

Séance 1

Introduction à l'édition électronique de sources anciennes en TEI

TD Antiquités numériques - 2ème partie

Université Lumière Lyon 2, site Berges du Rhône

Mardi 4/11/2025 - 12h00-14h00

Installations



Github

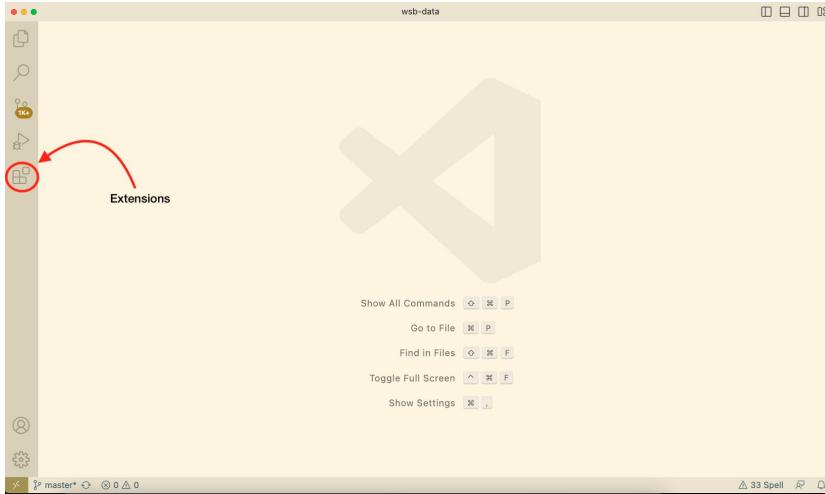
Gestion / partage / publication des fichiers

- Créez un compte sur la plateforme <https://github.com/>
- Créez un dossier appelé "Github" sur votre disque dur (ou tout autre nom)
- Faites une copie du dépôt
https://github.com/emmamorlock/cours_intro-edition-numerique-de-sources sous forme de "Fork" dans votre compte
- Copiez l'adresse de votre dépôt



Vs Code

Editeur de code modulables avec plugins pour XML/TEI



Installation

Télécharger le logiciel depuis
<https://code.visualstudio.com/download>

Choisir un thème pour modifier l'apparence

Installer les extensions suivantes :

[Scholarly XML](#)

[XML tools](#)

[tei-publisher-vscode](#)



Cheatsheet

Ouvrir la palette de commandes	F1 ou Ctrl+Shift+P (Windows et Linux) ou Shift + Cmd + p (Mac) ou click droit > Command palette
Indenter le code XML	Command Palette => XML Tools: formater as XML (ou clic droit + Format document)
Valider avec un schéma RNG	Command Palette =>"Scholarly XML: Validate XML with associated RELAX NG schema."
Pré-visualisation html	Clic sur l'icône TEI Publisher (en haut à droite de la fenêtre de code XML)

```
</teiHeader>
<text> >XML
<body>
</>
</text>
<TEI>
```

Scholarly XML: Validate XML with associated RELAX NG schema. recently used ⚙

XML Tools: Evaluate XPath

XML XPath: XML XPath: Search with XPath

XML Tools: Format as XML

Explorer: Focus on XML Document View

Scholarly XML: Wrap selection with element

XML Tools: Convert text to XML (<> -> <>)

XML Tools: Convert XML to text (<> -> <>)

XML Tools: Get Current XPath

XML Tools: Minify XML

XML XPath: XML XPath: Add Always-Include Tag From Cursor

XML XPath: XML XPath: Add Smart-Relative Identifying Child (from cursor)

XML XPath: XML XPath: Clear 'Don't Ignore After' Anchor

XML XPath: XML XPath: Clear Always-Include Tags

XML XPath: XML XPath: Clear Parent Tag

XML XPath: XML XPath: Clear Relative 'Don't Ignore After' Anchor

XML XPath: XML XPath: Clear Smart-Relative Identifying Children

XML XPath: XML XPath: Clear Virtual Root

XML XPath: XML XPath: Copy Universal XPath from Cursor

Introduction



Programme de la séance 1

- Présentation et objectifs du cours
- Qu'est-ce qu'une édition numérique de sources ?
- Pourquoi utiliser XML ?
- Installation des outils
- Exercices de modélisation et d'encodage



TD “Antiquités numériques” – 2e partie

5 séances de 2 heures

4/11/25	Présentation du cours Installations techniques Définitions et introductions : Editions numériques, langage XML, standard TEI.. Exercices de modélisation et d'encodage
25/11/25	Suite des exercices d'encodage Navigation dans l'arborescence : introduction à XPATH
2/12/25	Le modèle EpiDoc : balisages
9/12/25	Le modèle EpiDoc : outils de transformation et de publication, autres ressources communautaires
16/12/25	Le cycle de vie des projets et la curation des données - Mise en place des mini projets sur Github

(...) le support numérique introduit de nouvelles formes et de nouvelles dynamiques de représentation de l'information et des connaissances basées sur des manipulations calculatoires, ce que Bruno Bachimont (2007) qualifie de « raison computationnelle »

<https://projet.numerev.com/a-propos/agenda/99-revue-intelligibilite-du-numerique-n-7-le-4eme-paradigme-en-question/2024-12-08-23-59>



Objectif de la séance 1

- Comprendre la spécificité des éditions numériques de sources
- Comprendre le rôle du langage de balisage XML et du standard TEI
- Créer un compte sur Github et s'y connecter en ligne de commandes
- Installer et configurer l'éditeur gratuit VS Code
- Etre capable de modifier un document TEI à partir d'un modèle



Présentation d'exemples d'éditions numériques en ligne

- http://xtf.bvh.univ-tours.fr/xtf/view?docId=tei/M0275_01/M0275_01.tei.xml;chunk.id=M0275_01_n5.1.32;toc.depth=1;toc.id=M0275_01_n5.1;brand=default
- <https://shelleyqodwinarchive.org/sc/oxford/frankenstein/volume/ii/#/p41>
- https://www.unicaen.fr/puc/sources/depiscibus/consult/hortus_fr/FR.hs.4.47.html
- <https://vangoghletters.org/>
- <https://iospe.kcl.ac.uk/3.25.html>
- <http://hisoma.huma-num.fr/exist/apps/pyu/works/PYU001.xml?&odd=teipublisher.odd>
- https://www.iglouvre.net/exist/apps/iglouvre-tp/IGL_0958?debug=off&bibliocheck=off&view=single&odd=epidoc-iglouvre&panels=0.1.2
- http://evt.labcd.unipi.it/demo/evt_v1-3/dotr/#doc=VB_text_DOTR&page=VB_fol_105r
- <http://teicat.huma-num.fr/witnesses.php>
- https://peterstokes.org/elenaz/proust_prototype/
- <https://grammateus.unige.ch/document/133452>
- <https://tei.mittelalter.uni-tuebingen.de/exist/apps/salomon/Salomon5.xml>



La définition proposée par P. Sahle (2016)

- A scholarly edition is the critical representation of historic documents.
- A digitised edition is not a digital edition.
- A digital edition cannot be given in print without significant loss of content and functionality.

- (...) digital editions follow a **digital paradigm**, just as printed editions have been following a paradigm that was shaped by the technical limitations and cultural practices of typography and book printing

Sahle, Patrick. « 2. What Is a Scholarly Digital Edition? ». *Digital Scholarly Editing*, édité par Matthew James Driscoll et Elena Pierazzo, Open Book Publishers, 2016,
<https://books.openedition.org/obp/3397>.



Paradigme numérique ?

- Liens hypertextes,
- Visualisations interactives,
- Recherches multicritères et plein texte
- Affichage et modes de lectures et versions du texte multiples
- Annotations collaboratives...
- Le multimédia et la primauté du visuel
- La pluralité des textes
- L'édition comme processus, non comme produit
- L'édition comme entreprise collaborative
- Le principe de la source unique
- La “transmédialisation” (transcription plus proche de la réalité des sources)

Criteria for Reviewing Digital Editions and Resources

Criteria for Reviewing Scholarly Digital Editions



The catalogue of criteria for reviewing scholarly digital editions was established in order to facilitate both reviewers and creators of scholarly digital editions and related projects with a comprehensive list of crucial issues that arise in the context of said projects. We ask reviewers to pay close attention to this catalogue, but we do not require that each and every item mentioned in this catalogue is explicitly raised in the written version of the review. Keep in mind, however, that we do require that reviewers fill out a questionnaire that covers most of the catalogue, so becoming familiar with it before you start to evaluate the project under review is recommended.

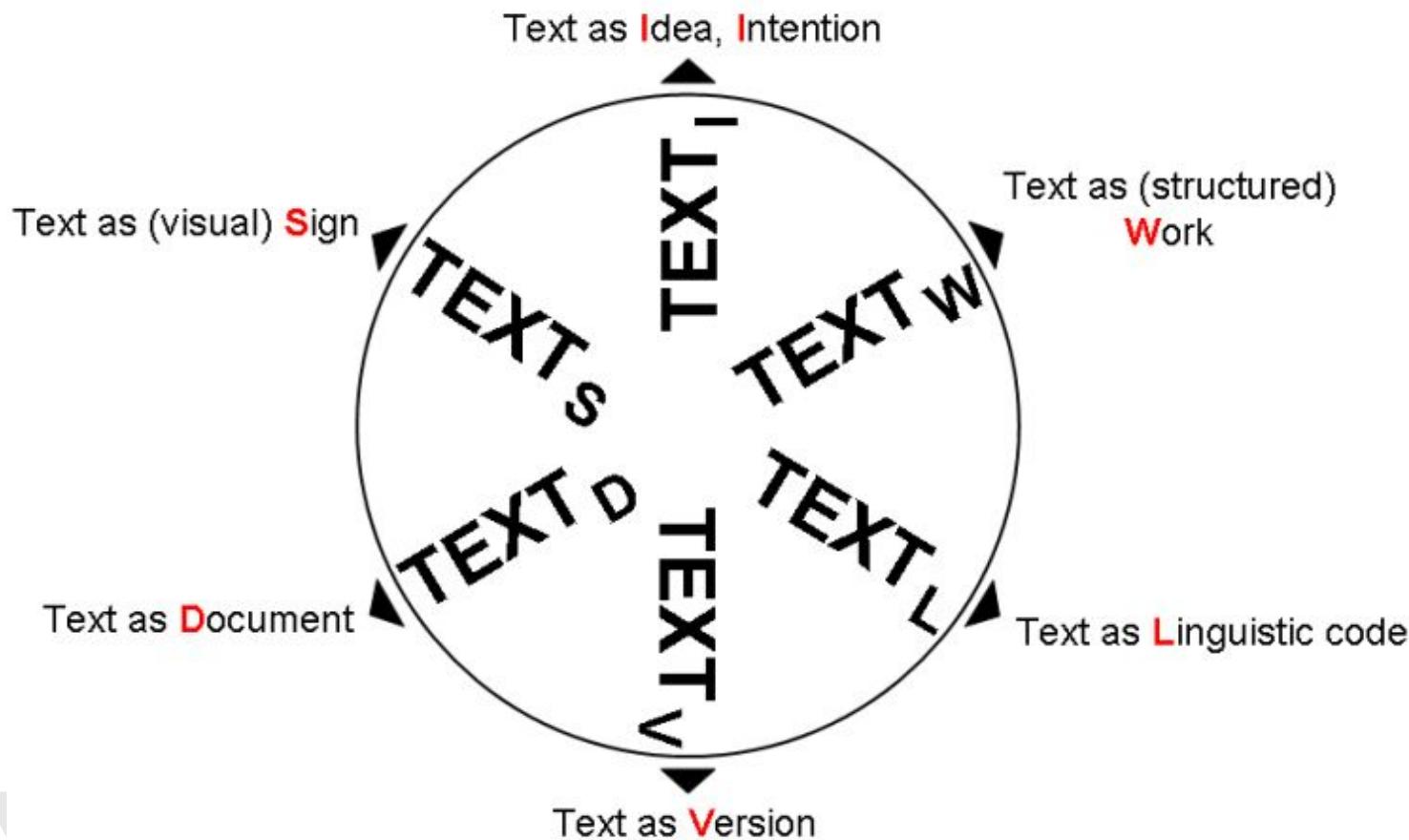
SOCIAL



META

First published: Apr 2014

Last updated: Sep 2023



The text wheel by Patrick Sahle (2016)

Qu'est-ce que l'encodage ?

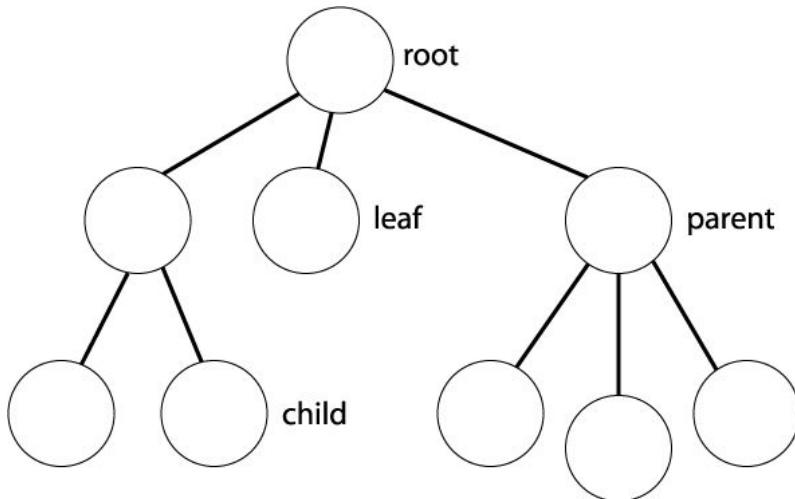


Les documents et leurs structures

- Le texte n'est pas un flux de mots, il communique des significations à l'aide d'une structure
- Il y a deux types de structures : textuelles et matérielles
 - Livre > chapitre > titre de chapitre + paragraphes
 - Poème > strophes > vers
 - Page > bloc > ligne
- Les structures des documents sont hiérarchisées
 - cf. plus haut
- Les structures textuelles sont ordonnées
 - Les paragraphes ne contiennent pas de chapitres...
 - Le titre d'un chapitre ne se place pas entre le 3e et le 4e paragraphe...



L'arbre XML



XML est un modèle formel basé sur hiérarchie ordonnée

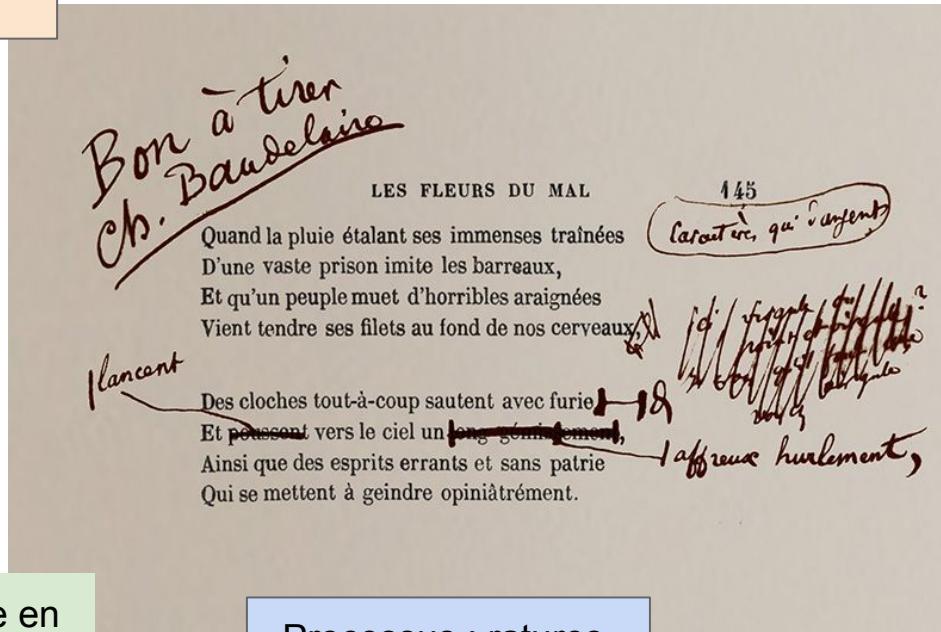
- Racine : contient tout le reste
- Noeuds :
 - ◆ composants et sous composants
 - ◆ peuvent contenir du texte ou d'autres noeuds (ou les deux)

Voir aussi D. J. Birnbaum "[What is XML and why should humanists care? An even gentler introduction to XML](#)"

structure du document (titres, paragraphes, strophes)

Métadonnées (auteur, date de création, provenance, genre)

Disposition, mise en page, analyse des formes graphiques et manuscrites



Processus : ratures, campagnes de correction

Structures linguistiques (parties du discours, phrases, mots...)

Références à des personnes ou objets du monde réel, des événements

Caractéristiques du document support (dimensions, état du support, couleur, type de papier...)

Annotations, notes de régie...



Avantages d'XML pour les sciences humaines

- eXtensible Markup Language (XML)
- L'arborescence XML permet d'enregistrer de manière analytique et documentée les textes et les documents, et de les annoter scientifiquement
- Rend **explicite** pour des machines ce qui est implicite pour des humains
- XML permet des traitements automatisés par les ordinateurs tout en demeurant lisible et éditables par les humains
- XML permet de **décrire les significations** sans avoir besoin de reproduire des apparences potentiellement ambiguës (WYSIWYG vs WYSIWYM)
- Single source publishing



Anatomie d'une balise 1

balise ouvrante



Contenu



balise fermante

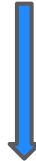


<balise>chaîne de caractères**</balise>**



Anatomie d'une balise 2

Attribut



```
<balise type="type-a">contenu</balise>
```



Valeur d'attribut



Eléments

Les éléments sont les blocs de construction fondamentaux de XML. Ils sont délimités par des balises ouvrantes et fermantes.

Exemple

```
<livre>  
  <titre>XML pour débutants</titre>  
  </livre>
```



Attributs

Les attributs fournissent des informations supplémentaires sur les éléments. Ils sont placés dans la balise ouvrante.

Exemple

```
<foreign xml:lang="la">  
    Alea jacta est...  
</foreign>
```



Texte

Le contenu textuel se trouve entre les balises d'ouverture et de fermeture d'un élément. C'est le contenu principal.

Exemple

```
<auteur>Marie Dupont</auteur>  
<prix>29.99</prix>
```



Commentaires

Les commentaires permettent d'ajouter des notes ou explications dans le code XML sans affecter le traitement.

Exemple

```
<!-- Ceci est un commentaire -->  
<section>Contenu</section>
```



Elément vide

Un élément vide est un élément qui ne contient ni texte, ni éléments enfants. Il peut cependant contenir des attributs.

Syntaxe complète

```
<image src="photo.jpg"></image>  
  
<saut></saut>  
  
<ligne></ligne>
```

Syntaxe auto-fermante

```
<image src="photo.jpg"/>  
  
<saut/>  
  
<ligne/> ou <ligne />
```



“Mixed content”

Le mixed content désigne un élément qui contient à la fois du texte et des éléments enfants, mélangés ensemble.

Pas de mixed content

```
<livre>
  <titre>XML Guide</titre>
  <auteur>Aurélie Martin</auteur>
</livre>
```

Mixed content

```
<paragraphe>
  Ceci est un texte avec du
  <gras>formatage</gras>
  et des
  <italique>styles</italique>.
</paragraphe>
```



Balisage sémantique vs présentationnel

- Le balisage présentationnel enregistre des apparences

<*i*>mens sana in corpore sano</*i*>

<*i*>Le roman de la Momie</*i*>

C'est <*i*>juste</*i*> une impression...

- Le balisage descriptif (ou sémantique) capture le sens, la fonction

<*foreign*>mens sana in corpore sano</*foreign*>

<*title*>Le roman de la momie</*title*>

C'est <*emph*>juste</*emph*> une impression...



XML en bref

- XML se contente de décrire, il ne fait rien (il n'affiche pas les données)
- Les balises ne sont pas pré-définies (*extensible*)
- Déclaration XML <?xml version="1.0" encoding="UTF-8"?>
- Sensible à la casse <book> vs <Book>
- Une racine unique
- Pas de chevauchements (possibilité d'utiliser des pointeurs pour représenter la structure matérielle au sein de la structure logique, ou inversement)
- Deux caractères "<" et "&" réservés, à remplacer par des entités
- Les références de caractères numériques permettent de représenter les caractères difficiles à afficher (espaces insécables par ex.)
- Bien distinguer l'arbre XML (le modèle abstrait) de sa sérialisation (le fichier texte représentation de cet arbre)



Chevauchements interdits

- <seg>C'est <emph>juste</seg></emph> une impression...</p>
- <seg>C'est <emph>juste</emph></seg> une impression...</p>
- Solution : combiner des balises avec contenu et pointeurs

```
<p>
  <lb/>C'est <emph>juste</emph>
  <lb/>une
  <lb/>impression... </p>
```



Processus de création et de traitement d'un fichier XML

- Analyse du document à représenter
- Création d'un schéma formel décrivant les éléments et attributs souhaités et leurs relations (imbrications autorisées, règles d'ordonnancements)
- Encodage (insertion des balises dans texte) en respectant les règles XML et le schéma
- Validation à l'aide du schéma
- Étapes itératives !
- Transformation vers les formats de publications

On crée un schéma XML en fonction de ce qu'on veut représenter

Text Encoding Initiative

The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world. We develop the Guidelines, which provide the infrastructure for developing machine-actionable cultural heritage texts. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation.

Want to become active in the TEI community?

- [Become a TEI Member](#)
- join a [special interest group](#)
- sign up for the [TEI-L mailing list](#)
- join a [Community Call](#)
- come to our [annual conferences and members' meetings](#)



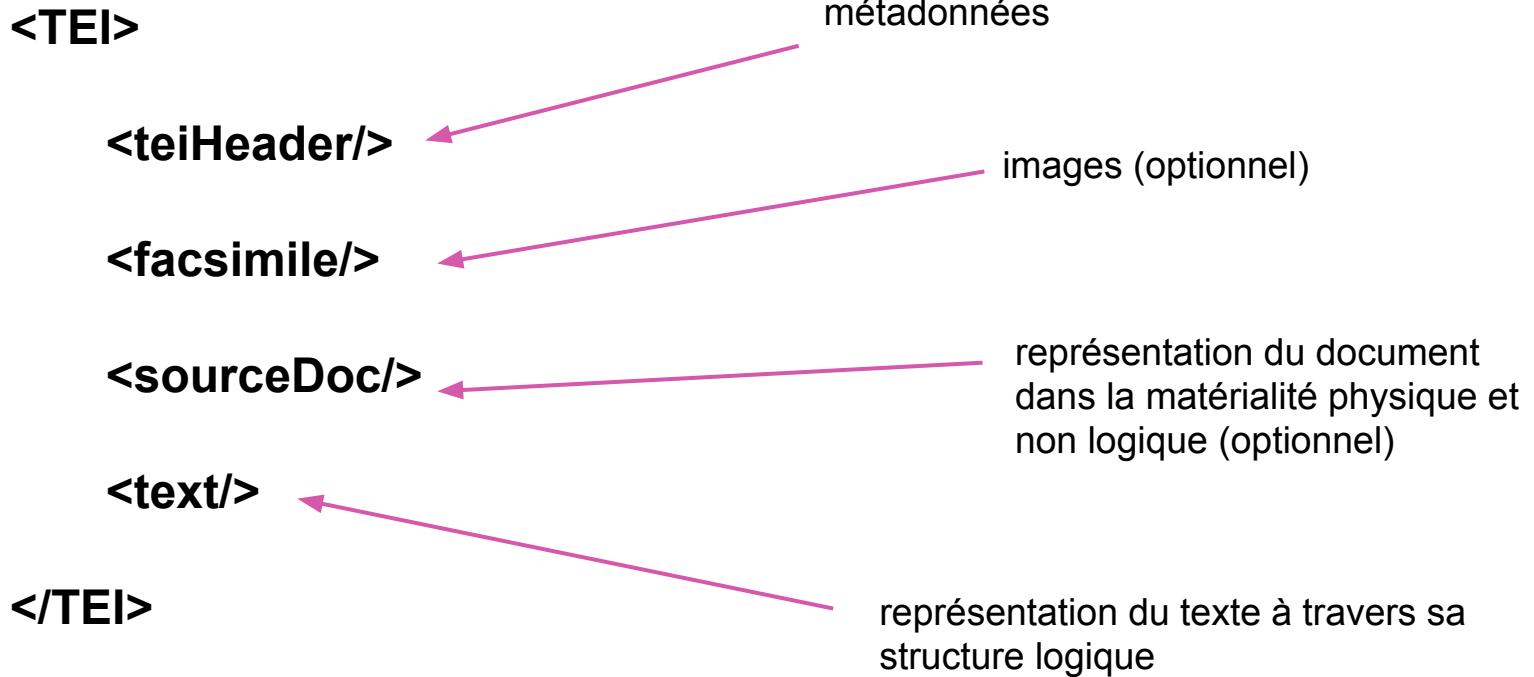


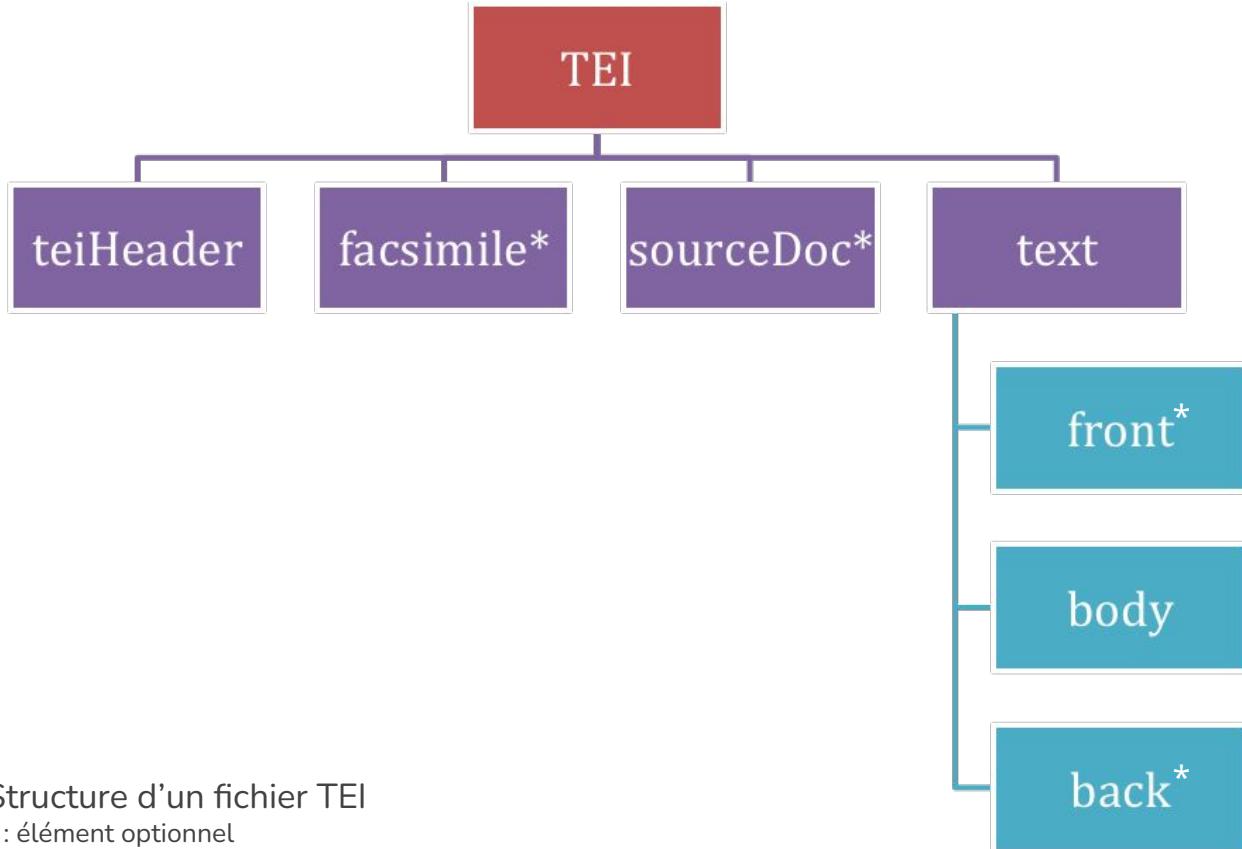
Pourquoi la TEI ?

- Acronyme : Text Encoding Initiative
- Standard communautaire (norme de fait, un critère de qualité et de pérennité)
- Des *Guidelines* décrivant plus de 600 balises
- Une architecture informatique permettant de gérer le schéma et de créer des “customizations” (personnalisations, sous-schéma spécialisés, comme EpiDoc utilisé par les disciplines de l'antiquité numérique)
- Richesse conceptuelle pour décrire les textes et flexibilité adaptée à la diversité et hétérogénéité des sciences humaines
- Garantie d'intelligibilité à long terme (perennité)
- Rend les textes partageables (I de TEI : aussi Interchange)



Structure d'un fichier TEI





Structure d'un fichier TEI

* : élément optionnel

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:space="preserve" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>title of document</title>
      </titleStmt>
      <publicationStmt>
        <authority/>
        <idno type="filename"/>
      </publicationStmt>
      <sourceDesc>
        <msDesc>
          <msIdentifier>
            <repository>museum/archive</repository>
            <idno>inventory number</idno>
          </msIdentifier>
          <physDesc>
            <objectDesc>
              <supportDesc>
                <support>description of object/monument (likely to include <material/> and <objectType/> information, <dimensions/>, etc.)</support>
              </supportDesc>
              <layoutDesc>
                <layout>description of text field/campus</layout>
              </layoutDesc>
            </objectDesc>
            <handDesc>
              <handNote>description of letters, possibly including <height>letter-heights</height></handNote>
            </handDesc>
          </physDesc>
          <history>
            <origin>
              <origPlace>Place of origin</origPlace>
              <origDate>Date of origin</origDate>
            </origin>
            <provenance type="found">Findspot and circumstances/context</provenance>
            <provenance type="observed">Modern location(s) (if different from repository, above)</provenance>
          </history>
        </msDesc>
        <sourceDesc>
          <teiHeader>
            <facsimile>
              <graphic url="photograph of text or monument"/>
            </facsimile>
          </teiHeader>
        </sourceDesc>
      </fileDesc>
    </teiHeader>
    <text>
      <body>
        <div type="edition">
          <ab>
            <lb n="1"/>Greek or Latin (etc.) text here
          </ab>
        </div>
        <div type="apparatus">
          <p>external apparatus criticus (if applicable)</p>
        </div>
        <div type="translation">
          <p>translation(s)</p>
        </div>
        <div type="commentary">
          <p>commentary</p>
        </div>
        <div type="bibliography">
          <p>bibliography of previous editions, discussion, etc.</p>
        </div>
      </body>
    </text>
  </TEI>
```

Structure d'un fichier EpiDoc

TEI

teiHeader

facsimile (images)

text

transcription

apparatus

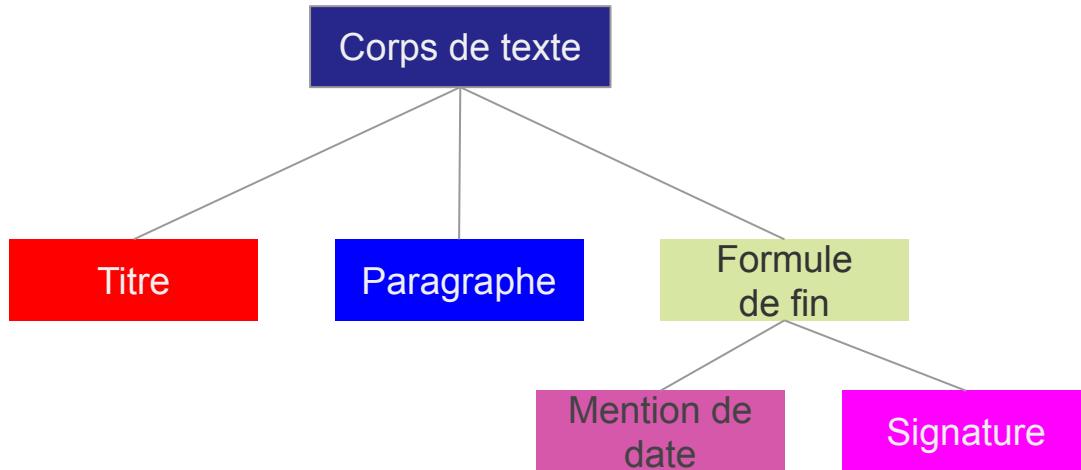
traduction

commentary

bibliography



Exemple : lettre (correspondance)



```
<body>
  <div>
    <head>Ceci est mon testament</head>
    <p>Je sous signe Francois Rebeyrotte
    institue Je legue à ma femme tout
    ce que la loi me permet de lui
    donner en Usufruit.</p>
    <p>Ecrit en entier de ma main
    à Versailles le 3 trois Aout
    mil neuf cent quatorze.</p>
    <p>Francois Rebeyrotte </p>
  </div></body>
```

```
<body>
  <div>
    <head>Ceci est mon testament</head>
    <p>Je sous signe Francois Rebeyrotte
    institue Je legue à ma femme tout
    ce que la loi me permet de lui
    donner en Usufruit.</p>
    <p>Ecrit en entier de ma main
    à Versailles le 3 trois Aout
    mil neuf cent quatorze.</p>
    <p>Francois Rebeyrotte </p>
  </div></body>
```

Paragraphes ???

```
<body>
  <div>
    <head>Ceci est mon testament</head>
    <p>Je sous signe Francois Rebeyrotte
    institue Je legue à ma femme tout
    ce que la loi me permet de lui
    donner en Usufruit.</p>
    <closer><dateline>Ecrit en entier de ma main
    à Versailles le 3 trois Aout
    mil neuf cent quatorze.</dateline>
    <signed>Francois Rebeyrotte</signed>
  </closer> </div></body>
```

Définitions TEI -1

<div> (division du texte)	Contient une subdivision dans le texte préliminaire, dans le corps d'un texte ou dans le texte postliminaire. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-div.html
<p> (paragraphe)	Marque les paragraphes dans un texte en prose. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-p.html
<ab> (bloc anonyme)	Contient une unité de texte quelconque, de niveau 'composant', faisant office de contenant anonyme pour une expression ou des éléments de niveau intermédiaire, analogue à un paragraphe mais sans sa portée sémantique. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-ab.html
<lb> (début de ligne)	Marque le début d'une nouvelle ligne (typographique) dans une édition ou dans une version d'un texte. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-lb.html



Définitions TEI - 2

<dateline> (mention de date)	Contient une brève description des lieux, date, heure, etc. concernant la production d'une lettre, d'un article de journal ou d'un autre texte qui, placée au début ou à la fin, lui est associée comme en-tête ou annonce de fin. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-datetime.html
<closer> (formule finale)	Regroupe une formule de politesse, une indication d'une date et d'autres expressions semblables figurant comme expression à la fin d'une division, en particulier à la fin d'une lettre. https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-closer.html
<signed> (signature)	Contient la dernière salutation, ajoutée à un avant-propos, à une dédicace ou à une autre division du texte. https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-signed.html

Comment utiliser les Guidelines de la TEI ?

The screenshot shows the homepage of the TEI Guidelines website. At the top, there is a blue header bar with the TEI logo and a menu icon. Below the header, the title "TEI: Guidelines for Electronic Text Encoding and Interchange" and the subtitle "P5 Version 4.10.2. Last updated on 4th September 2025, revision bcfa98f42" are displayed. There are language links for English, Deutsch, Español, Italiano, Français, 日本語, 한국어, and 中文. Below the header, there are two small icons: a person icon and a document icon.

Front Matter

- >Title
 - i. [Releases of the TEI](#)
 - ii. [Dedication](#)
 - iii. [Preface and Acknowledgments](#)
 - iv. [About These Guidelines](#)
 - v. [A Gentle Introduction to XML](#)
 - vi. [Languages and Character Sets](#)
- Back Matter
 - Appendix A [Model Classes](#)
 - Appendix B [Attribute Classes](#)
 - Appendix C [Elements](#)
 - Appendix D [Attributes](#)
 - Appendix E [Datatypes and Other Macros](#)
 - Appendix F [Bibliography](#)
 - Appendix G [Deprecations](#)
 - Appendix H [Prefatory Notes](#)
 - Appendix I [Colophon](#)

Text Body

- 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, Formulae, 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](#)

- Les *Guidelines* (Recommandations génériques) sont la documentation en ligne de la TEI en tant que Standard.
- La version en cours est la “P5” (parue en nov. 2007)
- Vous y trouverez :
 - L’explication globale de chaque module de balises (teiHeader, Drama, Names, Dates, People, and Places, etc.)
 - Des définitions (trad. française)
 - Le module d’appartenance
 - Les attributs possibles
 - Contenu par / Peut contenir
 - Des exemples
 - Les déclarations formelles du schéma XML

Exemple de la balise <q>

The screenshot shows the TEI Guidelines for Electronic Text Encoding and Interchange. At the top, there's a blue header bar with the TEI logo and a navigation menu. Below it, the title "TEI: Guidelines for Electronic Text Encoding and Interchange" and the subtitle "P5 Version 4.10.2. Last updated on 4th September 2025, revision bcfa98f42". A search bar is at the top right. The main content area has a sidebar on the left with sections like "Module" (core) and "Attributes". The main content area shows the definition of the <q> element, its attributes (att.global, att.ascribed, att.cmc), and its type (@type). The type is described as "(type) may be used to indicate whether the offset passage is spoken or thought, or to characterize it more finely." It includes status (Optional), datatype (teidata.enumerated), and suggested values (spoken, thought).

<q>

Attributes

- `att.global`: @xml:id, @n, @xml:lang, @xml:base, @xml:space
 - `att.global.analytic`: @ana
 - `att.global.change`: @change
 - `att.global.facs`: @facs
 - `att.global.linkage`: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select
 - `att.global.rendition`: @rend, @style, @rendition
 - `att.global.responsibility`: @cert, @resp
 - `att.global.source`: @source
- `att.ascribed.directed`: @toWhom
 - `att.ascribed`: @who
- `att.cmc`: @generatedBy

@type

(type) may be used to indicate whether the offset passage is spoken or thought, or to characterize it more finely.

Status	Optional
Datatype	<code>teidata.enumerated</code>
Suggested values	
include:	<code>spoken</code> (spoken) representation of speech <code>thought</code> (thought) representation of thought, e.g.

- Trouver une balise : via le moteur de recherche (en haut à droite)
- Lire la définition
- Regarder les attributs possibles et leur contraintes (texte libre ou format spécifique, par ex. "**att.datable.w3c**" ou à valeurs contrôlées ("**teidata.enumerated**")
- Regarder les exemples
- Regarder les contextes (parents et enfants possibles)
- Regarder les contraintes....

```
<body><pb/>
<div>
    <head><lb/>Ceci est mon testament</head>
    <p><lb/>Je sous signe Francois Rebeyrotte
    <lb/><del>institue</del> Je legue à ma femme tout
    <lb/>ce que la loi me permet de lui
    donner en Usufruit.</p>
    <closer><dateline><lb/>Ecrit en entier de ma main
    <lb/>à Versailles le 3 trois Aout
    <lb/>mil neuf cent quatorze.</dateline>
    <signed><lb/>Francois Rebeyrotte</signed>
</closer> </div></body>
```

```
<body><pb/>
<div>
    <head><lb/>Ceci est mon testament</head>
    <p><lb/>Je sous signe Francois Rebeyrotte
    <lb/>institue Je legue à ma femme tout
    <lb/>ce que la loi me permet de lui
    donner en Usufruit.</p>
    <closer><dateline><lb/>Ecrit en entier de ma main
    <lb/>à Versailles le 3 trois Aout
    <lb/>mil neuf cent quatorze.</dateline>
    <signed><lb/>Francois Rebeyrotte</signed>
</closer> </div></body>
```



Interventions scribales : suppression

```
<body>
  <pb/>
  <div>
    <head>Ceci est mon testament</head>
    <p>Je sous signe Francois Rebeyrotte
    <lb/><del>institue</del> Je legue à ma femme tout
    <lb/>ce que la loi me permet de lui
    <lb/>donner en Usufruit.</p>
    <closer><dateline>Ecrit en entier de ma main
    <lb/>à Versailles le 3 trois Aout
    <lb/>mil neuf cent quatorze.</dateline>
    <signed>Francois Rebeyrotte</signed>
    </closer>
  </div>
</body>
```

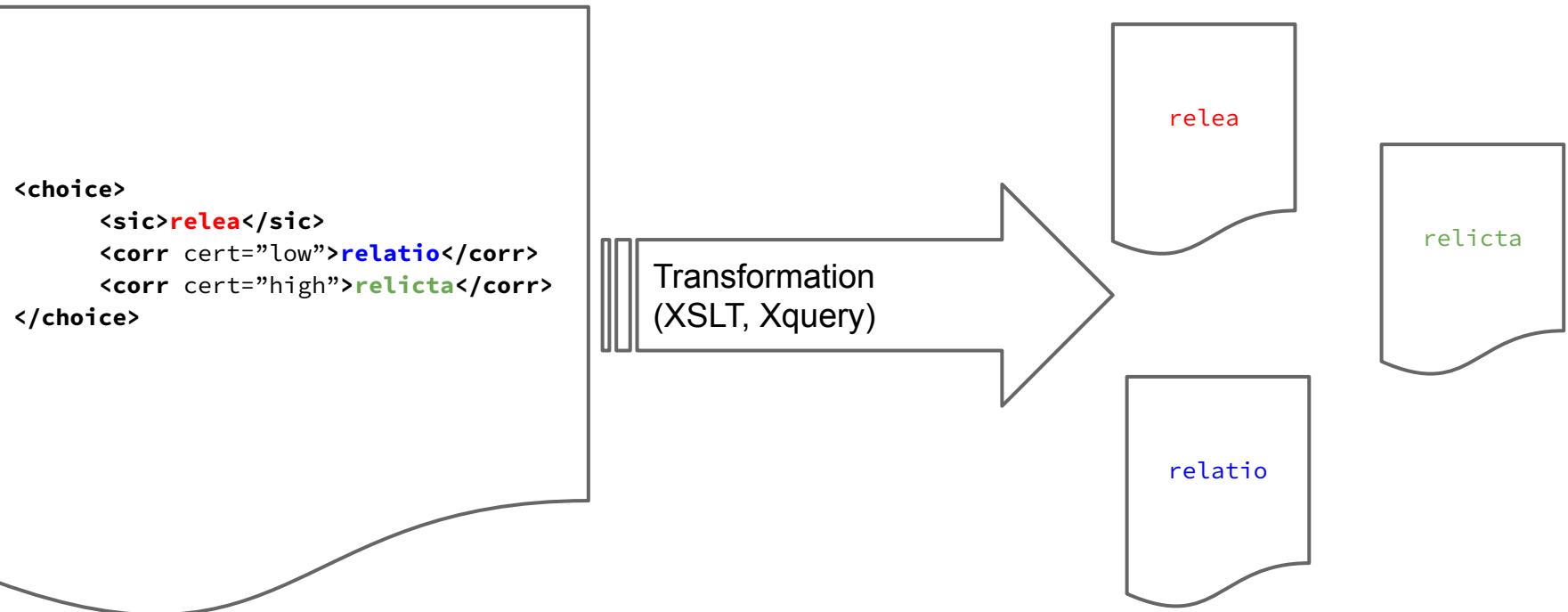


Intervention scribale : correction

```
<body>
  <pb/>
  <div>
    <head>Ceci est mon testament</head>
    <p>Je sous signe Francois Rebeyrotte
    <lb/><del>institue</del> Je legue à ma femme tout
    <lb/>ce que la loi me permet de lui
    <lb/>donner en Usufruit.</p>
    <closer><dateline>Ecrit en entier de ma main
    <lb/>à Versailles le 3 trois <choice><sic>Aout</sic><corr>août</corr></choice>
    <lb/>mil neuf cent quatorze.</dateline>
    <signed>Francois Rebeyrotte</signed>
  </closer>
</div>
</body>
```

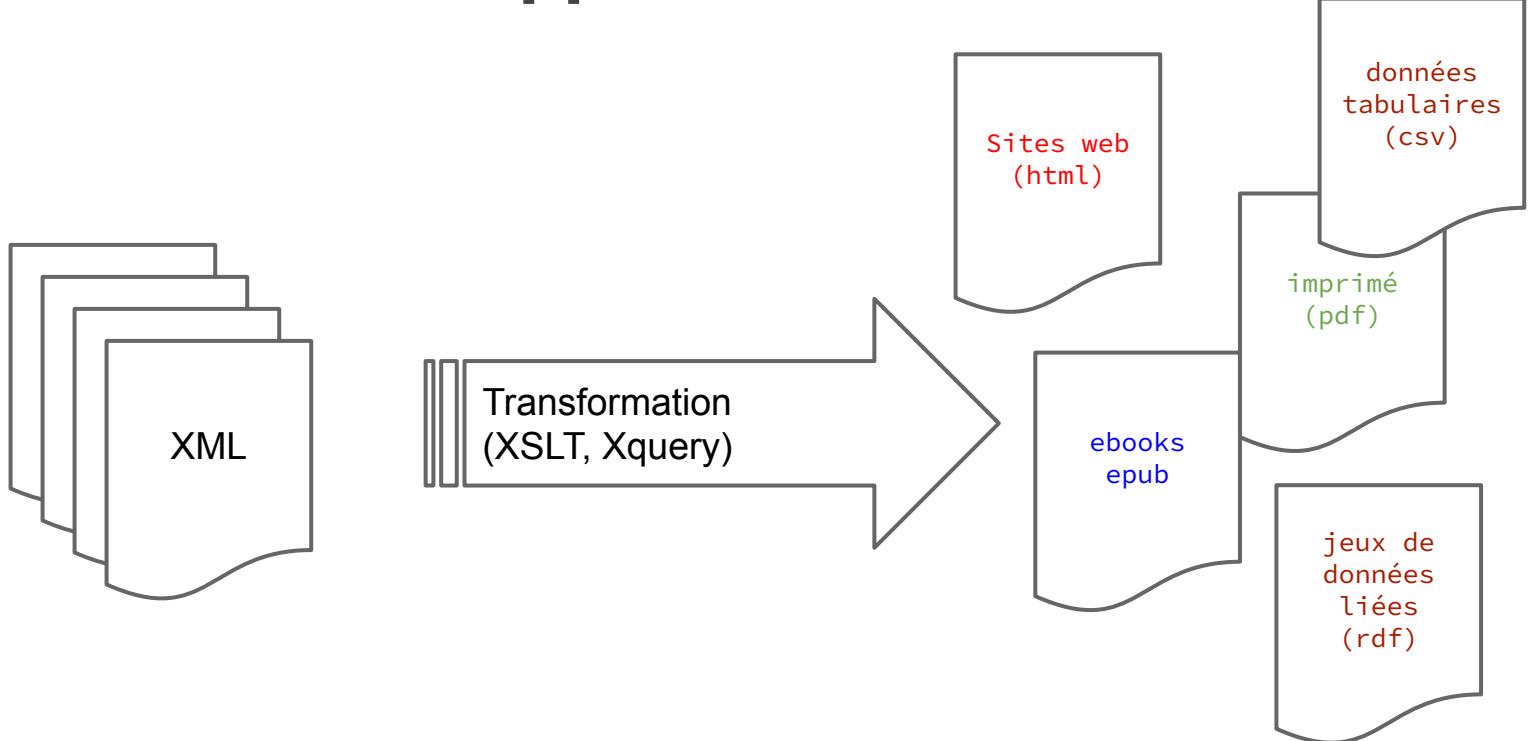


“Single source publishing”





Edition multisupports





Le processus de balisage

1. Analyse du document
2. Sélection des caractéristiques à représenter ou expliciter
3. Choix des balises TEI
4. Saisie des balises
5. Validation



Le processus de balisage

1. Analyse du document
2. Sélection des caractéristiques à représenter ou expliciter
3. Choix des balises TEI
4. Saisie des balises
5. Validation



Pour aller plus loin

Voir le cours vidéo de Marjorie Burghart et Elena Pierazzo :

[Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding](#)

(Dariah Teach)



Liens

- <https://francescaqiannetti.com/vscode-for-tei/>
- <https://teibyexample.org/exist/tools/TBEValidator.htm>
- <https://www.itisnotsound.com/2019/11/18/video-creating-a-tei-customization-with-the-new-roma/>
- <https://www.youtube.com/watch?v=avRO-b2Bw>
- <https://pittsburgh-neh-institute.github.io/Institute-Materials-2020/>
- <https://libripendis.eu/posts/13edition/>
- <https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>

Exercice 1 - Analyse préparatoire



1. Analyse d'un document

Dans le cadre d'un projet d'édition de correspondance, vous devez publier une collection de cartes postales.

- **Analyser ce "mini corpus" en commençant par identifier les catégories d'informations intéressantes pour le futur site web de cette édition.**

Liens vers les images sur la bibliothèque numérique Gallica

- <https://gallica.bnf.fr/view3if/ga/ark:/12148/btv1b10582689t>
- <https://gallica.bnf.fr/ark:/12148/btv1b10582693j/f1.item>
- <https://gallica.bnf.fr/ark:/12148/btv1b105826997/f2.item>
- <https://gallica.bnf.fr/ark:/12148/btv1b10582698s/f1.item#>



Questions pouvant guider votre réflexion

- Quelles sont les caractéristiques principales ?
- Quels sont les principaux blocs ou subdivisions ?
- Vous avez peu de temps : qu'allez-vous sélectionner ? (prenez en compte le lectorat visé, les usages de l'information, le rapport coût / bénéfice)
- Y a-t-il des singularités, des phénomènes rares ?
- Certains éléments paraissent-ils a priori problématiques pour l'encodage ?
- Quels aspects de la présentation (forme, mise en page) sont significatifs ? Lesquels devrait-on conserver
- Quels types de normalisation/modernisation du texte seraient utiles et appropriés ? Faudrait-il en garder la trace ou les faire directement sans les signaler ?



2. Structuration XML

Proposez une structuration XML en deux étapes :

1. Dessinez une arborescence
2. Transposez cette arborescence à l'aide de balises XML

Pour aller plus loin :

[Document Analysis Worksheet](#) (source : Women Writers Project)

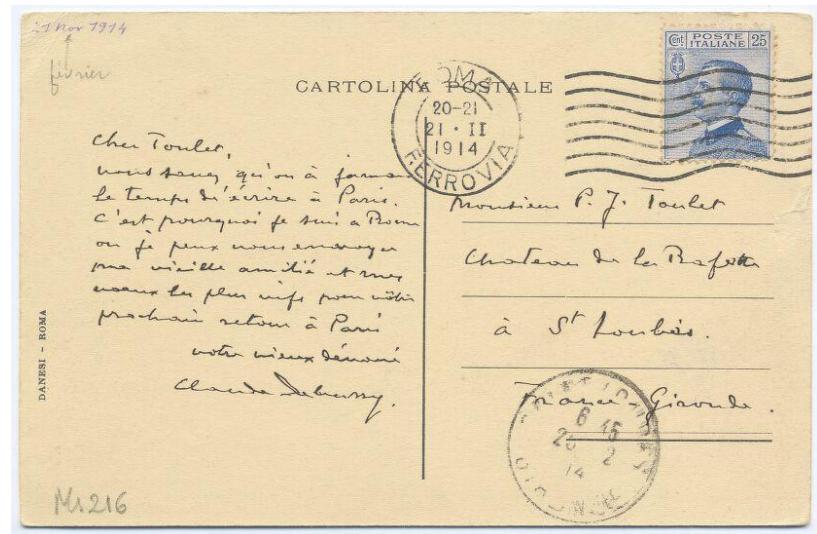
Corrigé 1



<https://gallica.bnf.fr/view3if/qa/ark:/12148/btv1b10582689t>



Le Grand Hôtel de Russie sur la Piazza del Popolo - Via del Babuino





Objets d'intérêts possibles

- L'image
- Le tampon postal
- Le texte du message écrit sur la carte
- l'adresse
- la légende
- Les informations sur l'édition, le traitement documentaires ,
d'autres métadonnées
- ...



Arborescence

- Carte postale
 - recto
 - image
 - légende
 - timbre
 - verso
 - message
 - adresse du destinataire
 - édition
 - timbre
 - éditeur



Structure XML ad hoc

```
<carte n="0010">

    <recto url="https://gallica.bnf.fr/ark:/12148/btv1b10582689t/f2.item">
        <legende>Le Grand Hôtel de Russie sur la Piazza del Populo - Via del Babuino</legende>
    <verso url="https://gallica.bnf.fr/ark:/12148/btv1b10582689t/f2.item">

        <obliteration>
            <date>21/11/1914</date>
            <lieu>Rome</lieu>
            <cachet>Roma Ferrovia</cachet>
        </obliteration>

        <message>
            <p>Cher Toulet</p>
            <p>Vous savez qu'on n'a jamais le temps d'écrire à Paris. C'est pourquoi (...) </p>
            <p>Votre vieux dévoué</p>
            <p>Claude Debussy</p>
        </message>

        <destinataire>
            <p>Monsieur P. J. Toulet</p>
            <p>Chateau de la Rafette à Saint-Loubès</p>
            <p>France - Gironde.</p>
        </destinataire>

    </verso>
</carte>
```



Remarques

- Limites du choix du français pour nommer les balises.
- Des choses sont laissées de côté (retours à la ligne, faute de syntaxe, mention sur l'édition, marques des bibliothécaires, structuration de l'adresse).
- On pourrait définir de nouvelles balises pour les encoder... ou bien utiliser un vocabulaire existant qui fait déjà très bien le travail, comme la TEI...