



Utiliser XML/TEI dans le cadre d'un travail d'ecdotique

Florence Clavaud

Déléguée à la formation informatique,
co-responsable de projets d'humanités numériques

École nationale des chartes

florence.clavaud@enc.sorbonne.fr



Pour commencer : définitions et concepts

Quels besoins TEI peut-elle satisfaire?

- Informatiser un travail d'édition critique à l'aide de TEI permettra :
 - de publier ce travail par différents canaux, notamment le Web (on produira un ensemble de fichiers, formant une matrice pour tous les formats de présentation)
 - d'indexer et de rendre recherchables les informations contenues dans ce travail ;
 - plus généralement, d'en fouiller le texte, de compter, rapprocher, relier des choses

TEI est en effet une ontologie XML du texte, et quand on l'utilise on produit des fichiers XML (**du texte balisé** en XML)

- Par la même occasion, on pourra rendre les fichiers obtenus compréhensibles, donc réutilisables, par une communauté large

En effet, **TEI est un standard**, et l'utiliser facilite les échanges, le partage des ressources et savoir-faire, l'interopérabilité entre systèmes informatiques

<titre>Concepts clé : XML</titre>

- **XML = Extensible Markup Language.** Un **métalangage** défini en **1998** par le W3C Consortium (<<http://www.w3.org/TR/1998/REC-xml-19980210>>) ; héritier du standard **SGML** (norme ISO 8879:1986) lui-même héritier d'un langage descriptif conçu en 1969 par des ingénieurs de la société IBM (Charles Goldfarb, Edward Mosher et Raymond Lorie)
- Maîtres mots : simplicité, clarté, universalité.
- **Utilisable pour baliser n'importe quel texte.**

Principe de base, simplissime : un segment de texte quel qu'il soit, du plus petit au plus important, peut être isolé au moyen de balises, et on peut en qualifier la nature ou en préciser le rôle.

Concrètement : ce segment de texte est le contenu textuel (ou le nœud fils) d'un **élément XML**, défini par un **nom** et syntaxiquement représenté par des **balises ouvrantes et fermantes**.

> la structure et le contenu d'un texte peuvent alors être marqués, explicités

> d'autres composants sont définis par la spécification XML, en particulier les **attributs**, paires nom-valeur qui sont associées à un élément, et saisies dans la balise ouvrante de l'élément

```
<diapo type="texte">
  <titre>définition des concepts clé</titre>
  <paragraphe>
    <phrase>Métalangage <!--etc.--></phrase>
  </paragraphe>
  <paragraphe><!--etc.--></paragraphe>
</diapo>
```

Concepts clé : XML, suite

- Fondé sur l'idée qu'*un texte est un système de hiérarchies ordonnées d'objets de contenu* (les **balises sont donc strictement imbriquées**, et un **élément, dit racine, englobe tout le texte**).
- Les fichiers XML sont des fichiers texte pur (non des fichiers binaires), utilisant simplement le standard **Unicode** pour coder les caractères et signes de tous les systèmes d'écriture du monde.
- **Métalangage universel et indépendant des plates-formes, qui peut être utilisé pour définir des langages adaptés à tel ou tel type de texte et tel ou tel besoin, par ex. :**
 - **EAD** (Encoded Archival Description, pour les instruments de recherche archivistiques)
 - **MarcXML** ou **MODS** pour les notices bibliographiques
 - **DocBook** pour la documentation technique
 - **MathML** pour les formules mathématiques
 - **TEI** pour les textes produits et/ou étudiés par les chercheurs en SHS
 - **XHTML** pour les pages Web
- Un langage XML donné se définit formellement au moyen de spécifications incluant au moins soit une DTD soit un schéma.
- Un fichier conforme aux règles de syntaxe du métalangage XML est dit bien formé ; un fichier XML conforme à un langage XML donné est dit valide

Concepts clé : XML, suite

- Ce métalangage est donc aujourd'hui utilisé partout en informatique, qu'il s'agisse de structurer des informations dont la durée de vie est importante ou d'échanger des informations entre applications, qu'on ait à s'occuper essentiellement de données, essentiellement de documents, ou d'une combinaison des deux.
- Sans parler des langages documentaires ou de métadonnées, de nombreux langages et protocoles informatiques ont été spécifiés depuis 1998, pour traiter et exploiter des fichiers XML. Ex. :
 - XPath (XML Path) 1.0 (1999) et 2.0 (2007)
 - XSLT (Extensible Stylesheet Language Transformations) 1.0 (1999) et 2.0 (2007)
 - XQuery (XML Query) (2007)
 - des protocoles d'échange entre applications tels que OAI-PMH, XML-RPC...

Concepts clé : TEI

- **TEI = Text Encoding Initiative.**

Au départ, un projet de recherche parrainé par 3 associations professionnelles anglo-saxonnes œuvrant dans le domaine des *digital humanities* (Association for Computers in the Humanities, Association for Literary and Linguistic Computing, Association for Computational Linguistics)

1^{ère} réunion en 1987 à New York : besoin d'une norme pour la structuration du texte

Plus de 50 chercheurs contribuent déjà > **TEI P1** (1990)

Création de groupes de travail par champ disciplinaire

> **TEI P3** (1994) (première version officielle des Guidelines, travail de 200 personnes)

Diffusion du modèle (séminaires, ateliers), tests, remarques des utilisateurs

Création d'un **consortium international en 1999**

TEI P4 (2002) (version XML de la P3)

Evolutions notables du modèle > publication de la **TEI P5 version 1.0** (novembre 2007) ; ... mars 2011 : TEI P5 version 1.9.1

TEI aujourd'hui

- En tant que consortium : une communauté active (utilisateurs, membres du consortium, conseil technique, bureau exécutif, groupes d'intérêt spécifiques)
- En tant que projet : un site Web, porte d'entrée pour tout trouver : <http://www.tei-c.org>; un wiki : http://www.tei-c.org.uk/wiki/index.php/Main_Page ; des listes d'outils : <http://www.tei-c.org/wiki/index.php/Category:Tools> ; une liste de discussion : <http://listserv.brown.edu/archives/cgi-bin/wa?SUBED1=tei-l&A=1>; une forge pour les sources (<http://sourceforge.net/projects/tei/files/>), une conférence annuelle
- En France : des acteurs de plus en plus nombreux et dynamiques, qui n'utilisent pas TEI que pour l'édition critique (cf. Revues.org, Persée, ou encore les travaux de l'ATILF), et qui ont cependant encore un peu de mal à agir en commun (noter l'existence des groupes suivants : groupe d'experts de l'AFNOR, réseau d'expertise francophone, futur réseau [MEET](#))

TEI aujourd'hui

- En tant que standard :
Les **Guidelines** consistent en :
 - une documentation sous la forme de discours denses et illustrés, synthétisant des recommandations consensuelles sur la manière de baliser les textes pour les différentes disciplines des SHS
 - un modèle abstrait qui définit l'ensemble des concepts et les relations entre ces concepts, basé sur une organisation en modules et classes
 - la possibilité (et des moyens pour le faire) d'exprimer des définitions formelles, une fois choisis les éléments et attributs dont on a besoin, sous la forme de DTD ou schémas XML (de définir en fait son propre modèle TEI)

Les modules principaux, obligatoires

- ***tei*** : définition des classes, macros et types de données
- ***textstructure*** : éléments de base pour structurer un texte de type livre
- ***core*** : éléments disponibles dans tous les documents TEI
- ***header*** : en-tête TEI (métadonnées du document TEI)

Autres modules

Modules de structures alternatives à la structure la plus fréquente d'un texte :

- *spoken* : transcription du discours oral

(<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TS.html>>)

- *dictionaries* : entrées d'un dictionnaire

(<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>>)

Modules propres à un type d'objet technique, à un point de vue ou à une discipline (en gras, les noms des modules qu'un travail d'ecdotique devrait logiquement employer en tout ou en partie, sans exclusive) :

- ***analysis*** (analyse, interprétation);

- ***certainty*** (niveaux de certitude et marquage de la responsabilité) ;

- *corpus* (corpus) ;

- *drama* (textes d'art dramatique) ;

- *figures* (tableaux, figures et formules) ;

- *gaiji* (caractères non standard et glyphes) ;

- *iso-fs* (structures de traits) ;

- *linking* (liens, segmentation, alignements) ;

- ***msdescription*** (description des manuscrits) ;

- ***namesdates*** (noms, dates, lieux) ;

- *nets* (graphes, réseaux, arbres) ;

- *tagdocs* (documentation) ;

- ***textcrit*** (apparat critique) ;

- ***transcr*** (transcription des sources primaires) ;

- ***verse*** (vers).

Le système des classes

Une classe regroupe des éléments en fonction de caractéristiques communes. Ces éléments **héritent** des propriétés de la classe à laquelle ils appartiennent mais aussi de la « super-classe » de cette classe, et ainsi de suite.

On peut distinguer :

- les classes sémantiques, qui regroupent les éléments par similarité de sens

Exemple : les divers types de noms. Ils font partie de la classe [model.nameLike](#)

- les classes structurelles, qui regroupent des éléments qui ont un comportement similaire dans la structure hiérarchique d'un document TEI

Exemple : les phrases ou éléments pouvant apparaître au même niveau qu'une phrase : classe [model.phrase](#)

- les classes attributives, qui fournissent des jeux d'attributs identiques pour certains éléments

Exemple : La classe [att.naming](#) fournit des attributs communs aux éléments qui concernent les personnes, lieux, organismes, etc. comme <forename>, <genName>, <nameLink>, <orgName>, <persName>, <roleName>, etc.

Pour en savoir plus sur les classes, voir la page

[<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STEC>](http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STEC)

Les attributs

Les attributs peuvent être :

- globaux, c'est-à-dire potentiellement applicables à tous les éléments TEI, par exemple :

xml:id fournit un identifiant unique

n fournit un numéro ou un label

xml:lang indique la langue du contenu textuel de l'élément (à l'aide du code ISO)

rend donne une indication sur l'apparence, le rendu

- définis au sein de classes attributives pour des groupes d'éléments (voir diapo précédente) ;

- définis au niveau des modules spécifiques

Moyens de personnalisation

- Pizza, puis **Roma** (<<http://www.tei-c.org/Roma/>>) (ou, si on veut utiliser le serveur de test du Consortium, <<http://tei.oucs.ox.ac.uk/Roma/>>) : application de génération de schémas TEI, permettant :
 - de choisir la langue de la documentation ;
 - de sélectionner les modules utiles ;
 - de sélectionner les éléments dans ces modules, les attributs des éléments ;
 - d'ajouter des éléments, de modifier des classes ;
 - de choisir la grammaire du schéma (Relax NG, XML W3C, DTD) ;
 - de générer le schéma demandé et sa documentation ;
 - de générer un fichier mémorisant les choix opérés et adaptations éventuelles réalisées (en fait, le fichier matrice de tout le reste, qui sera lui-même un fichier TEI utilisant le module tagdocs, et qu'on appellera fichier ODD, *One Document Does it all*).On pourra renvoyer à Roma, autant de fois que nécessaire, ce fichier ODD, pour revenir sur les choix faits et modifier/enrichir son modèle.

Bonnes pratiques de personnalisation

- Pour tout projet, après analyse et tests sur la base d'un schéma complet, il convient de sélectionner uniquement ce dont on a besoin en procédant si besoin est à des adaptations, pour obtenir un modèle dédié. Si on ajoute des éléments ou attributs à ce que propose TEI, les distinguer en les associant à un autre espace de noms que celui de TEI.
- Nécessité de documenter ses choix, pour permettre le partage, la réutilisation, l'interopérabilité.

Quelques documentations de projets disponibles en ligne :

- les Guidelines de la California Digital Library

(<<http://www.cdlib.org/groups/stwg/index.html>>) ;

- documentation technique de l'encodage du Glossarium de Du Cange, rédigée en 2007 (<<http://ducange.enc.sorbonne.fr/schema>>) ; documentation de certaines éditions électroniques publiées par l'Ecole des chartes (voir [ELEC](#)), guide de balisage pour le cartulaire de Saint-Denis, qui sera publié en juin 2011

- manuel d'encodage écrit pour le projet BVH du CESR

(<<http://www.bvh.univ-tours.fr/XML-TEI/index.asp>> ;

- documentation des schémas TEI utilisés dans le cadre du projet européen [Enrich](#) : <<http://tei.oucs.ox.ac.uk/ENRICH/>>

**Utiliser TEI dans un travail d'ecdotique :
aperçu des possibilités**

Balisateur de la structure du texte et de sa transcription

Henri Grady
Voyage à Gavarnie
les 21 & 22 août
1850

Le voyage de Gavarnie ^{présentait} des obstacles. Le temps, sans être ~~précisément~~ mauvais, n'offrait pas de ~~grandes garanties~~ ^{sécurité}. Les brouillards se levaient tous les jours sur la montagne et ~~menaçait~~ sans cesse.

[Extrait du Journal de voyage d'Henri Grady, 1850. Orig. autographe, AN, 181 AQ 94 n° 267]

<div type="transcription">

<head>Voyage à Gavarnie <lb/>les 21 <choice>
<abbr>&</abbr>
<expan>et</expan>
</choice> 22 août<lb/>1850</head>

<p>Le voyage de Gavarnie

<sub>

rencontrait

<add place="above">présentait</add>

</sub> des obstacles. Le<lb/>temps, sans être précisément mauvais, n'offrait pas de<lb/>

<sub>

grandes garanties

<add place="above">sécurité</add>

</sub> : les brouillards se levaient tous les jours<lb/> [...]</p>

<!-- autres paragraphes -->

</div>

(du texte inscrit sur son support, au texte structuré en XML, à autant de formats de sortie qu'on le souhaite)

A partir du code XML présenté, on peut par exemple faire une édition imitative :

Voyage à Gavarnie
les 21 & 22 Août
1850

présentait

Le voyage de Gavarnie ~~rencontrait~~ des obstacles. Le temps, sans être ~~précisément~~ mauvais, n'offrait pas de sécurité

~~grandes garanties~~ : les brouillards se levaient tous les jours

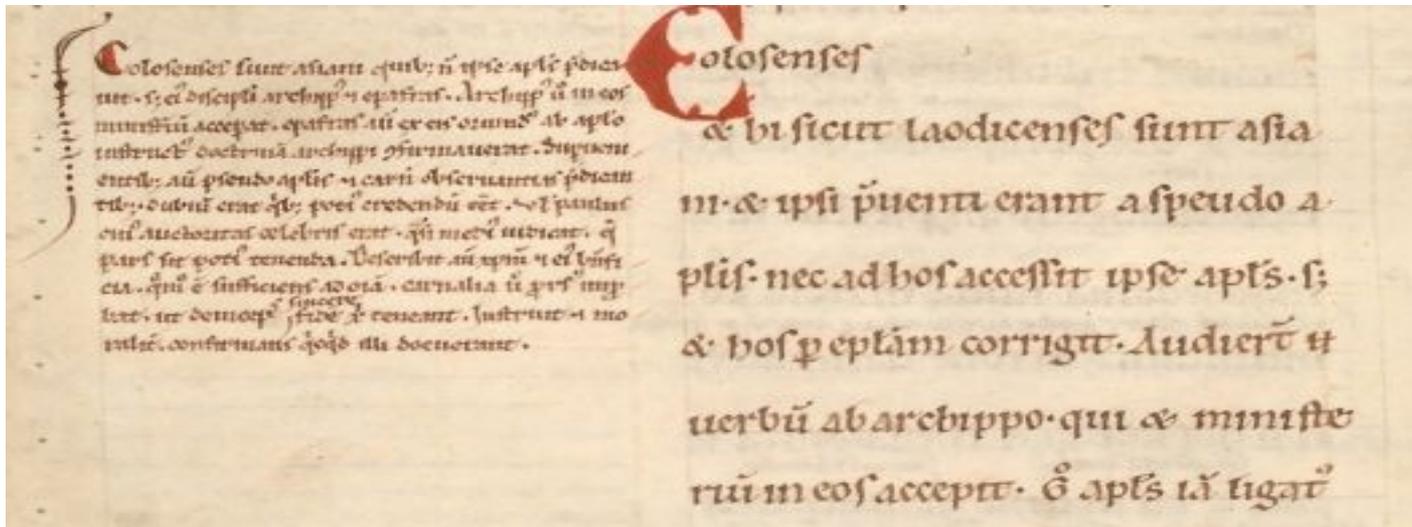
- Ou ceci (entre autres)

Voyage à Gavarnie les 21 et 22 août 1850

Le voyage de Gavarnie présentait¹ des obstacles. Le temps, sans être² mauvais, n'offrait pas de sécurité³ : les brouillards se levaient tous les jours sur la montagne

1. rencontrait
2. précisément
3. grandes garanties

Autre exemple



[Bible glosée d'Anselme de Laon, XII^e siècle. Bib. Mazarine, 123, fol. 94V^o. Edition électronique dans THELEME]

<div xml:id="transcription" facs="#fax" xml:lang="lat">

<p>

<hi rend="initiale">C</hi>olosenses <add place="margin-left" n="1">

Colosenses sunt Asiati quibus

Ap<ex>osto</ex>I<ex>u</ex>s p<ex>re</ex>dica-<lb/>-vit, s<ex>ed</ex>

ej<ex>us</ex> discip<ex>u</ex>li Archipp<ex>us</ex>

<ex>et</ex> Epafra. Archipp<ex>us</ex> v<ex>er</ex>o in

eos<lb/><!-- etc.-->

</add>

<lb n="2"/> et hi sicut Laodicensis sunt Asia- <lb n="3"/>-ni et ipsi

p<ex>re</ex>venti erant a pseudo a- <lb n="4"/> -postolis, nec ad hos

accessit ipse Ap<ex>osto</ex>I<ex>u</ex>s, s<ex>ed</ex>

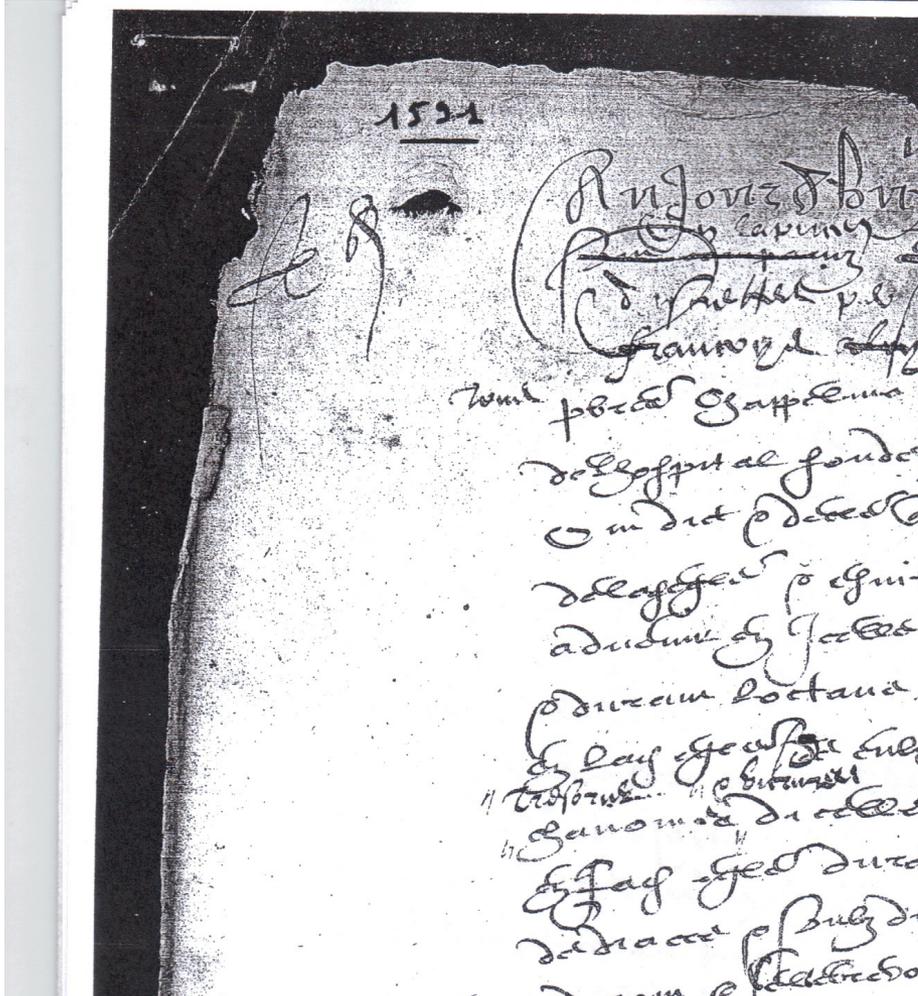
<lb n="5"/> et hos p<ex>er</ex> ep<ex>isto</ex>lam corrigit.

Audier<ex>un</ex>t <ex>enim</ex>

<lb n="6"/><!-- etc.-->

</p>

Métadonnées analytiques



```
<sourceDesc>
  <msDesc>
    <msIdentifier>
      <country>France</country>
      <institution>Archives nationales</institution>
      <collection>Minutier central</collection>
      <idno>ET//19</idno>
    </msIdentifier>
    <msContents>
      <summary><date when="1591-10">1591,
octobre</date>. —
      <placeName>Paris</placeName>. Accord entre
Georges Decourbel,
François Alavoine et Vincent André, prêtres
chapelains de
Saint-Jacques-de-l'Hôpital, pour éteindre un
procès engagée devant
l'archevêque de Sens à propos du port des chapes
et des offices de
diacre et sous-diacre des grandes
messes.</summary>
      <textLang>français</textLang>
    </msContents>
    <physDesc>
      <objectDesc>
        <supportDesc>
          <support>papier</support>
          <extent>
            <dimensions unit="mm">
              <height>266</height>
              <width>180</width>
            </dimensions>
          </extent>
        </supportDesc>
      </objectDesc>
    </physDesc>
  </msDesc>
</sourceDesc>
```

Apparat critique : plusieurs témoins, un texte.

1. Tableau de la tradition

```
<listWit>
  <witness xml:id="dugny-acte1-A" n="A">Arch. nat., L 846, n° 30.</witness>
  <witness xml:id="dugny-acte1-B" n="B">Cart. blanc, t. I, p. 425a-b, n° I, <quote xml:lang="lat" type="rubric">Carta hospitum
de Gargiis</quote>.</witness>
  <witness xml:id="dugny-acte1-C" n="C">Anc. cart. du grand prieur, p. 1-2, n° 1.</witness>
</listWit>
  <head>Indiqué</head>
  <witness xml:id="dugny-acte1-indiqué-i-noir">Anc. inv. noir,
    p. 119a-b, n° I : <quote xml:lang="lat">Hugonis
    abbatis Sancti Dyonisii, quomodo Balduinus de Andeliaco
    tradidit hospitibus suis de Gargiis quasdam terras que olim
    fuerant ad campipartem ad medietatem fructuum excolendas et
    jure hereditario possidendas. Qui tamen hospites non
    poterunt illas alicui vendere, dare, pignonari vel obligare
    nisi in Gargiis maneat <!-- etc. --></quote>.
  </witness>
  <witness xml:id="dugny-acte1-indiqué-ig-1">Inv. gén. I, n° 558, p. 570-571, d'après l'original.</witness>
</listWit>
```

[Extrait simplifié de l'édition critique du Cartulaire blanc de Saint-Denis, chapitre de Dugny – en cours ; voir édition d'autres chapitres ici <<http://saint-denis.enc.sorbonne.fr/>>

Voir l'image de l'acte Dugny 1 au cartulaire blanc,

<http://saint-denis.enc.sorbonne.fr/images/cartulaireblanc/tome1/voir.html?ns=FRCHANLL_LL1157_0287R_L.jpg>]

Apparat critique : plusieurs témoins, un texte.

2. Encodage des variantes (leçons) et encodage des notes associées

```
<div type="transcription" xml:lang="lat" >
  <p>Ego Hugo, Beati Dyonisii abbas Dei gratia, <app n="a">
    <lem wit="#dugny-acte1-A">totumque</lem>
    <rdg wit="#dugny-acte1-B">totum</rdg>
    <note><foreign xml:lang="lat">totum</foreign> B, corrigé avec A.</note>
  </app> ejusdem ecclesie capitulum. Noverint universi tam presentes quam
  futuri quod Baldoinus de Andiliaco hospitibus suis de Gargiis terras quas
  diu ad campipartem excoluerant et alias terras de culturis suis ad
  medietatem fructuum largitur excolendas et imperpetuum jure hereditario
  possidendas. Terras tamen illas prefati hospites alicui, nisi in Gargiis
  maneat, vendere, dare, pignori obligare nullatenus poterunt. Si autem
  terras illas per se vel per aliquem amicorum suorum, qui tamen in
  Gargiis maneat, predicti hospites aliquo casu excolere non potuerint,
  Baldoinus vel heredes ejus de terris illis quicquid voluerint facere
  poterunt. <!-- etc. → </p>
</div>
```

[Extrait simplifié de l'édition critique du Cartulaire blanc de Saint-Denis, chapitre de Dugny, acte Dugny – en cours ; voir édition d'autres chapitres ici <<http://saint-denis.enc.sorbonne.fr/>>

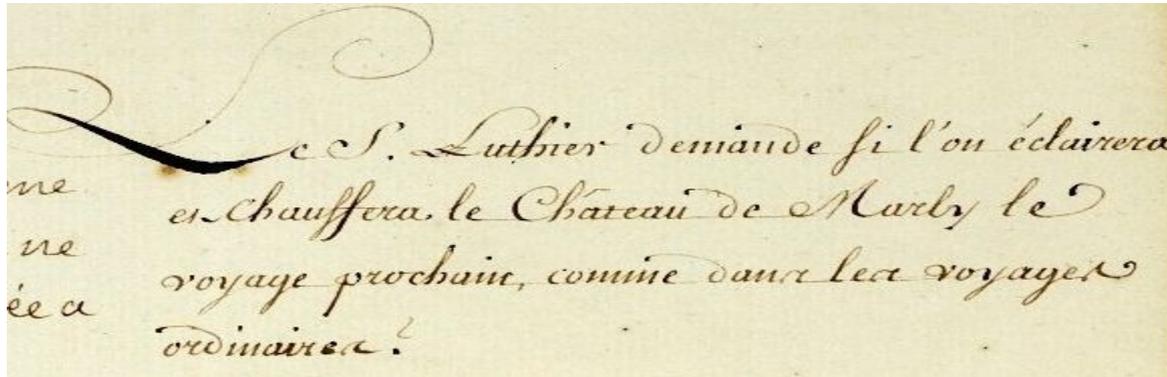
Voir l'image de l'acte Dugny 1 tel que copié au cartulaire blanc, tome 1 page 425,

<http://saint-denis.enc.sorbonne.fr/images/cartulaireblanc/tome1/voir.html?ns=FRCHANLL_LL1157_0287R_L.jpg>]

Apparat critique : autre exemple

```
<listWit>
  <witness xml:id="B">Copie collationnée du 4 novembre 1662, par Bauzit,
    secrétaire, sur ordre de Pierre Ducup, juge-mage, et de Jean Lanes,
    procureur du roi, dans la sénéchaussée de Lauragais, cahier de papier, 6
    ff., Archives du château de La Tour, liasse 94, d'après une copie
    collationnée sur l'original, le 16 septembre 1543, par Gueyne, notaire à
    Revel et conservée en 1662 aux archives de la sénéchaussée de
    Lauragais.
  </witness>
</listWit>
<!-- Plus loin : -->
<body>
  <div type="transcription">
    <p>A toutz ceux qui ces lettres verront, Jean seigneur de Folleville, chevalier,
      conseiller du roy notre sire, garde de la prevosté de Paris, salut. Sçavoir
      faisons que par devant Mile de Brueil et Guillaume Pardeur, notaires dudit
      seigneur au Chastelet de Paris, fut presant noble homme Jean de Voisins,
      escuier, seigneur de Coffolens<ref type="note" target="#note-1"/>, lequel advoa
      et par ces presantes advoue a tenir a une seule foy et en homage du roy notre
      dit seigneur les chasteaux [...]</p>
    <list>
      <item>Item lez le chastel, un jardin contenant environ une ay mine de
        terre.</item>
      <item>Item cinquante livres tournois de rente paiés a certain jour.</item>
      <item>Item un moulin <app>
        <rdg wit="#B"><sic>eau</sic></rdg>
      </app> appartenant audit chastel, qui peut valoir une annee par autre dix
      muis de blé mosturé.</item>
    </list>
  </div>
</body>
<back>
  <div type="notes">
    <note xml:id="note-1">Coffolentis, Aude, cant. et comm. Castelnaudary.</note>
  </div>
</back>
```

Annotation générique et indexation



[extrait de l'édition d'un acte royal de mai 1763, Arch. Nat. O/1/284, n° 525. . Edition électronique dans THELEME]

```
<div>
  <p><hi rend="bigger">L</hi>e <choice> <expan>s<ex>ieur</ex></expan>
    <abbr>s<am>.</am></abbr></choice> Luthier<ref type="note" target="#n-001"/> demande
    si l'on éclairera<lb/> et chauffera le <placeName ref="#marly">château de
    Marly</placeName> le <lb/> voyage prochain comme dans les voyages<lb/>ordinaires. </p>
</div>
<!-- plus loin dans le fichier XML -->
<div type="notes">
  <note xml:id="n-001">Louis-François Luthier était employé pour le bois et la bougie du château de Marly avec 1 000 l. de gages en 1759 (Arch. nat.,
O<hi rend="super">1</hi> 3913 bis, n° 2, état pour servir de certificat de service).</note>
<!-- etc. autres notes historiques-->
</div>
<div type="indexes">
<listPlace>
  <place xml:id="marly">
    <placeName type="complete">Marly, château de</placeName>
    <location>
      <district type="commune">Marly-le-Roi</district>
      <district type="departement">Yvelines</district>
      <region>Ile-de-France</region>
      <country>France</country>
      <geo>48,85 -2,023</geo>
    </location>
    <desc>Le château de Marly (aujourd'hui Marly-le-Roi) a été bâti sur les
    plans de Jules Hardouin-Mansart pour Louis XIV entre 1679 et 1685. Au
    XVIIIe siècle, il fait partie du domaine royal de Versailles.</desc>
    <bibl>Voir la <ref
      target="http://fr.wikipedia.org/wiki/Ch%C3%A2teau_de_Marly">page
      relative au château dans Wikipédia</ref></bibl>
    </bibl/>
  </place>
</div>
```

Et bien d'autres choses

- Segmentation analytique du texte
ex. : en parties du discours, voir THELEME ; balisage lexicographique ; balisage de glyphes
...
- Alignement d'un texte et de sa traduction, de portions de texte avec des zones d'image
- Etc... Les structures XML obtenues peuvent être très complexes
(à la mesure des objectifs éditoriaux)

Le balisage TEI vu comme un processus

TEI dans le processus d'édition : deux grands cas d'usage

- Conversion de fichiers préexistants

Cas **fréquent** : on dispose bien souvent de fichiers produits au traitement de texte (ou à l'aide d'une base de données) ou encore d'éditions imprimées.

Dans un tel cas, surtout ne pas repartir de zéro ! On peut faire beaucoup de choses automatiquement ; par ailleurs il est bien évidemment toujours possible de convertir manuellement.

Il faudra cependant relire et enrichir les résultats.

Et il est probable, si on n'a pas accès au manuscrit édité, qu'il subsiste des ambiguïtés, surtout dans les notes d'apparat critique.

Nota : des processus d'ingénierie inverse (de TEI vers fichier au traitement de texte) sont parfois nécessaires. Et tout aussi possibles.

- Production directe de fichiers TEI

Cas encore rare, et le plus favorable au chercheur finalement (sur le plan intellectuel, et en terme de résultats).

On manque d'outils confortables ; les plus génériques et performants sont des éditeurs XML, qu'on peut certes adapter.

TEI dans le processus d'édition : définir et faire vivre son modèle

- Dans tous les cas, le modèle TEI jouera un rôle crucial.
- Il conviendra de définir un premier modèle avant de commencer.
- Ce modèle évoluera pendant toute la durée du projet (ne serait-ce que parce que les hypothèses, questionnements ou même les besoins des chercheurs évolueront aussi). En fait il peut même se faire que plusieurs corpus numériques successifs puissent (doivent) être conservés.
Les éléments de macrostructure et les parties du modèle concernant la transcription du texte sont ceux qui ont le plus de chances de rester stables. Le paratexte bougera probablement, ainsi que tous les dispositifs de balisage sémantique ou analytique – index, grilles d'interprétation, liens.....
- Donc en fait, le processus de modélisation se fera en cycles itératifs

TEI dans le processus d'édition : définir et faire vivre son modèle (2)

- Attention à l'économie générale du projet : il faut reconnaître ce qui, parmi les options, est à la fois le plus important/intéressant qualitativement, tout en restant faisable dans le contexte et avec les moyens et outils du projet.
Donc : ne pas choisir un modèle trop complexe pour ses utilisateurs. Ne pas forcément non plus chercher à tout prix à se restreindre à TEI : on peut ajouter ou modifier des éléments/attributs à ceux apportés par TEI. Ce qui est très générique n'est pas toujours le plus parlant ou le plus simple.
- Au final, il est essentiel de documenter son modèle, dès le début.
- Et on peut le plus souvent **considérer le modèle documenté comme un des résultats d'un projet.**

TEI dans le processus d'édition : un compagnon pour le chercheur

- Une démarche qui s'inscrit pleinement dans le travail d'analyse du texte, et qui peut aider à ce travail :
 - il faut reconnaître et dénommer (pour faire un bon travail de balisage). L'à peu près et l'implicite doivent être réduits, le regard s'aiguise.
 - un modèle TEI doit pouvoir aider à ne rien oublier et à produire de manière cohérente et homogène
- Les gains :
 - cf. Introduction : expressivité, exploitabilité, évolutivité, réutilisabilité, des fichiers produits
 - et aussi réutilisabilité des savoir-faire acquis, peut-être des plates-formes logicielles
 - meilleure visibilité de l'édition si elle est en ligne
 - un point de non retour ?

Pendant et après le travail : les risques et contraintes

- Pendant :
 - le plus souvent une étape supplémentaire dans le travail
 - nécessité de déployer de nouveaux outils, et de se former ; les outils ergonomiques sont rares, l'apprentissage doit être suivi de mise en pratique
 - nécessité de développer un site Web ;
 - il faut finalement des équipes pluridisciplinaires (ou des personnes ayant une double culture, scientifique et informatique) pour arriver à l'édition électronique, pendant toute la durée du projet ;
 - investissement en temps et en moyens importants au début
- Après :
 - nécessité de maintenir (voire faire évoluer) le site Web, de veiller à la citabilité et à la pérennité des URLs
 - sur le plan scientifique, exposer ses sources (des images numériques, voire des fichiers XML) c'est bien évidemment s'exposer plus à la critique
 - la valeur accordée à ce genre de travail est (pour l'instant) moindre que celle de l'édition imprimée

Autrement dit, ce qui est en jeu :

- le rôle, la formation et les habitudes de travail du chercheur

L'auteur de l'édition critique devrait s'impliquer autant que possible ; c'est la maîtrise de sa production scientifique qui est en jeu, la fidélité à son interprétation, l'expression voire la dynamique de ses idées

- le renouvellement des chaînes de travail
- de nouvelles formes d'éditorialisation, et peut-être la construction de nouvelles règles éditoriales, ou tout du moins des évolutions du genre
- les méthodes de la recherche

Quelques références

Quelques exemples d'édition critique électronique utilisant TEI

- On verra en plus ici que décider de faire une édition électronique (une application Web, statique ou dynamique) notamment à partir de fichiers TEI, peut aussi être l'occasion de dépasser certaines limites de l'édition imprimée - autres que celles liées à son caractère non structuré ; parmi ces limites :
 - une volumétrie relativement faible
(voir par ex. [CBMA](#), les [chartes antérieures à 1121](#))
 - un caractère figé (voir par ex. [Hyperdonat](#), projet d'édition hypertexte du commentaire attribué à Aelius Donat aux comédies de Térence ; [chartes de Saint-Denis](#))
 - une présentation des textes monolithique (versus des empilements ou juxtapositions de vues : voir par ex. les [lettres de Van Gogh](#), le [ms de la quête du Saint Graal](#))
 - le faible nombre de reproductions des sources éditées (versus l'intégration de substituts numériques de l'intégralité des témoins principaux : Saint-Denis, quête du Saint Graal)
 - le caractère "fermé" (versus des applications référencées par des portails – tels [ISIDORE](#) – ou interrogeables par d'autres applications, comme [Du Cange](#), interrogeable depuis le [portail lexical du CNRTL](#))
 - une assez faible diversité de contenus (voir par contre, en ligne, des écrits du for privé, comme le [Journal de W. Godwin](#), des actes de la pratique, des correspondances)
...ce sans pour autant renoncer à publier sous cette forme plus classique (cf. Travaux des Presses universitaires de Caen ou de l'ENC)
- Un projet utile à connaître (dans le cadre du projet Bibliindex): [Sourcencyme](#)

Éléments de bibliographie

- sur XML :

L'article de Wikipédia en français sur XML, <http://fr.wikipedia.org/wiki/Extensible_Markup_Language>

- *A gentle introduction to XML* : <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>>

- *XML en concentré : manuel de référence* / Eliotte Rusty Harold & W. Scott Means; traduction de Philippe Ensarguet, Frédéric Laurent. - 3e éd. - Paris : O'Reilly, DL 2005. - 1 vol. (XX-760 p.) : couv. ill. en coul. ; 24 cm. - Index. - ISBN 2-84177-353-1 (br.) : 45 € - EAN 9782841773534

- sur TEI : le site Web du projet <<http://www.tei-c.org/index.xml>> ; pour apprendre : TEI by Example, et diverses ressources en français (actions et supports de formation), notamment les ressources mises en ligne après la formation organisée par l'ENC en mars 2011 (<<http://www.enc.sorbonne.fr/formationTEImars2011/>>) ;

- l'arrière-plan conceptuel : voir par ex. l'article suivant <<http://www.philo.at/mii/mii/node5.html>> (et le texte d'Allen Renear auquel il donne accès)

- les actes des conférences annuelles du Consortium TEI, du colloque annuel *Digital Humanities*, ceux du colloque annuel *Balisage*

- des revues spécialisées en humanités numériques, comme, pour le Moyen Âge, *Digital Medievalist* (<<http://www.digitalmedievalist.org/journal/>>); ou la revue *Literary and Linguistic Computing* de l'université d'Oxford (<<http://llc.oxfordjournals.org/>>), la revue *Digital Humanities Quarterly* (<<http://digitalhumanities.org/dhq/>>)

- *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford : Blackwell, 2004. <<http://www.digitalhumanities.org/companion/>> (en particulier le chapitre 22, "Electronic Scholarly Editing"). A noter, par ex., cette idée : "*Understanding the poetics and principles of electronic scholarly editing means understanding that the primary goal of this activity is not to dictate what can be seen but rather to open up ways*