pour la publication sur le Web ?
 - Est-ce un langage compréhensible pour la machine ?
- Pour échanger un texte, pourquoi NE PAS utiliser Word fait pour la publication papier ?
 - Peut-on distinguer les modifications d'origine, les interventions diverses, les codes ?
 - Comment suivre les différentes versions ?

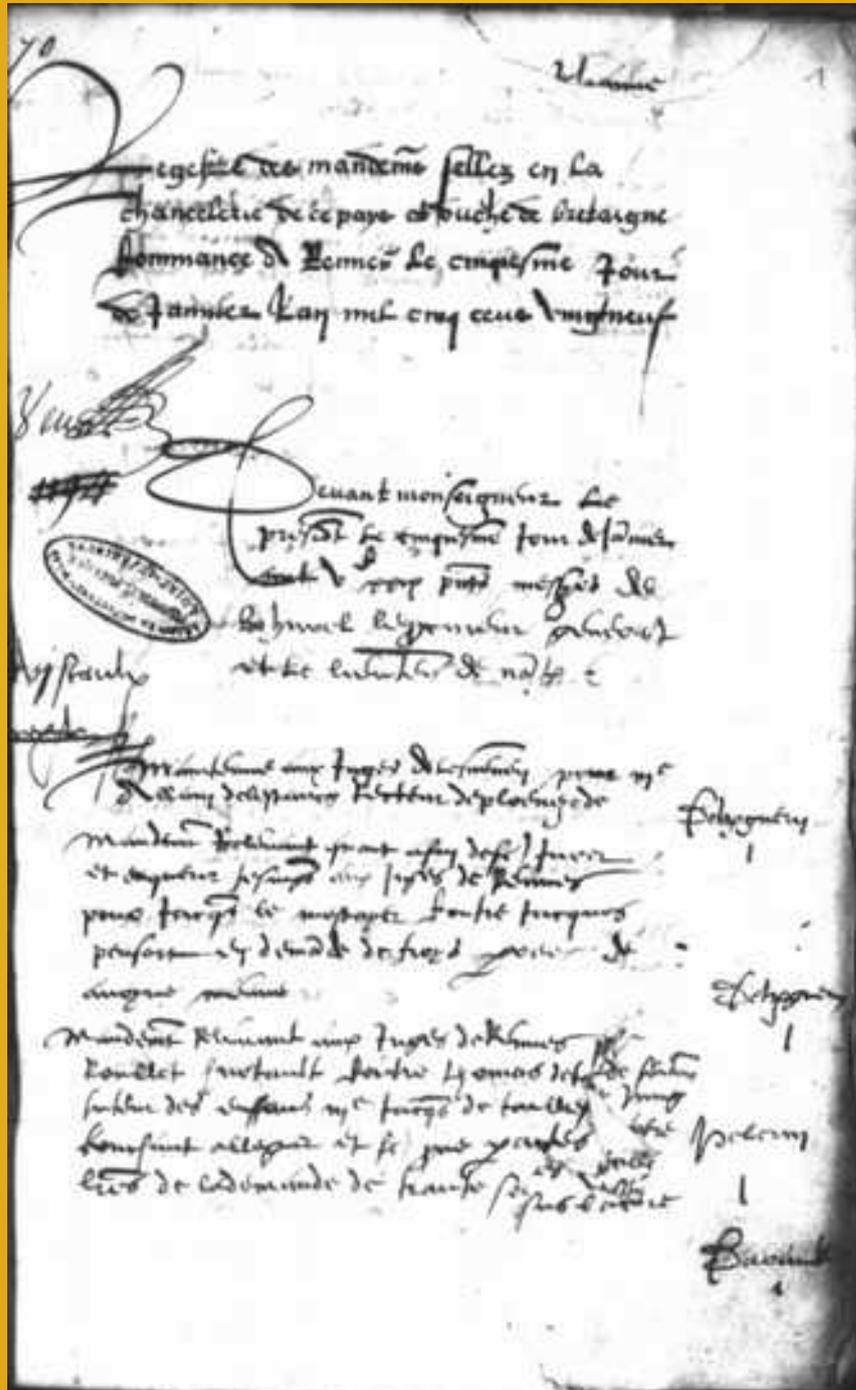
Les inconvénients des bases de données

Voir <http://www.lespetitescases.net/node/1088>

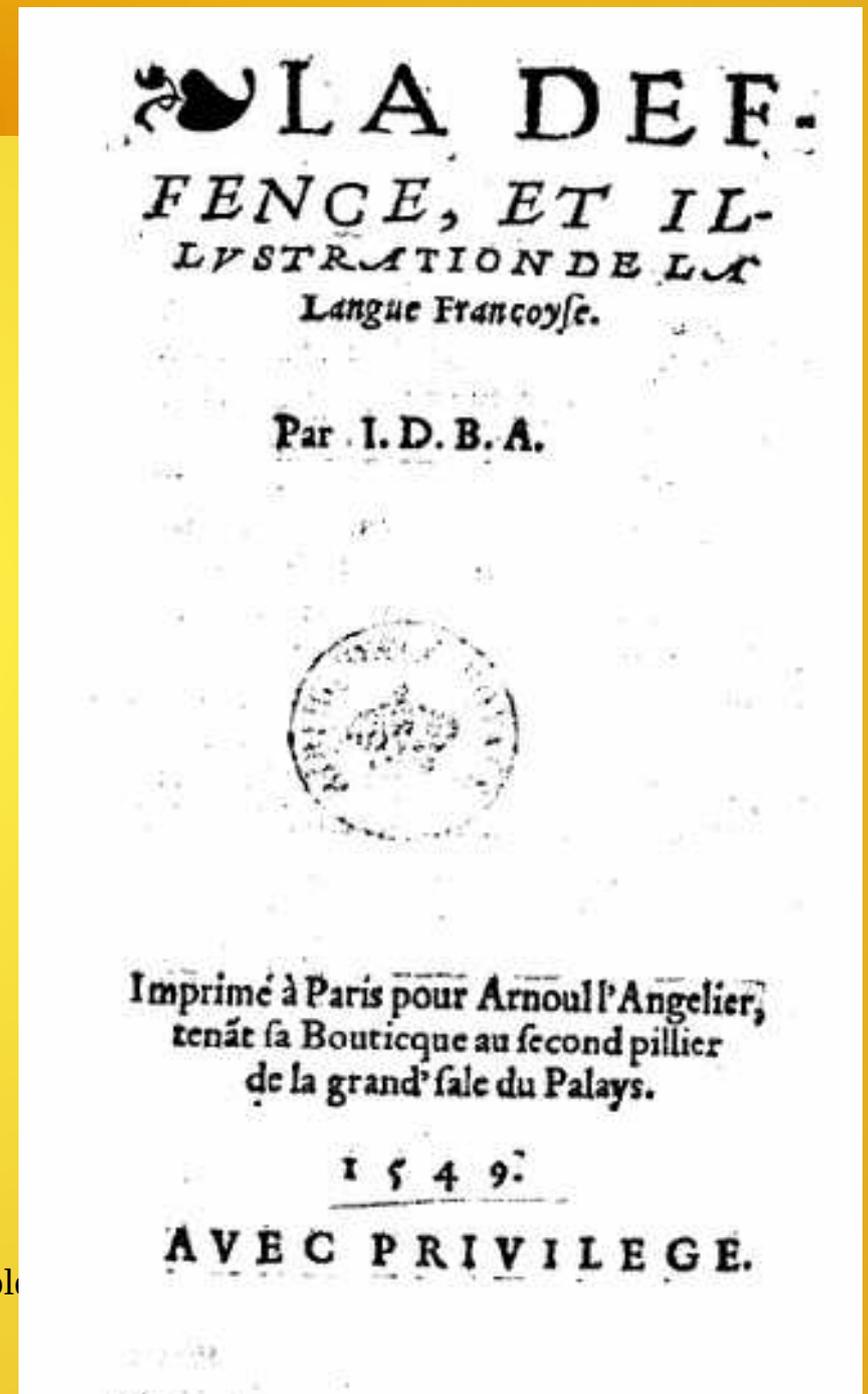
- Explosion de l'information en tables et en champs
 - *Atomisation du texte : décomposition*
- Reconstruction du texte par des requêtes
- Pas de fichier contenant l'ensemble du texte structuré
- *Pas d'édition incrémentale et cumulative*
- *Le retour au texte original est impossible*
- *Pas de réexploitation systématique possible*

Du document au texte

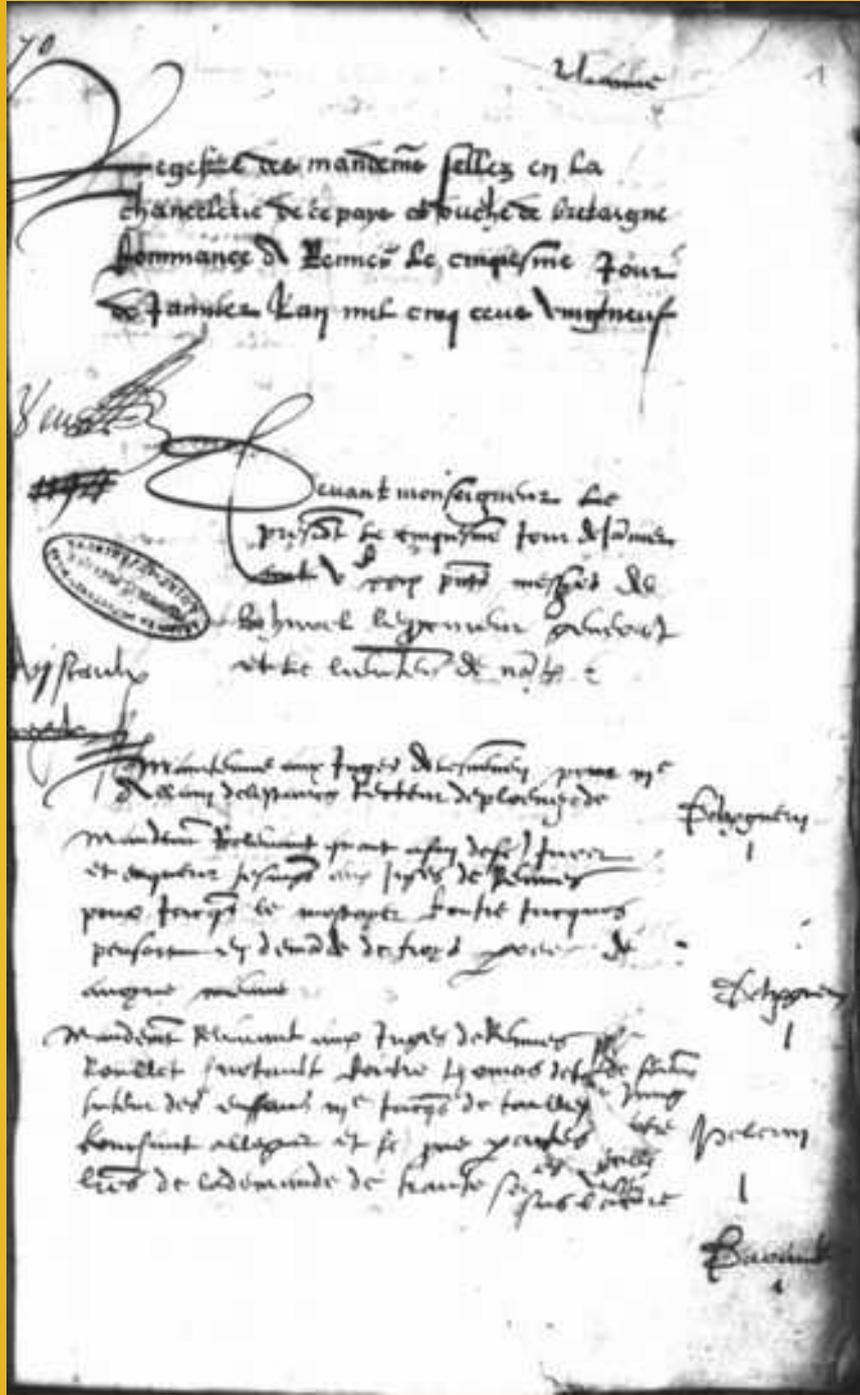
Un document manuscrit



Un document imprimé



PAGE DE TITRE d'un registre



Titre du registre

Titre de la division du registre

Corps du texte du registre

Structure d'un document : le manuscrit

- Structure du registre de la Chancellerie
 - Description du registre :
 - Critique externe : fabrication du registre, état du registre
 - Critique interne : le contenu du registre (durée, différence des pièces...)
 - Dresser la typologie des actes : évocation, mandement, placet, répit, sauvegarde, acte exécutoire, lettre d'office, relief, lettre de pardon, commutation de peine, abolition.
- Voir l'initiation à l'encodage des manuscrits par Lou Burnard

PAGE DE TITRE d'un imprimé

 LA DEF-
FENCE, ET IL-
LUSTRATION DE LA
Langue Francoyse.

Par I. D. B. A.



Imprimé à Paris pour Arnoul l'Angelier,
tenât sa Boutique au second pillier
de la grand' sale du Palays.

1 5 4 9.

AVEC PRIVILEGE.

Titre du livre

Auteur

Impression :
nom imprimeur
et date

Privilège

Structure d'un livre imprimé

- La page de titre (description bibliographique pour le catalogage et l'indexation)
 - Le titre
 - Auteur, date d'impression, etc.
- Les pièces liminaires
 - Privilège
 - Dédicace
 - Préface...
- Le texte
 - Le corps du texte
 - Un groupe de textes : recueils
- Les pièces annexes
 - Table des matières, postface, etc.

Structure du document

- Le corps du texte
 - Les divisions avec leur en-tête ou titre (chapitre, livre, partie, etc.)
 - Les paragraphes
 - Les strophes...
 - La mise en valeur typographique : le rendu
 - Les notes
 - Les marges
 - Les tableaux et les listes
 - Les illustrations

Structure du document

- Structure physique du document :
 - Le texte d'un document n'est pas linéaire
 - Feuillet ou la page
 - L'aspect physique : déchirure
 - Reliure

La structure physique interfère dans l'interprétation du texte

Comment structurer ?

Grâce à un format qui facilite l'échange :

« Un format d'échange s'exprime dans un encodage qui explicite une théorie concernant un aspect quelconque d'un document » Lou Burnard, Oxford University

Les limites de l'encodage :

- Tout encodage est une interprétation
- Aucun langage d'encodage ne peut prétendre être complet

Les atouts de l'encodage

- Permet l'exploration des données textuelles
- Donne des idées !

La **TEI** :
Text Encoding Initiative
ou
Text Encoding for Interchange

La TEI, un encodage parmi d'autres ?

Qui se sert de la TEI ?

- Les chercheurs en SHS et en lettres : linguistes, littéraires, historiens, sociologues, philosophes, etc.
- Les bibliothécaires
- Les archivistes
- Les éditeurs

Pourquoi s'en servent-ils ?

- Pour structurer leurs documents en vue de :
 - Les diffuser et les publier ou/et
 - Les échanger et les publier ou/et
 - Les analyser

La TEI est le seul format d'analyse qui permet l'encodage du contenu d'un texte !

Du document au texte

Projet e-LV

- TEI a une dimension sémantique
- L'encodage TEI est créatif
- Tension entre l'encodage linéaire et l'encodage génétique :
 - TEI a été créé pour l'encodage linéaire or le texte est rarement linéaire (voir les travaux d'Elena Pierazzo)
 - La matérialité du texte doit être encodée
 - L'encodeur interprête le texte quand il construit un autre texte en montrant le processus d'écriture

La TEI : un format d'échange pour des documents structurés

La TEI est un projet pluridisciplinaire pour uniformiser le codage de documents en vue de leur échange :

- recommandations pour un standard d'échange

La TEI a besoin d'un environnement :

- XML
 - XML est un meta-langage pour aider à la saisie d'un texte en le structurant.
 - XML/TEI : outil d'édition, de diffusion et d'analyse
- Eclipse, HTML, XSLT, Oxygen, etc.

Principes de balisage XML

- **L'élément** est la structure de base marquée par une balise de début et une balise de fin

- La balise ouvrante se marque avec des chevrons

`<...>`

- pour la balise fermante une barre oblique est ajoutée

`</...>`

- Exemple de balisage

`<...>...</...>`

- si le contenu est vide, les balises se raccourcissent :

`<.../>`

soit une barre oblique APRES le contenu

Principe de balisage

- L'**attribut** est un paramètre associé à un élément
 - Il fournit une description de l'élément
 - Exemple : @n @ désigne la marque de l'attribut
<pb n="1recto"/> 1recto étant la valeur de l'attribut
 - Autres attributs courant :
 - @type
 - @rend
- Certains éléments sont également des attributs:
 - Exemple : <name> et @name
 <label> et @label

Principes de balisage

- Deux règles essentielles dans XML :
 - La syntaxe des balises
 - Règle d'imbrication stricte : il est interdit de faire chevaucher les éléments
- ```
<title><p>Le Tiers Livre</title></p>
```

# Principes de balisage

- Hiérarchie : une certaine catégorie d'éléments est imbriquée dans une autre catégorie : exemple dans la TEI
  - `<addrLine>` est toujours imbriqué dans `<address>`
  - `<date>` est parfois utilisé seul ou parfois imbriqué dans d'autres éléments comme `<bibl>`

# Présentation du fichier

<TEI>

<teiHeader>...</teiHeader>



Les métadonnées

<text>

<front>...</front>



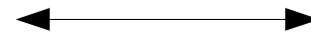
Les pièces liminaires

<body>

<div>

...

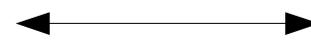
</div>



Le corps du texte

</body>

<back>...</back>



Les pièces annexes

</text>

</TEI>

# Un peu d'informatique !

- Un texte TEI est validé selon son schéma
  - La TEI est un système modulable
- Les éléments de la TEI sont réparties
  - en module : 20 modules en 2014
  - en classe
- Le choix des modules et des classes a des implications sur l'utilisation des éléments et des attributs
- Roma est une interface qui permet d'élaborer son propre schéma

# Un peu d'informatique !

- DTD (Document Type Definition)
  - Contraintes supplémentaires : indique quelles balises sont imbriquées dans d'autres balises :
    - Élément : `<div>` suivi de `<head>` suivi de `<p>`
  - Vient de SGML : obsolète !
- Schéma :
  - Plus flexible que la DTD qui est conservée pour la compatibilité : on exprime plus de contraintes plus précisément
    - Attribut : `date="an-mois-jour" time="h-mn-sec"`

# Environnement de la TEI en ligne sur <http://tei-c.org>

- Guidelines
- Les outils
  - ROMA
    - schémas
  - Feuille de style XSLT
  - Oxgarage conversion

→Attendre la seconde journée, initiation par Lauranne Bertrand et Jorge Fins

# Exercice de manipulation sous Oxygen

# Méthode de balisage avec la TEI

J-D. Fekete and N. Dufournaud. « Analyse historique de sources manuscrites : application de TEI à un corpus de lettres de rémission du XVI<sup>e</sup> siècle. » Document Numérique, numéro spécial “Les documents anciens”, 3(1-2):117-134, 1999.

- Balisage formel et typographique qui permet :
  - une transcription structurelle et logique :
    - <div> et <head>
    - <p>
  - Un respect de la mise en forme :
    - <lb/>
    - <fw> pour *Forme Work* pour les en-tête et pied de page
- Balisage sémantique : éléments analytiques succincts
  - <persName> <date> <geogName> <address>

# Processus d'encodage

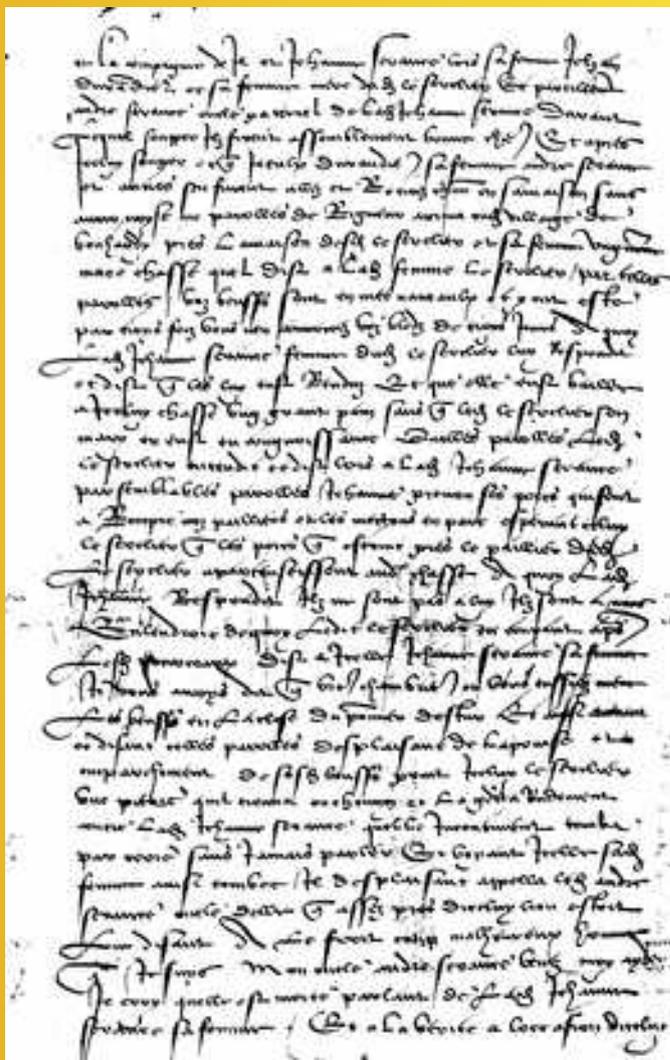
- Méthodologie : trois étapes
  - Saisie ou récupération de la transcription
  - Encodage formel et logique du texte
  - Enrichissement en TEI (noms, dates, lieux, typologies)

Encodage analytique en TEI possible avec liaison d'une grille de dépouillement



Exemple du travail d'encodage en maîtrise de l'intervenante

# Etape 0-Transcription diplomatique



- [11] en la compagnie de il et Jehanne Serance, lors sa femme, Jehan
- [12] durandiere et sa femme, mere dudit Leserclier, et pareillement
- [13] André Serance, oncle paternel de ladite Jehanne Serance, durant
- [14] lequel soupper ilz firent assemblement bonne chere et après
- [15] iceluy soupper et que iceulx Durandiere, sa femme, André Serance
- [16] et autres s'en furent allez et retirez chacun en sa maison sans
- [17] avoir noyse ne parolles de rigueur, arriva oudit villaige de
- [18] Bonhardy prés la maison desdits Leserclier et sa femme ung nommé
- [19] Mace Chassé qu'il dist a ladite femme Leserclier par celles
- [20] parolles : « Voz beuffs sont en mes naveaulx et y ont esté
- [21] par troys foiz ; vous n'en couvrirez voz bledz de troys jours ». A quoy
- [22] ladite Jehanne Serance, femme dudit Leserclier, luy respondit
- [23] et dist que les luy eust renduz et que elle eust baillé
- [24] a iceluy Chassé ung grant pain sans que ledit Leserclier, son
- [25] mary en eust eu congnoissance, quelles parolles ledit
- [26] Leserclier entendit et dist lors a ladite Jehanne Serance
- [27] par semblables parolles : « Jehanne, prenons ses porcs qui sont
- [28] a rompre noz pailliers et les mectons en parc », esperant celuy
- [29] Leserclier que les porcs qui estoient prés le paillier dudit
- [30] Leserclier apartenseissent audit Chassé. A quoy ladite
- [31] Jehanne respondit : « Ilz ne sont pas a luy, ilz sont a nous ».
- [32] En l'endroit de quoy ledit Leserclier en courant après
- [33] lesdits pourceaux, dist a icelle Jehanne Serance sa femme :
- [34] « je vous avoys dit que votre chambriere et vous, eussiez mené
- [35] les beuffs en la close du pommier dessus ». Et ainsi courant
- [36] et disant celles parolles desplaisant de la prinse et
- [37] emparchement de ses dits beuffs, print iceluy Leserclier
- [38] une pierre qu'il trouva ou chemyn et la getta rudement

# Etape 1-Codage diplomatique

- Annotations éditoriales:
  - <sic> pour des erreurs manifestes
  - <abbr> pour les abréviations
  - <unclear> pour les mots incertains
  - <del> pour les ratures
  - <add> pour les ajouts
  - <note> pour des notes
  - etc.

# Codage et balisage diplomatique TEI

[11] en la compagnie de il et Jehanne Serance, lors sa femme, Jehan  
[12] durandiere et sa femme, mere dudit Leserclier, et pareillement  
[13] André Serance, oncle paternel de ladite Jehanne Serance, durant  
[14] lequel soupper ilz firent assemblement bonne chere et après  
[15] iceluy soupper et que iceulx Durandiere, sa femme, André Serance  
[16] et autres s'en furent allez et retirez chacun en sa maison sans  
[17] avoir noyse ne parolles de rigueur, arriva oudit villaige de  
[18] Bonhardy prés la maison desdits Leserclier et sa femme ung nommé  
[19] Mace Chassé qu'il dist a ladite femme Leserclier par celles  
[20] parolles : « Voz beuffs sont en mes naveaulx et y ont esté  
[21] par troys foiz ; vous n'en couvrirez voz bledz de troys jours ». A quoy  
[22] ladite Jehanne Serance, femme dudit Leserclier, luy respondit  
[23] et dist que les luy eust renduz et que elle eust baillé  
[24] a iceluy Chassé ung grant pain sans que ledit Leserclier, son  
[25] mary en eust eu congnoissance, quelles parolles ledit  
[26] Leserclier entendit et dist lors a ladite Jehanne Serance  
[27] par semblables parolles : « Jehanne, prenon ses porcs qui sont  
[28] a rompre noz pailliers et les mectons en parc », esperant celuy  
[29] Leserclier que les porcs qui estoit prés le paillier dudit  
[30] Leserclier apartenseissent audit Chassé. A quoy ladite  
[31] Jehanne respondit : « Ilz ne sont pas a luy, ilz sont a nous ».  
[32] En l'endroit de quoy ledit Leserclier en courant après  
[33] lesdits pourceaux, dist a icelle Jehanne Serance sa femme :  
[34] « je vous avoys dit que votre chambriere et vous, eussiez mené  
[35] les beuffs en la close du pommier dessus ». Et ainsi courant  
[36] et disant celles parolles desplaisant de la prise et  
[37] emparchement de ses dits beuffs, print iceluy Leserclier  
[38] une pierre qu'il trouva ou chemyn et la gecta rudement

```
emars@SGRNA
Buffer Files Tools Edit Search Jede SGML Modfy Move Markup View DTD Help
<LB N="11">en la compagnie de il et Jehanne Serance, lors sa femme, Jehan
<LB N="12">durandiere; et sa femme, mere dudit Leserclier, et pareillement
<LB N="13">André Serance, oncle paternel de ladite Jehanne Serance, durant
<LB N="14">lequel soupper ilz firent assemblement bonne chere; et après
<LB N="15">iceluy soupper et que iceulx Durandiere, sa femme, André Serance
<LB N="16">et autres s'en furent allez et retirez chacun; en sa maison sans
<LB N="17">avoir noyse ne parolles de rigueur, arriva oudit villaige de
<LB N="18">Bonhardy prés la maison desdits Leserclier et sa femme ung nommé
e:
<LB N="19">Mace Chassé qu'il dist a ladite; femme Leserclier par celles
<LB N="20">parolles : <Q>Voz beuffs sont en mes naveaulx et y ont esté
<LB N="21">par troys foiz ; vous n'en couvrirez voz bledz de troys jours</Q>. \
A quoy
<LB N="22">ladite; Jehanne Serance, femme dudit Leserclier, luy respondit
<LB N="23">et dist que les luy eust renduz et que elle eust baillé
<LB N="24">a iceluy Chassé ung grant pain sans que ledit Leserclier, son
<LB N="25">mary en eust eu congnoissance, quelles parolles ledit;
<LB N="26">Leserclier entendit et dist lors a ladite; Jehanne Serance
<LB N="27">par semblables parolles : <Q>Jehanne, prenon ses porcs qui sont
<LB N="28">a rompre noz pailliers et les mectons en parc</Q>, esperant celuy
<LB N="29">Leserclier que les porcs qui estoit prés le paillier dudit;
<LB N="30">Leserclier apartenseissent audit Chassé. A quoy ladite;
<LB N="31">Jehanne respondit : <Q>Ilz ne sont pas a luy, ilz sont a nous</Q>\
.
<LB N="32">En l'endroit de quoy ledit Leserclier en courant après;
<LB N="33">lesdits pourceaux, dist a icelle Jehanne Serance sa femme :
<LB N="34"><Q>je vous avoys dit que votre chambriere et vous, eussiez mené
t vous, eussiez mené
<LB N="35">les beuffs en la close du pommier; dessus</Q>. Et ainsi courant
<LB N="36">et disant celles parolles desplaisant de la prise et
<LB N="37">emparchement de ses dits; beuffs, print iceluy Leserclier
<LB N="38">une pierre qu'il trouva ou chemyn et la gecta rudement
<LB N="39">contre ladite; Jehanne Serance quelle incontinent tomba
<LB N="40">par terre sans jamais parler. Et voyant icelle, ledite;
<LB N="41">femme, ainsi tombee, il desplaisant, appela ledit; André
<LB N="42">Serance, oncle d'elle, que s'assez; prés d'iceluy lieu estoit,
<LB N="43">luy disant : <Q>A le froit coup, malheureux homme
<LB N="44">que; je suys, mon oncle André Serance, venez moy aidé.
<LB N="45">Je croy qu'elle est morte</Q>, parlant de ladite; Jehanne
<LB N="46">Serance sa femme. Et a la veoire, a l'occasion d'iceluy
<PB N=3>
<LB N="47">coup, alla incontinent celle Jehanne Serance de vie a trespas.
<LB N="48">Et craignant ledit Leserclier rigueur de justice, a'est
-----Emars: Leserclier_sun.tei (8:SGML [tei.2])--L128--47%-----
Fontifying... done
```

# Etape 2 - les annotations sémantiques

- `<date>` pour les dates
- `<name>` pour les noms
- `<rs type="...">`
  - Toponyme pour les lieux
  - Arme pour les armes
  - Meubles, architecture, etc. suivant les centres d'intérêts
- `<signed>`

# Balisage sémantique TEI

```
emacs@SORNA
Buffers Files Tools Edit Search Jade SGML Moddy Move Markup View DTD Help
<LB N="11">en la compagnie de il et Jehanne Serance, lors sa femme, Jehan
<LB N="12">&durandiere; et sa femme, mere &duit; Leserclier, et pareilleent
<LB N="13">André Serance, oncle paternel de &ladite; Jehanne Serance, durant
<LB N="14">lequel soupper ilz firent assemblement bonne &chere; et après
<LB N="15">iceluy soupper et &que; iceulx Durandiere, sa femme, André Serance
<LB N="16">et autres s'en furent allez et retirez &chacun; en sa maison sans
<LB N="17">avoir noyee ne parolles de rigueur, arriva &duit; villaige de
<LB N="18">Bonhardy près la maison &desdits; Leserclier et sa femme ung &nomme\
&:
<LB N="19">Mace Chassé qu'il dist a &ladite; femme Leserclier par celles
<LB N="20">parolles : <Q>Uoz beuffs sont en mes naveaux et y ont esté
<LB N="21">par troys foiz : vous n'en couvrirez voz bledz de troys jours</Q>.\
A quoy
<LB N="22">&ladite; Jehanne Serance, femme &duit; Leserclier, luy respondit
<LB N="23">et dist &que; les luy eust renduz et que elle eust baillé
<LB N="24">a iceluy Chassé ung grant pain sans &que; &ledit; Leserclier, son
<LB N="25">mary en eust eu congnoissance, quelles parolles &ledit;
<LB N="26">Leserclier entendit et dist lors a &ladite; Jehanne Serance
<LB N="27">par seables parolles : <Q>Jehanne, prenon ses porcs qui sont
<LB N="28">a rompre noz pailliers et les sectons en parc</Q>, esperant celuy
<LB N="29">Leserclier &que; les porcs &qui; estoit près le paillier &duit;
<LB N="30">Leserclier apartenseissent &duit; Chassé, A quoy &ladite;
<LB N="31">Jehanne respondit : <Q>Ilz ne sont pas a luy, ilz sont a nous</Q>.\
.
<LB N="32">En l'endroit de quoy &ledit; Leserclier en courant &apres;
<LB N="33">&ledite; pourceaux, dist a icelle Jehanne Serance sa femme :
<LB N="34"><Q>je vous avoys dit &que; &votre; <unclear>chambriere</unclear> e\
t vous, eussiez mené
<LB N="35">les beuffs en la close du &pommier; dessus</Q>. Et ainsi courant
<LB N="36">et disant celles parolles desplaisant de la prise et
<LB N="37">esparchement de ces &dits; beuffs, print iceluy Leserclier
<LB N="38">une pierre qu'il trouva ou cheymn et la gecta rudement
<LB N="39">contre &ladite; Jehanne Serance quelle incontinent tomba
<LB N="40">par terre sans jamais parler, Et voyant icelle, &edit;
<LB N="41">femme, ainsi tombee, il desplaisant, appela &ledit; André
<LB N="42">Serance, oncle d'elle, &quessez; près d'iceluy lieu estoit,
<LB N="43">luy disant : <Q>A le froit coup, malheureux homme
<LB N="44">&que; je suy, mon oncle André Serance, venez moy aidé;
<LB N="45">Je croy qu'elle est morte</Q>, parlant de &ladite; Jehanne
<LB N="46">Serance sa femme. Et a la veoire, a l'occasion d'iceluy
(CPB N=3)
<LB N="47">coup, alla incontinent celle Jehanne Serance de vie a trespas.
<LB N="48">Et craignant &ledit; Leserclier rigueur de justice, s'est
----Emacs: leserclier_sun.tei (B:SGML [tei.2])--L126--47X-----
Fontifying...done
```

```
emacs@SORNA
Buffers Files Tools Edit Search Jade SGML Moddy Move Markup View DTD Help
<PB N=2>
<LB N="11">en la compagnie de il et <NAME KEY="JS">Jehanne Serance</NAME>, <S\
ANA="victime-statut-social">lors sa femme</S>, <NAME KEY="JD" REG="Jehan Dur\
andiere">Jehan
<LB N="12">&durandiere;</NAME> et sa femme, mere &duit; <NAME KEY="PL">Leser\
clier</NAME>, et pareilleent
<LB N="13"><NAME KEY="AS">André Serance</NAME>, oncle paternel de &ladite; <NAM\
E KEY="JS">Jehanne Serance</NAME>, durant
<LB N="14">lequel soupper ilz firent assemblement bonne &chere; et <S ANA="cr\
im-moment">après
<LB N="15">iceluy soupper</S> et &que; iceulx <NAME KEY="JD">Durandiere</NAME\
>, sa femme, <NAME KEY="AS">André Serance</NAME>
<LB N="16">et autres s'en furent allez et retirez &chacun; en sa maison sans
<LB N="17">avoir noyee ne parolles de rigueur, arriva &duit; villaige de
<LB N="18"><RS TYPE="toponyme" ana="crie-lieu">Bonhardy</RS> près la maison &\
desdits; <NAME KEY="PL">Leserclier</NAME> et sa femme ung &nomme;
<LB N="19"><NAME KEY="MC">Mace Chassé</NAME> qu'il dist a &ladite; femme <NAM\
E KEY="PL">Leserclier</NAME> par celles
<LB N="20">parolles : <Q>Uoz beuffs sont en mes naveaux et y ont esté
<LB N="21">par troys foiz : vous n'en couvrirez voz bledz de troys jours</Q>.\
A quoy
<LB N="22">&ladite; <NAME KEY="JS">Jehanne Serance</NAME>, femme &duit; <NAM\
E KEY="PL">Leserclier</NAME>, luy respondit
<LB N="23">et dist &que; les luy eust renduz et que elle eust baillé
<LB N="24">a iceluy <NAME KEY="MC">Chassé</NAME> ung grant pain sans &que; &l\
edit; <NAME KEY="PL">Leserclier</NAME>, son
<LB N="25">mary en eust eu congnoissance, quelles parolles &ledit;
<LB N="26"><NAME KEY="PL">Leserclier</NAME> entendit et dist lors a &ladite; \
<NAME KEY="JS">Jehanne Serance</NAME>
<LB N="27">par seables parolles : <Q>Jehanne, prenon ses porcs qui sont
<LB N="28">a rompre noz pailliers et les sectons en parc</Q>, esperant celuy
<LB N="29"><NAME KEY="PL">Leserclier</NAME> &que; les porcs &qui; estoit pré\
s le paillier &duit;
<LB N="30"><NAME KEY="PL">Leserclier</NAME> apartenseissent &duit; <NAME KEY\
="MC">Chassé</NAME>, A quoy &ladite;
<LB N="31"><NAME KEY="JS">Jehanne</NAME> respondit : <Q>Ilz ne sont pas a lu\
y, ilz sont a nous</Q>.
<LB N="32">En l'endroit de quoy &ledit; <NAME KEY="PL">Leserclier</NAME> en c\
ourant &apres;
<LB N="33">&ledits; pourceaux, dist a icelle <NAME KEY="JS">Jehanne Serance<\
/NAME> sa femme :
<LB N="34"><Q>je vous avoys dit &que; &votre; <unclear>chambriere</unclear> e\
t vous, eussiez mené
----Emacs: leserclier.tei (B:SGML [tei.2])--L114--44X-----
Fontifying...done
```

# Application : premiers exercices

- Remarques :
  - Codage de certains caractères  
$$\& = \&amp; ;$$
  - Commentaire XML  
$$\<! - - - \> \dots \<- - - - \>$$
  - Le caractère @ désigne un attribut

# Typologie des structures de texte

- Manuscrits et imprimés
  - Prose et vers
- Encodages singuliers
  - Correspondance
  - Poésie
  - Théâtre
  - Dictionnaire

# Structurer avec Oxygen

# Exercice de manipulation d'oxygène avec XML/TEI intro-dial.doc

## Dial de Saffré (Archives départementales de Loire-Atlantique)

INITIATION A L'ENCODAGE XML-TEI DES  
TEXTES PATRIMONIAUX  
— LES MANUSCRITS —

NICOLE DUFOURNAUD CESR – JANVIER 2014

Les éléments proposés :

```
<div> <head> <p> <lb/> <sic> <abbr> <unclear> <foreign
xml:lang="lat"> <gap/>
```

Source : *ADLA à Nantes*

Transcription : *Nicole Dufournaud*

---

Dial de Saffré

Commencant le premier jour de janvier 1570

En charge à Yves Papot

Seconde semaine

Le ixme jour de janvier 1570 arriva

Louys Paquier de la Rochelle

apporta lettres et bonnes nouvelles

de Monsieur lequel partit le

vime de septembre 1568 pour

aller trouver messieurs les

princes en la compagnie de

monsieur [?] d'Andelot

6 1

# Exercice de manipulation de TEI intro-dial.doc

But de l'exercice : récupérer un texte saisi sous un éditeur graphique et le transformer en un fichier structuré avec XML/TEI

- Créer un nouveau fichier dans Oxygen
- Copier et coller le texte de word
- Encoder le texte en utilisant :
  - Éléments : `<body><div>` `<head>` `<lb/>` `<p>` `<sic>`  
`<gap/>` `<unclear>` `<del>` `<foreign>` `<abbr>`
  - Attributs : `@xml:lang`

# Correction

Le Dial de Saffré (Archives départementales de Loire-  
Atlantique)  
intro-dial\_correction.xml



# Exercice de manipulation de TEI

## intro\_jeanne-de-remy.doc

Procès en appel de Jeanne la devineresse (BM d'Orléans)

Utiliser les fichiers en format .png et .doc

Créer un nouveau fichier dans Oxygen

- Récupérer le texte saisi sous un éditeur graphique
- Éléments : <body><div> <pb/> <lb/> <p>  
<unclear> <del> <expan> <sic> <foreign>  
<gap/>
- Attributs : @n @xml:lang

# Correction

Procès en appel de Jeanne la devineresse (BM d'Orléans)  
intro\_jeanne-de-remy\_correction.xml

Structurer pour éditer sur le Web ?  
OUI mais pas seulement !

# Le but ultime de la structuration

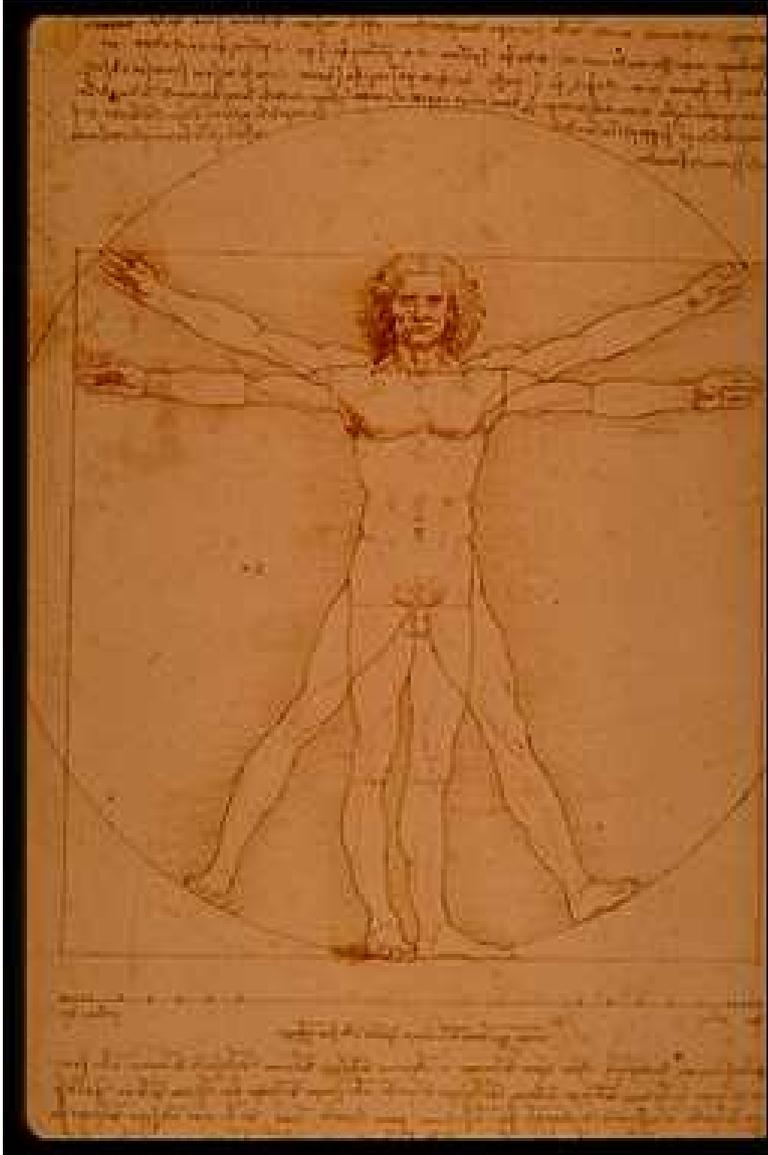
- Structurer pour structurer
  - Classer comme Buffon, le naturaliste
  - Pour une aide intellectuelle, une meilleure compréhension : structurer la pensée
- Pour éditer les textes
  - Comment éditer en ligne ? Comment mettre en ligne une édition critique, une édition savante
- Pour analyser les textes
  - Comment explorer les données et comment les exploiter ?

**Beaucoup d'éléments structurés et homogènes !**

**Une analyse possible à partir d'une vue globale !**

**Comment les parcourir sur tout le corpus, vite et bien ?**

# Visualisation



l'œil, « fenêtre de l'âme est la principale voie par laquelle notre intellect peut apprécier pleinement l'œuvre infinie de la nature »

Leonard de Vinci  
(1452-1519)

# Visualisation d'informations

<http://www.aviz.fr/>

- Lorsqu'on a des données et qu'on cherche des questions intéressantes, on doit **explorer**.
- **Principe de l'analyse exploratoire :**
  - utiliser nos capacités visuelles pour détecter des phénomènes inattendus
- Spécialité ancienne, la visualisation est une aide à l'analyse (voir les travaux de J. Bertin).
  - L'interactivité l'enrichit énormément.
- Inconvénient : difficile à utiliser sur des données non structurées.
- Objectifs : vue d'ensemble, construction d'hypothèses et analyses.

# Visualisation de données

- Grandes bases documentaires
- Données scientifiques massives (séquençage de génome...)
- Réseaux dont les généalogies
- Cartes
- Bases de données
- Etc.

# Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. MARTEL, Ingénieur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes prisent sont exprimés par les longueurs des lignes colorées à raison d'un millimètre pour dix mille hommes, et sont le plus écrit en traits des lignes. Le tracé indique les hommes qui restent en Russie, le noir ceux qui se retirent. Les arrangements qui ont été à travers la carte ont été prisés dans les ouvrages de M. de Clapiers, de Ligny, de Foy, de Choisey et le journal inédit de Napoléon, pharaon de l'Armée depuis le 25 Octobre. Pour avoir fait usage à l'égard de l'armée, j'ai supposé que le camp de Smolensk en de Marschal (Dawson) qui consistait de 200000 hommes en Russie et de 100000 en son retour vers Ostrowo, Wislitz, avaient toujours marché vers l'est.

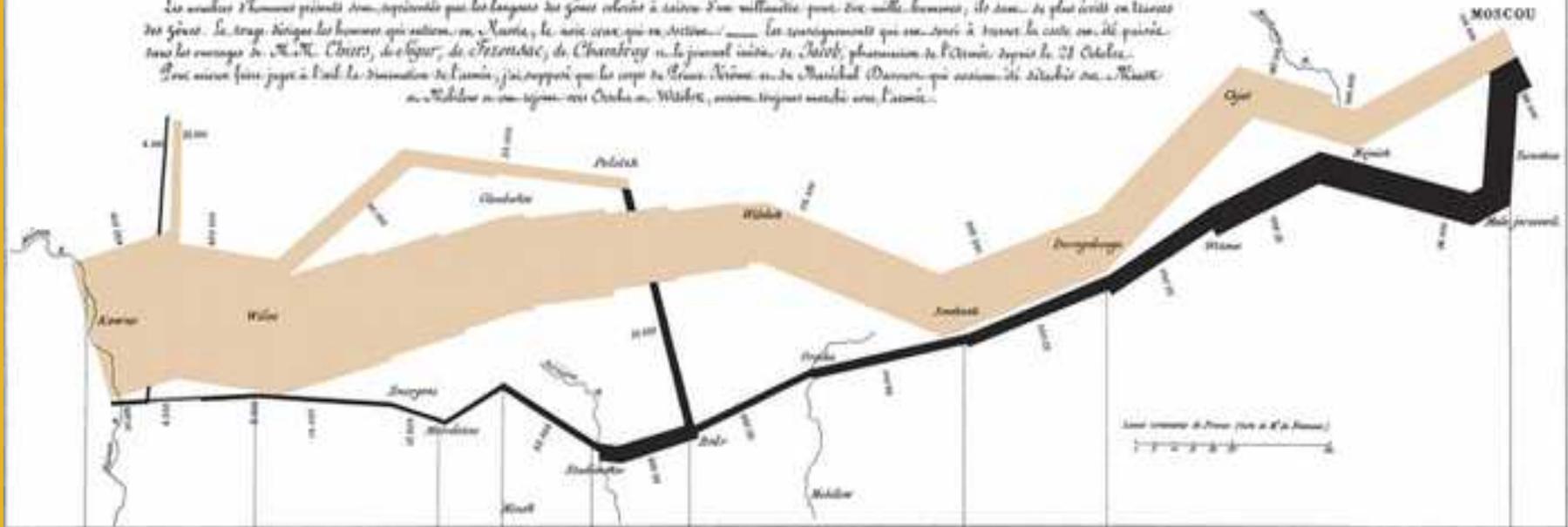


TABLEAU GRAPHIQUE de la température en degrés de thermomètre de Réaumur au dessous de zéro.



*A Gift to the World*  
 in the memory of  
 his gift to the world

Gifted by  
 His Majesty  
 King George  
 VI in 1944  
 to the  
 British  
 Library  
 as a  
 memorial  
 to the  
 author  
 of the  
 book  
 "The  
 World  
 as It  
 Was"  
 by  
 H. G. Wells

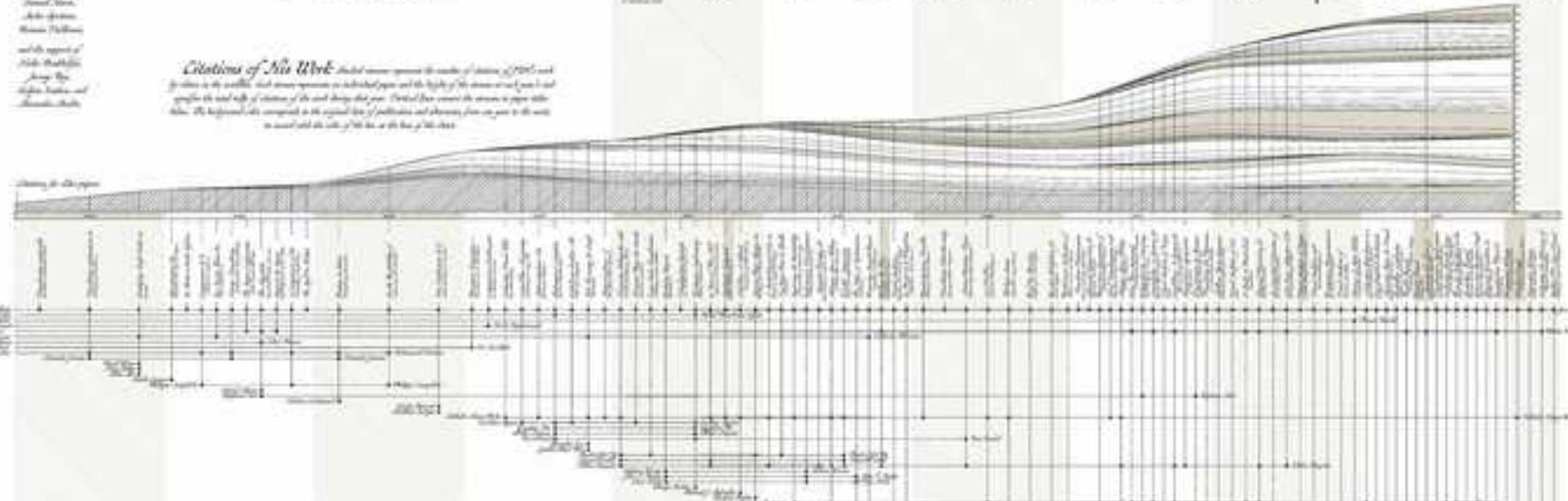
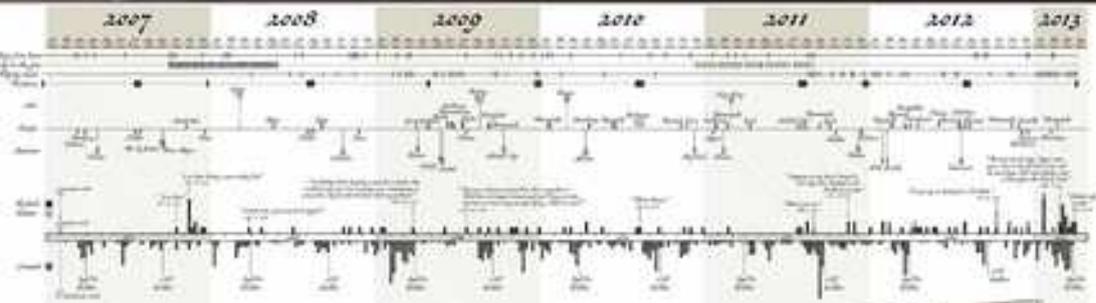
**Calendar** *Each year, the present year begins with a calendar page from the year 1900 and ends with a calendar page from the year 2013. The calendar pages are arranged in a grid with the year 1900 at the top and the year 2013 at the bottom.*

**Years Undertaken** *Each year is marked by a number which shows the number of years undertaken. The number of years undertaken is the number of years.*

**Employment of Social Media** *These represent the number of years in which the author has used social media. Facebook (represented by a square), Twitter (represented by a circle), LinkedIn (represented by a triangle), and YouTube (represented by a diamond) have all been used in the past and throughout the course of his career.*

**Contributions to Repertory** *These represent the number of repertory contributions made by the author. The number of contributions is the number of contributions.*

**Citations of His Work** *These represent the number of citations of his work in the world. The number of citations is the number of citations.*



*Chart of  
 Jean-Daniel Tékete's Biography*

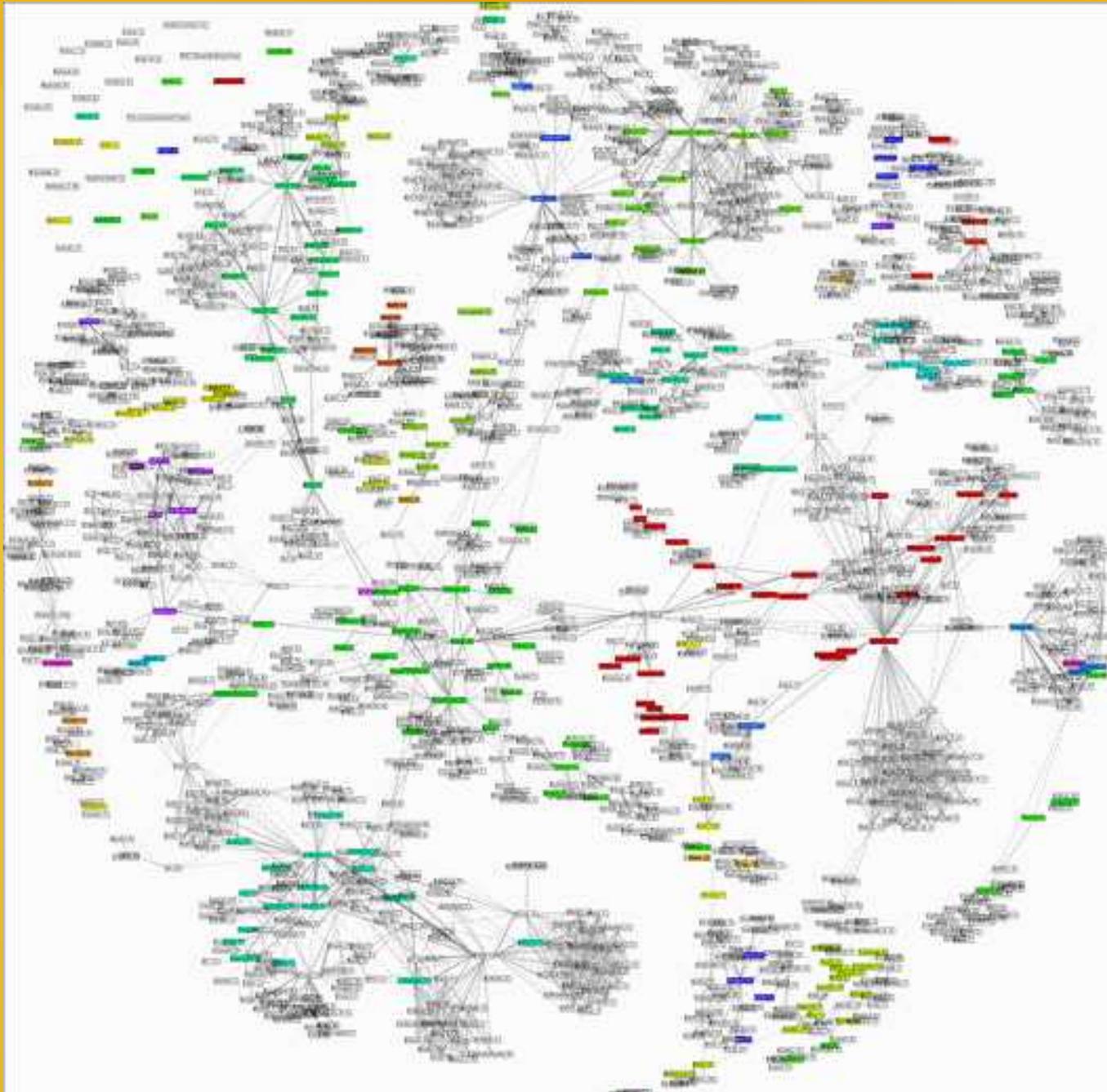
*From the Year 4 before the A.D. Era  
 to the present Year 2013.*

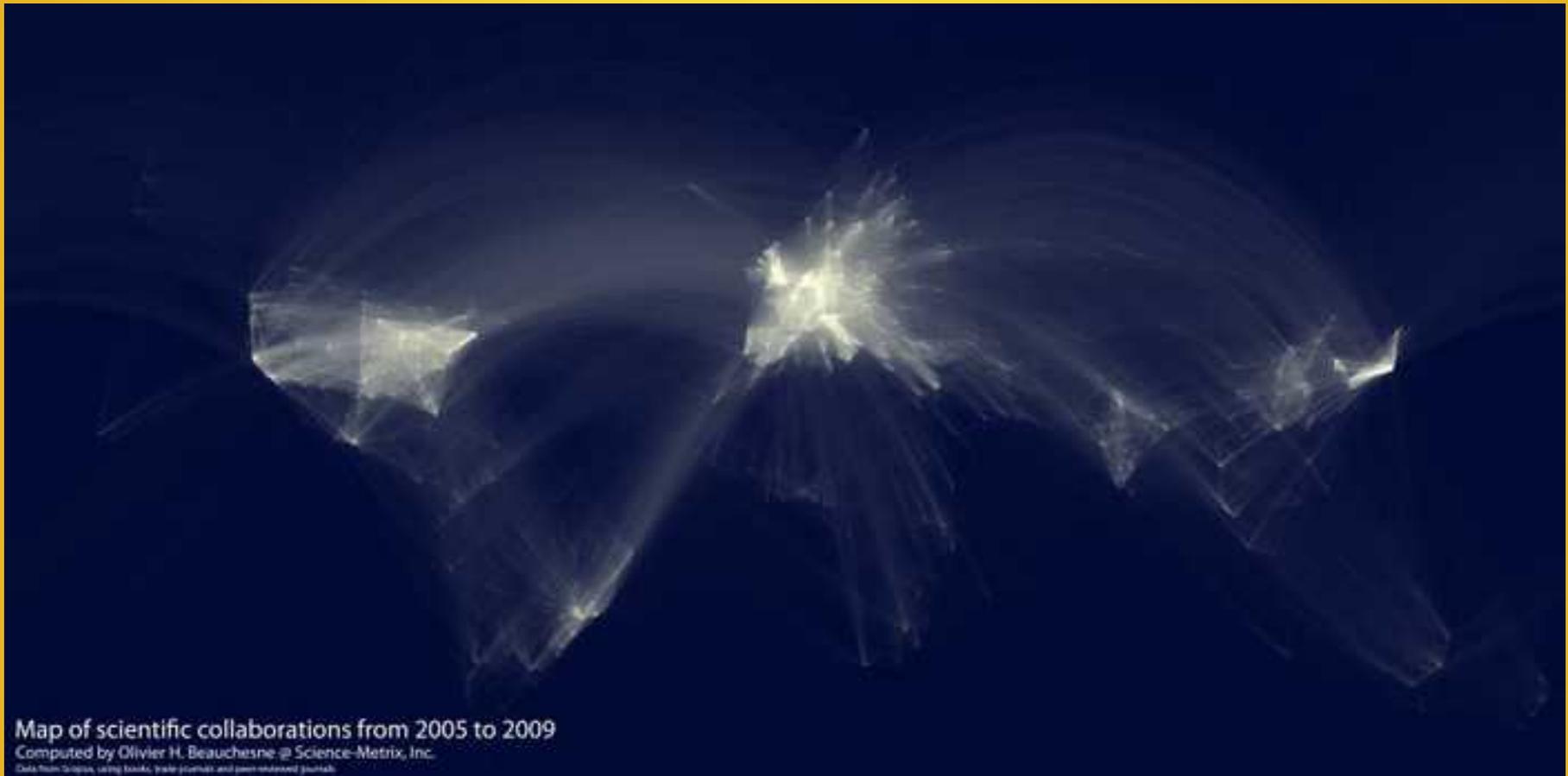
*Being a year of Two Thousand Two Hundred and Sixty*

**Collaborations & Published Works** *These represent the collaborations and published works of the author. The number of collaborations and published works is the number of collaborations and published works.*



# Co-publication INRIA Saclay (04-08)





<http://flowingdata.com/2011/01/27/map-of-scientific-collaboration-between-researchers/>

# Mapping America: Every City, Every Block

Find something interesting? Share this view on [Facebook](#) [Twitter](#) [LinkedIn](#) [Google+](#)

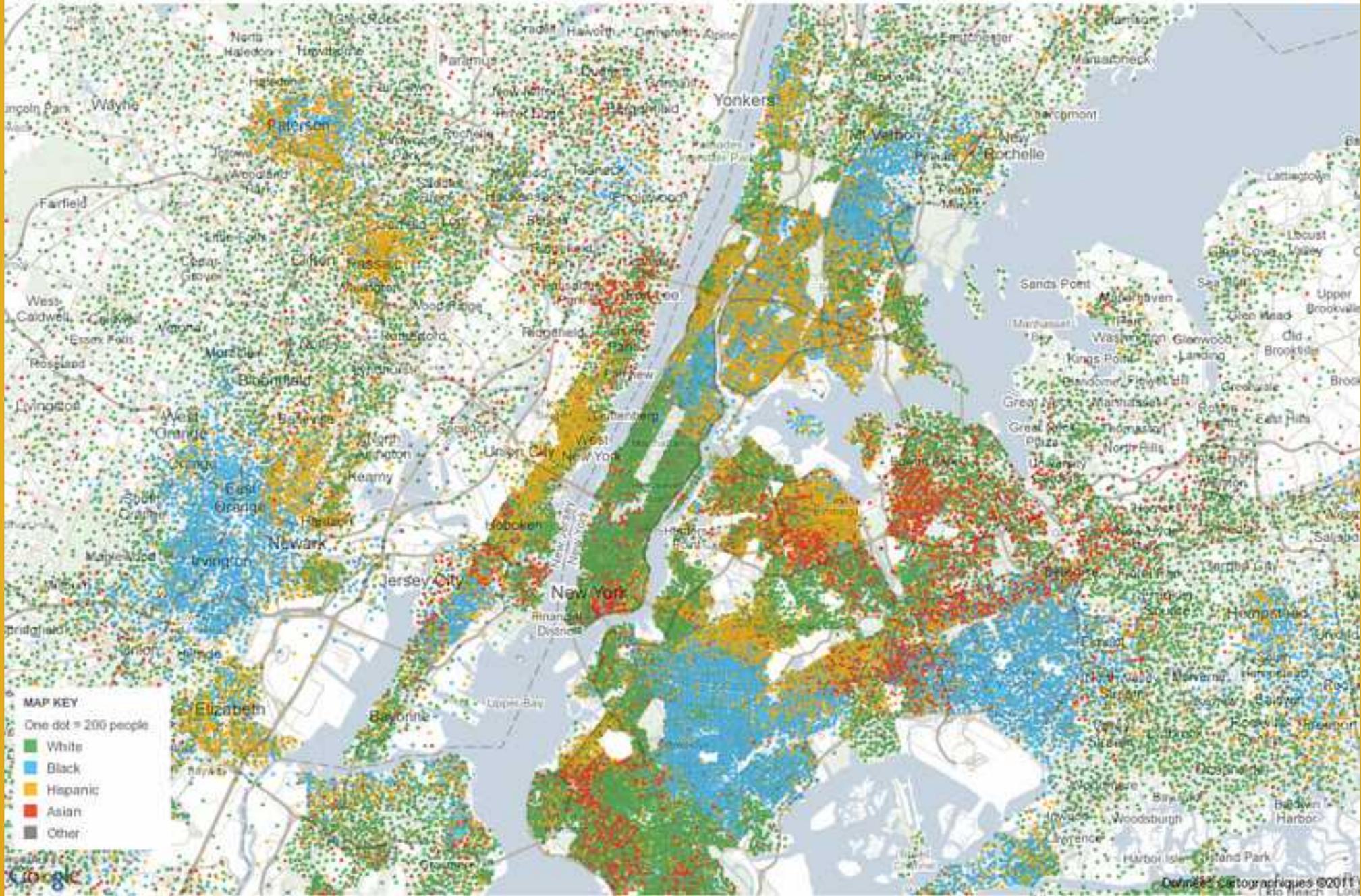
Browse local data from the Census Bureau's American Community Survey, based on samples from 2005 to 2009. Because these figures are based on samples, they are subject to a margin of error, particularly in places with a low population, and are best regarded as estimates.

## Distribution of racial and ethnic groups

[View More Maps](#)

Address, ZIP code or city

Go



## German defense minister's plagiarized PhD dissertation visualized

March 2, 2011 | Visualization | 13 comments | [Share on Twitter](#)



<http://flowingdata.com/2011/03/02/german-defense-ministers-plagiarized-phd-dissertation-visualized/>



# The Making of a Law

This visualization shows the version history of the German law on political parties. Since its publication in 1967 the law was changed several times. Sometimes entire sections had been re-written, sometimes only small parts had been corrected. Each column represents a version of the law, with the boxes representing the individual paragraphs. Changed, added and deleted articles are highlighted.



## Imprint / Source

The revisions are hand-scraped from [individual change-laws](#), as published in the [Bundesgesetzblatt](#).

Visualization & Scraping: [Gregor Aisch](#)

[German version](#)

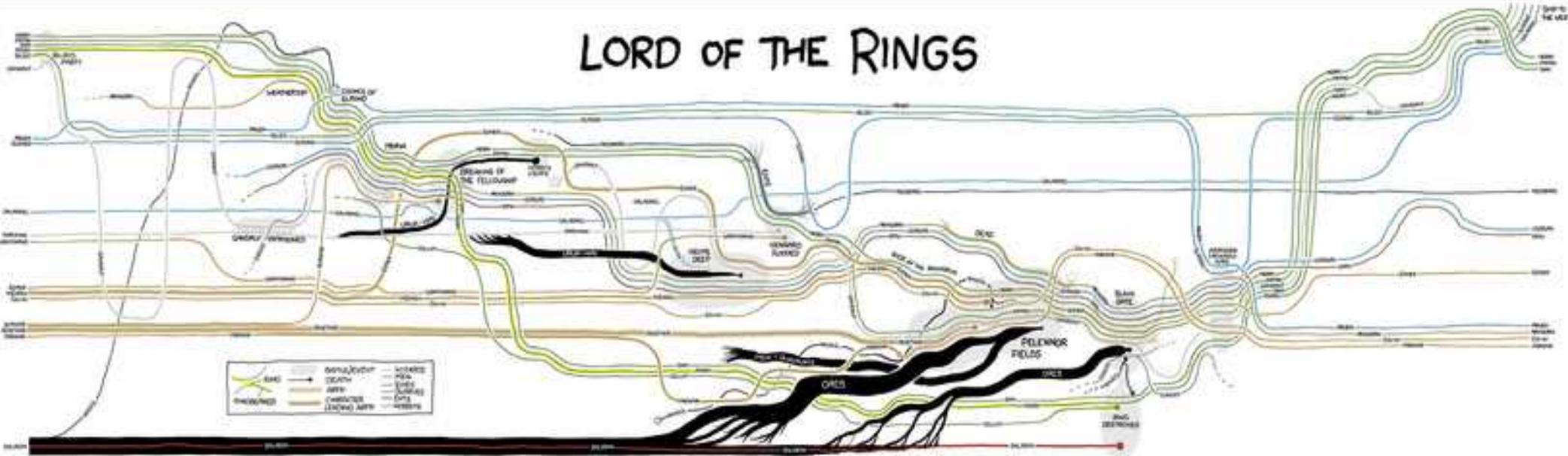
## License

You may use screenshots under terms of [CC-BY-SA](#) license.



THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS. /  
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE  
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.

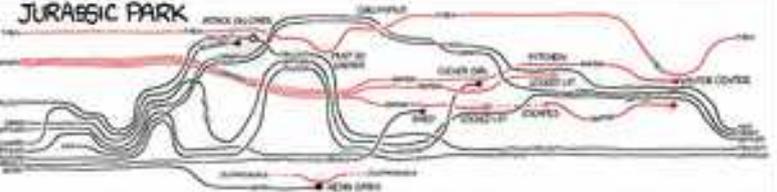
# LORD OF THE RINGS



# STAR WARS (ORIGINAL TRILOGY)



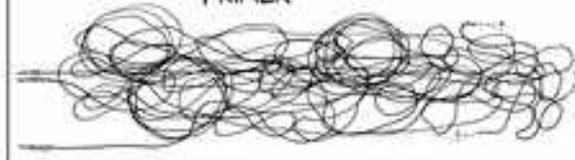
# JURASSIC PARK



# 12 ANGRY MEN



# PRIMER



# Visualisation et TEI

- La TEI est une norme qui structure du contenu d'un document ; elle ne permet pas de montrer graphiquement des analyses car la TEI ne s'intéresse qu'au caractère
- La Visualisation est une autre technologie issue de l'IHM
- Associer TEI avec la visualisation permet :
  - de faire une édition graphique des analyses structurelles d'un texte
  - de réexploiter une grande quantité de données pour un autre traitement

# Projet Millefeuille (2007)

The screenshot displays the 'Millefeuille' software interface. The main window shows a hierarchical tree structure of administrative units, organized into four vertical columns. Each unit is represented by a small icon and a text label. The units are color-coded by column: red, yellow, green, and blue. The right-hand side of the interface features a 'Detail' panel with a menu bar (Filtres, Visual, Liens Visual, Layout, Exemples, Filtrages, Param) and a search field. Below the search field, there is a list of 'Attributs' (Attributes) and a 'Filtres' (Filters) section. The 'Filtres' section is currently expanded, showing a list of administrative units and their associated attributes.

**Detail** Filtres Visual Liens Visual Layout Exemples Filtrages Param

310278  
Niveau: 07  
Niveau: 2  
Nom: 07185  
Attribution: La correspondance avec les Préfets de province administratifs mentionnés dans les  
textes  
Général de l'Etat et départements provinciaux

**Attributs**  
niveau: 2

**Filtres**

**Le budget du clergé**  
L'Institut  
Musées, écoles de peinture, sculpture, architecture et de décoration  
Ordonnancement des dépenses d'administration générale  
Ordonnancement des dépenses d'administration générale de des établissements de bienfaisance,  
Théâtres, fêtes et cérémonies publiques  
[Ordonnancement des dépenses] des cultes protestants  
[Ordonnancement des dépenses] des établissements de bienfaisance, des secours aux colons, de  
l'administration des cultes  
Examens et discussions des actes d'administration générale des préfets et des conseils de préfecture  
Examens et discussions des actes d'administration générale des préfets et des conseils de préfecture  
Grande et petite voirie  
L'administration des communes, des hospices des départements  
L'administration des communes, des hospices des départements, des actuels  
L'administration générale  
L'administration générale  
La Correspondance avec les Préfets  
Les institutions et mouvements des ingénieurs du Corps Royal des Ponts et Chaussées et du Corps  
Les institutions et mouvements des ingénieurs du Corps Royal des Ponts et Chaussées et du Corps  
Ordonnancement des dépenses d'administration générale  
Ordonnancement des dépenses d'administration générale de des établissements de bienfaisance,  
Ordonnancement des dépenses des Ponts et Chaussées et mines, et des travaux d'utilité générale  
Ordonnancement des dépenses des Ponts et Chaussées et mines, et des travaux d'utilité générale  
Ordonnancement des dépenses des cultes, des ponts et chaussées, des mines et des travaux de l'  
Ordonnancement des dépenses des ponts et chaussées, des mines et des travaux d'utilité générale  
Ordonnancement des dépenses des ponts et chaussées, des mines et des travaux de Paris

# Visualisations

<http://nicole.dufournaud.net/>

Nombre de lettres de grâce par année

Délais entre le crime et la rémission au 16<sup>e</sup> siècle en Bretagne

Compus et la visualisation XML

Réseaux sociaux

Arbres généalogiques

Carte des lieux mentionnés

Diagrammes de flux liés à la généalogie

# La ligne de temps : timeline

www.simile-widgets.org/timeline/examples/monet/monet.html

ant Message WebMail Radio People Yellow Pages Download Calendar Channels RealPlayer Home Page

ette page est en anglais Voulez-vous la traduire? Traduire Non Options

LE Widgets » Timeline » Exemples » The Life of Monet

## The Life of Monet

Sources:

- <http://en.wikipedia.org/wiki/Monet>
- <http://www.accents-n-art.com/artists/claude-monet-biography.html>

Timeline version 2.3.1 (with Ajax lib 2.2.1).



The timeline displays key events in Claude Monet's life from 1812 to 1926. A tooltip is shown for the event 'Monet's Garden at Argenteuil' on January 1, 1873. The tooltip contains the following text: 'Monet's Garden at Argenteuil', 'Some of Claude Monet's best known work was done in Argenteuil, including one of his famous garden scenes Monet's Garden at Argenteuil which he painted in 1873.', and 'Wed, 01 Jan 1873 00:00:00 GMT'. A small image of the painting 'Monet's Garden at Argenteuil' is also included in the tooltip. The timeline events include: Birth (1812), Family moved to Le Havre, Normandy (1825), Mother died (1826), Entered Le Havre secondary school (1831), Joined studio of Charles Gleyre in Paris (1859), Lived at Argenteuil, near Paris (1871-1878), Took refuge in England (1870), Married Doncleux (1874), Married Hoschedé (1875), Second son Michel born (1878), The Artist's Garden in Giverny (1890), Painted landscapes and seascapes (1890-1900), 'Series' painting (1890-1900), Water Lily Pond (1914), and Durand-Ruel gallery (1918).

Monet's Garden at Argenteuil

Some of Claude Monet's best known work was done in Argenteuil, including one of his famous garden scenes Monet's Garden at Argenteuil which he painted in 1873.

Wed, 01 Jan 1873 00:00:00 GMT

Birth Family moved to Le Havre, Normandy Mother died Entered Le Havre secondary school Joined studio of Charles Gleyre in Paris Lived at Argenteuil, near Paris Took refuge in England Married Doncleux Married Hoschedé Second son Michel born The Artist's Garden in Giverny Painted landscapes and seascapes 'Series' painting Water Lily Pond Durand-Ruel gallery

# Et pour conclure !



# Un travail pérenne !

- TEI est flexible, TEI s'adapte ! C'est un outil
  - qui permettra l'exploration d'une quantité de données en augmentation constante
- Structuration des données textuelles pour rendre explicite le sens du texte
  - Même s'il y a des limites
  - Même si on la remet en cause
  - Progrès fondamental ! Semble acquis !

# Des recommandations

- Les formats de structure :
  - Plusieurs formats possibles : EAD, Marc, DublinCore, EAC
    - Choix selon les besoins
  - Nouveauté à la mode : RDF
  - Réflexions actuelles :
    - Ontologies
    - Entités nommées : temporalité, identités et les lieux

# Des recommandations

- Les difficultés :
  - Les finances et les choix politiques
  - La gestion de projet
  - Les outils
  - Les compétences et le travail interdisciplinaire en collaboration

**BON COURAGE !**