

Séance 4 : Structurer un fichier TEI

Tout document TEI a au moins 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>
```

```
    <!-- métadonnées décrivant un texte -->
```

```
  </teiHeader>
```

```
  <text>
```

```
    <!-- une représentation du texte lui-même -->
```

```
  </text>
```

```
</TEI>
```

Le **teiHeader** minimal comporte les trois sections suivantes :

- `<titleStmt>` : informations identifiant le document lui-même ;
- `<publicationStmt>` : informations sur la façon dont il est distribué ou publié ;
- `<sourceDesc>` : indications sur ses origines.

L'élément `<text>` contient les trois parties suivantes :

- `<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.

Le teiHeader

L'en-tête TEI possède quatre composants principaux :

- `<fileDesc>` : description bibliographique du document, **attention : il est obligatoire ;**
- `<encodingDesc>` : description de l'encodage ;
- `<profileDesc>` : description détaillée des aspects non bibliographiques;
- `<revisionDesc>` (revision description, description des révisions) résume l'historique des révisions pour un fichier.

NB : Si je déclare une langue dans le profileDesc, j'utilise la norme iso 639 pour renseigner la valeur de `@xml:lang` :

https://fr.wikipedia.org/wiki/ISO_639

Le fileDesc

Il se compose de trois parties obligatoires :

- `<titleStmt>` : Déclaration du titre ;
- `<publicationStmt>` : Déclaration de la publication ;
- `<sourceDesc>` : Description de la source.

Exemple :

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Titre de l'œuvre</title>
    </titleStmt>
    <publicationStmt>
      <p>Informations sur la publication de l'œuvre.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informations sur la source dont est tirée l'œuvre.
</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Allez plus loin...

Le teiHeader est extrêmement complexe, car il est conçu pour s'adapter à des usages scientifiques très divers.

Le header doit garantir la qualité scientifique du document, permettre l'échange des données et leur conservation. On peut donc, la plupart du temps, chercher à reproduire des modèles qui ont déjà fait leurs preuves comme le *Dublin Core*.

L'objectif du Dublin Core est de fournir un socle commun d'éléments descriptifs pour améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré.

Le *Dublin Core* prévoit 15 éléments tous facultatifs et tous répétables, qui portent sur la description :

- du contenu : Title, Subject, Description, Source, Language, Relation, Coverage ;
- de la propriété intellectuelle : Creator, Contributor, Publisher, Rights ;
- de l'instanciation : Date, Type, Format, Identifier.

Pour en savoir plus :

http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html

Certaines informations du *Dublin Core* ne sont pas comprises en TEI dans le `<fileDesc>` comme la langue de l'œuvre ou sa date de création. Ces informations primordiales sont à ajouter dans le `<profileDesc>` .