



**HAL**  
open science

## **TEI, un support méthodologique pour l'encodage du patrimoine textuel. Données, structuration et visualisation.**

Nicole Dufournaud

### ► To cite this version:

Nicole Dufournaud. TEI, un support méthodologique pour l'encodage du patrimoine textuel. Données, structuration et visualisation.. Le patrimoine à l'ère du numérique : structuration et balisage Colloque international (Université-MRSH de Caen), Dec 2009, Caen, France. halshs-00512530

**HAL Id: halshs-00512530**

**<https://shs.hal.science/halshs-00512530>**

Submitted on 30 Aug 2010

**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.

« TEI, un support méthodologique pour l'encodage du patrimoine textuel.  
Données, structuration et visualisation. »

Nicole Dufournaud

*Le patrimoine à l'ère du numérique : structuration et balisage*  
*Colloque international (Université-MRSH de Caen) 10 et 11 décembre 2009*

## Résumé en français

Traditionnellement, le travail de dépouillement historique de manuscrits est fastidieux et non réutilisable : un historien paléographe transcrit son corpus, l'analyse et produit une étude qui est la forme finale du travail. Rarement, les transcriptions sont jointes en annexe à l'analyse. Dans le cadre de mes travaux universitaires sur les rôles et les pouvoirs des femmes au XVI<sup>e</sup> siècle dans la France de l'Ouest, j'ai choisi de reconstituer des itinéraires personnels dont le sens profond n'apparaît que lorsqu'on les ajoute les uns aux autres, même si, pour certains, j'avais peu de données. Grâce à l'informatique, j'ai utilisé une méthode qui crée des liens entre les sources et leurs interprétations : l'avantage est la traçabilité et son inconvénient majeur une certaine austérité à l'analyse qui peut être compensé par le recours à la visualisation. Cette méthode, rigoureuse, repose sur XML/TEI (Text Encoding Initiative), une grammaire de description des fichiers XML. Le travail préparatoire est long et fastidieux, mais payant : le résultat augmente le pouvoir d'analyse. Pour l'historien, cette méthode est une véritable aide à établir des interprétations qui se vérifient par les textes de sources primaires.

## Résumé en anglais

Traditionally, exploiting historical manuscripts is tedious and non-reusable. A paleographer-historian transcribes a corpus, analyses it and writes an article that becomes the ultimate form of the work; the transcriptions being seldom provided in appendix. In the context of my work on the role and power of women in the 16th century in Western France, I have chosen to reconstruct personal life itineraries which reveal a deeper meaning once they are aggregated; even when, for some itineraries, I had very few data. By using a computer supported method, I have been able to link the transcribed source and its interpretations, offering a simplified tracking of inferences at the cost of a more austere analysis that can be improved by using visualization. This rigorous method relies on the XML/TEI (Text Encoding Initiative) file format, a XML document description grammar. It requires a long and tedious preparatory work but it ends-up being rewarding: it results in an increase of analysis capabilities. This method is an effective support for historians, helping them to establish interpretations of sources that can be checked easily and directly.

## Mots-clés

histoire, humanités numériques, manuscrits, corpus historique, transcription, XML/TEI, analyse de textes historiques, visualisation, outils

## Keywords

history, digital humanities, manuscripts, historical corpus, transcription, XML/TEI, historical text analysis, visualization, tools

## Introduction

Une des tâches de l'historien consiste à dépouiller les archives. Dans les années 1990, l'historien paléographe transcrivait partiellement son corpus, l'analysait grâce à une grille de dépouillement, produisait son étude et pouvait archiver définitivement la quantité de données papiers non réutilisables.

Cet article présente l'expérimentation informatique que j'ai menée pour mes travaux universitaires dans ses aspects méthodologiques et pluridisciplinaires : histoire et informatique. De 1998 à 2007, de la maîtrise à la thèse, j'ai travaillé avec Jean-Daniel Fekete, chercheur en informatique à l'Institut national de recherche en informatique et automatique (INRIA)<sup>1</sup>, qui m'a conseillée, guidée et fourni les instruments innovants dont j'avais besoin. Grâce à cette coopération, ce travail a été rendu possible.

Le but de cette expérimentation était de trouver une nouvelle méthode pour aider les historien(ne)s à établir des interprétations qui se vérifient par les textes de sources primaires. Le travail initial consistait à encoder et à structurer ces documents pour autoriser des traitements informatiques permettant la navigation et l'extraction des données. À partir de sources manuscrites du XVI<sup>e</sup> siècle que j'ai transcrites exhaustivement, j'ai élaboré une méthode qui repose sur XML/TEI (Text Encoding Initiative) : une grammaire de description des fichiers XML. Enfin, de nouveaux outils sont apparus que j'ai utilisés pour encoder puis pour indexer : une analyse informatisée des textes en a été facilitée : je présenterai quelques logiciels. Enfin, je montrerai des exemples de visualisation de certaines données qui a facilité l'exploration visuelle. Les résultats de mes travaux sont visibles sur mon site personnel<sup>2</sup>.

## De la maîtrise à la thèse

En 1998, mon travail de maîtrise en histoire à l'université de Nantes portait sur les « Comportements et relations sociales en Bretagne vers 1530, d'après les lettres de pardon »<sup>3</sup>. Ces lettres étaient recopiées dans des registres de la Chancellerie de Bretagne pour être enregistrées. Les registres sont en bon état de conservation aux Archives départementales de la Loire-Atlantique à Nantes. Leur examen permet de comprendre les structures physique et logique, de la fabrication du registre à son contenu. Les actes sont divers : actes d'évocation, de sauvegarde, des mandements, des placets, des lettres d'office, des lettres de rémission, de commutation de peine et d'abolition, etc. Comme le montre l'illustration de la première page du registre de 1529-1530<sup>4</sup>, la structure est la suivante (1) :

- le titre du registre :

*Registre des mandemens sellez en la chancellerie de ce pays et duche de bretagne comance a Rennes le cinqiesme jour de janvier lan mil cinq cens vingt neuf*

- le titre de la session :

*Devant monseigneur le president le cinqiesme jour de janvier mil v c xxix presents*

1 En 1998, Jean-Daniel Fekete était à l'École des mines de Nantes ; depuis, il est devenu le directeur de l'équipe AVIZ à l'INRIA. Voir les sites <http://www.inria.fr/> et <http://www.aviz.fr/>

2 Les mémoires de maîtrise et de DEA ainsi que la thèse sont en ligne avec l'ensemble des sources sur <http://nicole.dufournaud.net/> Les publications sont dans la bibliographie choisie, en annexe.

3 Nicole Dufournaud, *Comportements et relations sociales en Bretagne vers 1530, d'après les lettres de grâce royale*, mémoire de maîtrise d'histoire, Université de Nantes, 1999, 2 vol., (dactyl.).

4 Les registres de la Chancellerie sont en ligne sur le site des Archives départementales de la Loire-Atlantique. Pour le registre B32, voir [http://www.archinoe.net/cg44v2/visu\\_affiche.php?PHPSID=61a065f90820778d8a682298fd55a55e&param=visu&page=1](http://www.archinoe.net/cg44v2/visu_affiche.php?PHPSID=61a065f90820778d8a682298fd55a55e&param=visu&page=1)

messieurs de Kerhurel le procureur general et le lieutenant de nantes

- les trois actes enregistrés du jour de la session : ici, trois mandements

Plusieurs signatures en marges de gauche et de droite ainsi que des ratures et le tampon des Archives départementales de la Loire-Atlantique complètent cette description structurale.

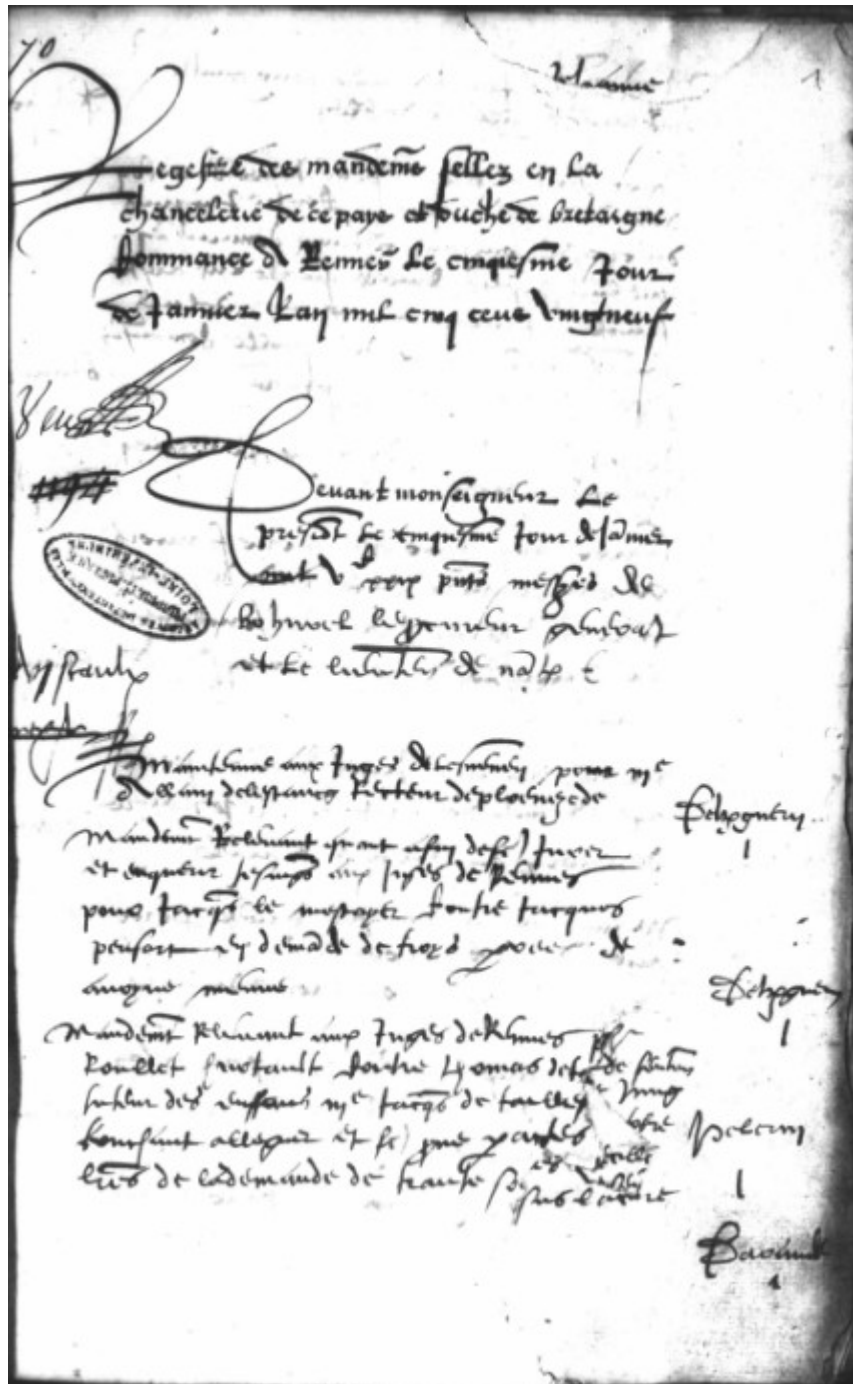


Illustration 1: Première page du registre B32, ADLA Nantes

Ces registres comprennent entre autres les lettres de pardon enregistrées par les greffiers : elles étaient souvent recopiées dans leur totalité. Les lettres de pardon sont aussi appelées lettres de grâce ; les lettres de rémission en forment un sous-ensemble.

L'historienne Claude Gauvard a étudié les lettres de rémission sous Charles VI. Elle en donne la définition suivante :

« ...La lettre de rémission est un acte de la Chancellerie par lequel le roi octroie son pardon à la suite d'un crime ou d'un délit, arrêtant ainsi le cours ordinaire de la justice, qu'elle soit royale, seigneuriale, urbaine ou ecclésiastique... »<sup>5</sup>

Mon corpus était composé de lettres de pardon comprenant des lettres de rémission, des abolitions et des commutations de peine sous François Ier<sup>6</sup>. L'illustration suivante est une reproduction du début d'une de ces lettres (2).

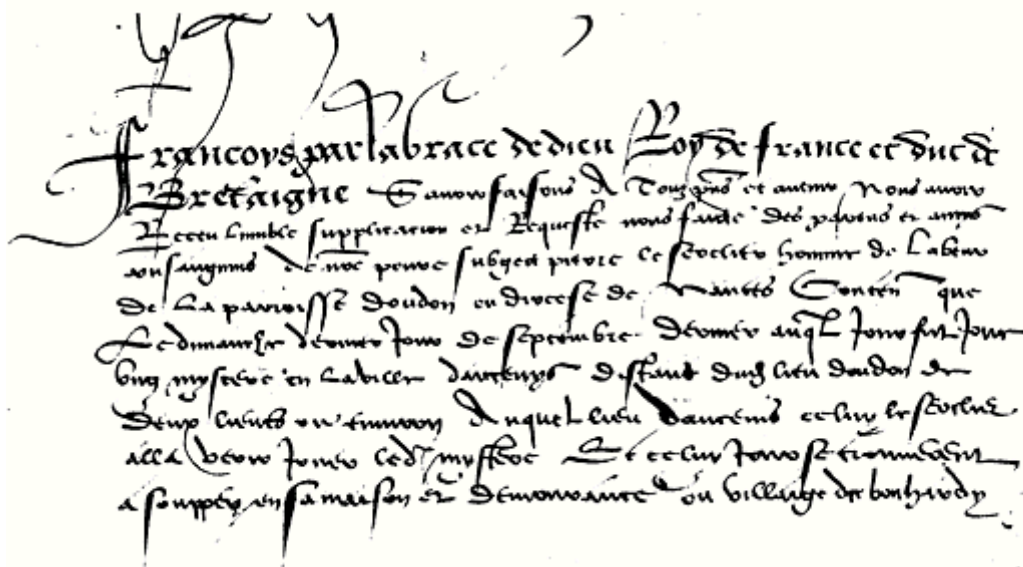


Illustration 2: Lettre de rémission de Pierre Leserlier, registre B25, ADLA Nantes

Source difficile à lire, à comprendre et à analyser, les lettres de pardon permettent des études qualitatives et quantitatives.

A partir de la transcription des lettres de pardon, mon intuition était d'utiliser des outils informatiques pour les tâches suivantes :

- encoder la transcription exhaustive des lettres ;
- diffuser les sources transcrites ;
- dépouiller les lettres de façon analytique c'est-à-dire qualifier des segments de textes à des fins analytiques ;
- et analyser les données informatiquement pour une exploitation automatique des résultats.

Il s'agissait alors de structurer les données.

Dans le langage informatique, que signifie la structuration ? C'est convertir des données – ici 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. Un exemple bien connu d'un format compris par la machine est HTML, utilisé par tous les navigateurs Web.

La recherche d'un format de codage de texte dit « structuré » comme support a abouti à choisir TEI

<sup>5</sup> Claude Gauvard, « *De grace especial* » *Crime, Etat et société en France à la fin du Moyen Age*, Paris, Publications de la Sorbonne, 1991, 2 volumes.

<sup>6</sup> Avec le dépouillement des lettres lors de mon DEA, mon corpus comprend 150 lettres de pardon.

(Text Encoding Initiative)<sup>7</sup>. TEI est une convention commune pour diffuser et échanger des textes qui repose sur un langage compréhensible par l'ordinateur en vue de traitements informatiques<sup>8</sup>. En 1998, le langage était SGML ; aujourd'hui TEI repose sur XML.

TEI semblait idéal pour rendre la transcription exhaustive exploitable, analysable et réutilisable. Utiliser TEI permettait de garder l'aspect diplomatique des lettres à l'inverse des bases de données qui obligeaient à « atomiser » les données : le résultat devait permettre une publication simplifiée lors de la mise en ligne. Au fur et à mesure de l'expérimentation, la mise en relation des données entre l'analyse et les éléments textuels est apparue comme un bonne méthode de travail. Enfin le travail final pouvait être vérifié grâce à la liaison entre les analyses et les sources.

Après quelques tâtonnements, il est apparu que la méthode était composée de trois étapes :

1. la saisie de la transcription exhaustive avec un encodage typographique et formel en TEI (SGML/TEI version P3) ;
2. l'enrichissement de l'encodage en TEI pour les noms de personne, les noms de lieux et les dates mais aussi pour des typologies reposant sur des catégories pré-déterminées comme les armes, les vêtements, etc ;
3. l'encodage analytique en TEI de la lettre de pardon en lien avec un autre fichier où était encodée une grille de dépouillement.

Je détaille ci-après chaque étape.

### ***Encodage typographique et formel***

Pour structurer, il faut baliser ! Pour baliser, il faut utiliser des éléments et des attributs. L'élément est la structure de base marquée par une balise de début et une balise de fin. L'attribut est un paramètre associé à un élément.

La première étape consiste à établir une transcription diplomatique en prenant note de tous les éléments supplémentaires au texte comme les erreurs manifestes, les abréviations, les ratures, les ajouts, etc.

Voici quelques exemples de codes :

<abbr>	Abréviation
<add>	Ajout
<del>	Rature
<lb>	Changement de ligne
<note>	Note
<pb>	Changement de page
<sic>	Erreur
<unclear>	Mot incertain

Et quelques exemples plus explicites :

<lb n="36">	Pour un changement de ligne avec son numéro.
-------------	--

<sup>7</sup> Voir le site <http://www.tei-c.org/>

<sup>8</sup> TEI avait fait l'objet d'une expérimentation en France dans le cadre du projet Emma auquel j'avais participé (1996-1997).

<code>&lt;del&gt;present&lt;/del&gt;</code>	Pour une rature dans le texte
---	-------------------------------

Ce premier travail proche d'une analyse lexicale est de loin le plus fastidieux et reste laborieux ; il doit être fait obligatoirement devant la source.

### **Encodage sémantique**

La seconde étape permet une analyse mot à mot afin de mettre en évidence les dates et les noms et établir des index. S'y ajoutent les thèmes d'étude choisis.

Voici quelques exemples de codes :

<code>&lt;date&gt;</code>	Dates
<code>&lt;name&gt;</code>	Noms
<code>&lt;signed&gt;</code>	Signatures
<code>&lt;rs&gt;</code>	Pour la typologie

Et quelques exemples plus explicites :

<code>&lt;date value="1572-08-24"&gt;</code>	Pour régulariser une date
<code>&lt;name reg="François Ier"&gt;</code>	Pour régulariser un nom propre

Une typologie a été dressée pour les lieux<sup>9</sup>, les armes et d'autres thèmes en utilisant l'élément TEI `<rs>` pour *referecing string*.

<code>&lt;rs type="toponyme"&gt;</code>	Pour mettre en évidence les noms de lieux.
<code>&lt;rs type="arme"&gt;</code>	Pour repérer les armes dans le texte.

Les noms du XVI<sup>e</sup> siècle sans orthographe ou dans leur forme de l'époque pouvaient être régularisés grâce à l'attribut `@reg` :

```
<rs type="toponyme" reg="Vannes">Vennes</rs>
```

Ce second balisage que l'on peut appeler « sémantique » permet à l'historien de découvrir toute la signification du texte. C'est un début d'enrichissement du texte.

### **Encodage analytique**

TEI propose un mécanisme permettant de décrire une grille de dépouillement. Notre grille est inspirée de celle établie par l'historienne Claude Gauvard<sup>10</sup>.

La difficulté de manipulation de la grille, très fréquemment relatée par les historiens, a été levée instantanément : dans un premier temps, nous avons modifié la grille, puis nous l'avons fait évoluer peu à peu tout au long de l'analyse en abandonnant progressivement la grille d'analyse de Claude Gauvard ; nous en avons gardé les classifications suivantes :

- la diplomatique,

9 En TEI version P3, les éléments comme `<geogName>` et `<placeName>` n'étaient pas encore créés.

10 Claude Gauvard, « *De Grace Especial* » Crime..., tome 1, pp. 76 et suiv.

- la déclinaison d'identité du criminel et de la victime, ainsi que leurs liens,
- le sort du coupable après le crime,
- les participants du crime,
- le déroulement du crime,
- l'octroi de la rémission.

Elle se présente sous la forme de catégories d'interprétation dont voici un exemple :

```
<interpGrp type="sexe">
  <interp id="sexe-masculin"/>
  <interp id="sexe-feminin"/>
</interpGrp>
```

Dans le fichier, l'annotation était encodée grâce à l'attribut @ana :

```
<name ana="criminel-sexe-masculin">Jehan Basset</name>
<rs ana="sortcriminel-condamnation-prison">prisonnier</rs>
```

Cette grille pouvait être enrichie au fur et à mesure de l'évolution du travail de transcription et de l'évolution de la réflexion. Les tables d'analyses qui en ont découlé sont devenus de véritables instruments de travail comme le montre l'exemple suivant :

#### **Interprétation**

Intercesseur (Personnage)

[parens et amys](#)

Victime (Personnage)

[Jehanne Serance](#)

Relation criminel/intercesseur (Personnage)

[consanguins de notre povre subject](#)

Nom du criminel (Personnage)

[Pierre Leserclier](#)

Statut social du criminel (Statut)

[homme de labeur](#)

Statut social de la victime (Statut)

[lors sa femme](#)

Moment du crime (Date)

[après iceluy soupper](#)

Motif du crime (Motif)

[parolles desplaisant](#)

Lieu du crime (Lieu)

[Bonhardy](#)

Circonstance atténuante (Circonstance)

[qui est chargé de petiz enfans orphelins avoit auparavant bien et honnestement vescu et entretenu icelle sadite femme et enfans sans jamais avoir esté accusé ni convancu dautre cas reprouchable](#)

Date du crime (Date)

[dimanche, dernier jour de septembre dernier](#)



Date du jugement (Date)  
[moys d'octobre l'an mil cinq cens vingt](#)  
 Type du crime (Crime)  
[de vie a trespas](#)  
 Élément volé (Objet)  
 None

Cette table est devenue un instrument de contrôle de l'encodage. Par exemple, si la victime n'apparaît pas, il y a forcément une omission ou une erreur dans l'encodage de la lettre.

## **Glossaire externalisé**

Pour finir cette première expérimentation, j'ai dressé un glossaire que j'ai externalisé du fichier afin de ne pas polluer le texte. Seul l'élément TEI <index> avec son attribut @level permettait le lien entre les deux fichiers.

- Dans le fichier TEI :

```
<lb n="19">et ce pour raison de labourer et recrue<index
level1="recrue"> que faisait...
```

- Dans le fichier glossaire :

```
<div type="glossaire">
  <list type="gloss">
    <label id="abrevier"><term>Abrévier</term></label>
    <item><gloss>abréger, raccourcir, résumer.</gloss></item>...
    <label id="recrue"><term>Recrue</term></label>
    <item><gloss>aménagement d'une parcelle cultivée conquise sur un
terrain inculte</gloss></item>
...</list>
</div>
```

A la mise en ligne du texte, un astérisque de couleur guide l'internaute : il clique sur le symbole qui lui permet d'atteindre le glossaire dans une autre fenêtre HTML.

## **Les outils et les traitements informatiques**

Mes connaissances de l'éditeur Emacs et de certains outils informatiques m'ont permis d'utiliser directement un environnement informatique presque identique à celui utilisé par des gros projets informatiques. Se sont ajoutés les outils mis au point par Jean-Daniel Fekete. Il a mis au point un traducteur avec Perl pour mettre en page les données TEI, et pour lier automatiquement les lettres à leurs images ; puis il a créé une feuille de style en XSL, devenu depuis XSLT ; enfin il a traduit les fichiers SGML/TEI version P3 en XML/TEI version P4. Pour les traitements, il a créé également un testeur d'intégrité sémantique et un indexeur, un générateur de table et des filtres pour retirer l'encodage pour une utilisation ultérieure.

Le travail d'encodage qui suivait les conventions du Consortium a permis une automatisation des traitements informatiques. Il ne s'agissait plus, si je peux m'exprimer ainsi, que de créer les outils au fur et à mesure des nouveaux besoins provenant des résultats innovants de la méthode.

## **Les premiers résultats**

En 1999, les premiers résultats ont été très satisfaisants et prometteurs. La production automatique de tables d'analyses sous un tableur a permis qu'elles deviennent des instruments de travail :

- les tables se sont transformés spontanément en études statistiques ;
- la production de multiples index croisés ont permis de faciliter le travail d'écriture, de tester l'intégrité de l'encodage et la visualisation des données.

A la fin de la maîtrise, l'expérimentation s'est avérée positive. TEI permettait:

- de garder une source unique de la transcription jusqu'à l'analyse, l'interprétation et la publication ;
- la réutilisation d'un corpus déjà encodé.

Il est apparu que les traitements informatiques autorisaient la vérification ou falsification de mes hypothèses. La démarche devient réfutable, critère de scientificité selon Karl Popper<sup>11</sup>.

J'ai continué la même expérimentation pour le DEA<sup>12</sup>.

## **Thèse**

L'expérimentation s'étant avérée concluante, j'ai utilisé la même méthode pour ma thèse qui portait sur les « Rôles et pouvoirs des femmes au XVIe siècle dans la France de l'Ouest ». Sujet à controverse, son étude posait d'autres problèmes que la première expérimentation.

Mon corpus était composé de sources disparates : minutes notariales<sup>13</sup>, titres de famille, registres paroissiaux, visites pastorales, etc. Les difficultés principales étaient les suivantes : elles étaient difficiles à lire et à comprendre ; elles ne permettaient pas une étude quantitative par manque de données sérielles ; elles avaient une structure différentes les unes des autres. Le nombre élevé de textes ajoutait à la complexité : 1000 textes à saisir, à encoder et analyser. Enfin, le sujet méritait une refonte complète de la grille de dépouillement.

Ma volonté était de bénéficier encore plus de l'aide de l'informatique. Cette nouvelle étape a abouti à une mise en relation directe entre le texte de la thèse et les références historiques ainsi qu'à la visualisation exploratoire de données et l'extraction de données : de nombreux arbres généalogiques complexes, un réseau social et des cartes. Ma collaboration avec Jean-Daniel Fekete est ici maintenant présentée.

Plusieurs obstacles sont apparus. Ce sont tout d'abord les arbres généalogiques qui ont posé problème. Les logiciels de généalogie, créés par les informaticiens du XXe siècle, ne permettent pas de présenter les renchaînements d'alliance très nombreux au XVIe siècle ni les alliances hors normes comme celle d'une veuve se remariant avec son gendre veuf. Pour afficher ces généalogies, j'ai dû de nouveau utiliser des outils réservés aux informaticiens avec des commandes informatiques peu attrayantes et des programmes faits sur mesure. L'extraction des généalogies en format image a été une aide pour l'analyse des textes ; puis j'ai inséré les images dans ma thèse.

Une seconde difficulté était de dresser les index de noms de personnes et de lieux. Devant la complexité, j'ai écarté avec regret l'index automatique des noms de lieux et de terre.

<sup>11</sup> Karl Popper, *La logique de la découverte scientifique*, Paris, Payot, 1973.

<sup>12</sup> Cette première expérimentation a fait l'objet de trois publications (voir la bibliographie) et les résultats sont mis en ligne sur mon site personnel <http://nicole.dufournaud.net/remission/>

<sup>13</sup> Une minute notariale est l'acte original notarié, signé par les parties en présence ainsi que par les notaires ; la minute est conservée chez le notaire à la différence de la « grosse » qui est la copie remise aux parties.

Enfin, l'interprétation des textes a été réalisée avec un nouveau logiciel apparu vers 2003 appelé « Infozoom »<sup>14</sup>. Le logiciel permet de naviguer dans les 1000 textes, en particulier grâce aux différents groupes et analyses issues automatiquement de la grille dépouillement. Un simple clic permet d'afficher en HTML la source choisie et de récupérer la partie textuelle à insérer dans le mémoire de thèse.

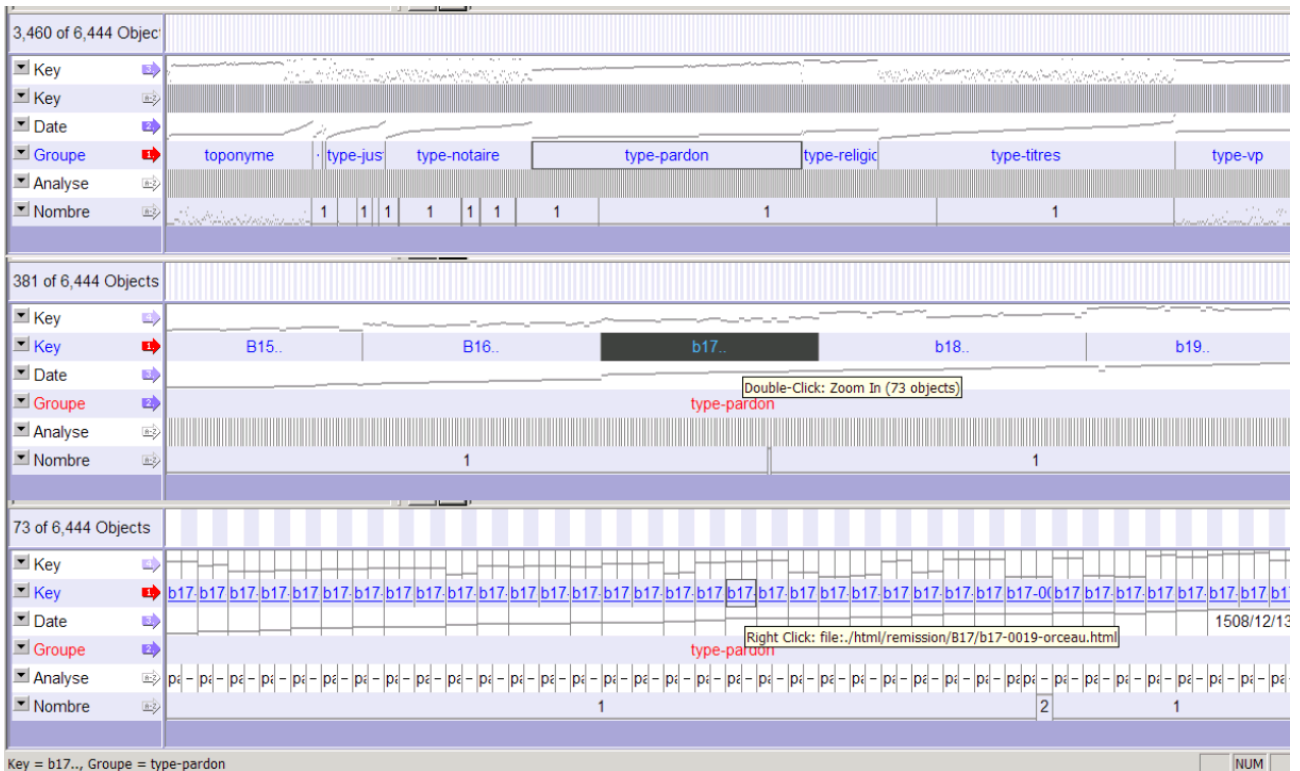


Illustration 1: Exemple d'une extraction de données avec Infozoom

Ici l'exploration de données avec Infozoom est conjuguée à l'extraction de données. Nous entrons dans le domaine de la Visualisation.

## Visualisation<sup>15</sup>

La Visualisation est une science ancienne. Au XIXe siècle, la carte figurative de M. Minard sur la campagne de Russie de Napoléon est riche de données ; elle donne visuellement des informations qui n'ont pas besoin d'être écrites pour être comprises : la taille de l'armée, sa progression, ses batailles, etc.

14 Voir le site <http://www.infozoom.com/>

15 Voir la notice Wikipedia en anglais [http://en.wikipedia.org/wiki/Visualization\\_\(computer\\_graphics\)](http://en.wikipedia.org/wiki/Visualization_(computer_graphics))

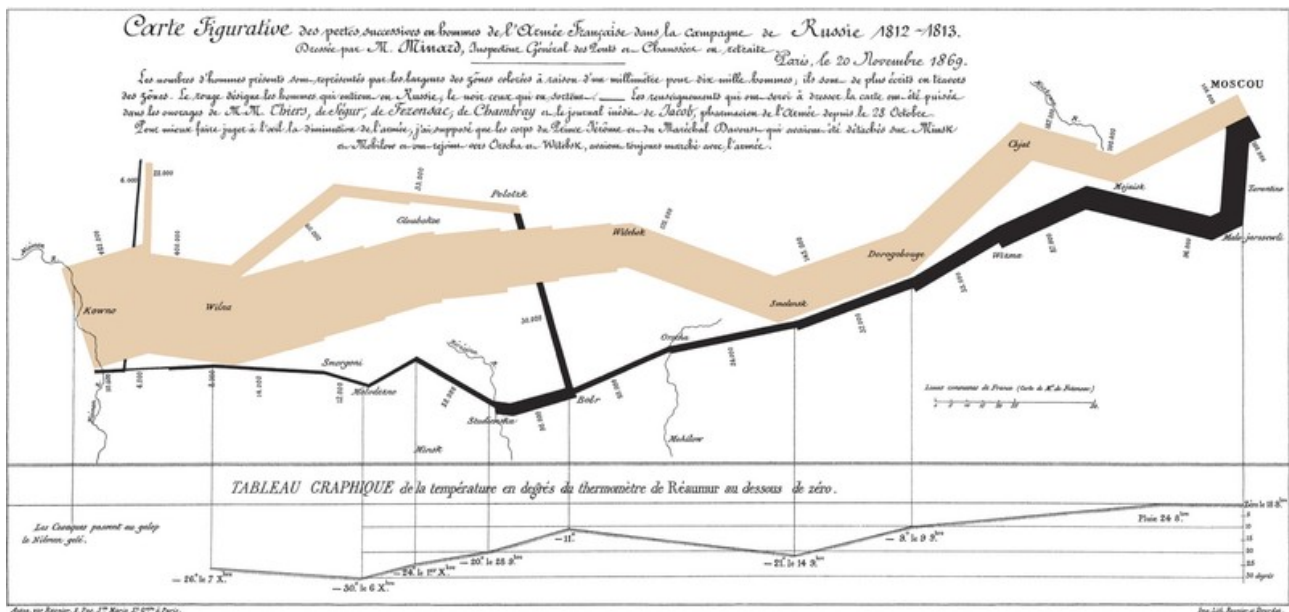


Illustration 2: Carte figurative de C. Minard

Le principe de la visualisation peut se résumer ainsi : lorsqu'on a des données et qu'on cherche des questions intéressantes, on doit explorer. Il s'agit d'utiliser nos capacités visuelles pour détecter des phénomènes inattendus. La visualisation est une aide à l'analyse. Si la spécialité est ancienne<sup>16</sup>, la nouveauté réside dans l'interactivité qui l'enrichit énormément. Pour utiliser pleinement l'informatique, les données doivent être structurées.

Quel est le rapport entre la visualisation et TEI ? TEI est une norme qui structure du contenu d'un document mais ne permet pas de montrer graphiquement des analyses : TEI ne s'intéresse qu'au caractère. La visualisation est une autre technologie. Associer TEI avec la visualisation permet de faire une édition graphique des analyses structurées du texte.

Revenons à Infozoom qui est un logiciel de visualisation puissant : il permet une vue d'ensemble, de construire des hypothèses et d'analyser les textes ; l'exploration est rendu possible grâce aux données structurées en TEI ; enfin, le logiciel extrait automatiquement des tables. Par rapport à la première expérimentation, ces premiers outils qui arrivent sur le marché permettent une autonomie du chercheur – même si elle est limitée – vis-à-vis des informaticiens.

Ma volonté était d'afficher deux types de visualisation : les cartes et les graphes. Pour ces derniers, Jean-Daniel Fekete a utilisé les programmes de Graph'Viz distribués gratuitement par la société AT&T. Parmi les algorithmes de calcul de placement de graphes proposés, il a choisi *dot* pour les graphes orientés placés en niveaux, *neato* pour les graphes non orientés placés par champs de forces, *twopi* et *circo* pour des placements de graphes circulaires.

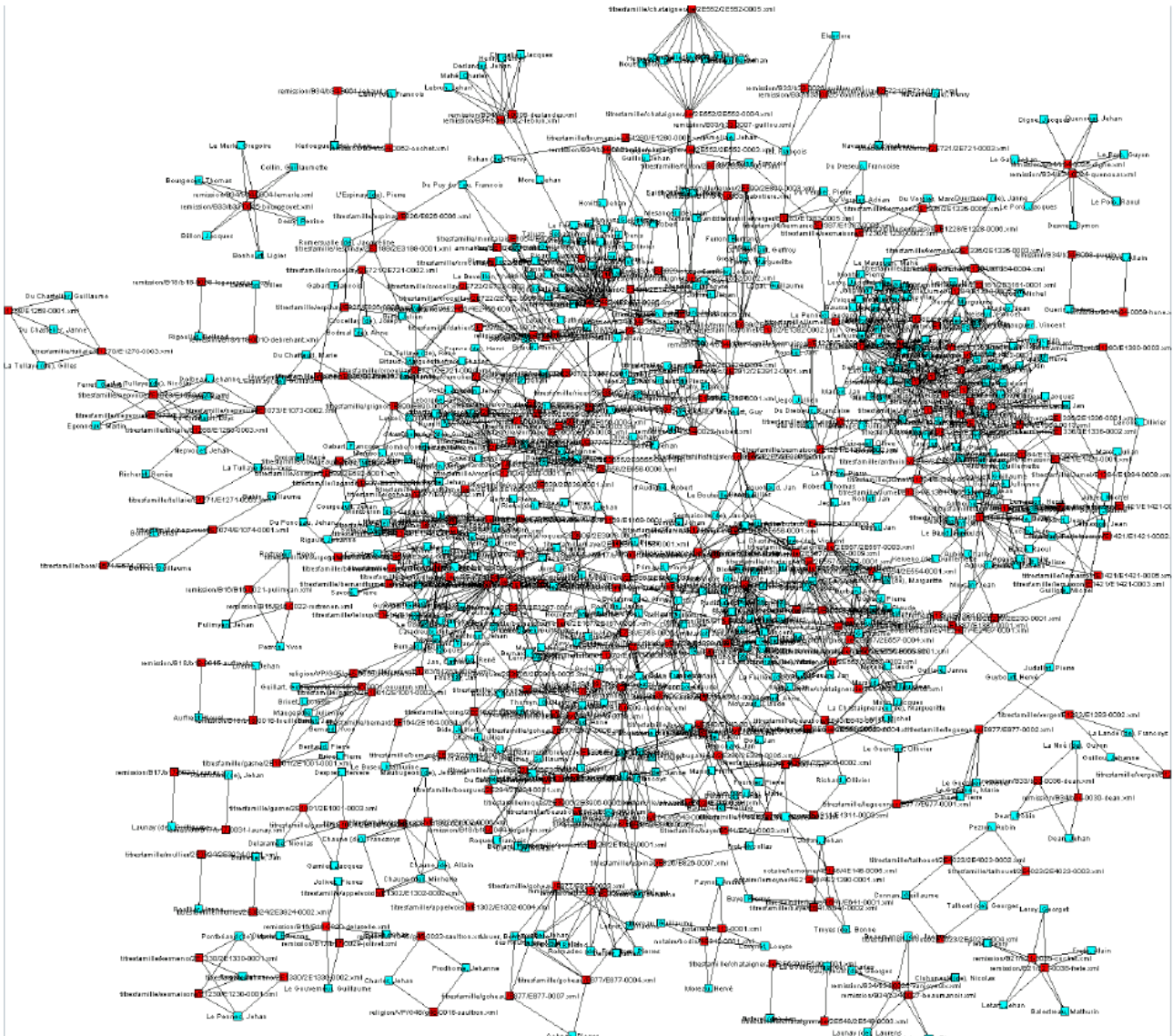
Parmi toutes les visualisations insérées dans ma thèse, en voici une sélection :

- réseau social
- arbre généalogique
- carte des lieux mentionnés
- diagramme de flux liés à la généalogie

La première visualisation est le réseau social qui est extrait des 1000 textes structurés. Les critères

<sup>16</sup> Les travaux actuels se basent sur ceux de Jacques Bertin (voir bibliographie).

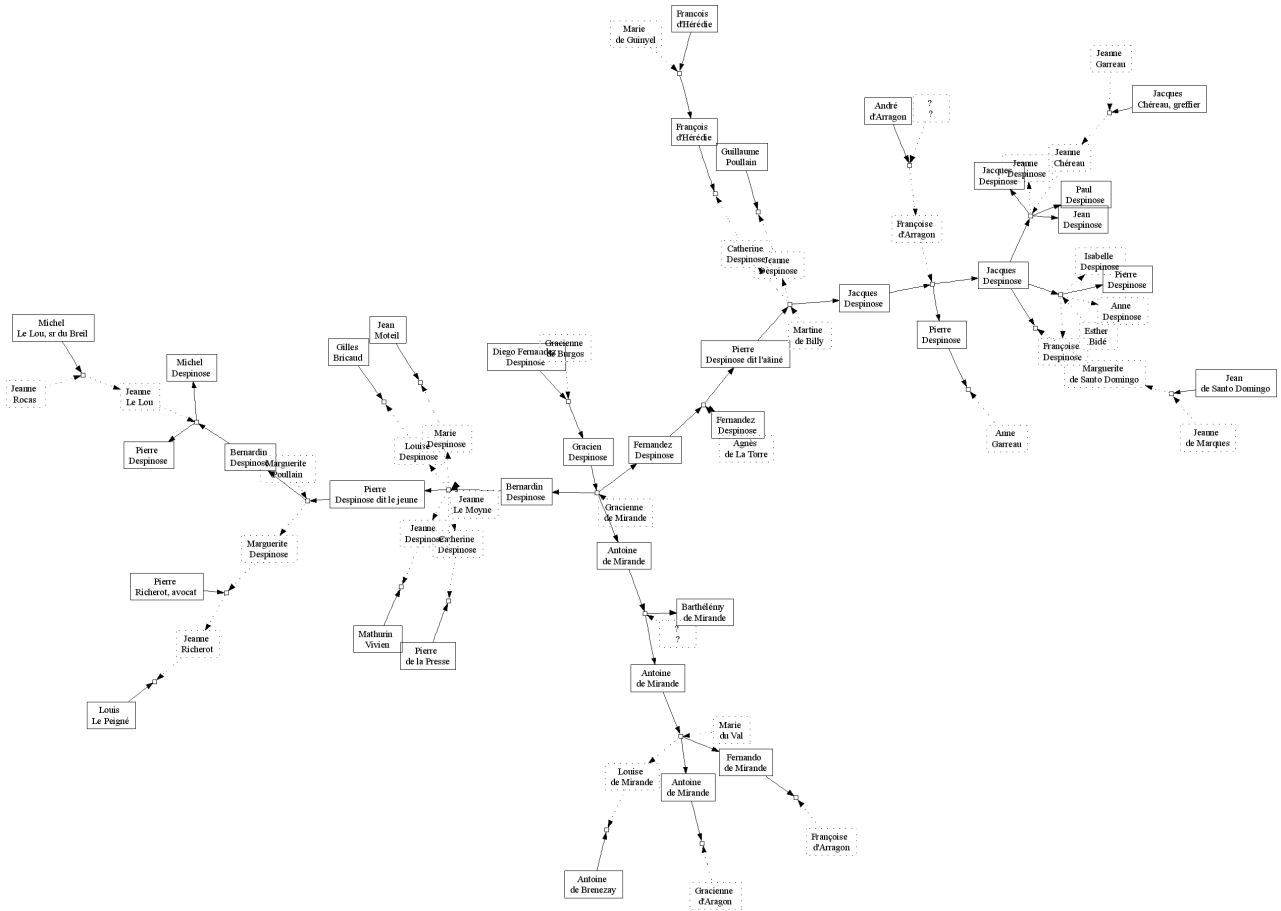
pour le construire sont simples : seules les structures composées des noms de personnes par fichier sont retenues.



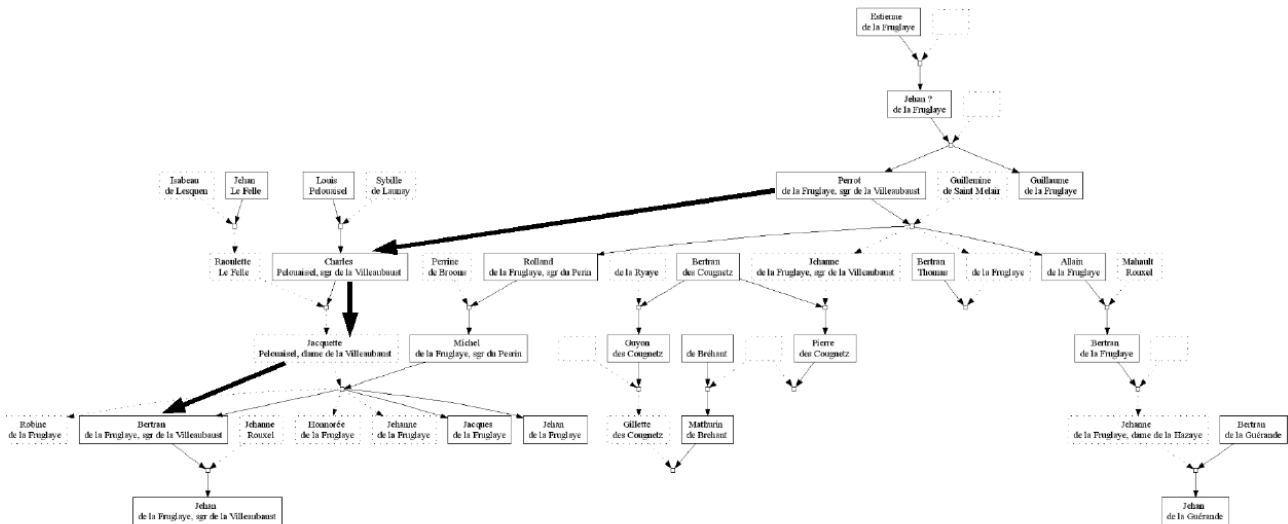
Sur un ordinateur, il est bien entendu possible de naviguer et de zoomer dans le réseau.

Le second exemple de visualisation de graphe repose sur les arbres généalogiques. Je propose deux cas :

- le premier est l'arbre généalogique d'une famille avec trois branches : la famille Despinoze à Nantes ;



- le second est l'apport de la généalogie dans la recherche du mécanisme de la dévolution des biens : ici, nous voyons que le transfert de la terre passe une femme.





Enfin, les troisième et quatrième exemples sont d'apparence plus classiques : ce sont des cartes <sup>17</sup>:

- la première carte est celle des lieux référencés ; grâce à la longitude et la latitude, il est possible d'extraire facilement les données et de les apposer sur un fonds de carte ;



- la seconde carte visualise le flux migratoire de la famille Menard d'Angers ;

<sup>17</sup> Le fonds de carte choisi est celui de Google



Ces visualisations ont permis une meilleure exploitation des sources, ont aidé à l'écriture de ma thèse et en améliorent la qualité de la publication. Les données structurées résultant d'un travail d'encodage fastidieux mais nécessaire ont rendu possible ces extractions et ces traitements par un informaticien.

## Conclusion

Cette méthode de travail en histoire est une véritable aide pour l'historien(ne) à établir des interprétations qui se vérifient par les textes de sources primaires. Utiliser TEI, c'est aller au-delà d'une simple publication pour laquelle un logiciel tel que Word suffirait. TEI a facilité le dépouillement, l'analyse, la production de résultats et a permis de référencer les analyses et de les vérifier. Enfin, il est envisageable de réutiliser le même processus pour d'autres analyses ultérieures comme je l'ai fait pour la thèse : j'ai repris 150 lettres de pardon dépouillées lors de la maîtrise et du DEA.

Après l'ère de la mise en page, nous sommes entrés dans celle du contenu : structurer un texte rend explicite le sens du texte ! Car TEI est aussi un instrument d'analyse et un outil qui permettra l'exploration d'une quantité de données en augmentation constante.

Des outils d'analyse et d'exploration sont mûrs mais il manque encore la connexion entre les données et les outils. Pour y remédier, changeons nos méthodes de travail et utilisons une seule convention commune.



## Bibliographie choisie

Jacques Bertin, *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes*, Paris, La Haye, Mouton, Gauthier-Villars, 1967. 2e édition : 1973, 3e édition : 1999, EHESS, Paris.

Nicole Dufournaud, *Manuel d'encodage XML-TEI Renaissance et temps modernes Imprimés – manuscrits : Recommandations pour une application TEI. Projet BVH, Bibliothèques Virtuelles Humanistes*, <http://www.bvh.univ-tours.fr/XML-TEI.>, Juillet 2009.

Jean-Daniel Fekete et Nicole Dufournaud, « Comparaison d'outils pour la visualisation de sources historiques codées en XML/TEI », *Document Numérique*, 9(2), Hermès – Lavoisier, 2006, p. 37.

Jean-Daniel Fekete and Nicole Dufournaud, « Compus: visualization and analysis of structured documents for understanding social life in the 16th century », in *Proceedings of the fifth ACM conference on Digital libraries*, ACM Press, 2000, p. 47.

Jean-Daniel Fekete et Nicole 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), Hermès – Lavoisier, 1999, p. 117.

Jean-Daniel Fekete and Nicole Dufournaud, « Utilisation de TEI comme support méthodologique au dépouillement de sources manuscrites », in *Actes du Vième Colloque National de L'Association Française pour l'Histoire et l'Informatique (AHI)*, november 1998.

Jean-Daniel Fekete. « Expérience de codage de document à intérêt graphique à l'aide de TEI », in *Actes du congrès Eurotex 98*, Saint Malo, mars 1998, p. 131.

Claude Gauvard, « *De grace especial* » *Crime, Etat et société en France à la fin du Moyen Age*, Publications de la Sorbonne, Paris, 1991, 2 volumes.

Karl Popper, *La logique de la découverte scientifique*, Paris, Payot, 1973.

## Sites Internet

<http://nicole.dufournaud.net>

<http://inria.fr>

<http://www.aviz.fr>