



UNIL | Université de Lausanne

Faculté des lettres

**Centre NUCLEUS**



Formation de deux jours

## **Initiation à l'encodage XML-TEI**

Lausanne, février 2025

Sonia Solfrini et Simon Gabay

(Université de Genève)



## Plan du cours 3 : le teiHeader

- Le <teiHeader>
- Le <fileDesc>
- Quelques exercices

Le **<teiHeader>**, l'en-tête du fichier, contient les métadonnées du texte et la granularité de son encodage dépend de chaque projet. Voici un teiHeader minimal :

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Titre du texte électronique.</title>
    </titleStmt>
    <publicationStmt>
      <p>Informations concernant la publication ou la distribution d'un texte électronique.
        Pas forcément dans la balise "p", mais au moins une balise est attendue.</p>
    </publicationStmt>
    <sourceDesc>
      <p>La ou les sources à partir desquelles un texte électronique a été dérivé ou généré.
        Pas forcément dans la balise "p", mais au moins une balise est attendue.</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Le TEI Header se compose de 4 parties :

- <fileDesc> : donne la description bibliographique complète d'un fichier électronique → obligatoire ;
- <encodingDesc> : précise les méthodes et les principes éditoriaux qui ont permis la transcription du texte et l'élaboration du fichier → optionnel ;
- <profileDesc> : concerne les informations non bibliographiques décrivant un document → optionnel ;
- <revisionDesc> : comprend la description des modifications apportées dans le fichier TEI et alimente un journal des modifications → optionnel.

<teiHeader>

<fileDesc>

<encodingDesc>

<profileDesc>

<revisionDesc>

</teiHeader>

## Quelques exemples

- Le TEI Header du projet *Gallicorpora* :  
<https://github.com/Gallicorpora/TEI>
- Le TEI Header du projet *Démêler le corbel* :  
<https://github.com/DesenrollandoElCorbel/Varios-TEI-files/tree/main/xml-files>
- Le TEI Header du projet SETAF :  
<https://github.com/SETAFDH/TEI-SETAF>

# Le <teiHeader> : le <fileDesc> du projet SETAF

N.B. : Le <fileDesc> est obligatoire !

Le **<fileDesc>** (file description) donne la description bibliographique complète d'un fichier électronique. Notre <fileDesc> contient les éléments suivants : **<titleStmt>**, <extent>, **<publicationStmt>** et **<sourceDesc>**.

```
<fileDesc>  
  <titleStmt>  
  <extent>  
  <publicationStmt>  
  <sourceDesc>  
</fileDesc>
```

# Le <teiHeader> : le <fileDesc> du projet SETAF

N.B. : Le <fileDesc> est obligatoire !

Le <**fileDesc**> (file description) donne la description bibliographique complète d'un fichier électronique. Notre <fileDesc> contient les éléments suivants : <titleStmt>, <extent>, <publicationStmt> et <sourceDesc>.

```
<fileDesc>  
  <titleStmt>  
  <extent>  
  <publicationStmt>  
  <sourceDesc>  
</fileDesc>
```

## Le <teiHeader> : le <fileDesc> du projet SETAF

Le **<titleStmt>** (title statement) regroupe les informations relatives au titre électronique d'un ouvrage et aux responsables de son contenu. Notre <titleStmt> contient les éléments suivants : <title> et <respStmt>.

**<title>** désigne le titre du fichier TEI selon les règles suivantes:

- Titre bref. Lieu : imprimeur, date.
- Si lieu, imprimeur ou date ne sont pas connus : Titre bref. [s.l.] : [s.n.], [s.d.].
- Si lieu, imprimeur ou date ne sont pas indiqués mais connus : Titre bref. [Lieu] : [imprimeur], [date].



## Le <teiHeader> : le <fileDesc> du projet SETAF

**<respStmt>** donne le nom du responsable de la collection électronique, ainsi que ceux de l'encodage, etc. Notre <respStmt> contient les éléments <resp> et <persName>.

- <resp> indique les tâches dont une personne est responsable (e.g. l'encodage des métadonnées).
- <persName> inclut l'attribut xml:id dont la valeur sont les initiales d'une personne donnée en majuscules. Il contient les éléments <forename>, <surname> et <ptr>. <ptr> inclut la valeur orcid pour l'attribut type et la valeur x (une variable) pour l'attribut target.

# Le <teiHeader> : le <fileDesc> du projet SETAF

## Modèle de base du <titleStmt> :

```
<titleStmt>
  <title>Titre bref. Lieu : imprimeur, date.</title>
  <respStmt>
    <resp>Tâche(s).</resp>
    <persName xml:id="INITIALES">
      <forename>Prénom</forename>
      <surname>Nom</surname>
      <ptr type="orcid" target="x"/>
    </persName>
  </respStmt>
</titleStmt>
```

## Exemple :

```
<titleStmt>
  <title>Plusieurs belles et bonnes chansons. [Neuchâtel] : [Pierre de Vingle], 1533.
</title>
  <respStmt>
    <resp>Metadata encoding.</resp>
    <persName xml:id="SS">
      <forename>Sonia</forename>
      <surname>Solfrini</surname>
      <ptr type="orcid" target="000900097367048X"/>
    </persName>
  </respStmt>
</titleStmt>
```

# Le <teiHeader> : le <fileDesc> du projet SETAF

N.B. : Le <fileDesc> est obligatoire !

Le <fileDesc> (file description) donne la description bibliographique complète d'un fichier électronique. Notre <fileDesc> contient les éléments suivants : <titleStmt>, <extent>, <publicationStmt> et <sourceDesc>.

```
<fileDesc>
  <titleStmt>
  <extent>
  <publicationStmt>
  <sourceDesc>
</fileDesc>
```

## Le <teiHeader> : le <fileDesc> du projet SETAF

Le **<publicationStmt>** (publication statement) donne les informations sur l'organisme responsable de la publication. Notre <publicationStmt> contient les éléments suivants : <publisher>, <authority>, <address> (avec deux <addrLine>), <availability> et <date>.

<availability> décrit la restriction juridique pour la distribution de la publication. Il inclut les attributs status="restricted" et n="CC-BY", et il contient les éléments <licence> et <p>. <licence> inclut l'attribut target="<https://creativecommons.org/licenses/by/4.0/>".

<date> indique la date de création du fichier TEI : année-mois-jour.

# Le <teiHeader> : le <fileDesc> du projet SETAF

## Exemple du <publicationStmt> :

```
<publicationStmt>
  <publisher>Projet SETAF</publisher>
  <authority>Institut Histoire de la Réformation (IHR),
    université de Genève</authority>
  <address>
    <addrLine>22 boulevard des Philosophes</addrLine>
    <addrLine>CH-1211 Genève</addrLine>
  </address>
  <availability status="restricted" n="CC-BY">
    <licence
      target="https://creativecommons.org/licenses/by/4.0/">
      <p>Creative Commons Attribution 4.0 CC BY 4.0</p>
    </availability>
  <date when="0000-00-00"/>
</publicationStmt>
```

# Le <teiHeader> : le projet *Démêler le cordel*

## Exemple de <publicationStmt> :

```
<publicationStmt>  
  <publisher>Herederos de Juan Jolis</publisher>  
  <pubPlace ref="https://www.geonames.org/3128760/barcelona.html">Barcelona</pubPlace>  
  <date cert="low" when="1800-01-01">[s.a.]</date>  
</publicationStmt>
```

# Le <teiHeader> : le <fileDesc> du projet SETAF

N.B. : Le <fileDesc> est obligatoire !

Le <fileDesc> (file description) donne la description bibliographique complète d'un fichier électronique. Notre <fileDesc> contient les éléments suivants : <titleStmt>, <extent>, <publicationStmt> et <sourceDesc>.

```
<fileDesc>
  <titleStmt>
  <extent>
  <publicationStmt>
  <sourceDesc>
</fileDesc>
```

## Le <teiHeader> : le <fileDesc> du projet SETAF

Le **<sourceDesc>** (source description) donne toutes les informations sur la source originale. Il contient l'élément **<msDesc>** (manuscript description) qui concerne à l'origine la description de manuscrits mais qu'on peut étendre aux imprimés anciens. Notre <msDesc> contient les éléments suivants : **<msIdentifier>**, **<msContents>**, <physDesc> et <additional>.

- **<msIdentifier>** donne les informations pour retrouver la source. Notre <msIdentifier> inclut l'attribut `xml:id="x"` dont la variable x se construit de la manière suivante: `codeBibliothèque_coteBibliothèque` (par ex. `xml:id="ZB_1327"`). Ensuite, notre <msIdentifier> contient les éléments suivants : <country>, <settlement>, <institution> et <idno> avec l'attribut `type="shelfmark"`.



# Le <teiHeader> : le <fileDesc> du projet SETAF

## Modèle de base du <msIdentifier> :

```
<msIdentifier xml:id="x">  
  <country>Pays</country>  
  <settlement>Ville</settlement>  
  <institution>Bibliothèque</institution>  
  <idno type="shelfmark">Cote</idno>  
</msIdentifier>
```

## Exemple :

```
<msIdentifier xml:id="BGE_Bd_1475">  
  <country>Suisse</country>  
  <settlement>Genève</settlement>  
  <institution>BGE</institution>  
  <idno type="shelfmark">Bd 1475</idno>  
</msIdentifier>
```

## Le <teiHeader> : le <fileDesc> du projet SETAF

- **<msContents>** (manuscript contents) décrit le contenu intellectuel d'un ouvrage. Notre <msContents> contient l'élément <p> qui contient <biblStruct> qui, à son tour, contient <monogr>. Notre <monogr> comprend les principales informations bibliographiques de l'ouvrage et contient les éléments suivants : **<author>**, **<title>** et **<imprint>**.

```
<msContents>
  <p>
    <biblStruct>
      <monogr>
        <author>
        <title>
        <imprint>
      </monogr>
    </biblStruct>
  </p>
</msContents>
```

# Le <teiHeader> : le <fileDesc> du projet SETAF

## Modèle de base :

```
<author>
  <persName ref="isni:x">
    <surname>Nom</surname>
    <forename>Prénom</forename>
  </persName>
</author>
<title type="complete_title">Titre entier</title>
<title type="short_title">Titre bref</title>
<imprint>
  [...]
</imprint>
```

## Exemple :

```
<author>
  <persName role="presumed_author" ref="isni:0000000083926271">
    <surname>Malingre</surname>
    <forename>Matthieu</forename>
  </persName>
</author>
<title type="complete_title">S'ensuivent plusieurs belles et bonnes chansons/ que les
chrestiens peuvent chanter en grande affection de cuer/ pour et affin de soulager
leurs esperitz et de leur donner repos en dieu/ au nom duquel elles sont composees
par rithmes/ au plus pres de l'esperit de Jesus Christ/ contenu es saintes
escriptures.</title>
<title type="short_title">Plusieurs belles et bonnes chansons</title>
<imprint>
  [...]
</imprint>
```

# Le <teiHeader> : le <fileDesc> du projet SETAF

## Modèle de base :

```
<imprint>
  <pubPlace ref="geonames:x">Lieu de
publication</pubPlace>
  <respStmt>
    <resp>Imprimeur</resp>
    <persName ref="isni:x">
      <surname>Nom</surname>
      <forename>Prénom</forename>
    </persName>
  </respStmt>
  <date>Date de publication</date>
</imprint>
```

## Exemple :

```
<imprint>
  <pubPlace cert="medium"
ref="geonames:2659496">Neuchâtel</pubPlace>
  <respStmt>
    <resp>Imprimeur</resp>
    <persName role="presumed_printer"
ref="isni:000000007752492X">
      <surname>Vingle</surname>
      <forename>Pierre de</forename>
    </persName>
  </respStmt>
  <date cert="high">1533</date>
</imprint>
```

## Plan du cours 3 : le teiHeader

- Le <teiHeader>
- Le <fileDesc>
- Quelques exercices

Suivant le modèle indiqué, encodez les métadonnées de l'ouvrage suivant :

<https://www.ville-ge.ch/musinfo/bd/bge/gln/notice/?no=GLN-4499>.