

# Introduction à l'édition numérique de sources anciennes

TD M2 Antiquités numériques, 2ème partie

# 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

# Présentation du cours

2e partie du TD “Antiquités numériques” - 5 séances de 2h le mardi de 12:00 à 14:00

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	Exploiter l'encodage : navigation dans l'arborescence, transformations
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 »

# Objectifs de la séance

- 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

Qu'est-ce qu'une  
édition numérique de sources ?

- [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](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://shelleygodwinarchive.org/sc/oxford/frankenstein/volume/ii/#/p41>
- [https://www.unicaen.fr/puc/sources/depiscibus/consult/hortus\\_fr/FR.hs.4.47.html](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](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://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/elena/proust\\_prototype/](https://peterstokes.org/elena/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
- 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 "transmedialisation"  
(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.

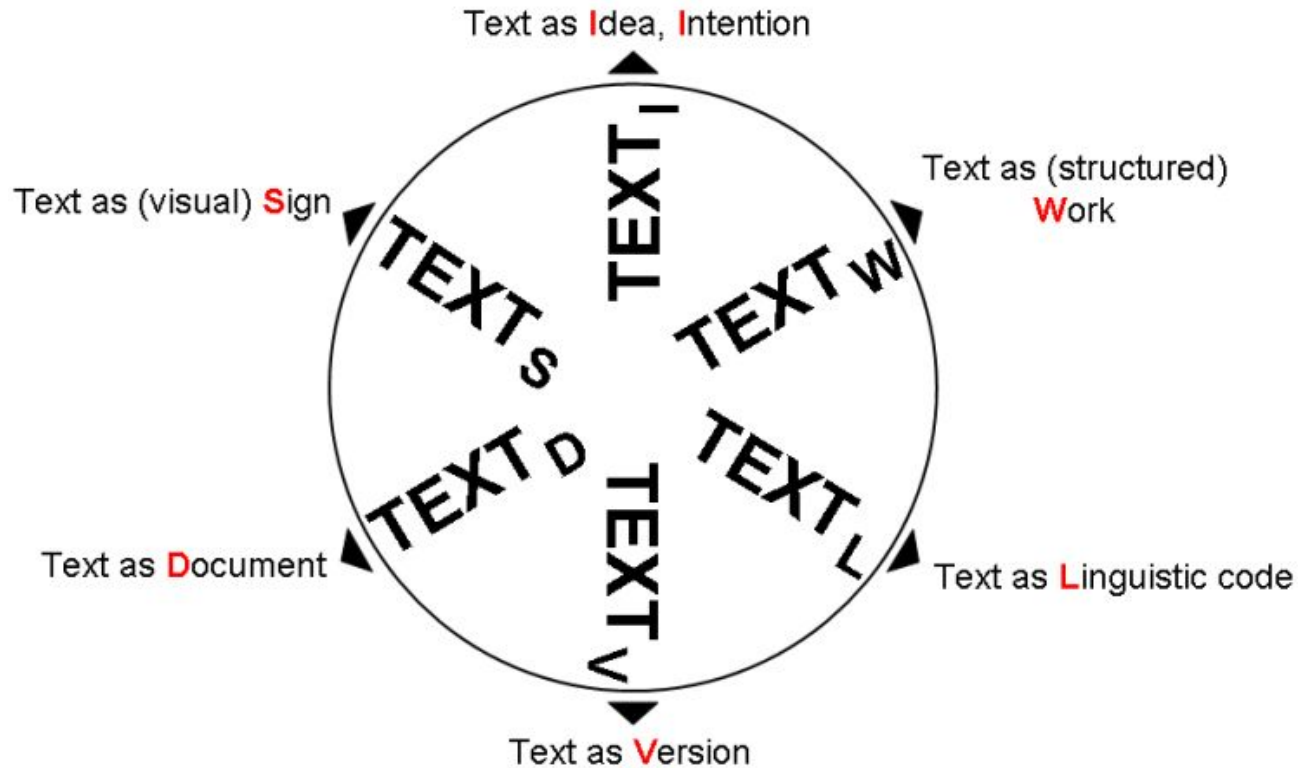
### SOCIAL



### META

First published: Apr 2014

Last updated: Sep 2023



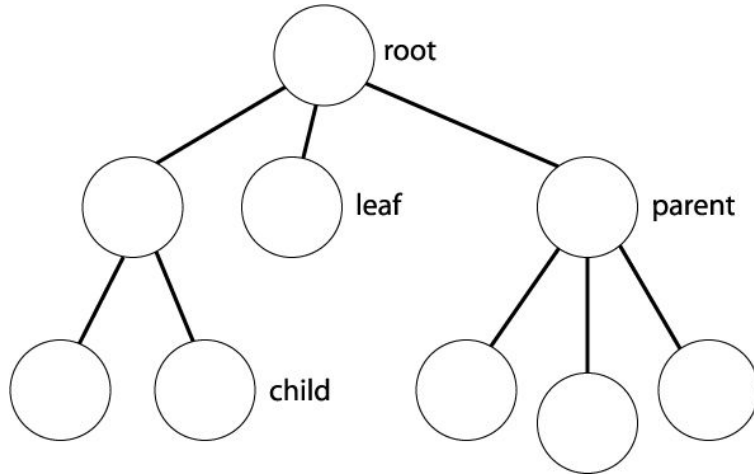
The text wheel by Patrick Sahle (2016)

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)

structure du document (titres, paragraphes, strophes)

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

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

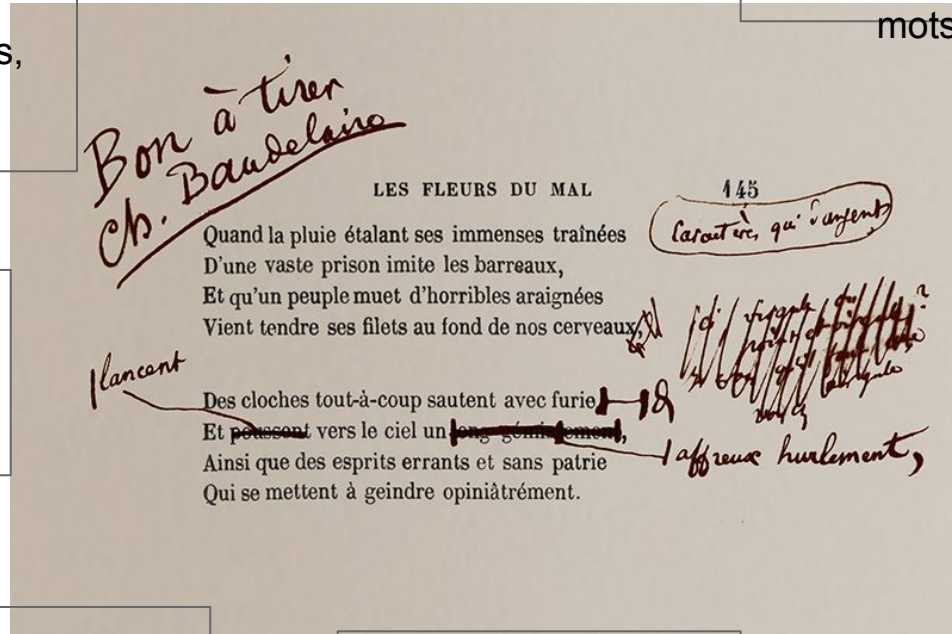
Métadonnées (auteur, date de création, provenance, genre)

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

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

Processus : ratures, campagnes de correction

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 éditable 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

balise ouvrante



Contenu



balise fermante



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

# Anatomie d'une balise

Attribut



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



valeur d'attribut

# Balisage sémantique et non 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...

# Caractéristiques à connaître

- 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
- 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)

# 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

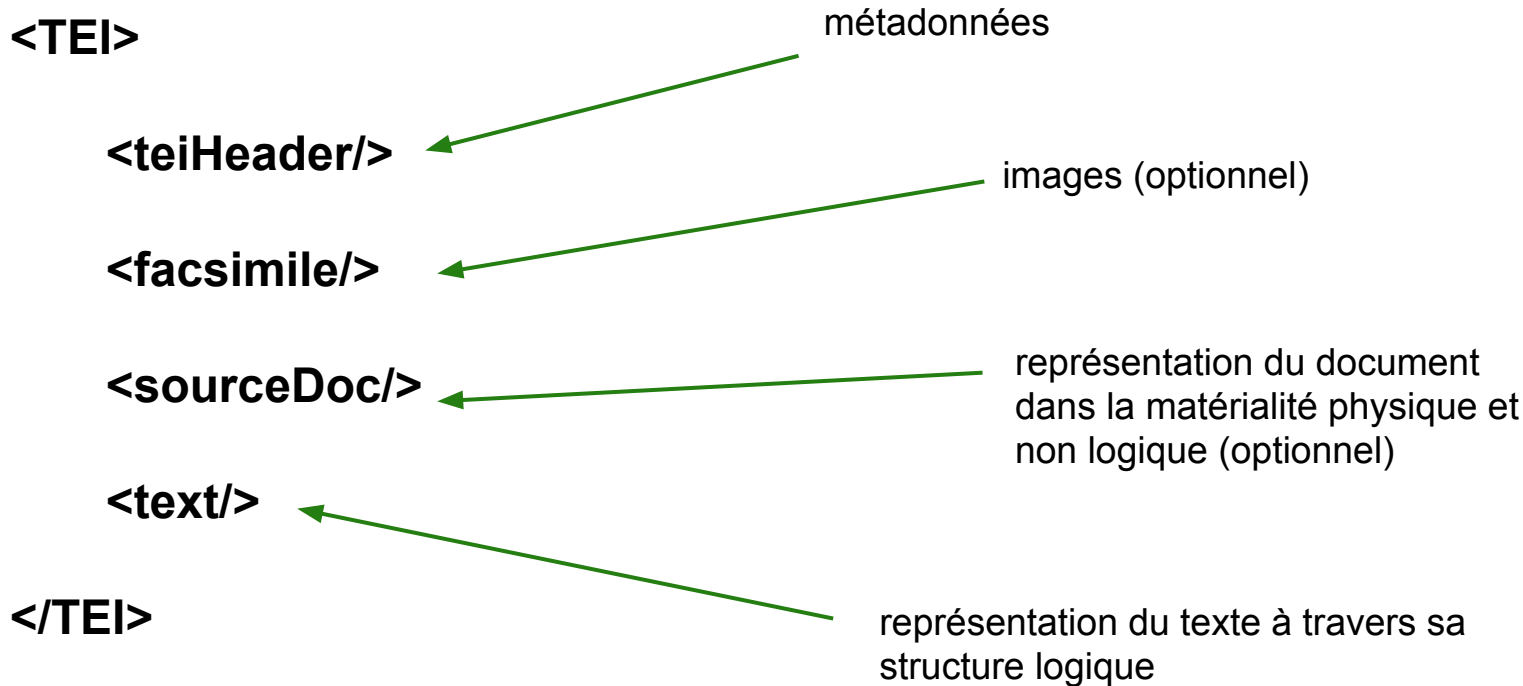
Develop your schema with your goals in mind !

# 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 EpiDoc

TEI

teiHeader

facsimile (images)

text

transcription

apparatus

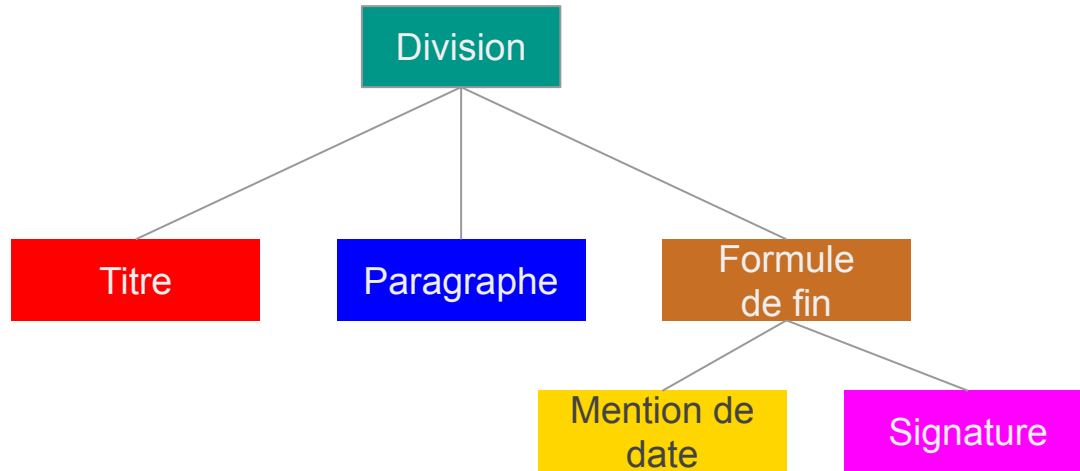
traduction

commentary

bibliography

```
<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>
          </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>
    </fileDesc>
  </teiHeader>
  <facsimile>
    <graphic url="photograph of text or monument"/>
  </facsimile>
  <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 logique d'une 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

## **<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-dateline.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://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-signed.html>

```
<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>
```

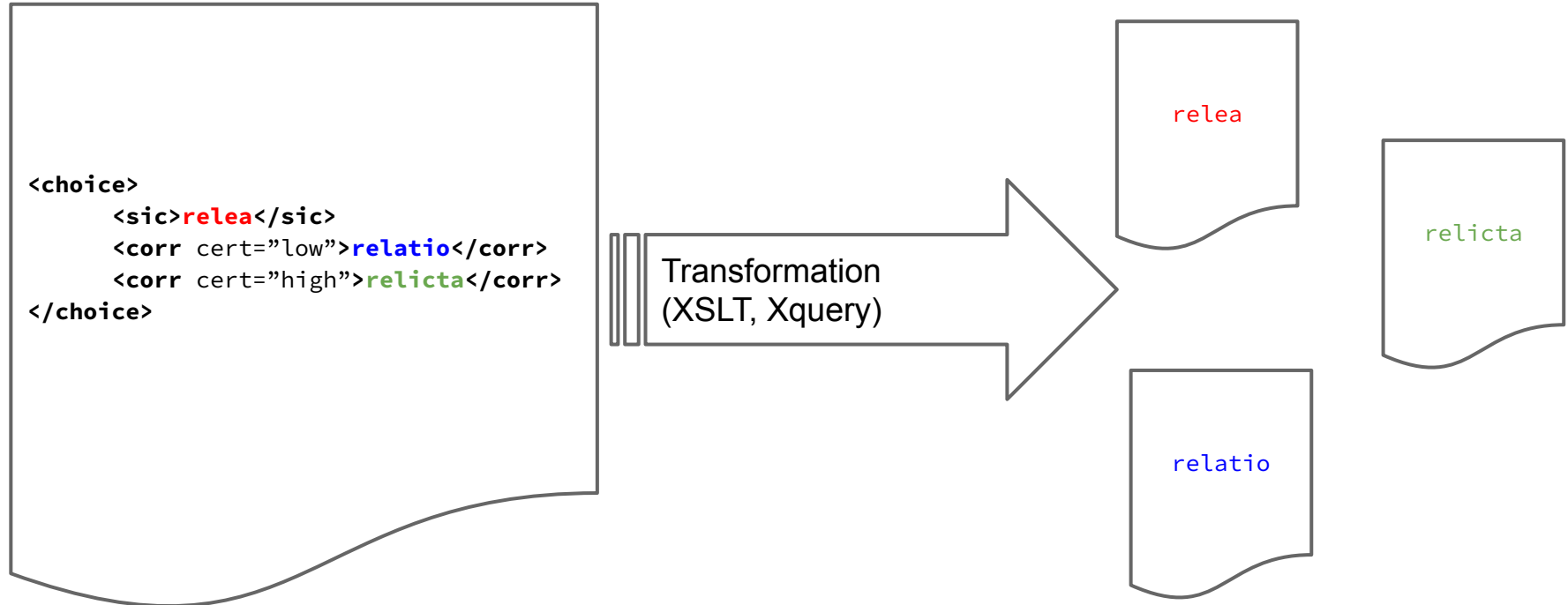




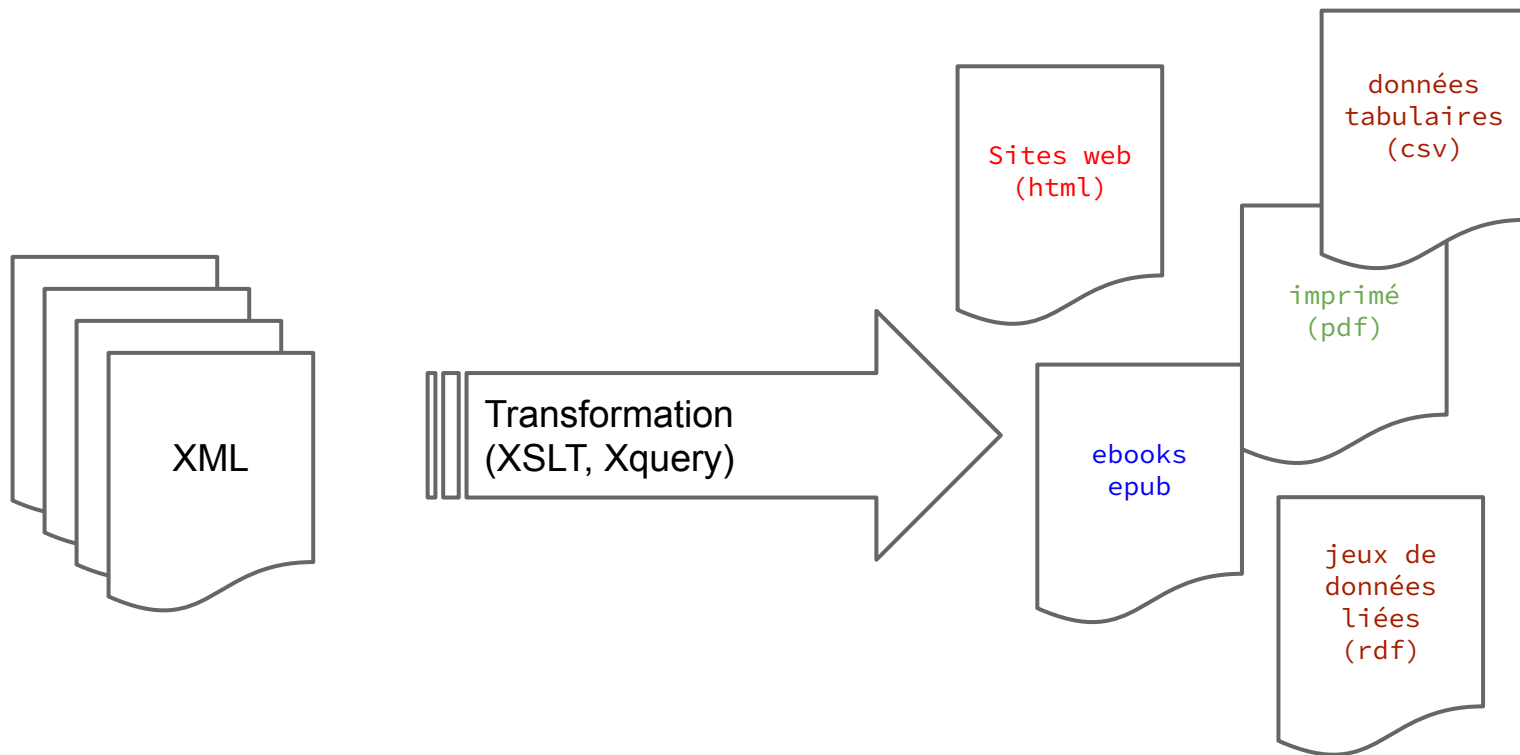
# Intervention scribale : correction

```
<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>
```

# “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