

Atelier 2 - Encodage et annotation du texte

Atelier "Créer une édition scientifique numérique"

Floriane Chiffolleau
Ingénieure en humanités numériques à l'ObTIC
DataLab - BNF (Salle 70)
14 mars 2025



INTRODUCTION À LA XML-TEI : POURQUOI ENCODER SES CORPUS ?

HISTORIQUE DE LA CRÉATION DE LA XML-TEI

Historique de la création de la XML-TEI

- ❑ Novembre 1987 : Création de la Text Encoding Initiative (TEI)
 - Pallier la prolifération de systèmes divers et incompatibles pour la représentation numérique de textes
- ❑ Mai 1994 : Publication de la première version officielle des TEI Guidelines (P3)
 - Règles produites par le travail combiné de nombreux ateliers de réflexions et par les révisions et extensions faites à la version P1 produite en 1990
 - Utilisation du SGML pour rédiger en TEI
- ❑ Janvier 1999-Janvier 2001 : Création du consortium TEI
 - Avoir une organisation officielle qui maintient, développe et promeut la TEI
- ❑ Juin 2002 : Publication de la version P4 des TEI Guidelines
 - Introduction du format XML (créé en 1999), comme nouveau format de rédaction de la TEI
- ❑ Novembre 2007 : Publication de la dernière version des TEI Guidelines (P5)
 - Version révisée de P4 avec de nouveaux développements pour un certain nombre de domaines
 - Abandon du SGML pour un passage complet au format XML

QU'EST-CE QUE LE FORMAT XML ?



XML ou eXtensible Markup Language (langage de balisage extensible)

- ❑ Langage pour le partage de données entre différents systèmes grâce à un format souple
- ❑ Langage pour le balisage de documents textuels, dont les chaînes de caractères sont délimitées par des étiquettes appelées *élément* qui sont encadrées par des chevrons (<, >)
- ❑ Langage extensible qui n'est pas prédéfini et qui permet la création de son propre jeu d'éléments, qui peut toujours être étendu

Comment ça fonctionne ?

La XML se structure sous forme d'**arbre** :

- ❑ Un élément racine
- ❑ Des balises imbriquées que l'on désignera comme des parents et des enfants
- ❑ Des balises, à la suite, désignées comme des frères/sœurs

Prologue de fichier : déclaration du fichier XML et de l'encodage des caractères

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

Comment ça fonctionne ?

Plusieurs **règles** importantes à respecter dans la rédaction d'un arbre XML

- ❑ **Balise ouvrante/fermante** : Une balise ouvrante devra toujours être suivie dans l'arbre par sa balise fermante

Exemple : `<titre>Présentation de personnes</titre>`

- ❑ **Sensibilité à la casse** : l'écriture de la balise fermante devra être identique à celle de la balise ouvrante

Exemple correct : `<titre>Présentation de personnes</titre>`

Exemple incorrect : `<titre>Présentation de personnes</Titre>`

Comment ça fonctionne ?

Plusieurs **règles** importantes à respecter dans la rédaction d'un arbre XML

- ❑ **Pas de chevauchement** : il faut bien faire attention à l'imbrication des balises

Exemple correct :

```
<personne><nom>Martin</nom></personne>
```

Exemple incorrect :

```
<personne><nom>Martin</personne></nom>
```

- ❑ **Attributs** : les éléments peuvent être enrichis d'attributs et leurs valeurs

Exemple :

```
<personne n="1" type="homme"  
  profession="professeur">  
  <nom>Martin</nom>  
  <prenom>Jean</prenom>  
</personne>
```

- ❑ Les **valeurs** des attributs doivent toujours être mises entre **guillemets ou apostrophes**

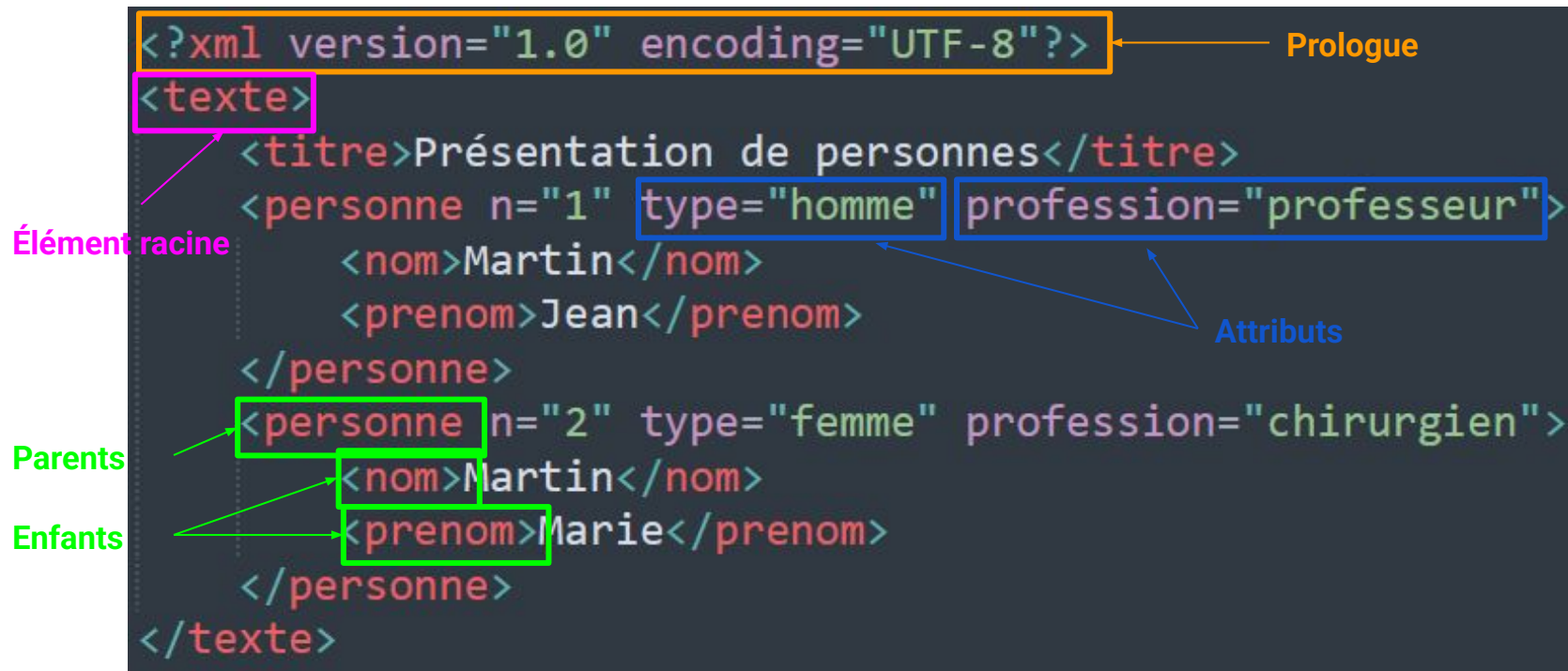
Exemple correct :

```
<personne n="1" type="homme">
```

Exemple incorrect :

```
<personne n="1" type=homme>
```

Comment ça fonctionne ?



QU'EST-CE QUE LA TEI ?

Les TEI Guidelines

- ❑ **Règles** définies pour l'encodage de tous les textes sous leur format numérique et lisibles par machine
- ❑ Une liste longue, détaillée et en **constante évolution**
- ❑ De **nombreuses balises** pour encoder des corpus de type, langue et structure différentes (romans, pièces de théâtre, poème, lettres, rapport officiel, etc.)
- ❑ Page unique pour **chaque balise et attribut**, afin d'avoir toutes les informations sur la manière dont ils s'utilisent

Corps du texte

- ❑ 1 [The TEI Infrastructure](#)
- ❑ 2 [The TEI Header](#)
- ❑ 3 [Elements Available in All TEI Documents](#)
- ❑ 4 [Default Text Structure](#)
- ❑ 5 [Characters, Glyphs, and Writing Modes](#)
- ❑ 6 [Verse](#)
- ❑ 7 [Performance Texts](#)
- ❑ 8 [Transcriptions of Speech](#)
- ❑ 9 [Computer-mediated Communication](#)
- ❑ 10 [Dictionaries](#)
- ❑ 11 [Manuscript Description](#)
- ❑ 12 [Representation of Primary Sources](#)
- ❑ 13 [Critical Apparatus](#)
- ❑ 14 [Names, Dates, People, and Places](#)
- ❑ 15 [Tables, Formulæ, Graphics, and Notated Music](#)
- ❑ 16 [Language Corpora](#)
- ❑ 17 [Linking, Segmentation, and Alignment](#)
- ❑ 18 [Simple Analytic Mechanisms](#)
- ❑ 19 [Feature Structures](#)
- ❑ 20 [Graphs, Networks, and Trees](#)
- ❑ 21 [Non-hierarchical Structures](#)
- ❑ 22 [Certainty, Precision, and Responsibility](#)
- ❑ 23 [Documentation Elements](#)
- ❑ 24 [Using the TEI](#)

Un consortium collaboratif

jTEI : Journal de la TEI :

<https://journals.openedition.org/jtei/>

- ❑ Journal officiel du consortium TEI
- ❑ Articles sur l'état de l'art, des innovations en matière de TEI et des exploitations dans des projets
- ❑ Articles rigoureusement évalués avant publication sur la plateforme Open Edition

Conference/Annual meeting :

<https://members.tei-c.org/Events/meetings/>

- ❑ Partage et découverte des projets en cours sur la TEI
- ❑ Rencontre avec les autres membres de la communauté
- ❑ Suivi officiel du fonctionnement du bureau du consortium

Un consortium collaboratif

Special Interest Groups (SIGs) :
<https://tei-c.org/activities/sig/>

- ❑ Groupe de réflexion, d'échanges et de débats sur des sujets spécifiques liés à la TEI
- ❑ Aide au développement et à l'évolution de la TEI
- ❑ Trois exemples de SIGs :
 - ❑ Computer-Mediated Communication SIG
 - ❑ Correspondence SIG
 - ❑ East Asian/Japanese SIG

TEI-L mailing-list :
<https://tei-c.org/support/#tei-l>

- ❑ Ouverte à toute la communauté TEI, pour poser des questions ou y répondre
- ❑ Partage d'expertise ou d'expérience
- ❑ Archivage de tous les problèmes précédemment rencontrés

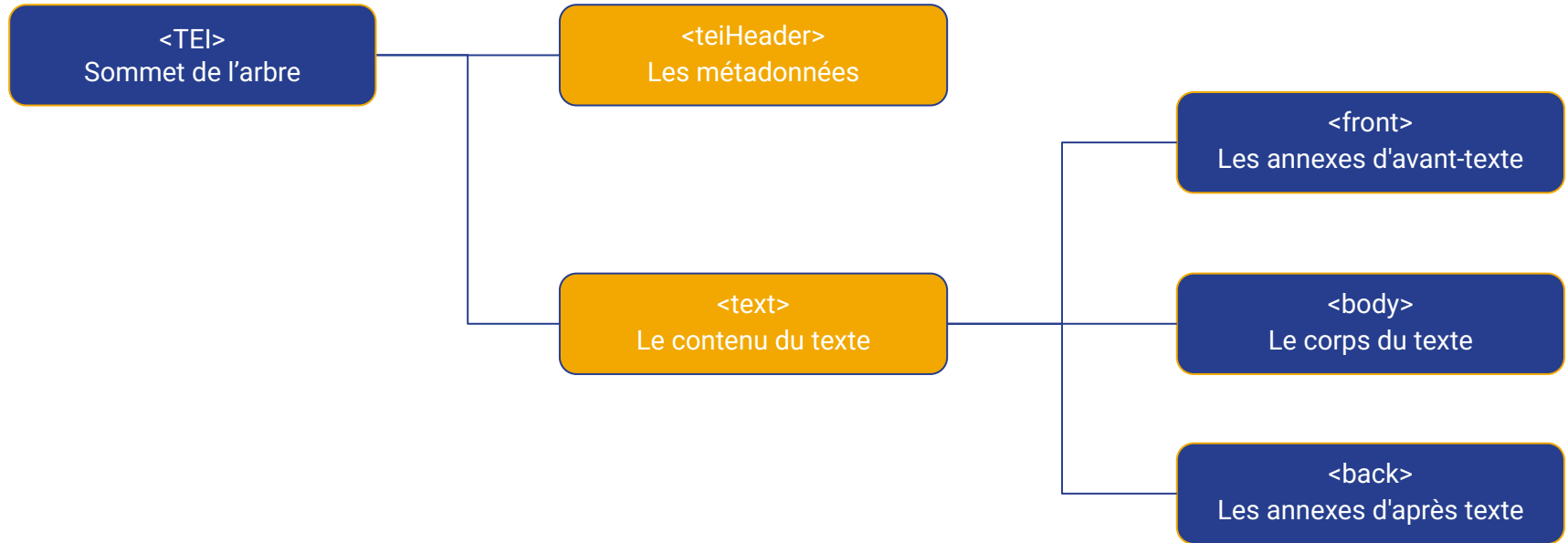
CRÉER SON ÉDITION SCIENTIFIQUE NUMÉRIQUE

GitHub :


**[https://github.com/FloChiff/AtelierObTIC-creer-une-
edition-scientifique-numerique](https://github.com/FloChiff/AtelierObTIC-creer-une-edition-scientifique-numerique)**

LA TEI AU SERVICE DE SON ÉDITION

Contenu de l'arbre XML-TEI



Oxygen XML Editor

- ❑ Éditeur pour tout type de document utilisant le format XML : <https://www.oxygenxml.com/>
- ❑ Autocomplétion → si *namespace* déclaré et vocabulaire défini, suggestions d'utilisation de balises, selon la hiérarchie imposée par la TEI
- ❑ Quelques petits raccourcis pour l'utilisation d'Oxygen
 - ❑ Balisage rapide : Cmd + E (Mac) ; Ctrl + E (Linux)
 - ❑ Indenter : Bouton  ou Cmd + Shift + P (Mac) ; Ctrl + Shift + P (Linux)
 - ❑ Recherche/Remplace : Cmd + F (Mac) ; Ctrl + F (Linux)

Créer son fichier XML-TEI

- ❑ Ouvrir Oxygen XML Editor
- ❑ Créer un nouveau document et choisir “Document XML”
- ❑ Enregistrer le document sous le nom de “atelier_obtic.xml”
- ❑ Créer une nouvelle balise appelée **TEI** et mettre l’attribut suivant `xmlns=` <http://www.tei-c.org/ns/1.0>
- ❑ Ajouter les deux balises enfants **teiHeader** et **text**

À ce niveau, qu’observez-vous ?

LES MÉTADONNÉES - PRÉSENTATION

Les métadonnées : le <teiHeader>



Les métadonnées : le <teiHeader>

Chapitre 2 The TEI Header

<https://tei-c.org/release/doc/tei-p5-doc/fr/html/HD.html>

Le <fileDesc>

- ❑ **<titleStmt>** (mention de titre) regroupe les informations sur le titre d'une œuvre et les personnes ou institutions responsables de son contenu intellectuel.
- ❑ **<editionStmt>** (mention d'édition) regroupe les informations relatives à l'édition d'un texte.
- ❑ **<publicationStmt>** (mention de publication) regroupe des informations concernant la publication ou la diffusion d'un texte électronique ou d'un autre type de texte.

Les métadonnées : le <teiHeader>

Le <fileDesc>

- ❑ <seriesStmt> (mention de collection) regroupe toute information relative à la collection (si elle existe) à laquelle appartient une publication.
- ❑ <notesStmt> (mention de notes) rassemble toutes les notes fournissant des informations sur un texte, en plus des informations mentionnées dans d'autres parties de la description bibliographique.
- ❑ <sourceDesc> (description de la source) décrit la source à partir de laquelle un texte électronique a été dérivé ou produit, habituellement une description bibliographique pour un texte numérisé ou une expression comme "document numérique natif " pour un texte qui n'a aucune existence précédente.

Les métadonnées : le <teiHeader>

L'<encodingDesc>

- ❑ Décrire l'encodage du fichier électronique : description du projet <projectDesc>, déclaration d'échantillonnage <samplingDecl>, déclaration des pratiques éditoriales <editorialDecl> et autres types de déclarations (balisage, classifications, coordonnées géographiques, etc).

Le <profileDesc>

- ❑ Fournir la description détaillée des aspects non bibliographiques : date de création du texte <creation>, langues utilisées dans le texte <langUsage>, etc.

Le <revisionDesc>

- ❑ Fournir un résumé de l'historique de révisions d'un fichier. Encodage avec la balise <change> et à l'intérieur les attributs @who (responsable de la modification) et @when (date de la modification).
- ❑ Encodage à faire dans un ordre chronologique inversé (plus récent en premier)

Les métadonnées : le <teiHeader>

Le <fileDesc>

❏ <titleStmt>

```
<titleStmt>
  <title type="main">Orgueil et préjugé</title>
  <title type="sub">Par l'auteur de Raison et Sensibilité</title>
  <title type="sub">Traduit de l'anglais</title>
  <author>Jane Austen</author>
  <respStmt>
    <resp>Transcription by</resp>
    <persName>Floriane Chiffolleau</persName>
  </respStmt>
  <respStmt>
    <resp>Encoded by</resp>
    <persName>Floriane Chiffolleau</persName>
  </respStmt>
</titleStmt>
```

Les métadonnées : le <teiHeader>

Le <fileDesc>

❏ <publicationStmt>

```
<publicationStmt>
  <publisher>J. J. Paschoud, libraire</publisher>
  <pubPlace>Paris</pubPlace>
  <date>1822</date>
</publicationStmt>
```

❏ <seriesStmt>

```
<seriesStmt>
  <title type="main">Littérature anglaise</title>
</seriesStmt>
```

❏ <sourceDesc> → nativement numérique

```
<sourceDesc>
  <p>Born digital</p>
</sourceDesc>
```

❏ <sourceDesc> → extrait d'un manuscrit

```
<sourceDesc>
  <msDesc>
    <msIdentifier>
      <country>France</country>
      <settlement>Paris</settlement>
      <institution>BNF</institution>
      <collection>Réserve des livres rares</collection>
      <idno>RES P-Y2-2802</idno>
    </msIdentifier>
  </msDesc>
</sourceDesc>
```

Les métadonnées : le <teiHeader>

L'<encodingDesc>

❑ <projectDesc>

```
<projectDesc>
  <p>L'encodage de ce document s'est fait dans le cadre d'une formation d'introduction à la TEI.</p>
</projectDesc>
```

❑ <samplingDecl>

```
<samplingDecl>
  <p>Seuls les textes de littérature anglaise de début du XIXème siècle ont été choisis.</p>
</samplingDecl>
```

❑ <editorialDesc>

```
<editorialDecl>
  <correction>
    <p>Aucune correction</p>
  </correction>
  <hyphenation eol="hard" rend="sh">
    <p>Toutes les coupures de mots pour fin de ligne (indiquées principalement avec un tiret simple) ont été conservées.</p>
  </hyphenation>
</editorialDecl>
```

Les métadonnées : le <teiHeader>

Le <profileDesc>

❏ <creation>/<langUsage>

```
<creation when="1813"/>
<langUsage>
  <language ident="fr"/>
  <!-- <language ident="fr" usage="75">Français</language>
  <language ident="en" usage="25">Anglais</language>-->
</langUsage>
```

Le <revisionDesc>

```
<revisionDesc>
  <change when-iso="2021-02-17" who="#floriane.chiffoleau">Added a particDesc and settingDesc in the profileDesc</change>
  <change when-iso="2021-02-16" who="#floriane.chiffoleau">Encoding of the file</change>
</revisionDesc>
```

LES MÉTADONNÉES - TRAVAUX PRATIQUES

Encoder les métadonnées : Exercice

En utilisant les métadonnées de l'ouvrage Le Tour du monde en quatre-vingts jours présent dans le document metadonnees.md dans le dossier GitHub de l'atelier 2 et ce qui figure dans la présentation et les exemples précédents, encoder votre <teiHeader>

Encoder les métadonnées : Résultat

À RETROUVER UNE FOIS L'ATELIER PASSÉ

LE CORPS DU TEXTE - PRÉSENTATION

Le corps du texte : le <body>

Chapitre 3 Elements Available in All TEI Docs <https://tei-c.org/release/doc/tei-p5-doc/fr/html/C0.html#C0PU>

Chapitre 4 Default Text Structure <https://tei-c.org/release/doc/tei-p5-doc/fr/html/DS.html>

Encodage du <body> à l'aide de divisions et subdivisions en utilisant la balise <div>. Trois attributs peuvent principalement être utilisés avec :

- ❑ **@xml:id** → donner un attribut bien spécifique à la division pouvant être référencé par la suite.
- ❑ **@type** → spécifier le type de divisions que l'on a (par exemple, tome, chapitre, poème, lettre, etc.)
- ❑ **@n** → numérotation et associé généralement au type que l'on aura défini avant

Possible de mettre un en-tête en début de division (titre de section, intitulé de liste, description de manuscrit, etc.) avec la balise <head>.

Le corps du texte : le <body>

Chapitre 3 Elements Available in All TEI Docs <https://tei-c.org/release/doc/tei-p5-doc/fr/html/CO.html#COPU>

Chapitre 4 Default Text Structure <https://tei-c.org/release/doc/tei-p5-doc/fr/html/DS.html>

Deux autres balises importantes à retenir, des balises autofermantes :

- ❑ Saut de ligne **<lb/>** → à mettre en fin de ligne pour faire respecter la structure
 - ❑ **@break="no"** (le mot se poursuit dans la ligne suivante)
- ❑ Saut de page **<pb/>** → à mettre lorsque le document passe à la page suivante
 - ❑ **@n** → Numéro de page
 - ❑ **@facs** → information sur l'image sur lequel figure le texte transcrit

Le corps du texte : le <body>

Diverses balises en fonction du type de texte encodé :

- ❑ **<p>** → marquer les paragraphes dans des textes en prose
- ❑ **<lg>** (groupe de vers) et **<l>** (vers) → encoder les poèmes ou les pièces de théâtre
- ❑ **<list>** (liste) et **<item>** (composant) → encoder une liste et ses composants
- ❑ **<table>** (tableau), **<row>** (ligne) et **<cell>** (cellule) → encoder le contenu d'un tableau
- ❑ **<figure>** (figure) → encoder une illustration. Elle s'accompagne généralement d'une balise **<graphic>** qui contiendra le lien URL vers l'image. Elle peut aussi être accompagnée d'un **<figureDesc>** qui servira à décrire l'image.

Bien d'autres balises existent pour encoder les textes, de manière plus ou moins détaillée, en fonction du type de texte sur lequel on travaille (une pièce de théâtre, un roman, un poème, un discours, etc.). Quelques chapitres des guidelines sont dédiés exclusivement à certains de ces textes.

Le corps du texte : le <body>

Exemples

Un extrait de roman : <div> imbriqués contenant des attributs et composés d'une balise paragraphe, de sauts de ligne et d'un saut de page :

```
<div type="book" n="3">
  <div type="chapter" n="9">
    <p>Elisabeth avoit arrangé dans sa tête<lb/> que Mr. Darcy lui amèneroit sa sœur le<lb/> lendemain de son
      arrivée à Pemberley ,<lb/> et en conséquence elle étoit bien déci-<lb break="no"/>dée à ne pas s'éloigner
      de l'auberge de<lb/> toute la matinée; mais elle avoit mal<lb/> calculé, car il l'amena à Lambton le<lb/>
      jour même. Elisabeth et sa tante qui<lb/> s'étoient promenées près de là avec quel-<lb break="no"
      />ques-uns de leurs nouveaux amis, ren-<lb break="no"/>troient à l'auberge pour faire leur toi-<lb
      break="no"/>lette et aller dîner, lorsque le bruit d'un<lb/> équipage les attira vers la fenêtre, et
      elles<lb/> virent un monsieur et une dame dans un<lb/> Carriole qui s'arrêta à leur porte. Elisabeth,<lb/>
      reconnoissant la livrée, ne causa pas à son<lb/> oncle et à sa tante une légère surprise,<pb n="113"
      /><note type="foliation" place="top(right)">113</note> en leur annonçant la visite qu'elle atten-<lb
      break="no"/>doit. Jusqu'alors ils n'avoient eu aucun<lb/> soupçon de ce qui se passoit ; mais com-<lb
      break="no"/>ment expliquer l'embarras d'Elisabeth, et<lb/> toutes les attentions de Darcy? Il
      falloit<lb/> qu'il eût du penchant pour leur nièce.</p>
  </div>
</div>
```

Le corps du texte : le <body>

Exemples

Un poème

```
<div type="poem">
  <head>Exemple de poèmes</head>
  <lg type="sonnet">
    <lg type="quatrain">
      <l>Vers 1</l>
      <l>Vers 2</l>
      <l>Vers 3</l>
      <l>Vers 4</l>
    </lg>
    <lg type="quatrain"> [5 lines]
    <lg type="tercet">
      <l>Vers 1</l>
      <l>Vers 2</l>
      <l>Vers 3</l>
    </lg>
    <lg type="tercet"> [4 lines]
  </lg>
</div>
```

Un tableau

```
<div type="table">
  <head>Exemple de tableau</head>
  <table>
    <head>Informations personnelles</head>
    <row>
      <cell/>
      <cell>Genre</cell>
      <cell>Profession</cell>
    </row>
    <row> [4 lines]
    <row>
      <cell>Marie Martin</cell>
      <cell>Femme</cell>
      <cell>Chirurgien</cell>
    </row>
  </table>
</div>
```

Le corps du texte : le <body>

Exemples

Une figure

```
<div type="figure">
  <head>Exemple de figure</head>
  <figure>
    <graphic url="images/logo_obtic.jpg" width="2cm"/>
    <head>Logo de l'ObTIC</head>
    <figDesc>L'image montre le logo de l'ObTIC,
      écrit en bleu sur fond blanc.</figDesc>
  </figure>
</div>
```

Une liste


```
<div type="list">
  <head>Exemple de liste</head>
  <list>
    <head>Liste d'objets</head>
    <item>Une table</item>
    <item>Une chaise</item>
    <item>Un lit</item>
  </list>
</div>
```

LE CORPS DU TEXTE – TRAVAUX PRATIQUES

Encoder le corps du texte : Exercice

En reprenant le document XML que vous avez créé et en utilisant ce qui figure dans la présentation et les exemples précédents, encoder votre <body> avec le texte présent dans le document body.txt dans le dossier GitHub de l'atelier 2, afin de représenter tous les éléments que l'on peut trouver dans cet extrait de l'ouvrage (numéro de chapitre, titres, pagination, figures, etc.)

Encoder le corps du texte : Astuce

- ❑ Encoder d'abord les paragraphes (**<p>**) du texte en mettant les balises entre les fins et débuts de paragraphe
- ❑ Ajouter un premier **<lb/>**, copier le, puis coller à chaque fin de ligne
- ❑ Encoder le reste (**<pb>**, **<figure>**, etc.)
- ❑ Indenter ()

Encoder le corps du texte : Résultat

À RETROUVER UNE FOIS L'ATELIER PASSÉ

LES ÉLÉMENTS SPÉCIFIQUES – PRÉSENTATION

Les éléments spécifiques

3 Elements Available in All TEI Docs <https://tei-c.org/release/doc/tei-p5-doc/fr/html/CO.html>

14 Names, Dates, People, and Places <https://tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html>

16 Language Corpora - Contextual Information <https://tei-c.org/release/doc/tei-p5-doc/fr/html/CC.html#CAH>

Encodage d'entités nommées, soit tout ce qui a trait aux personnes, lieux et organisations :

❑ Encodage générique

- ❑ **<rs>** chaîne de référence
- ❑ **<name>** nom ou nom propre

} S'accompagne d'un attribut **@type** (person, place ou org) et **@ref**

❑ Encodage précis

- ❑ **<persName>** nom de personne
- ❑ **<placeName>** nom de lieu
- ❑ **<orgName>** nom d'organisation

} À accompagner d'un attribut **@ref**

Les éléments spécifiques

3 Elements Available in All TEI Docs <https://tei-c.org/release/doc/tei-p5-doc/fr/html/CO.html>

14 Names, Dates, People, and Places <https://tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html>

16 Language Corpora - Contextual Information <https://tei-c.org/release/doc/tei-p5-doc/fr/html/CC.html#CCAH>

Encodage d'informations contextuelles sur les entités nommées :

- ❑ **<particDesc>** → description des personnages et organisations référencés
 - ❑ **<listPerson>/<listOrg>**
 - ❑ **<person>/<org>**
 - ❑ **<persName>/<orgName>**
- ❑ **<settingDesc>** → description des lieux référencés
 - ❑ **<listPlace>**
 - ❑ **<place>**
 - ❑ **<placeName>**

Les éléments spécifiques : le corps du texte

Exemple d'utilisation de balises pour entités nommées génériques et spécifiques

```
<div type="named_entities">
  <div type="generic">
    <p><name type="person" ref="#p0001">Elisabeth</name> avait arrangé dans sa tête que <name type="person"
      ref="#p0002">Mr. Darcy</name> lui amènerait <rs type="person" ref="#p0003">sa sœur</rs> le lendemain
      de son arrivée à <name type="place" ref="#l0001">Pemberley</name> , et en conséquence elle était bien
      décidée à ne pas s'éloigner de <rs type="place" ref="#l0003">l'auberge</rs> de toute la matinée; mais elle
      avait mal calculé, car il l'amena à <name type="place" ref="#l0002">Lambton</name> le jour même. <name
        type="person" ref="#p0001">Elisabeth</name> et <rs type="person" ref="#p0004">sa tante</rs> qui
        s'étaient promenées près de là.</p>
    </div>
    <div type="specific">
      <p><persName ref="#p0001">Elisabeth</persName> avait arrangé dans sa tête que <persName ref="#p0002">Mr.
        Darcy</persName> lui amènerait <rs type="person" ref="#p0003">sa sœur</rs> le lendemain de son arrivée
        à <placeName ref="#l0001">Pemberley</placeName> , et en conséquence elle était bien décidée à ne pas
        s'éloigner de <rs type="place" ref="#l0003">l'auberge</rs> de toute la matinée; mais elle avait mal
        calculé, car il l'amena à <placeName ref="#l0002">Lambton</placeName> le jour même. <persName ref="#p0001"
          >Elisabeth</persName> et <rs type="person" ref="#p0004">sa tante</rs> qui s'étaient promenées près de
          là.</p>
    </div>
  </div>
```

Les éléments spécifiques : les métadonnées

Le **<profileDesc>**

❏ **<particDesc>/<settingDesc>**

```
<particDesc>
  <listPerson>
    <person xml:id="p0001">
      <persName>
        <forename>Elisabeth</forename>
        <surname>Bennet</surname>
      </persName>
    </person>
    <person xml:id="p0002"> [5 lines]
    <person xml:id="p0003"> [5 lines]
    <person xml:id="p0004"> [2 lines]
  </listPerson>
</particDesc>
```

```
<settingDesc>
  <listPlace>
    <place xml:id="l0001" type="house">
      <placeName>Pemberley</placeName>
      <region>Derbyshire</region>
      <country>England</country>
    </place>
    <place xml:id="l0002" type="city"> [4 lines]
    <place xml:id="l0003" type="inn"> [5 lines]
  </listPlace>
</settingDesc>
```


LES ÉLÉMENTS SPÉCIFIQUES – TRAVAUX PRATIQUES

Encoder les éléments spécifiques : Exercice

En reprenant le document XML que vous avez créé et en utilisant ce qui figure dans la présentation et les exemples précédents (choisir balises spécifiques plutôt que génériques), encoder dans le <body> les entités nommées mentionnées dans le document entitesnommees.md dans le dossier GitHub de l'atelier 2, et ajouter les ensuite dans le <teiHeader>

Encoder les éléments spécifiques : Résultat

À RETROUVER UNE FOIS L'ATELIER PASSÉ

LES ANNEXES - PRÉSENTATION

Les annexes : les <front> et <back>

Chapitre 4 Default Text Structure <https://tei-c.org/release/doc/tei-p5-doc/fr/html/DS.html>

Éléments contenus dans le <front>

- ❑ Une **préface** ou un avant-propos adressé aux lecteurs par l'auteur
- ❑ Des **remerciements** rédigés par l'auteur
- ❑ Une ou des **dédicace(s)** faites par l'auteur pour des personnes en particulier ou une/des institution(s)
- ❑ Un **résumé** ou court extrait du contenu du texte
- ❑ Une **table des matières**, qui présente le contenu et la manière dont il est structuré au sein du texte
- ❑ Un **frontispice** avec la description, s'il y en a une, de l'illustration présente avec la page de titre

Les annexes : les <front> et <back>

Chapitre 4 Default Text Structure <https://tei-c.org/release/doc/tei-p5-doc/fr/html/DS.html>

Éléments contenus dans le <back>

- ❑ Une **annexe** contenant des sections auxiliaires du texte avec des informations complémentaires
- ❑ Un **glossaire** avec une liste de termes associée à leurs définitions
- ❑ Un regroupement de **notes** textuelles ou autre genre de notes contenues dans le texte
- ❑ Une liste des **citations bibliographiques** qui se retrouvent dans le texte
- ❑ Toute forme d'**index** associé à l'œuvre
- ❑ Un **colophon**, soit une note finale précisant les conditions physiques de production de l'ouvrage

LES ANNEXES - TRAVAUX PRATIQUES

Encoder les annexes : Exercice

En reprenant le document XML que vous avez créé et en utilisant ce qui figure dans la présentation et les exemples précédents, encoder dans un <front> la table des matières (tableofcontents.txt) et le frontispice (frontispice.jpeg) présents dans le dossier GitHub de l'atelier 2

Encoder les annexes : Résultat

À RETROUVER UNE FOIS L'ATELIER PASSÉ

**ALLER PLUS LOIN : CONSTRUIRE ET
DOCUMENTER SON SCHÉMA
D'ENCODAGE (ET S'Y TENIR !)**

Construire son schéma d'encodage

- ❑ Déclarer le jeu de balises de son encodage
- ❑ Quatre types de déclaration de classes TEI :
 - ❑ La classe n'est absolument pas utilisée → pas de déclaration et tous ses éléments seront exclus de l'arbre
 - ❑ La classe est un peu utilisée → déclaration et intégration des éléments utilisés avec @include
 - ❑ La classe est beaucoup utilisée → déclaration et exclusion des éléments non utilisés avec @except
 - ❑ La classe est entièrement utilisée → déclaration et aucune exclusion
- ❑ Possibilité de personnalisation du schéma TEI
 - ❑ Addition/Suppression/Changement d'éléments
 - ❑ Personnalisation d'attributs et de leurs valeurs

Créer son schéma d'encodage avec Roma





- ❑ Aller sur Roma : <https://romabeta.tei-c.org/>
- ❑ Choisir "TEI ALL" et "START"
- ❑ Mettre comme titre "ODD Formation TEI" et mettre votre nom dans "Author" puis cliquer sur "--> CUSTOMIZE ODD"
- ❑ Classer les éléments par module avec "↕ by module"
- ❑ Désélectionner *"analysis"*, *"certainty"*, *"corpus"*, *"dictionaries"*, *"drama"*, *"gaiji"*, *"iso-fs"*, *"linking"*, *"nets"*, *"spoken"*, *"tagdocs"* et *"textcrit"*
- ❑ Avec l'outil de recherche, taper *"div"* et désélectionner les div numérotés
- ❑ Avec l'outil de recherche, taper *"name"* et désélectionner l'élément *"name"*
- ❑ Avec l'outil de recherche, taper *"del"*, cliquer sur l'élément puis *"Attributes"*, modifier l'attribut *"rend"*, mettre l'option "Closed" et ajouter les valeurs *"erasure"*, *"overwritten"* et *"strikethrough"*
- ❑ Cliquer sur "Download" et choisir "Customization as ODD"
- ❑ Modifier l'extension du fichier pour mettre un ".xml"

```

<schemaSpec ident="tei_all" start="TEI teiCorpus" prefix="tei_" targetLang="en"
  docLang="en">
  <moduleRef n="01" key="tei"/>
  <moduleRef n="02" key="header"/>
  <moduleRef n="03" key="core" except="name"/>
  <moduleRef n="04" key="textstructure"
    except="div1 div2 div3 div4 div5 div6 div7"/>
  <moduleRef n="06" key="verse"/>
  <moduleRef n="10" key="msdescription"/>
  <moduleRef n="11" key="transcr"/>
  <moduleRef n="13" key="namesdates"/>
  <moduleRef n="14" key="figures"/>
  <moduleRef key="corpus" include="particDesc settingDesc"/>
  <elementSpec ident="del" mode="change">
    <attList>
      <attDef ident="rend" mode="change">
        <valList type="closed" mode="change">
          <valItem mode="add" ident="erasure"/>
          <valItem mode="add" ident="overwritten"/>
          <valItem mode="add" ident="strikethrough"/>
        </valList>
      </attDef>
    </attList>
  </elementSpec>
</schemaSpec>

```

Appliquer son ODD sur un fichier XML-TEI

- ❑ Ouvrir l'ODD nouvellement créée et opérer la transformation en schéma Relax NG (qui est un type de fichier utilisé pour valider la conformité à la TEI)
 - ❑ Appuyer sur le bouton "Appliquer les scénarios de transformation" 
 - ❑ Choisir "TEI ODD to RELAX NG XML"
 - ❑ Appuyer sur le bouton "Appliquer"
- ❑ Dans le dossier contenant votre fichier XML, ouvrir le fichier
 - ❑ Appuyer sur le bouton "Associer schéma" 
 - ❑ Dans "URL", utiliser le bouton "Naviguer"  et choisir le fichier RNG nouvellement créé et situé dans le dossier "out"
 - ❑ Dans "Type de schéma", sélectionner "RELAX NG XML Syntax" puis appuyer sur "OK"
- ❑ Appuyer sur le bouton  pour vérifier que le fichier est valide et corriger les erreurs selon les recommandations qui vous sont données pour permettre de rendre le fichier conforme

Pourquoi documenter son schéma d'encodage ?

Pourquoi documenter ?

- ❑ Préciser et expliquer les choix faits,
- ❑ Garantir l'homogénéité de son encodage
- ❑ Aider à la réutilisation du type d'encodage fait

Deux types de documentation :

- ❑ Documentation partielle, présentation des balises dont l'utilisation peut différer par rapport à ce que définit les TEI guidelines
- ❑ Documentation exhaustive, présentation de l'arbre de A à Z, avec description de l'utilisation des éléments/attributs dans l'arbre

Comment documenter son schéma d'encodage ?

Documentation écrite avec les balises traditionnelles pour l'écriture en prose dans le <body> :

- ❑ <div> : pour la hiérarchie des parties
- ❑ <head> : pour les titres de parties
- ❑ <p> : pour le contenu de la présentation et des descriptions

Balises spécifiques pour les composants de la structure TEI :

- ❑ <gi> : pour les éléments
- ❑ <att> : pour les attributs
- ❑ <val> : pour la valeur des attributs
- ❑ <egXML xmlns="http://www.tei-c.org/ns/Examples"> : pour les exemples tirés de l'arbre XML

Documenter son schéma d'encodage : Exercice

Dans l'ODD que vous avez généré avec Roma et avec les informations données précédemment à propos de l'écriture d'une documentation d'ODD, faites un paragraphe qui présente votre choix d'utiliser les balises <persName>, <placeName> et <orgName> pour les entités nommées, plutôt que <name> avec un attribut @type qui contient les valeurs "person", "place", ou "organisation", et fournissez un exemple tiré du texte de démonstration Orgueil et Préjugés.

Documentation d'encodage

```
<div>
  <head>Documentation</head>
  <p>Dans le cadre de l'encodage des entités nommées, nous avons décidé de pas utiliser la balise <gi>name</gi>
    accompagnée de l'attribut <att>type</att> et des valeurs <val>person</val>, <val>place</val> ou
    <val>organisation</val>. Nous avons préféré à cela l'utilisation des balises <gi>persName</gi>,
    <gi>placeName</gi> et <gi>orgName</gi>, comme dans l'exemple suivant: <egXML
    xmlns="http://www.tei-c.org/ns/Examples">
    <persName ref="#p0001">Elisabeth</persName> avoit arrangé dans sa tête que <persName ref="#p0002">Mr.
      Darcy</persName> lui amèneroit <rs type="person" ref="#p0003">sa sœur</rs>
    </egXML></p>
</div>
```

RESSOURCES

Ouvrages

- ❑ HAROLD, Elliotte Rusty, MEANS, W. Scott, ENSARGUET, Philippe [et al.], *XML en concentré*, Paris, O'Reilly, 2005.
- ❑ BURNARD, Lou et BURGHART, Marjorie, *Qu'est-ce que la Text Encoding Initiative ?*, 2015 [en ligne : <http://books.openedition.org/oep/1237>].

Ressources électroniques

- ❑ TEI guidelines :
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- ❑ Journal de la Text Encoding Initiative :
<https://journals.openedition.org/jtei/>
- ❑ TEI Special Interest Groups (SIGs):
<https://tei-c.org/Activities/SIG/>
- ❑ Archives des réunions annuelles précédentes :
<https://members.tei-c.org/Events/meetings/>
- ❑ Oxygen XML Editor :
<https://www.oxygenxml.com>
- ❑ Documentation ODD :
<https://tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>

Merci de votre attention

Des questions ?

[https://github.com/FloChiff/
AtelierObTIC-creer-une-edition-scientifique-numerique](https://github.com/FloChiff/AtelierObTIC-creer-une-edition-scientifique-numerique)
[chiffolleau.floriane\[at\]gmail.com](mailto:chiffolleau.floriane[at]gmail.com)