

TEI, Schéma et ODD

- Pour avoir un fichier TEI valide, il faut :
 - Respecter la syntaxe XML;
 - Suivre le schéma ;
 - Respecter la sémantique des éléments TEI.
- ODD : One Document Does it all.
- Notion au cœur de la TEI, l'ODD est un document qui contient :
 - La documentation (sur les éléments utilisés, les choix éditoriaux, etc.), dans plusieurs langues et plusieurs formats;
 - Le schéma, contenant les éléments et la manière de les utiliser.
- Pourquoi ?
 - Repose sur un format XML bien établi, qui fait corps avec la TEI;
 - Peut être échangée avec d'autres projets ;
 - Est pérenne, stable et standardisée.

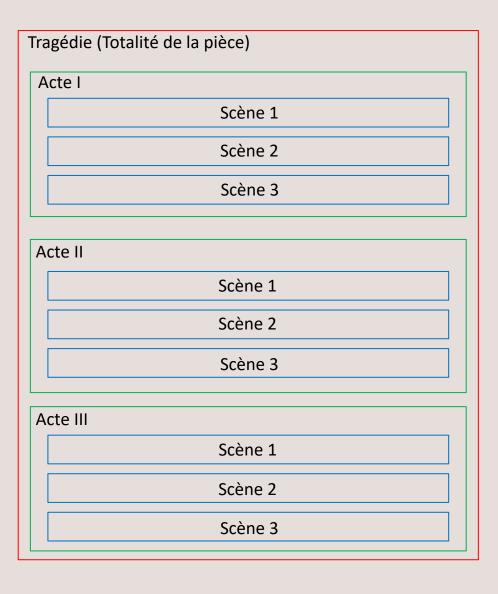
Structure de base d'un fichier XML

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>...</teiHeader>
                                                      Métadonnées
  <text>
                                                     Contenus préliminaires
          <front>...</front>
                                                     (Page de titre, préface, dédicaces, etc.)
          <body>
            <div>...</div>
                                                      Corps du texte
          </body>
          <back>...</back>
                                                      Annexes
                                                      (tables, index, etc.)
   </text>
</TEI>
```

Un texte, plusieurs unités

- La TEI considère qu'un texte se compose de plusieurs unités, appelées divisions : livres, parties, chapitres, tomes, poèmes, pièces, actes, etc.
- Ces divisions structurent le texte en plusieurs unités logiques avec <div>.
- Plusieurs attributs :
 - @n: numéro d'une division;
 - @type : type de la division.
- Le titre d'une division est encodé avec <head>.
- Avant d'encoder, il est important de réfléchir à la structure de votre texte en détail : cela vous évitera de nombreux inconvénients !

Bérénice est divisée en acte ; les actes, en scènes.



```
<div type="tragedie">
  <head>BERENICE</head>
  <div type="acte" n="1">
      <head>Acte I</head>
      <div type="scene" n="1">
          <head>Scène 1</head>
          Texte de la scène 1
      </div>
      <div type="scene" n="2">
          <head>Scène 2</head>
          Texte de la scene 2
      </div>
  </div>
  <div type="acte" n="2">
     <head>Acte II</head>
  </div>
</div>
```

Let's try: Les Misérables (1)

Enfin, ajouter les attributs @n et @type.



L'encodage de la prose (1)

- La structure d'une page :
 - <pb/> <pb/>(Page Beginning) : indique le début d'une page ;
 - <fw> (Form Work): indique les titres courants, les numéros de page, etc.
 - @type: pageNum, header, sig, catch
- Les blocs de texte :
 - : indique les paragraphes ;
 - <lb/> (Line Beginning) : indique le début d'une ligne.

L'encodage de la prose (2)

Les italiques :

- <hi> (Highlight): Élément générique, qui vous permet d'encoder n'importe quel élément différent du reste du texte (italique, gras, majuscules, petites capitales, etc.).
 - @rend: italic, bold, uppercase, lowercase, etc.
- <emph> (Emphasized) : Passage mis en évidence pour des raisons linguistiques ou rhétoriques. Ex. : C'est <emph>le</emph> livre de l'année !
- <foreign> : Passage dans une autre langue.
 - @xml:lang : Pour indiquer la langue utilisée.
- <title>: Titre d'une œuvre.
- <mentioned>: Mot autonyme. Ex. : <mentioned>Toujours</mentioned> prend toujours un « s ».

Zoom sur @rend

- Attribut passe-partout qui indique la manière dont un élément est présenté dans le texte source (gras, italique, souligné, en couleur, etc.);
- S'emploie avec la plupart des éléments TEI : attribut global.

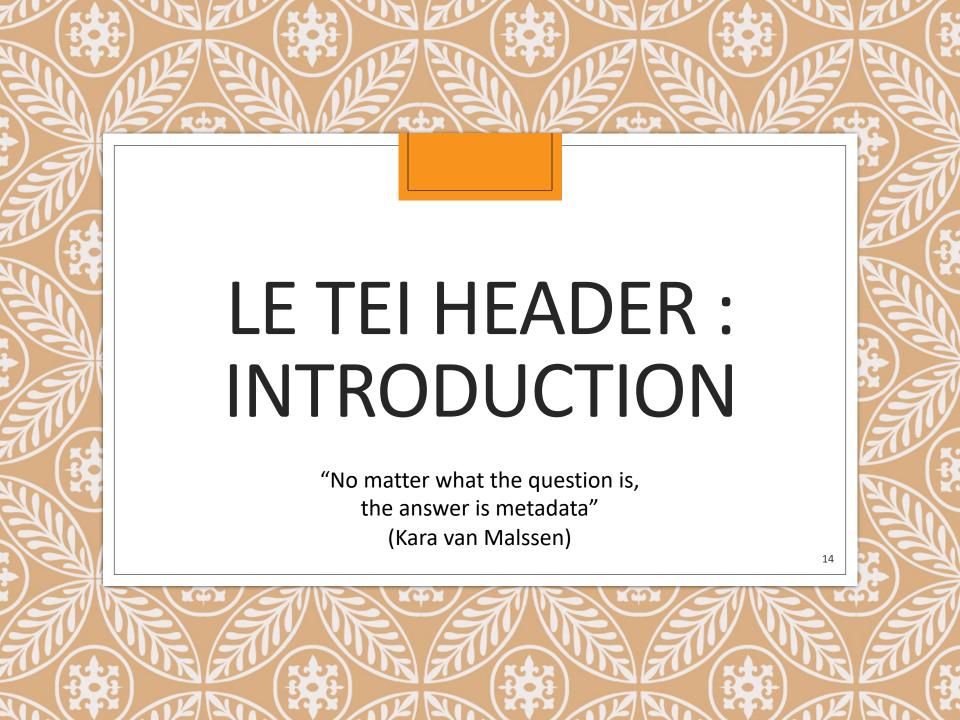
```
<head rend="smallCaps">Le mariage de Figaro</head>
<hi rend="italic">Le Barbier de Séville</hi>
Nul n'aurait pu le dire ; [...]
<c rend="drop">I</c>l était une fois [...]
```

L'encodage de la prose (3)

- Les citations :
 - <q>: Passage séparé du texte par des guillemets (générique).
 - <said>: Discours ou pensée par une personne réelle.
 - Ex: Il a dit qu'<said aloud="true" direct="false">il aimait les pommes</said>.
 - <quote> : Citation (référence à un autre auteur).
 - Ex : On rapporte que Xénocrate, philosophe du premier ordre, interrogé sur l'avantage que ses disciples retiraient de ses leçons, répondit : <quote>« Ils apprennent à faire par leur choix ce que les lois leur ordonnent de faire. »</quote>

Quid du théâtre ? De la poésie ?

- La TEI propose des éléments spécifiques pour encoder les vers, les strophes, les répliques d'une pièce de théâtre, la correspondance ou encore les entrées d'un dictionnaire.
- o Pour en savoir plus sur la poésie et le théâtre, rendez-vous sur TEI By Example :
 - Cours sur le théâtre : https://teibyexample.org/tutorials/TBED05v00.htm
 - Cours sur la poésie : https://teibyexample.org/tutorials/TBED04v00.htm
- Pour la correspondance : https://encoding-correspondence.bbaw.de/v1/index.html
- Pour les dictionnaires : https://dariah-
 eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html



Le TEI Header : généralités

- Contient les métadonnées descriptives, administratives, bibliographiques et techniques de votre fichier TEI.
- Les métadonnées sont partout et permettent de :
 - Décrire une ressource numérique ;
 - Faciliter la recherche de données;
 - Gérer des collections numériques ;
 - **Préserver** et conserver les ressources numériques.

Le TEI Header: structure

- Le TEI Header se compose de 4 parties :

 - <encodingDesc> : Description du projet et des choix éditoriaux (corrections, particularités de l'encodage...) → Optionnel ;
 - ∘ contexte o contexte, de la langue, du sujet... → Optionnel;
- Pour aller plus loin : https://teibyexample.org/tutorials/TBED02v00.htm

Le TEI Header – FileDesc (1)

 3 éléments principaux : <titleStmt> : Titre du fichier TFL <title> : Titre du fichier électronique (Obligatoire). <author> : Auteur (Facultative). <respStmt> : Autres responsabilités (Facultative). <resp>: Travail effectué (transcription, encodage, révision...); <name> : Nom du responsable de l'action ; <date>. <publicationStmt> : Cadre de diffusion du fichier (responsable de la publication, gestion des droits). <authority> : Organismes responsables de l'œuvre numérique ; <availability>: Aspects juridiques (copyright...). OU

Le TEI Header – FileDesc (2)

- < sourceDesc> : Description de la source physique. Plusieurs solutions:
 - > : Description non-structurée.
 - <bibl> : Description semi-structurée.

Le TEI Header – FileDesc (3)

```
    <bibl/>biblFull> : Données structurées.

   <titleStmt> : Titre et auteur(s) de la source physique.
       <title>;
       < <author>;
       <editor> (traducteur, imprimeur...);
       <resp> (Autres rôles : illustrateur...).

    <editionStmt> : Description de l'édition.

    <edition> : Caractéristiques principales (Nouvelle édition...).

   • <publicationStmt> : Information sur la publication.
       o <pubPlace> : Lieu de publication ;
       o <publisher> : Imprimeur ;
       <date>: Date de publication;
       <idno>: Identifiant (cote).

    <notesStmt> : Informations complémentaires sur la source.

       < note>;
       <relatedItem>.
```