



Séance 4

Structure du document TEI

Anatomie d'un encodage XML TEI

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Structure globale

Composants d'un document TEI

Composants minimaux

Tout document TEI contient a minima deux parties :

- Un **en-tête**, représenté au moyen d'un élément `<teiHeader>` contenant des métadonnées décrivant le document ;
- le **texte** lui-même, représenté par un élément `<text>`.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title></title>
      </titleStmt>
      <publicationStmt>
        <p></p>
      </publicationStmt>
      <sourceDesc>
        <p></p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p></p>
    </body>
  </text>
</TEI>
```

```
<teiCorpus version="3.3.0" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!--[en-tête du corpus]-->
  </teiHeader>
  <TEI>
    <teiHeader>
      <!--[en-tête du premier texte]-->
    </teiHeader>
    <text>
      <!--[premier texte du corpus]-->
    </text>
  </TEI>
  <TEI>
    <teiHeader></teiHeader>
    <text></text>
  </TEI>
</teiCorpus>
```

TEI

teiHeader

text

facsimile standOff
sourceDoc fsdDecl

teiHeader minimal

<fileDesc>

<titleStmt>

informations
identifiant le
document lui-même

<publicationStmt>

informations sur la façon
dont il est distribué ou
publié

<sourceDesc>

indications sur la source
du document

Contenu de l'élément text

<text>



<front>

pour les préfaces, et
tous les éléments
liminaires du texte

<body>

pour le corps du texte
proprement dit

<back>

pour tous les
appendices, épilogues,
postfaces, etc.

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom square is a dark blue, and the top square is a lighter blue, shifted slightly to the right and up.

TEI header

Structuration des métadonnées du document

Multiplicité de l'objet texte

Document source

Document physique qui a été numérisé et sur lequel s'appuie l'encodage

Document numérisé

Le document image issu de la numérisation du document source

Document encodé

Le fichier numérique en lui-même au format XML TEI

Convention de nommage

_Stmt

contenu structuré en
sous-éléments

_Desc

description en prose,
pouvant admettre
des sous-éléments

_Decl

Déclarations associant
sous-éléments et
description en prose

teiHeader

fileDesc

profileDesc

encodingDesc

revisionDesc

xenoData

fileDesc

OBLIGATOIRE

Description bibliographique du document numérique : ce qu'on pourrait renseigner dans une notice de catalogue (sauf sa langue et date de création).

<titleStmt>

Déclaration du titre et de la responsabilité de la ressource numérique

<publicationStmt>

Informations sur la diffusion du fichier numérique/du document source et ses droits d'utilisation

<sourceDesc>

Description du document source

profileDesc

Description détaillée des aspects non bibliographiques : informations contextuelles (sujet, contexte de production, langue, etc.)

<settingDesc>

Information sur le contexte, notamment les lieux mentionnés dans le texte

<particDesc>

Information sur les personnes citées dans le texte

<creation>

Information sur la date et le contexte de création du texte (≠ publication)

<langUsage>

Information sur les langues présentes dans le texte

encodingDesc

Description des choix d'encodage

<projectDesc>

Description de l'objectif visé dans l'encodage et de la manière dont il a été construit ou recueilli

<editorialDecl>

Précisions sur les pratiques et les principes éditoriaux appliqués au cours de l'encodage du texte.

<variantEncoding>

Déclaration de la méthode de constitution de l'apparat critique

<schemaRef>

Décrit la relation de l'encodage à n schéma

revisionDesc

Information sur les modifications apportées à l'encodage depuis sa version initiale

<listChange>

Liste de modifications effectuées

<change>

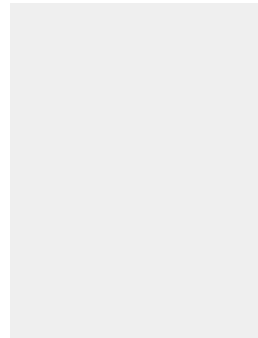
Description d'une des modifications effectuées avec sa date, son auteur, sa nature, etc.

Modules du *header*

Le *header* peut être étendu par différents modules optionnels en fonction de la nature du texte encodé : `msdescription` pour les manuscrits, `textcrit` pour l'édition critique, etc.

Exercice

Compléter un
teiHeader



Consigne

En vous appuyant sur le [template](#) fourni dans le *repository*, compléter au maximum le `teiHeader` pour correspondre à un encodage fictif des [Mystères de Paris d'Eugène Sue](#) en vous appuyant sur la notice de [catalogue](#), en supprimant les éléments non pertinents pour ladite édition

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Divisions du document

Contenu de l'élément <text>

Qu'est-ce que le text

Renferme le contenu encodé en tant que tel : texte de toute nature, qu'il soit unitaire ou composite, par exemple un poème ou un drame, un recueil d'essais, un roman, un dictionnaire ou un échantillon de corpus.

text

front

body

back

...

Structuration générique

Balise Utilisation

div Division générique (peut inclure des **div**)

div1 Division de niveau 1

div2 Division de niveau 2

div3 Division de niveau 3

div4 Division de niveau 4

... **div7** Division de niveau 7

Structuration de la prose

Balise Utilisation

p Paragraphes

seg Segment de texte (division par défaut)

list Liste contenant des **item**

quote Citations simples, (**q** pour ce qui est entre guillemets)

cit Citations avec référence (**bibl**)

table Tableaux

figure Images

Structuration matérielle

Balise Utilisation

- pb** Changement de page (*page beginning*)
avec @n pour le n° de page qui commence
- cb** Changement de colonne (*column beginning*)
- lb** Changement de ligne (*line beginning*)

Exercice

A decorative graphic on the right side of the slide, consisting of a light gray square positioned above a blue square, both of which are partially cut off by the right edge of the frame.

**Éditer numériquement
une pièce de théâtre**

Consigne

À partir de la [transcription du Cid](#) au format `txt` fourni dans le *repository*, structurer grossièrement les grandes parties de la pièce et de sa préface et compléter le `teiHeader` en vous appuyant sur le [chapitre des Guidelines](#) consacré au théâtre