



**XML TEI**

**M2 TNAH**



**Segolene-Albouy/XML-TEI\_M2TNAH**

[segolene.albouy\[at\]gmail.com](mailto:segolene.albouy[at]gmail.com)

**[tinyurl.com/M2TNAH-2023](https://tinyurl.com/M2TNAH-2023)**

# Déroulé des séances

<b>4 octobre</b>	Encoder le texte : les langages à balises
<b><u>10 octobre ?</u></b>	Contraindre le XML : la <i>Document Type Definition</i>
<b>18 octobre</b>	Décrire les documents patrimoniaux : la TEI
<b>25 octobre</b>	Anatomie d'un document TEI : parties et composants
<b>8 novembre</b>	Les TEI <i>guidelines</i> : modules, modèles, macros, etc.

# Déroulé des séances

<b>15 novembre</b>	Éditer un texte : initiation à l'édition scientifique
<b>22 novembre</b>	Éditer un texte : description des sources manuscrites
<b>29 novembre</b>	Éditer un corpus : création d'une chaîne de traitement
<b>6 décembre</b>	Localisation dans le document : XPath
<b>13 décembre</b>	Les schémas XML : initiation à l'ODD
<b>20 décembre</b>	Les schémas XML : ODD personnalisations avancées
<b>10 janvier</b>	Les schémas XML : Mise en pratique

# Évaluation

## Devoir sur table

**15 novembre**

Vérification des bases de la structure XML et des principes de la TEI

*Coefficient 1*

## Projet individuel

**À rendre avant le 31 janvier**

Édition en XML TEI  
accompagnée d'une ODD  
documentée

*Coefficient 3*

# Objectifs

**Maîtriser la  
structure XML**

**Créer une édition  
numérique**


**Décrire un  
document textuel**

**Documenter ses  
choix d'encodage**

# Oxygen

[oxygenxml.com/xml\\_editor](https://oxygenxml.com/xml_editor)





# Séance 1

# Encoder le texte

Les langages à balises

# Le texte informatisé



## Texte brut

Le contenu textuel,  
dépourvu de toute  
structure et mise en  
forme



## Texte balisé

Le contenu textuel,  
structuré à l'aide de  
balises



## Texte formaté

Le texte mis en forme à  
l'aide d'un éditeur de  
texte

# Les formats textuels

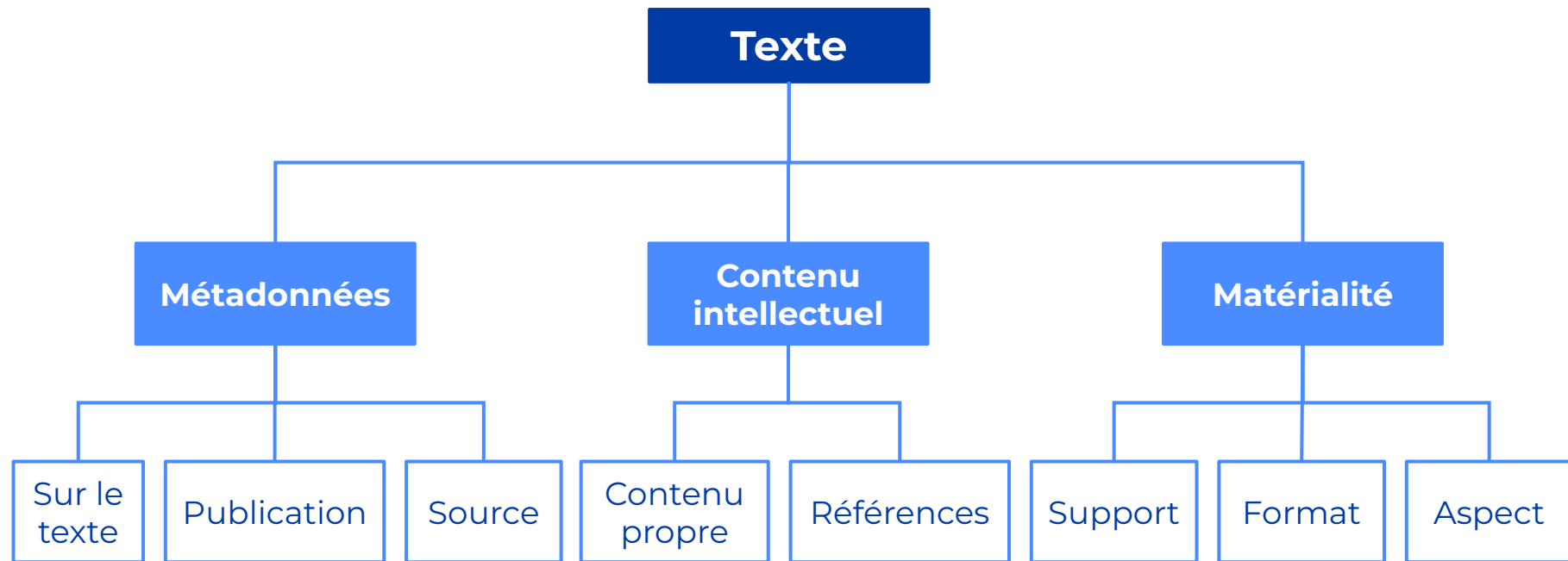


A decorative graphic on the left side of the slide consisting of two overlapping squares: a larger dark blue square at the bottom and a smaller light blue square on top of it, partially offset to the right.

# Qu'est-ce qu'un “texte” ?

Quelles peuvent être ses caractéristiques ?  
Comment les retranscrire numériquement ?

# Caractéristiques d'un texte



# Les langages à balises

```
{poeme}  
  {meta}  
    {type}Haïku{type}  
    {auteur}Takahama Kyoshi{auteur}  
  {meta}  
  {contenu}  
    {vers}Le serpent s'esquiva{vers}  
    {vers}Mais {gras}le regard{gras} qu'il me lança{vers}  
    {vers}Resta dans l'herbe.{vers}  
  {contenu}  
{poeme}
```

# Différents type d'encodage

	Typographique	Sémantique
LaTeX	<code>\emph{ad hoc}</code>	<code>\selectlanguage{latin}{ad hoc}</code>
HTML	<code>&lt;i&gt;ad hoc&lt;/i&gt;</code>	<code>&lt;i lang="la"&gt;ad hoc&lt;/i&gt;</code>
TEI	<code>&lt;hi rend="i"&gt;ad hoc&lt;/hi&gt;</code>	<code>&lt;foreign xml:lang="la"&gt; ad hoc&lt;/foreign&gt;</code>

# Exercice

## Relever les différences d'encodage

[./Exemples\\_encodage/Mon\\_reve\\_familier](#)



A decorative graphic on the left side of the slide consisting of two blue squares. The top square is a lighter shade of blue and is positioned above the bottom square, which is a darker shade of blue. They are aligned to the left of the main text area.

# Caractéristiques du XML

Format de données simple, documenté  
Conçu pour la description des documents textuels  
Pas de jeu de balise prédéfini

# Standard international

Depuis 1998, XML est un langage libre et documenté. C'est un standard respectant les recommandations du W3C, il facilite :

- la **lisibilité** par les machines ou par l'œil humain ;
- l'**échange** de données ;
- la **migration** vers d'autres plateformes, d'autres logiciels, d'autres formats.

A decorative graphic on the left side of the slide consisting of two blue squares. The top square is light blue and the bottom square is a darker blue, both positioned to the left of the main text area.

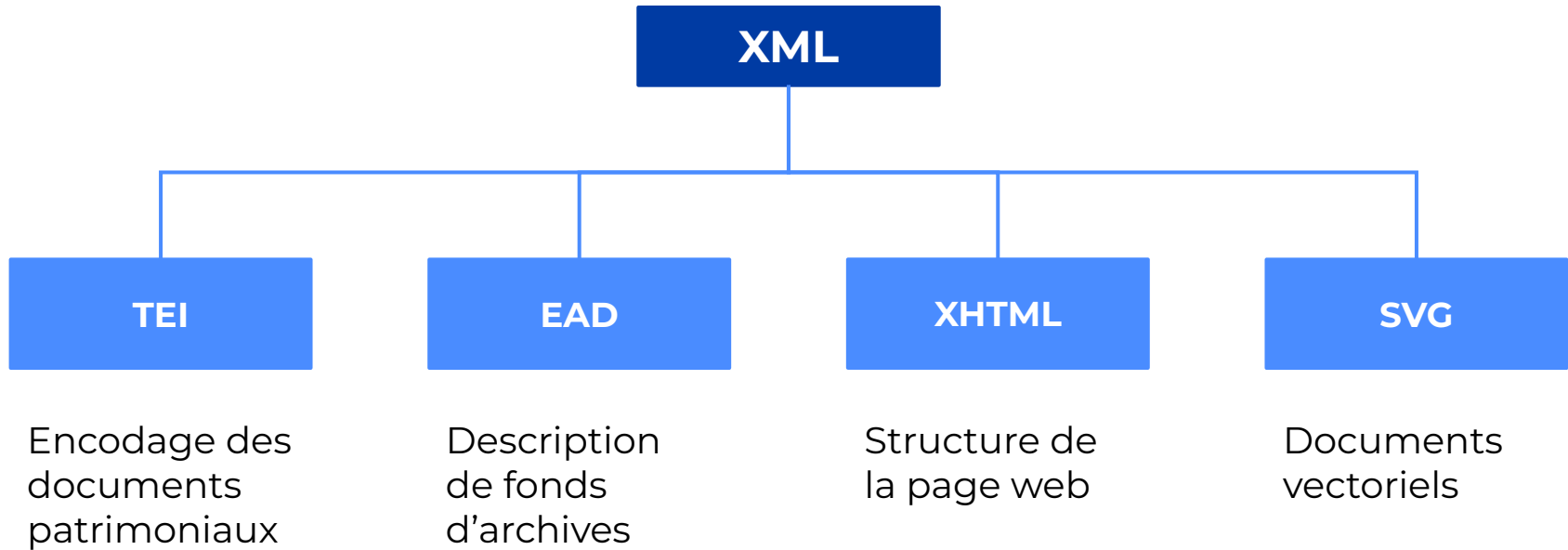
# Historique

**1970** : SGML, *Standard Generalized Markup Language* ;

**1989-1992** : HTML, *HyperText Markup Language* ;

**1998** : XML, *eXtensible Markup Language*.

# Les langages XML



# Principes fondamentaux

## Balises

Les données textuelles sont repérées de part et d'autres par des "étiquettes"

## Imbrication

Les éléments sont contenus les uns dans les autres et ne se chevauchent pas

## Héritage

Les éléments enfants héritent des éléments parents

# Frankenstein Volume I Draft in Chapter Sequence

Author(s) : Mary Wollstonecraft Shelley

Date Written : [August or September]-[?December] 1816

Title/Literary Work : Frankenstein

State : draft

Institution : Bodleian Library, University of Oxford

Hand(s) : Mary Shelley, Percy Shelley

Shelfmark : MS. Abinger c. 56

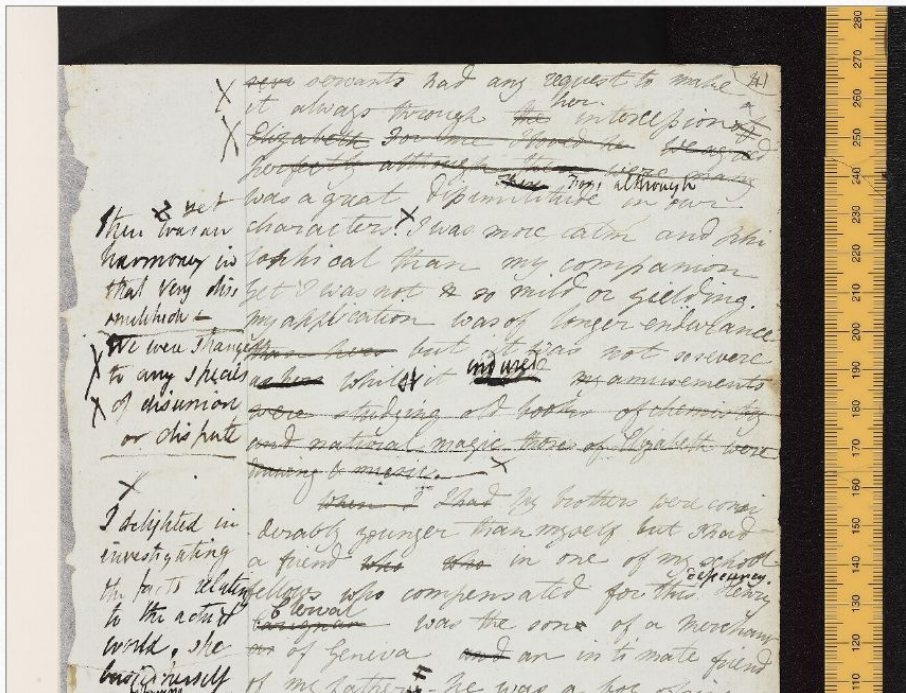
Folio : 4r

Transcription Status: ● ● ●  
Metadata Status: ● ● ●



LIMIT VIEW:

☒ All ☐ Mary Shelley ☐ Percy Shelley



X  
X

X yet  
there was an  
harmony in  
that very diss  
imilitude-

X We were strangers  
to any species  
X of disunion  
or dispute

41

4

sevr servants had any request to make

it always through <sup>her</sup> the intercession of  
Elizabeth For me I loved he We agreed  
perfectly although there were many

<sup>There For, although</sup>  
was a great dissimilitude in our  
characters. X I was more calm and phi  
losophical than my companion  
Yet I was not <sup>a</sup> so mild or yielding.  
My application was of longer endurance  
than hers but it was not so severe

<sup>endured</sup>  
as hers whilst it lasted my amusements  
were studying old books of chemistry  
and natural magic those of Elizabeth were  
<sup>X</sup>  
drawing & music.

<sup>When I had</sup> My brothers were consi  
derably younger than myself but I had  
a friend who ~~who~~ in one of my school  
fellows who compensated for this. Henry

<sup>Clerval</sup>  
Carignan was the sons of a merchant  
an of Geneva and an intimate friend  
of my father's - he was a boy of singu



Antwerp M 15.6 SHF 3A-306	<a href="#">Close</a> <a href="#">Preferences</a> <a href="#">Show Page Image</a>	BerlIn Redhiger 3 SHF 3A-306	<a href="#">Close</a> <a href="#">Preferences</a>	Bern A14 SHF 3A-306	<a href="#">Close</a> <a href="#">Preferences</a>
------------------------------	---	---------------------------------	--	------------------------	--

[SHF 3A-306 sync](#)

Le [duc de Berri](#), madame Jehanne d'Erminac trespassee, sa premiere femme, avoit grant ymagination et bien le monstra que secondement il feust marié. Car si tost come il peut veoir qu'il avoit failly il ne fut onques avisé ne eut sejour. Mais il mist clers en oeuvre et messagiers pour envoyer devers le [conte Gascon de Foix](#), qui avoit en garde la fille au conte Jehan de Bouloingne et avoit eue plus de neuf ans. Et pour tant que le [duc de Berri](#) ne pouoit parvenir a ce mariage fors par le dangier du conte de Foix. Car au fort le dit conte ne pour pere, ne pour mere, ne pour pape, ne pour amy que la damoiselle eust, il n'en eust riens fait s'il ne lui fut bien venu a plaisir. Il en parla au roy de France, son nepveu, et au [duc de Bourgoingne, son frere](#), et leur pria affectueusement qu'ilz s'en voulsissent charger et emboisngnier avecques lui. Le roy de France en eut bon ris pour tant que le [duc de Berry](#) estoit ja tout ancien et lui dit: "Bel oncle, que feréz vous d'une fillette qui n'a que XII ans et vous en avéz LX? Par ma foy, il me semble grant folie de vous de penser a telle chose. Faictes en parler pour Jehan, beau cousin vostre filz, qui est jeune et a venir, la chose est mieulx pareille a lui que a vous." "Monseigneur," respondit le [duc de Berri](#), "on en a parlé mais le conte de Foix n'y vault entendre pour la cause de ce que mon filz descend de [ceulx d'Armignac](#) et ilz sont en guerre et en hayne et ont esté long temps ensemble. Se la fille est jeune, je l'espargneray trois ou quatre ans tant qu'elle sera femme parfaite et formee." "Voire," dit le roy, "mais elle ne vous espargnera pas." Et puis dist le roy tout en riant: "Bel oncle, puis que nous veons que tant grande affection vous y avéz, nous y entendrons volentiers." Et depuis ne demoura guerres de temps que le roy ordonna le sire de la Riviere, messire Brueil, son souverain et maistre d'ostel et chambrelain pour aller en ce voyage et avecques lui le conte d'Acy. Et le [duc de Bourgoingne](#) y ordonna pour lui et en son nom l'evesque d'Authun et messire Guillaume de la Trimouille. Et le [duc de Berry](#) y ordonna en pria le conte Jehan de Xancerre ung grant saige et vaillant chevalier. Ces cinq seigneurs pour venir vers le conte de Foix et pour requerir ou nom [p341 r](#) et pour le [duc de Berry](#) ceste jeune dame a mariage, se partirent de leurs lieux et se devoient trouver en [Avignon](#) ainsi comme ilz firent deléz le [pape Clement](#) bien XV jours, qui cousin germain estoit a la damoiselle et environ la Chandeller ilz se partirent d'[Avignon](#) et prindrent le chemin de [Nîmes](#) et de [Montpellier](#) et chevaucherent a petites journées et a grans fraiz et passerent [Besiers](#) et vindrent a [Carcassonne](#) ou ilz trouverent messire Loys de Xancerre, mareschal de France, qui les recueillit grandement. Ilz se partirent de la et vindrent a [Thoulouse](#) et de la s'arrestèrent et envoyerent leurs messagiers devant devers le conte de Foix, qui se tenoit a [Orthaix en Biene](#). Si s'attenterent les traittiéz de ce mariage, mais ilz y furent moult longuement, car du commencement le conte de Foix fut moult froit pour ce que le duc de Lancastre, qui se tenoit pour le temps a [Bourdeaulz](#) ou a

[SHF 3A-306 sync](#)

Le [duc de Berry](#), madame Jehanne d'Armeignach sa premiere femme trespassee, avoit grant ymagination. Et bien le monstra que secondement il fust remarié. Car tantost que il pot veoir comment il avoit failly a la fille de Castille, il ne fut onques aise ne n'ot repos. Mais mist en oeuvre clers et messagiers pour envoyer devers le conte [Gascoing de Foix](#), qui avoit en garde la fille au conte Jehan de Boulongne et avoit ja eue plus de nouef ans. Et pour tant que le [duc de Berry](#) ne pouoit a ce mariage parvenir fors par les dangiers du conte de Foix. Car au fort le dit conte ne pour pere ne pour mere ne pour pape ne pour amy que la damoiselle eust, il n'en eust riens fait, se bien [p371 r](#) ne luy feust venu a sa plaisir. Il en parla au roy de France, son nepveu, et au [duc de Bourgoingne, son frere](#). Et leur requist moult affectueusement que ilz s'en voulsissent charger, avecques luy et emboisngnier. Le roy de France en ot bon ris, pour tant que le [duc de Berry](#) estoit ja tout ancien, et luy dist: "Beaulx oncles, que feréz vous de une si jenne femme? Elle n'a que douze ans et vous en avéz soixante. Par ma foy, c'est grant folie pour vous de penser a telle besoigne. Faictes ent parler pour Jehan beau cousin vostre filz, qui est jenne et a venir. La chose est trop mieulx pareille a

[p341 r](#) luy que a vous." "Monseigneur," respondi le [duc de Berry](#), "on en a parlé, mais le conte de Foix n'y a voulu entendre pour la cause de ce que mon filz descend de [ceulx d'Armeignach](#). Et ilz sont en guerre et en hayne et ont esté long temps ensemble. Se la fille est jenne, je la espargneray trois ou quatre ans, voire si longuement qu'elle sera femme parfaite et fourmee." "Voire," respondi le roy, "beaulx oncles, mais elle ne vous espargnera pas." Et puis dist le roy tout en riant: "Or ça, beaulx oncles, puisque nous veons que si tres grande affection vous y avéz nous y entendrons moult volentiers."

Depuis ne demoura gaires de temps que le roy ordonna le sire de la Riviere, messire Burel, son souverain chevalier et maistre d'ostel et chambellan, pour aler en ce voyage. Et avecques luy le conte d'Assy. Et le [duc de Bourgoingne](#) y ordonna pour luy et en son nom l'evesque d'Authun et messire Guillemme de la Trimouille et le [duc de Berry](#) y ordonna et en pria le conte Jehan de Xancerre ung grant, sage et vaillant chevalier.

¶ Ces cinq seigneurs pour venir devers le conte de Foix et requerir en nom de mariage pour le [duc de Berry](#) ceste jenne dame, se partirent de leurs lieux et se devoient trouver en [Avignon](#) deléz le pape comme ilz firent. Ce [pape Clement](#), qui cousin estoit du pere a la damoiselle, les retint la bien quinze jours. Et environ la Chandelleur ilz se departirent tous d'[Avignon](#) et prindrent le chemin de Nîmes et de [Montpellier](#). Si chevauchierent a petites journées et a grans frais. Si passerent [Besiers](#) et vindrent a



Le [duc de Berry](#) madame Jehenne d'Armignac trespassee sa premiere avoit grant ymagination et bien le monstra que secondement il fut marié. Car si [p333 v](#) trestost come il peut veoir qu'il avoit failli il ne fut onques aise ne eust sejour mais il mist clers en euvre et messagiers pour envoyer devers le [conte Gascon de Foix](#) qui avoit en garde le [filz](#) au conte Jehan de Bouloungne et avoit eue plus de IX ans. Et pour tant que le [duc de Berry](#) ne pouoit a ce mariage parvenir fors par le dangier du conte de Foix. Car au fort le dit conte ne pour pere, ne pour mere, pour [pape](#), ne pour amy que la damoiselle eust il n'en eust riens fait s'il ne lui fust bien venu a plaisir. Il en parla au roy de France, son nepveu, et au [duc de Bourgoingne, son frere](#), et leur pria affectueusement qu'ilz s'en voulsissent charger et embesongnier avecques lui. Le roy de France en eut bon ris pour tant que le [duc de Berry](#) estoit ja tout ancien et lui dist: "Bel oncle, que feréz vous d'une telle fillette? Elle n'a que XII ans et vous en avéz LX. Par ma foy, c'est grant folie pour vous de penser a telle chose. Faictes ent parler pour Jehan, beau cousin vostre filz, qui est jenne et a venir. La chose est mieulx pareille a lui que a vous." "Monseigneur," respondit le [duc de Berry](#), "on en a parlé mais le conte de Foix n'y vault entendre pour la cause de ce que mon filz descend de [ceulx d'Armignac](#) et ilz sont en guerre et en hayne et ont esté longtemps ensemble. Se la fille est jenne, je l'espargneray trois ou quatre ans tant qu'elle sera femme parfaite." "Voire," respondit le roy, "bel oncle, mais elle ne vous espargnera pas." Et puis dist le roy tout en riant: "Bel oncle, puis que nous veons que si grande affection vous y avéz nous y entendrons volentiers."

¶ Depuis ne demoura gaires de temps que le roy ordonna le sire de la Riviere, messire Brueil, son souverain maistre d'ostel et chambrelain, pour aler en ce voyage et avecques lui le conte d'Acy. Et le [duc de Bourgoingne](#) y ordonna pour lui et en son nom l'evesque d'Autun et messire Guillaume de la Trimouille. Et le [duc de Berry](#) y ordonna et en pria le conte Jehan de Xancerre, un grant saige et vaillant chevalier.

¶ Ces cinq seigneurs pour venir vers le conte de Foix et requerir ou nom de mariage pour le [duc de Berry](#) ceste jenne dame se partirent de leurs lieux et se devoient trouver en [Avignon](#) aussi comme ilz firent la deléz le [pape Clement](#) bien XV jours, qui cousin germain estoit a la damoiselle, et environ la Chandeleur ilz se departirent d'[Avignon](#) et prindrent le chemin de Nîmes et de [Montpellier](#) et chevaucèrent a petites journées et a grans fraiz et passerent [Besiers](#) et vindrent a [Carcassonne](#) [p334 r](#) ne et trouverent messire Loys de Xancerre, mareschal de France, qui les recueillit grandement, et parla a eulx assés du conte de Foix de son estat et de son affaire. Car il y avoit esté n'avoit point

Passer la souris sur le fac-similé pour faire apparaître la transcription





```
<desc>
<fileDesc>
  <langUsage>
    <language ident="fre">français (langue de l'édition,<0xa0>langue par défaut)</lang>
    <language ident="frm">français moyen (langue principale du document édité)</language>
    <language ident="lat">latin (langue utilisée dans le document édité)</language>
  </langUsage>
</fileDesc>
<revisionDesc>
  <change who="#KF #FC">
    <date when="2009-07">juillet 2009</date>
    <desc>nouvelle conversion en TEI,<0xa0>intégrant notamment un index enrichi,<0xa0>
      corrections manuelles</desc>
  </change>
  <change who="#FC">
    <date when="2010-01">août 2009-janvier 2010</date>
    <desc>enrichissement manuel du document TEI,<0xa0>en particulier pour l'indexation
  </change>
  <change who="#KF #DB #FC">
    <date when="2010-05">mai 2010</date>
    <desc>corrections dans les index,<0xa0>établissement de nouveaux liens et renvois
  </change>
  <change who="#KF #FC">
    <date when="2010-12">décembre 2010</date>
    <desc>nouvelle conversion en TEI à partir d'une édition au traitement de texte incluant
      un index complet de toutes les entités nommées,<0xa0>et intégrant les dernières
      corrections de transcription</desc>
  </change>
  <change who="#KF #FC">
    <date when="2014">2014</date>
    <desc>relecture et corrections de l'édition et des index,<0xa0>intégration de
      l'introduction</desc>
  </change>
  <change who="#FC">
    <date when="2015-04">avril 2015</date>
    <desc>corrections de l'édition,<0xa0>développement du site web définitif<0xa0>

```

# Observer un document XML

[./Christofle-septembre2016.xml](#)

## L'année 1437 dans la pratique de Pierre Christofle, notaire du Châtelet d'Orléans

ACCUEIL | INTRODUCTION | ÉDITION DES NOTES | TYPES D'ACTES | LIEUX ET PERSONNES



Reproduction de Pierre Christofle sur un document  
portant le sceau de la prévôté d'Orléans  
ADL, H Dépôt2, 1B54, 1<sup>er</sup> août 1442)

Pierre Christofle fut notaire royal à Orléans de 1423 à 1450, attaché à la prévôté pour qui il rédigeait des contrats portant le sceau de l'institution. Comme ses confrères, Pierre Christofle inscrivait dans un registre et en une forme abrégée les conventions qu'il attestait. Ces notes, au nombre de 387, rédigées entre le 1<sup>er</sup> janvier et le 29 décembre 1437, conservées aux Archives départementales du Loiret sous la cote 3E 10144, sont ici éditées dans leur ensemble. L'édition, indexée et prochainement téléchargeable, permet l'enquête sur le lexique des actes. Elle accompagne et complète une version imprimée, publiée sous le titre *Promettre, confesser, s'obliger devant Pierre Christofle, notaire royal d'Orléans [1437]*, dans la collection « Mémoires et documents de l'École des chartes », qui propose un essai interprétatif de cette documentation.

La mise en ligne autorise l'accès rapide au contenu par le biais de plusieurs outils de consultation, une vue en parallèle des images numérisées du registre et de leur transcription afin de mieux comprendre (voire d'étudier) la mise en scène de l'écrit notarié, et, bientôt, la possibilité de télécharger l'ensemble des notes.

L'édition électronique intégrale des notes du notaire Pierre Christofle pour l'année 1437 est un projet d'abord mené à l'université d'Ottawa par Kouky Fianu, professeur à l'université d'Ottawa, avec la collaboration d'Anne Fortier, étudiante de maîtrise de l'université d'Ottawa, puis développé dans sa version électronique par Florence Clavaud, conservateur en chef du patrimoine

# Observer le résultat d'une édition XML

[elec.enc.sorbonne.fr/christofle/index.html](http://elec.enc.sorbonne.fr/christofle/index.html)

# Exercice

**Relever les différents  
“composants” du XML**



# Les “composants” du XML

## Éléments

Les balises ainsi que leur contenu

## Attributs

Les “caractéristiques” d’une balise

## Entités

Appel à un caractère ou une séquence prédéfinie

## Commentaires

Pour extraire des passages du XML

# Document XML

```
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <element><!-- comment -- ></element>
  <empty/>
  <el attribute="value">&lt;/el>
</root>
```

# Le prologue

```
<?xml version="1.0" encoding="UTF-8"?>
```

Il indique la version de la recommandation XML qui sert de base à l'écriture du fichier (il n'y a eu à cette date qu'une seule version, la version 1.0), ainsi que l'encodage de caractères utilisé.

# Les éléments

```
<element>texte</element>
```

```
<elementVide/>
```

## Dénomination

Un élément doit être entouré de deux balises de même nom (sensible à la casse)

## Clôture

Un élément ouvert doit être fermé obligatoirement

## Imbrication

Un élément ne peut pas en chevaucher un autre

# Les attributs

```
<miseEnValeur position="centrePage">  
    texte  
</miseEnValeur>
```

## Décrire

Utilisés pour décrire le contenu d'un élément

## Situation

Placé les uns à la suite des autres dans la balise ouvrante

## Répétabilité

Un même attribut ne peut pas être réutilisé dans la même balise



# Les entités

&entité;

## Caractère

Est utilisée à la place d'un caractère interdit en XML comme le “&”

## Remplacement

Mise à la place d'une séquence de code définie dans une DTD

## Conversion

[hapax.qc.ca/conversion.fr](http://hapax.qc.ca/conversion.fr)

# Les commentaires

```
<!-- texteCommentaire -->
```

## Ajout

Destinés à ajouter une note généralement pour préciser le contexte à un lecteur humain

## Hors XML

Pour baliser une chaîne de caractère non contrainte par le standard XML

# Règles à retenir

**Balise ouvrante =  
balise fermante**

**1 seul élément racine**

**Respect de  
l'imbrication**

**1 attribut du même  
nom par balise**

= encodage

**bien formé**

# Exercice

**Encoder un court  
passage**



# Encoder

À l'aide des balises suivantes, structurer le texte proposé :

```
<texte>  
<chapitre>  
<titreChapitre>  
<paragraphe>  
<nomPersonnage>  
<noteDeBasDePage>
```

## Blanche-Neige mange la pomme empoisonnée

Il était une fois Blanche-Neige<sup>1</sup> que sa belle-mère détestait. [...]

## Blanche-Neige attend le prince Charmant

Un jour Charmant<sup>2</sup> arriva sur son cheval blanc et la sauva. [...]

<sup>1</sup> Elle porte ce nom, car son visage est blanc comme la neige

<sup>2</sup> Nom du prince qui sauve Blanche-Neige.

# Exercice

**Bien formé ou mal  
formé ?**



# Bien formé ?

```
<paragraphe>du texte</paragraphe>
```

```
<paragraphe><article>du</article><nom>texte</nom></paragraphe>
```

```
<paragraphe><article>du <nom></article>texte</nom></paragraphe>
```

```
<paragraphe><article>du <nom>texte</nom></article></paragraphe>
```

```
<paragraphe type="texte">du texte</paragraphe>
```

```
<paragraphe type=texte>du texte</paragraphe>
```

```
<paragraphe type="texte">du texte<paragraphe/>
```

```
<paragraphe type="texte">du texte<nom>nom</paragraphe>
```

```
<paragraphe type="texte">du texte</Paragraphe>
```

```
<segment type="texte" type="nombre">du texte</paragraphe>
```