Introduction au langage XML pour l'analyse de texte

une série d'ateliers

Les mardis de 10 h 30 à 12 h (FR)
Les mardis de 13 h 30 à 15 h (EN)

Avec Alice Defours et
Constance Crompton



Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique
4. Conseils avec Oxygen XML Editor
5. Types de documents
6. Exercice pratique (Suite)
7. Fiches d’apprentissage



Atelier II

24 mars

2026

1. Anatomie du XML



1. Anatomie du XML

Qu'est-ce que le XML ?

Cas d'utilisation du XML dans la vie numérique de tous les jours :

- KML (cartes)
- TEI (éditions textuelles)
- XHTML / HTML5 (pour l'affichage dans les navigateurs)
- WordprocessingML (Microsoft Word)
- ePub (édition électronique, iBook)
- RSS (syndication web ou flux web)
- RDF (pour la recherche avancée sur le web basée sur l'inférence)



XML signifie
eXtensible Mark-up Language
(langage de balisage extensible)



1. Anatomie du XML

Le XML est :

1. Indépendant de toute plateforme
2. Conçu pour décrire des données, et non pour les transformer
3. Lisible par l'humain et par l'ordinateur
4. Extensible

Le XML n'est pas conçu pour tout faire !
Il capture les données dans des balises afin qu'un ordinateur puisse les manipuler.
(Autrement dit, il est conçu pour l'encodage, et non pour le codage.)



1. Anatomie du XML

<location type="university">uOttawa</location>

balise ouvrante — balise fermante

élément

1. Anatomie du XML





1. Anatomie du XML

</br> et <br/> sont équivalents

attention à la barre oblique /



1. Anatomie du XML

✓ **<persName>**<forename>Anne</forename>
</persName>

✗ **<persName>**<forename>Anne**</persName>**
</forename>



1. Anatomie du XML

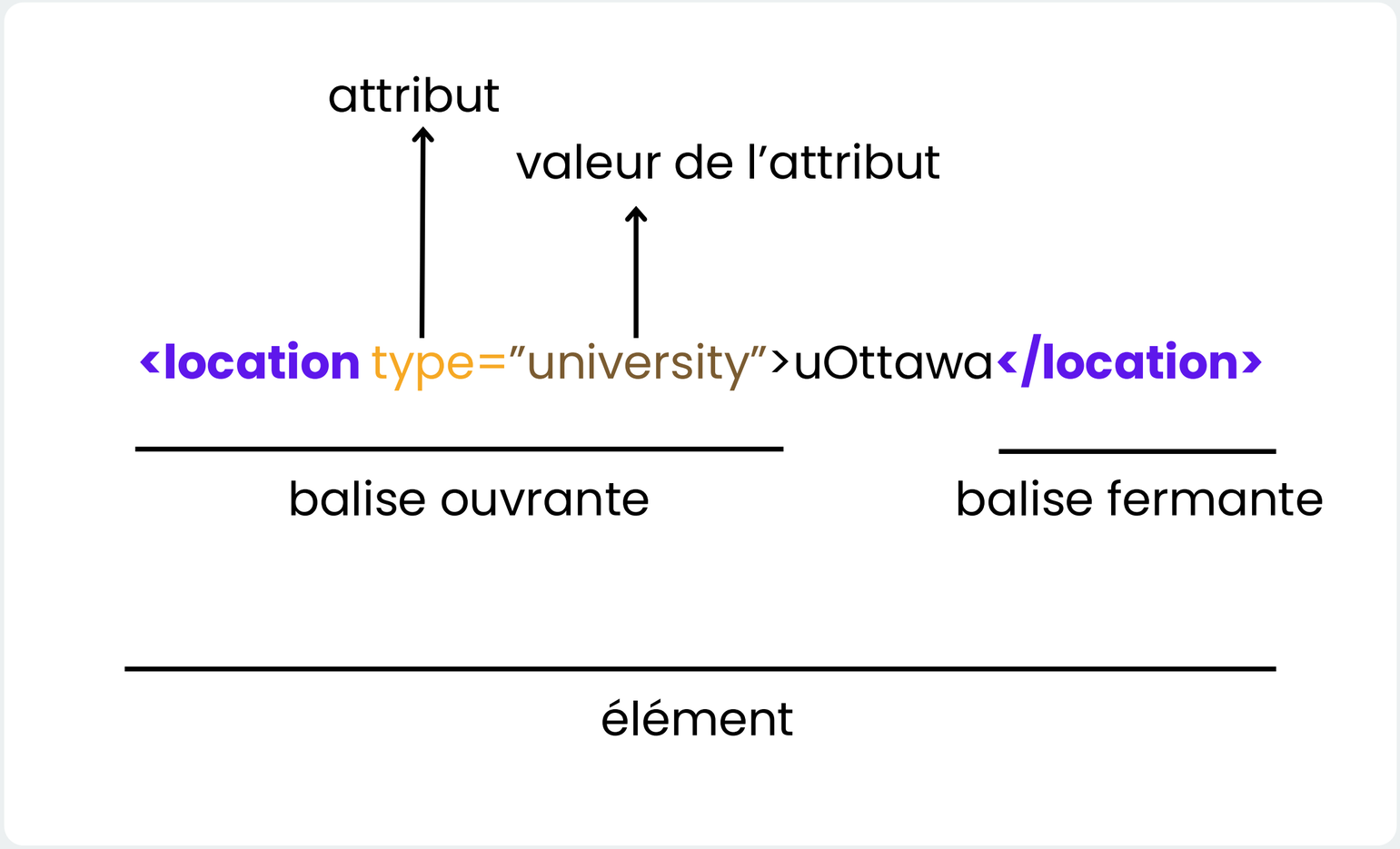


Parmi les erreurs les plus fréquentes :

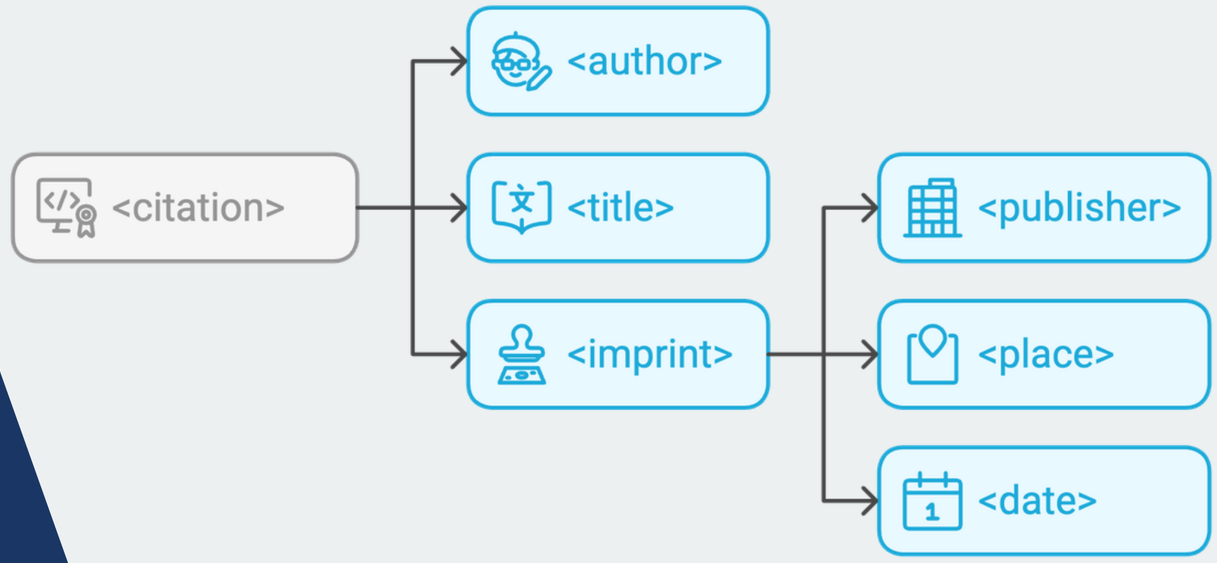
1. Oublier de fermer un élément
2. Chevauchement d'éléments
3. Oublier la hiérarchie d'un élément



1. Anatomie du XML



1. Anatomie du XML




```
<?xml version="1.0" encoding="UTF-8"?>
<citation>
    <author>Katherine Hayles</author>
    <title>Writing Machines</title>
    <imprint>
        <publisher>MIT Press</publisher>
        <place>Cambridge, MA</place>
        <date>2002</date>
    </imprint>
</citation>
```



1. Anatomie du XML

Document bien formé

Un document XML a besoin de plusieurs choses pour être “bien formé” :

1. L'espace de noms
    a. indique à l'ordinateur le type de XML (par exemple, TEI) utilisé
    b. <TEI xmlns="http://www.tei-c.org/ns/1.0"**>**
2. Un seul élément à la racine
3. Imbrication
    a. pas de chevauchement !
4. Guillemets tricks :
    a. " … " ou ' … ', et non " … '



1. Anatomie du XML

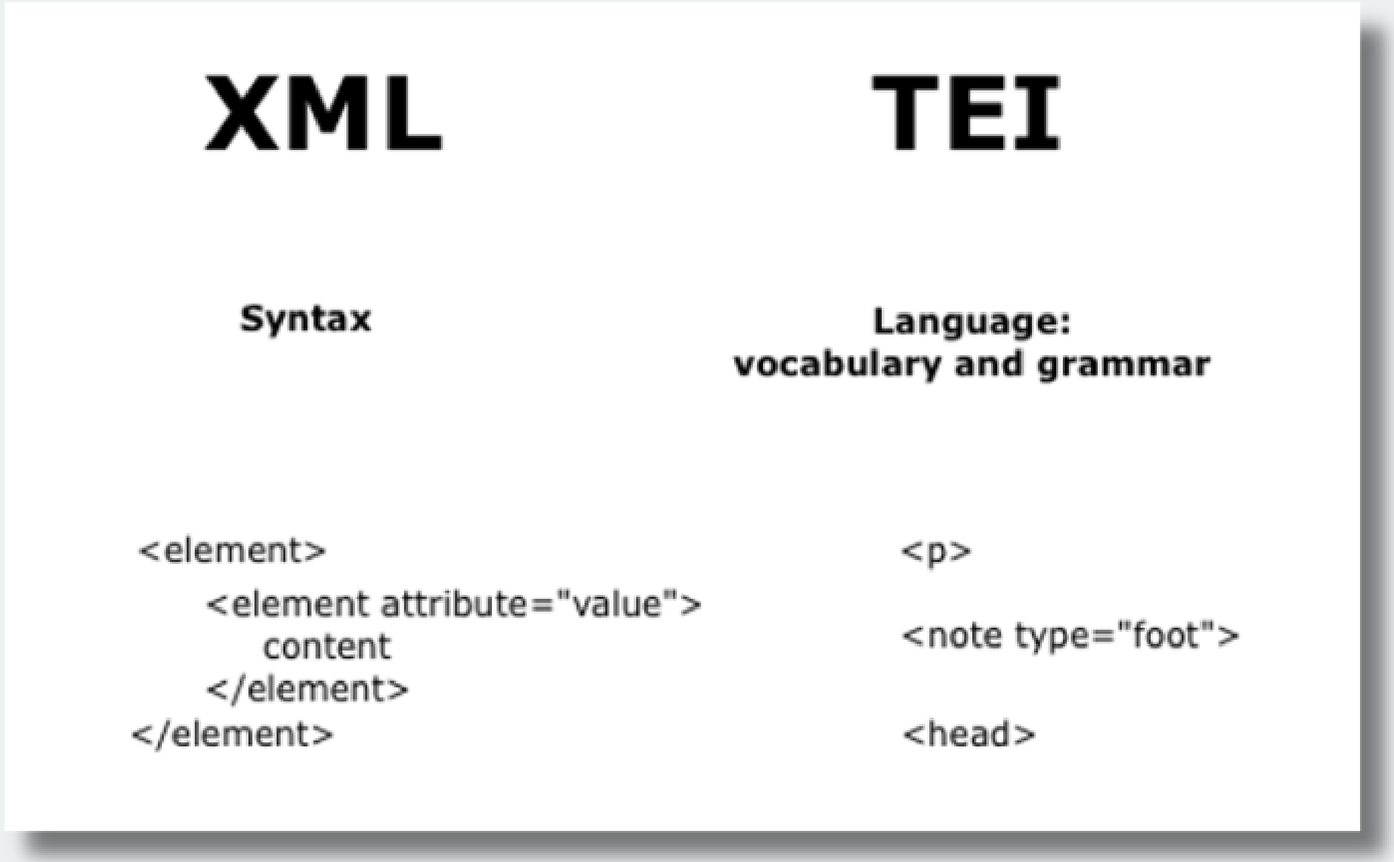
Validité

Un document XML valide utilise un **vocabulaire** correct :
Il ne contient que les éléments et attributs spécifiés par les règles du langage (par exemple, TEI, XHTML, WordProcessingML).

Un document XML valide respecte la **grammaire** :
Les éléments sont correctement positionnés et dans le bon ordre.

Les **règles** sont définies dans un fichier séparé appelé **schéma :**
Les applications de traitement de texte, les navigateurs et autres logiciels vérifient la validité de notre document XML par rapport à ce schéma.

Concept-Clé



XML façonne une structure de balisage
TEI donne une signification sémantique à un balisage

XML

- Syntaxe utilisée pour le stockage et l'échange de données.
- Fournit une structure mais ne transmet pas de sens
- Personnalisable (balises et structure)
- Structure simple
- Utilisée dans les domaines +++ (métadonnées, partage de données)
- Syntaxe de base pour les vocabulaires et autres langages de balisage comme la TEI



TEI

- Langage spécialisé pour l'encodage de textes savants
- Fournit un vocabulaire normalisé pour la signification sémantique des textes.
- Doit se conformer aux lignes directrices, mais peut autoriser des vocabulaires personnalisés (inscriptions latines).
- Riche, complexe, spécialisé pour le balisage scientifique
- Utilisé en DH pour les textes littéraires et historiques.
- Conforme à XML mais spécifiquement adapté à la recherche textuelle.

Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?



2. La TEI : Comment ça fonctionne ?

Grands éléments : Livres - Chapitres
Moyens éléments : Paragraphes - Strophes
Petits éléments : Phrases - Mots

2. La TEI : Comment ça fonctionne ?

Grands éléments : Livres - Chapitres
Moyens éléments : Paragraphes - Strophes
Petits éléments : Phrases - Mots



- **Grands éléments**

= les principales divisions du document :
‹div› : segmentation structurelle de base, pouvant s'imbriquer récursivement (elle signifie « division ». Vous pouvez avoir autant de divisions imbriquées que nécessaire pour votre texte).

2. La TEI : Comment ça fonctionne ?

Grands éléments : Livres - Chapitres
Moyens éléments : Paragraphes - Strophes
Petits éléments : Phrases - Mots

- **Grands éléments**

= les principales divisions du document :
** ** : segmentation structurelle de base, pouvant s'imbriquer récursivement (elle signifie « division ». Vous pouvez avoir autant de divisions imbriquées que nécessaire pour votre texte).

- **Eléments Moyens**

Éléments structurels de niveau intermédiaire pouvant s'insérer dans une division principale et pouvant contenir contenant des petits éléments.
<head>: <head>The Wonderful Wizard of Oz**</head>**
<p>: paragraphes
<list>: listes (<list> doit contenir <items> pour être valide)
<item>: contenu par <list>

2. La TEI : Comment ça fonctionne ?

Grands éléments : Livres - Chapitres
Moyens éléments : Paragraphes - Strophes
Petits éléments : Phrases - Mots

- **Grands éléments**

= les principales divisions du document :
** ** : segmentation structurelle de base, pouvant s'imbriquer récursivement (elle signifie « division ». Vous pouvez avoir autant de divisions imbriquées que nécessaire pour votre texte).

- **Eléments Moyens**

Éléments structurels de niveau intermédiaire pouvant s'insérer dans une division principale et pouvant contenir contenant des petits éléments.
<head>: <head>The Wonderful Wizard of Oz**</head>**
<p>: paragraphes
<list>: listes (<list> doit contenir <items> pour être valide)
<item>: contenu par <list>

- **Petits éléments**

Petits éléments représentant des segments de la taille d'une phrase ou d'un mot :
persName: <persName>Baron Olivier of Brighton**</persName>**
placeName: <placeName>New York**</placeName>**
orgName: <orgName>Podunk Sewing Club**</orgName>**
title: <title>Hamlet**</title>, <title>**La Gioconda**</title>**

Les éléments non-spatiaux représentent des objets absents :

- **<lb/>** : Début de ligne
- **<pb/>** : Début de page
- **<cb/>** : Début de colonne
- **<anchor/>** : Point non spécifié

On en parlera plus tard. Pour l'instant, retenez simplement qu'il existe des éléments représentant des objets textuels qui n'occupent aucun espace physique.

Eléments non-spatiaux

Eléments génériques

Eléments génériques :

Avec plus de 500 éléments, la TEI contient déjà surement tout ce dont vous avez besoin pour décrire votre texte, mais au cas où du cas contraire, vous pouvez utiliser des éléments génériques :

- ** ** division : balise générique pour les grands blocs de texte
- **<ab>** bloc anonyme : pour les blocs de texte de taille moyenne ou de paragraphe
- **<seg>** segment : pour les divisions de la taille d'une phrase ou d'un mot
- **<anchor/>** ancre : pour marquer un point sur la page

Vous pouvez ajouter des attributs et des valeurs pour mieux décrire l'élément :

Exemple : **<ab** type="step"**>**1. Préchauffer le four à 200 °C**</ab>**

2. La TEI : Comment ça fonctionne ?
Format d'un document TEI

Chaque document TEI possède un élément racine "**<TEI>**"

Cet élément racine contient toujours un "**<teiHeader>**"
(qui regroupe les informations relatives au fichier : qui l'a encodé, d'où provient le texte, etc.)
et un élément "**<text>**"

L'élément "**<text>**" inclut toujours "**<body>**", e
t éventuellement "**<front>**" et "**<back>**" pour contenir le corps du texte.

Le corps est généralement divisé en balises "** "**

Voyons cela de plus près...

```
<text>
  <front>
     <div type="introduction"><!-- contenu dans les balises <p> --></div>
  </front>

  <body>
     <!-- Corps du chapitre ici dans les balises <p> -->
  </body>

  <back>
     <div type="notes"><!-- contenu dans les balises <p> --></div>
     <div type="glossary"><!-- contenu dans les balises <p>--></div>
  </back>

</text>
```

Chaque document TEI possède un élément racine “**<TEI>**”

Cet élément racine contient toujours un “**<teiHeader>**”
(qui regroupe les informations relatives au fichier : qui l’a encodé, d’où provient le texte, etc.)
et un élément “**<text>**”

L’élément “**<text>**” inclut toujours “**<body>**”, e
t éventuellement “**<front>**” et “**<back>**” pour contenir le corps du texte.

Le corps est généralement divisé en balises “** ”**

Un document TEI valide et bien formé

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
    <teiHeader>
        <!-- Plus de trucs sur l'en-tête le jour 3. -->
        <fileDesc>
            <titleStmt>
                <title/>
            </titleStmt>
            <publicationStmt>
                <ab/>
            </publicationStmt>
            <sourceDesc>
                <ab/>
            </sourceDesc>
        </fileDesc>

        <profileDesc/>

    </teiHeader>
    <text>
        <body>
            <div type="book">
                <head n="1">LA VÉNUS D'ILLE</head>

                <p>Je descendais le dernier coteau du Canigou, et, bien que le soleil fût déjà
                    couché, je distinguais dans la plaine les maisons de la petite ville d'Ille,
                    vers laquelle je me dirigeais.</p>
                <p><said> Vous savez</said> dis-je au Catalan qui me servait de guide depuis la veille</p>
                <p><said>vous savez sans doute où demeure M. de Peyrehorade?</said></p>
                <p><said>Si je le sais !</said> s'écria-t-il,</p>
                <p><said>je connais sa maison comme la mienne; et s'il ne faisait pas si noir, je
                        vous la montrerais. C'est la plus belle d'Ille. Il a de l'argent, oui, M. de
                        Peyrehorade; et il marie son fils à plus riche que lui encore.</said></p>

            </div>
        </body>
    </text>
</TEI>
```

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
    <teiHeader>
        <!-- Plus de trucs sur l'en-tête le jour 3. -->
        <fileDesc>
            <titleStmt>
                <title/>
            </titleStmt>
            <publicationStmt>
                <ab/>
            </publicationStmt>
            <sourceDesc>
                <ab/>
            </sourceDesc>
        </fileDesc>

        <profileDesc/>

    </teiHeader>
    <text>
        <body>
            <div type="book">
                <head n="1">LA VÉNUS D'ILLE</head>

                <p>Je descendais le dernier coteau du Canigou, et, bien que le soleil fût déjà
                    couché, je distinguais dans la plaine les maisons de la petite ville d'Ille,
                    vers laquelle je me dirigeais.</p>
                <p><said> Vous savez</said> dis-je au Catalan qui me servait de guide depuis la veille</p>
                <p><said>vous savez sans doute où demeure M. de Peyrehorade?</said></p>
                <p><said>Si je le sais !</said> s'écria-t-il,</p>
                <p><said>je connais sa maison comme la mienne; et s'il ne faisait pas si noir, je
                        vous la montrerais. C'est la plus belle d'Ille. Il a de l'argent, oui, M. de
                        Peyrehorade; et il marie son fils à plus riche que lui encore.</said></p>

            </div>
        </body>
    </text>
</TEI>
```

Essayez de lire de l'extérieur vers l'intérieur
Commencez par l'ouverture et trouvez la fermeture
la TEI se lit comme une boîte

2. La TEI : Comment ça fonctionne ?
Autres caractères spéciaux

Les esperluettes (&) déclencheront un message d'erreur, car le symbole & en XML est un caractère réservé aux instructions de traitement.

Pour conserver une représentation fidèle de votre texte source, si besoin, vous pourrez remplacer :

- & par &
- < par <
- > par >

Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique



3. Exercice pratique

Téléchargez **Oxygen XML Editor et demandez la licence gratuite**

On peut aussi utiliser l’option gratuite VS Code, avec l'extension Scholarly XML. Scholarly XML vérifiera la conformité de votre document au schéma TEI et s'assurera de sa validité et de sa bonne formation. Ils permettent de réaliser gratuitement des projets TEI très robustes

Oxygen Editor, est propriétaire (et cher!) mais souvent utilisé par les chercheurs pour créer et traiter du XML. Lorsque vous aborderez le traitement du TEI et l'écriture de transformations XML en XSLT, l'achat d'Oxygen se justifiera (environ 139 $ pour un usage académique et environ 1 500 $ pour un usage commercial).

VS Code et Scholarly XML sont parfaits pour des projets simples et robustes (comme ce cours !). Oxygen Editor, quant à lui, est suffisamment puissant pour des projets d'envergure (comme d'autres cours ! des projets complexes s'étalant sur plusieurs décennies !).

L’atelier en français utilisera Oxygen XML Editor et sa licence d’essai d’un mois
L’atelier en anglais utilisera VS Code et Scholarly XML !

3. Exercice pratique

Téléchargez **Oxygen XML Editor et demandez la licence gratuite**
ET bit.ly/TEI-Vide

On peut aussi utiliser l’option gratuite VS Code, avec l'extension Scholarly XML. Scholarly XML vérifiera la conformité de votre document au schéma TEI et s'assurera de sa validité et de sa bonne formation. Ils permettent de réaliser gratuitement des projets TEI très robustes

Oxygen Editor, est propriétaire (et cher!) mais souvent utilisé par les chercheurs pour créer et traiter du XML. Lorsque vous aborderez le traitement du TEI et l'écriture de transformations XML en XSLT, l'achat d'Oxygen se justifiera (environ 139 $ pour un usage académique et environ 1 500 $ pour un usage commercial).

VS Code et Scholarly XML sont parfaits pour des projets simples et robustes (comme ce cours !). Oxygen Editor, quant à lui, est suffisamment puissant pour des projets d'envergure (comme d'autres cours ! des projets complexes s'étalant sur plusieurs décennies !).

L’atelier en français utilisera Oxygen XML Editor et sa licence d’essai d’un mois
L’atelier en anglais utilisera VS Code et Scholarly XML !

3. Exercice pratique

1. Ouvrez TEI-modèle-vide.xml avec Oxygen
2. Essayez de supprimer les balises <p>.
    a. Vous obtiendrez immédiatement un message d'erreur.
    b. Que dit-il ? Pourquoi ?
3. Ouvrez vos données avec Oxygen !
    a. Constatez-vous des messages d’erreur ?

Atelier II

24 mars

2026

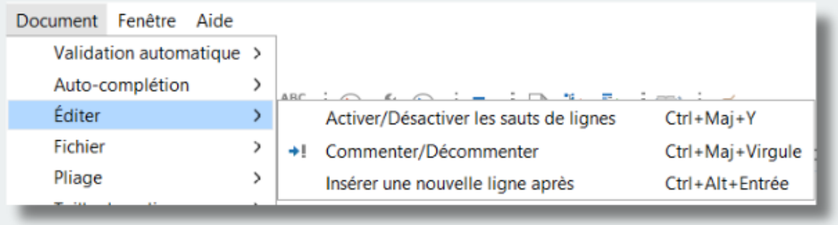
1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique
4. Conseils avec Oxygen XML Editor



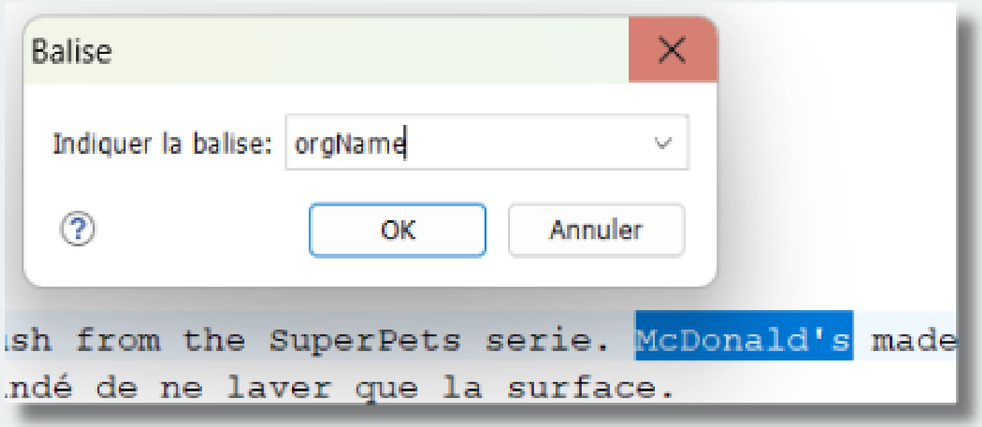
4. Conseils avec Oxygen XML Editor

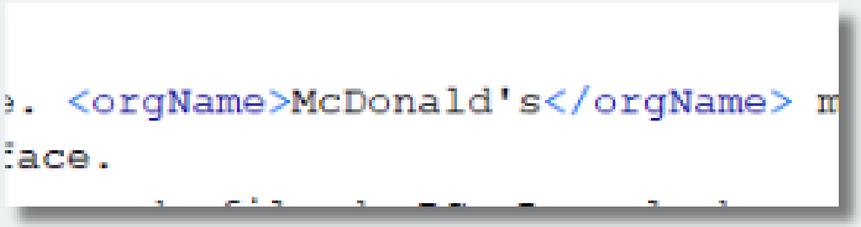


Gardez un œil sur la case du validateur – **vous voulez la garder verte**
Apprenez à lire les messages d'erreur d'Oxygen. Les premiers cahractères n'auront pas nécessairement de sens, mais la prose qui suit est très informative.



Allez dans Documents > Editer > **Activer les sauts de lignes** pour condenser les paragraphes et rendre vos fenêtres lisibles





Utilisez CTRL+E ou CMD+E
pour insérer automatiquement les balises ouvrantes et fermantes de l’élément que vous avez choisi

Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique
4. Conseils avec Oxygen XML Editor
5. Types de documents



5. Types de documents

- Prose
- **Poésie**
- Lettres
- Manuscrits
- Images
- Pièces de théâtre

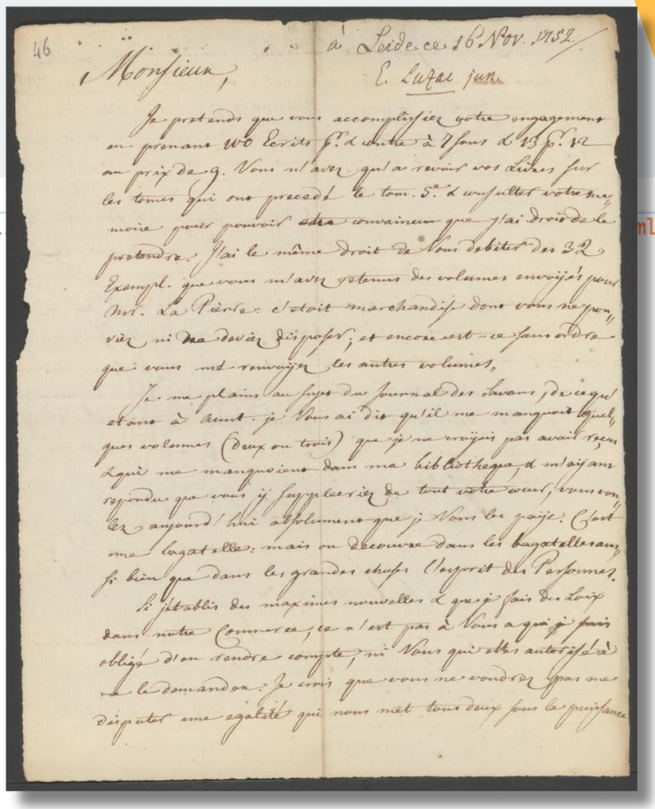
```
<div type="verse">
  <head>1</head>
  <lg type="stanza" rhyme="ABABCCDDA">
    <l>On the top of the Crumpetty Tree</l>
    <l>The Quangle Wangle sat</l>
    <l>But his face you could not see</l>
    <l>On account of his Beaver Hat.</l>
    <l>His Beaver Hat was twelve feet wide</l>
    <l>With Ribbons and Bibbons on every side</l>
    <l>And Bells, and Buttons, and Loops, and Lace</l>
    <l>So that nobody ever could see the face<lb/>Of the Quangle Wangle wee.</l>
  </lg>
</div>
```



The Quangle Wangle's Hat
Par Edward Lear

5. Types de documents

- Prose
- Poésie
- **Lettres**
- Manuscrits
- Images
- Pièces de théâtre



TEI text body div table row cell

```
1  <TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.tei-c.org/ns/1                    ml:space="preserve" xml:id
1  ="Rey17521116">
2      <teiHeader>
3          <fileDesc>
4              <titleStmt>                                                                                        [22 lines]
27             <publicationStmt>                                                                                  [8 lines]
36             <sourceDesc>                                                                                       [4 lines]
41         </fileDesc>
42         <profileDesc>
43             <creation>
44                 <date when="1752">1752</date>
45             </creation>
46         </profileDesc>
47     </teiHeader>
48     <facsimile>                                                                                                [3 lines]
52     <text>
53         <body>
54             <div>
55                 <p><date><add>à Leide ce 16 nov. 1752.</add></date></p>
56                 <p>E. Luzac jun<supplied>ior</supplied>.</p>
57
58                 <p>Monsieur,</p>
59
60                 <p>Je pretends que vous accomplissiez votre engagement en prenant 100 Ecrits p<supplied>our</supplied> et contre à 7 sous et 13 p<supplied>our</supplied>
60  12 au prix de 9<note><title>Écrits pour et contre les Immunités prétendues par le Clergé de France</title> (<placeName type="edition">La Haye</placeName>, [Elie
60 Luzac], 1751-1752, 8°, 7 vol.) qui contient diverses « Lettres supprimées ». Rey en avait renvoyé cent exemplaires sur les deux cents reçus des premiers volumes et
60 Luzac lui rappelait son engagement de prendre cent exemplaires des volumes suivants (<ref target="Rey17521013"/>, <ref target="Rey17510308"/>). Par ailleurs, cet
60 ouvrage est l'objet d'une querelle sur le prix d'achat depuis l'année précédente. Voir <ref target="Rey17510308"/>, <ref target="Rey17510430"/>, <ref
60 target="Rey17510806"/>.</note>. Vous n'avez qu'à revoir vos livres sur les tomes qui ont precedé le tom. 5e. et consulter votre memoire pour pouvoir etre convaincu que
60 j'ai droit de le pretendre. J'ai le même droit de vous debiter des 32 exempl<supplied>aires</supplied> que vous m'avez retenus des volumes envoyés pour <persName>Mr.
60 La Pierre<note><persName>M. La Pierre</persName> est mentionné dans la lettre du 21 janvier 1752, Rey lui devait 60 florins (<ref target="Rey17520121"/>). Les lettres
60 de Rey à Luzac de cette période ne nous étant pas parvenues, nous n'avons pas d'informations sur les exemplaires en question ici.</note></persName> : c'etoit
60 marchandise dont vous ne pouviez ni ne deviez disposer; et encore est-ce sans ordre que vous me renvoyez les autres volumes.</p>
61
62                 <p>Je me plains au sujet du <title>Journal des Savans<note>Rey propose une édition hollandaise du <title>Journal des Savants</title> depuis septembre
62 1749. </note></title>, de ce qu'etant à <placeName>Amst<supplied>erdam</supplied></placeName> je vous ai dit qu'il me manquait quelques volumes (deux ou trois) que je
62 ne croyois pas avoir reçus et qui me manquoient dans ma bibliotheque, et m'aÿant repondu que vous ÿ suppleeriez de tout votre cœur, vous voulez aujourd'hui absolument
62 que je vous les paye. C'est une bagatelle : mais on decouvre dans les bagatelles aussi bien que dans les grandes choses l'esprit des personnes.</p>
63
```

MMR, Rey17521116

5. Types de documents

- Prose
- Poésie
- Lettres
- Manuscrits
- Images
- **Pièces de théâtre**
- **Scénarios**

```
<body>
  <div type="act">
    <head>ACT I</head>
    <div type="scene">
      <head>SCENE 1: HELLO WORLD</head>
      <stage>user, computer</stage>
      <stage type="setting">In a clean room without windows, a table stands in the middle, carrying a computer screen.
      The screen blinks, <sound type="soundEffect">the computer emits random beeps</sound>.</stage>
      <stage type="technical">[<view>A big screen on the left hand side of the stage displays the word <caption>H
      ELLO</caption>, and will be used to project the contents of the computer screen.</view> <view>A big screen on the
      right hand stage of the stage displays the word <caption>WORD</caption>, until the action starts. Then, the user
      's face is filmed in <camera type="angle">close up</camera> and projected onto this screen.</view>]</stage>
      <stage type="entrance"><move who="#user" type="entrance" where="C"/> (user enters room, walks towards the table,
      takes a seat, and reads what's on the screen)</stage>
      <tech type="light">[spotlight follows user]</tech>
      <sp>
        <speaker>COMP</speaker>
        <p>Hello.</p>
      </sp>
      <stage type="gesture">(user looks around and starts typing)</stage>
      <sp>
        <speaker>USER</speaker>
        <p>What is on the table?</p>
      </sp>
      <sp>
        <speaker>COMP</speaker>
        <p>The table does not exist.</p>
      </sp>
      <stage type="gesture">(user frowns)</stage>
      <sp>
        <speaker>USER</speaker>
        <p>There is a table in the room.</p>
      </sp>
```

TEI Tutorials, Module 5: Drama.

Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique
4. Conseils avec Oxygen XML Editor
5. Types de documents
6. Exercice pratique (Suite)



6. Exercice pratique (Suite)

1. Ouvrez TEI-modèle-vide.xml avec Oxygen
2. Essayez de supprimer les balises <p>.
    a. Vous obtiendrez immédiatement un message d'erreur.
    b. Que dit-il ? Pourquoi ?
3. Ouvrez vos données avec Oxygen !
    a. Constatez-vous des messages d’erreur ?
4. **Aujourd’hui on va commencer à encoder la structure de notre document !**
    a. **Encodez la structure du matériel que vous avez apporté !**

6. Exercice pratique (Suite)

1. Ouvrez TEI-modèle-vide.xml avec Oxygen
2. Essayez de supprimer les balises <p>.
    a. Vous obtiendrez immédiatement un message d'erreur.
    b. Que dit-il ? Pourquoi ?
3. Ouvrez vos données avec Oxygen !
    a. Constatez-vous des messages d’erreur ?
4. **Aujourd’hui on va commencer à encoder la structure de notre document !**
    a. **Encodez la structure du matériel que vous avez apporté !**
    b. **Essayez d’ajouter des <seg> et de leur donner un type**
    c. **“D’accoooord….?”**



Atelier II

24 mars

2026

1. Anatomie du XML
2. La TEI : Comment ça fonctionne ?
3. Exercice Pratique
4. Conseils avec Oxygen XML Editor
5. Types de documents
6. Exercice pratique (Suite)
7. Fiches d’apprentissage

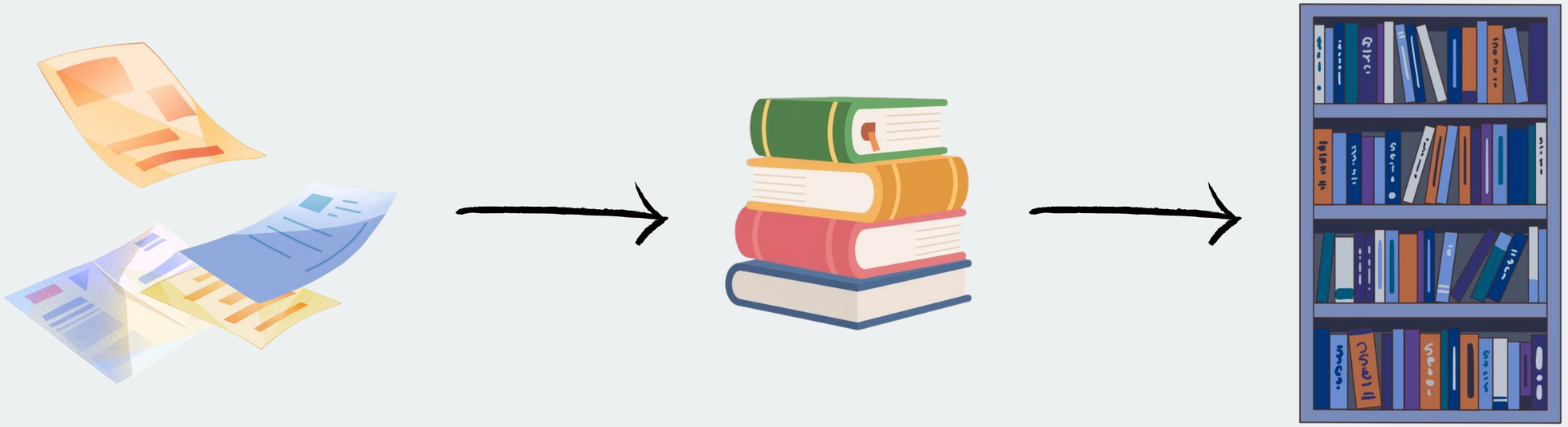


7. Fiches d’apprentissage

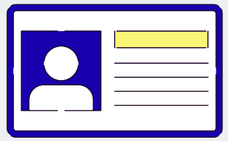
POUR VOUS AIDER

VOICI DES FICHES D’APPRENTISSAGE PAS-À-PAS

Composées de 5 niveaux



La TEI c’est pas compliqué, c’est juste une superposition de choses simples les unes sur les autres



NIVEAU 1 – CRÉER LE ‹TEIHEADER›



NIVEAU 2 – STRUCTURE PHYSIQUE



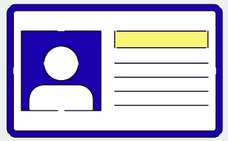
NIVEAU 3 – STRUCTURE DU CONTENU



NIVEAU 4 – CONTENU (ENTITÉS ET MÉTADONNÉES)



NIVEAU 5 – DETAILLER LE CONTENU PAR DES ATTRIBUTS



NIVEAU 1 - CRÉER LE <TEIHEADER>

C'est quoi? C'est comme la couverture d'un livre qui contient le péritexte éditorial (auteurs, publication...)

<**teiHeader**>: Conteneur principal pour le péritexte éditorial
<**titleStmt**>: Contient les informations de l'édition numérique en cours de description, et non de l'œuvre originale
<**title**> : Titre du document, ex: "The Time Machine, la version numérique!"
<**author**> : Vos noms et prénoms à vous (puisque vous êtes les créateurs de l'édition numérique)
<**editor**> : Éditeur, si applicable (mais ici, vous n'en avez pas)

<**publicationStmt**>: Détails de publication numérique
<**publisher**> : Editeur (ex: disons que Atelier TEI est notre éditeur dans notre cas)
<**pubPlace**> : Lieu de publication
<**date**> : Date de publication (date de création du document, ou de publication finale)

<**sourceDesc**>: Description de la source de l'oeuvre originale
<**bibl**> : Référence bibliographique complète
<**title**>: Titre de l'oeuvre originale
<**author**>: Auteur ou artiste
<**publisher**>: Editeur
<**date when="2019-04-12**"/>:Date de publication
<**biblScope unit="pages">29-40</biblScope**>: Envergure physique couverte par l'édition numérique



NIVEAU 2 – STRUCTURE PHYSIQUE

C'est quoi? La division de l'ouvrage en chapitres, les sauts de page, tous les éléments qui physiquement structurent un contenu textuel écrit ou une liste de choses.

<**text**> : Place principale pour le contenu
<**body**>: Elément obligatoire pour faire comprendre à l'ordinateur qu'il lit le corps du contenu désormais

<**div**> : Divisions textuelles ou artistiques auxquelles on peut donner des attributs @types
@type : Type de division (ex: indiquer si la division est un "chapitre", une "section", etc.)

<**head**> : Titre du chapitre ou de l'œuvre suivant la présentation du document ou de la liste
<**p**> : Paragraphes de prose ou descriptions écrites
<**lg**> : Groupes de lignes, surtout utilisé en poésie pour les strophes
<**l**> : Lignes individuelles, surtout utilisé en poésie pour les vers

<**pb n="20"/**> : Sauts de page auxquelles on doit donner un numéro, celui de la page, dans un attribut @n
<**lb**> : Sauts de ligne

Ces fiches de niveaux
sont des références de base pour vous aider à démarrer.

En vérité, chaque élément accepte énormément d'attributs,
bien plus que ce qui est présent dans ces fiches.

Allez consulter les TEI Guidelines
pour avoir une idée de toutes les choses que vous pouvez faire.

Introduction au langage XML pour l'analyse de texte

une série d'ateliers

Les mardis de 10 h 30 à 12 h (FR)
Les mardis de 13 h 30 à 15 h (EN)

Avec Alice Defours et
Constance Crompton

