20 – Page de titre électronique

Chaque texte TEI a un en-tête qui fournit des informations analogues à celles que l'on trouve sur une page de titre imprimée. L'en-tête est introduite par l'élément *<teiHeader>* et comporte quatre parties majeures :

<fileDesc>
: contient une description bibliographique complète d'un fichier électronique ;

<encodingDesc>
: documente le rapport entre un texte électronique et la source ou les sources dont il dérive ;

<profileDesc>
: fournit une description détaillée des aspects non bibliographiques d'un texte, spécifiquement les langues et les sous-langues employées, les circonstances de sa production, les participants, et leur environnement ;

<revisionDesc>
: résume l'historique des révisions du fichier.

Un corpus ou une collection de textes, qui partagent beaucoup de caractéristiques, peuvent avoir un en-tête pour le corpus et des en-têtes individuels pour chaque composant du corpus. Dans ce cas l'attribut *type* indique le type d'en-tête.

```
<teiHeader type=corpus>
```

introduit l'en-tête pour les informations concernant le corpus.

Certains des éléments de l'en-tête contiennent du texte libre, codé sous forme d'un ou plusieurs *<p>*. D'autres sont groupés :

– les éléments dont les noms se terminent par *Stmt* (pour *statement*) englobent habituellement un groupe d'éléments représentant certaines informations structurées ;

– les éléments dont les noms se terminent par *Decl* (pour *declaration*) comportent des informations sur les règles de codage spécifiques mises en œuvre dans le document ;

– les éléments dont les noms se terminent par *Desc* (pour *description*) contiennent une description en texte libre.

20.1. Description du fichier

L'élément *<fileDesc>* est obligatoire. Il contient une description bibliographique complète du fichier, à l'aide des éléments suivants :

<titleStmt>
: regroupe des informations concernant le titre d'une œuvre, ainsi que les parties responsables de son contenu intellectuel ;

<editionStmt>
: regroupe des informations relatives à une édition donnée d'un texte ;

<extent>
: décrit la taille approximative du texte électronique tel que stocké sur un certain type de support, spécifié dans n'importe quelles unités communément employées ;

<publicationStmt>
: regroupe des informations concernant la publication ou la diffusion d'un texte de type électronique ou autre ;

<seriesStmt>
: regroupe des informations concernant la collection ou série, s'il y en a une, à laquelle une publication appartient ;

<notesStmt>
: regroupe toute annotation comportant des informations venant s'ajouter à celles qui sont enregistrées dans d'autres parties de la description bibliographique ;

<sourceDesc>
: comprend une description bibliographique du ou des textes originaux à partir desquels un texte électronique a été dérivé ou généré.

Un en-tête minimal aura la structure suivante :

```
<teiHeader>
    <fileDesc>
         <titleStmt>... </titleStmt>
         <publicationStmt>... <publicationStmt>
         <sourceDesc>... <sourceDesc>
    </fileDesc>
</teiHeader>
```

20.1.1. Mention du titre

Un élément *<titleStmt>* peut contenir les éléments suivants :

<title> contient le titre d'une œuvre, que ce soit un article, un livre, un journal ou une série, y compris tous les sous-titres ou titres alternatifs ;

<author>
dans une référence bibliographique, contient le nom du ou des auteur(s) d'une œuvre, que ce soit une personne ou une institution ; équivaut à la mention de responsabilité principale pour toute unité bibliographique ;

<sponsor>
précise le nom d'une institution ou organisation de parrainage ;

<funder>
précise le nom d'un individu, d'une institution, ou d'une organisation assurant le financement d'un projet ou d'un texte ;

<principal>
fournit le nom du chercheur principal responsable de la création d'un texte électronique ;

<respStmt>
fournit une mention de responsabilité nommant un responsable quant au contenu intellectuel d'un texte, une édition, un enregistrement ou une série, lorsque les éléments spécialisés prévus pour les auteurs, les rédacteurs, etc. ne suffisent pas ou ne s'appliquent pas.

Il est souhaitable que le titre permette de distinguer le fichier électronique du texte source, par exemple :

```
[title of source]: a machine readable transcription
[title of source]: electronic edition
A machine readable version of: [title of source]
```

L'élément *<respStmt>* regroupe les sous-composants suivants :

<resp> contient une expression décrivant la nature de la responsabilité intellectuelle d'une personne ;

<name>
contient un nom propre ou une expression substantive.

Exemple :

```
<titleStmt>
     <title>Maximes</title>
     <author>La Rochefoucauld, (1613--1680).
     <respStmt><resp>Texte établi, avec introduction ...</resp>
     <name>Jean Truchet</name></respStmt>
</titleStmt>
```

20.1.2. *Mention d'édition*

L'élément *<editionStmt>* regroupe des informations relatives à une édition d'un texte (l'expression *édition* étant employée avec son sens usuel en bibliographie) et peut inclure les éléments suivants :

<edition>
décrit les particularités d'une édition d'un texte ;

<respStmt>
fournit une mention de responsabilité nommant un responsable pour le contenu intellectuel d'un texte, une édition, un enregistrement ou une série, lorsque les éléments spécialisés pour les auteurs, les rédacteurs etc., ne suffisent pas ou ne s'appliquent pas.

Exemple :

```
<editionStmt>
     <edition n=U2>Troisième version, revue et corrigée
     <date>1987</date>
     </edition>
 </editionStmt>
```

La détermination exacte de ce qui représente une nouvelle édition d'un texte électronique est laissée à l'appréciation du codeur.

20.1.3. Mention relative à la taille du fichier

La mention <*extent*> décrit la taille approximative d'un fichier.

Exemple :

```
<extent>4500 bytes</extent>
```

20.1.4. Mention de publication

L'élément <*publicationStmt*> est obligatoire. Il peut contenir une simple description en texte libre ou des groupes composés à partir des éléments ci-dessus :

<*publisher*>
: fournit le nom de l'organisation responsable pour la publication ou la diffusion d'un article bibliographique ;

<*distributor*>
: fournit le nom d'une personne ou d'un organisme responsable de la diffusion d'un texte ;

<*authority*>
: fournit le nom d'une personne ou d'un autre organisme responsable de la fourniture d'un fichier électronique, quand cette personne ou cet organisme n'est ni un éditeur ni un distributeur.

Au moins un de ces trois éléments doit être présent, à moins que la totalité de la mention de publication ne soit en texte libre. Ces éléments peuvent à leur tour contenir les éléments suivants :

<*pubPlace*>
: contient le nom de l'endroit où une unité bibliographique a été publiée ;

<*address*>
: contient l'adresse (postale ou d'autre type) d'un éditeur, d'une organisation, d'un individu, etc. ;

<*idno*> fournit tout chiffre standard ou non standard employé pour identifier une unité bibliographique ; parmi les attributs, citons :

- *type* catégorise le nombre, par exemple en tant qu'ISBN ou tout autre numéro d'identification normalisé ;

<availability>
: fournit des informations concernant la disponibilité d'un texte, par exemple pré sence de restrictions sur son emploi ou sa diffusion ; son statut au plan du copyright, etc. ; parmi les attributs possibles, citons :

  status fournit un code identifiant la disponibilité actuelle du texte ; exemples : `restricted`, `unknown` et `free`.

<date> contient une date dans n'importe quel format.

Exemple (Céline, *Romans, iv*, La Pléiade) :

```
<publicationStmt>
     <publisher>Éditions Gallimard</publisher>
     <pubPlace>Paris</pubPlace> <date>1993</date>
     <idno type=ISBN> 2-07-011336-1</idno>
     <idno type=numero edition>64107</idno>
     <idno type=numero impression>I3-1903</idno>
     <idno type=depot legal>octobre 1993</idno>
     <availability>Copyright:  Éditions Gallimard,
      Féerie pour une autre fois I, 1952 ;
      Féerie pour une autre fois II, 1954 ;
      Entretiens avec le professeur Y, 1955 ;
      Appendices, text, préface et apparat critique, 1993.
      </availability>
     </publicationStmt>
```

20.1.5. Mentions de série et de notes

L'élément *<seriesStmt>* regroupe les informations concernant la collection ou série, si elle existe, à laquelle une publication appartient. Il peut contenir *<title>*, *<idno>*, ou des éléments *<respStmt>*.

L'élément *<notesStmt>*, s'il est employé, contient un ou plusieurs éléments *<note>* qui contiennent une note ou annotation. Certaines des informations que l'on trouve dans la zone des notes dans les descriptions bibliographiques conventionnelles sont représentées par des éléments spécifiques dans la TEI.

20.1.6. Description de la source

L'élément *<sourceDesc>* est un élément obligatoire qui enregistre des détails concernant la ou les sources à partir desquelles le fichier électronique a été dérivé. Il peut contenir du texte libre ou une citation bibliographique, au moyen d'un ou plusieurs des éléments suivants :

<bibl> contient une citation bibliographique structurée de façon très lâche, dont les composants peuvent être balisés explicitement ou non ;

<biblFull>
contient une citation bibliographique complètement structurée, dans laquelle tous les composants de la description d'un fichier selon la TEI sont présents ;

<listBibl>
contient une liste de citations bibliographiques de toute nature.

Exemples :

```
<sourceDesc>
     <bibl>The first folio of Shakespeare, prepared by
          Charlton Hinman (The Norton Facsimile, 1968)</bibl>
</sourceDesc>
```

```
<sourceDesc>
     <scriptStmt id=CNN12>
     <bibl><author>CNN Network News
          <title>News headlines
          <date>12 Jun 1989
     </bibl>
     </scriptStmt>
 </sourceDesc}
```

20.2. Description relative au codage

L'élément *<encodingDesc>* précise les méthodes et les principes éditoriaux qui on régi la transcription du texte. Son emploi est fortement recommandé. Il peut être une description en texte libre, ou se composer d'éléments pris dans la liste suivante :

<projectDesc>
contient une description détaillée de la raison pour laquelle un fichier électronique a été créé ainsi que toute autre information pertinente concernant le procédé par lequel le fichier a été constitué ;

<samplingDecl>
regroupe une description en texte libre des critères et des méthodes employés pour sélectionner des textes lors de la création d'un corpus ou d'un recueil ;

<editorialDecl>
: fournit des détails des pratiques et des principes éditoriaux appliqués pendant le codage d'un texte ;

<tagsDecl>
: fournit des informations détaillées concernant le balisage appliqué à un document SGML ;

<refsDecl>
: précise la manière dont les références canoniques sont construites pour ce texte ;

<classDecl>
: regroupe une ou plusieurs taxonomies définissant l'ensemble des codes de classification employés ailleurs dans le texte.

20.2.1. Description du projet et des critères de sélection

Exemples d'utilisation des éléments *<projecfDesc>* et *<samplingDesc>* :

```
<encodingDesc>
     <projectDesc>Texte saisi d'après le fond du Musée d'Art Brut
                  de Lausanne, Suisse, juillet 1996.
     </projectDesc>
</encodingDesc>

<encodingDesc>
     <samplingDecl>Samples of 2000 words taken from the
          beginning of the text
     </samplingDecl>
 </encodingDesc}
```

20.2.2. Mentions éditoriales

L'élément *<editorialDecl>* regroupe une description en prose des pratiques employées lors du codage du texte. Typiquement, cette description devrait couvrir des sujets tels que ceux qui sont indiqués ci-après. Pour chacun de ces thèmes, il conviendrait de créer un paragraphe séparé :

correction
: comment et dans quelles circonstances les corrections ont été apportées au texte ;

normalisation
: le degré de régularisation ou de normalisation apporté à la source originale ;

citation
: le traitement réservé aux marques de citation dans l'original – ont-ils été retenus ou remplacés par des références d'entité ? les guillemets ouvrants et fermants sont-ils différenciés, etc. ;

division des mots
: la façon dont les tirets (surtout ceux en fin de ligne) ont été traités dans l'original : ont-ils été retenus ou remplacés par des références d'entité ?

segmentation
: la façon dont le texte a été segmenté, par exemple en phrases, unités de ton, strates graphèmiques, etc. ;

interprétation
: les informations interprétatives ou analytiques éventuellement ajoutées au texte.

Exemple :

```
<editorialDecl>
      <p> L'analyse du récit oral de la section 4 a été
          faite à la main et n'a pas été vérifiée ;
      <p> La transcription a été vérifiée à l'aide du
          « speller » d'Unix, mais avec le dictionnaire de
          l'Université de Paris-7 ;
      <p> Tous les mots américains sont écrits en suivant
          l'orthographe du Webster's 9th Collegiate dictionary.
      <p> Ici, nous utilisons le codage Iso-8859 (Latin1).
</editorialDecl>
```

20.2.3. Déclarations relatives au balisage, au système de référence et à la classification utilisés

L'élément <*tagsDecl*> est employé pour fournir des informations détaillées concernant les balises SGML paraissant effectivement à l'intérieur d'un texte. Il peut contenir une simple liste d'éléments à employer, avec un compteur pour chacun, au moyen des éléments suivants :

<tagUsage>
: fournit des informations concernant l'usage d'un élément spécifique au sein de l'élément <*text*> le plus à l'extérieur d'un document conforme à la TEI ; parmi les attributs citons :

  gi
  : le nom (identifiant générique) de l'élément indiqué par la balise ;

occurs précise le nombre d'occurrences de cet élément au sein du texte.

L'élément <*rendition*> est employé pour documenter les manières différentes dont les éléments sont rendus dans le texte source.

<*rendition*>
: fournit une information concernant l'interprétation prévue d'un ou plusieurs éléments ;

<*tagUsage*>
: fournit une information concernant l'usage d'un élément spécifique à l'intérieur d'un élément <*text*> ; parmi les attributs possibles, citons :

  occurs précise le nombre d'occurrences de cet élément au sein du texte ;

  ident précise le nombre d'occurrences de cet élément au sein du texte qui portent une valeur distincte pour l'attribut global *id* ;

  render précise l'identifiant d'un élément <*rendition*> qui définit comment cet élément doit être rendu.

Par exemple :

```
<tagsDecl>
 <tagUsage gi=text occurs=1>
 <tagUsage gi=body occurs=1>
 <tagUsage gi=p occurs=12>
 <tagUsage gi=hi occurs=6>
 </tagsDecl>
```

Cette déclaration (imaginaire) de balisage conviendrait pour un texte contenant douze paragraphes dans son corps, parmi lesquels six éléments <*hi*> ont été marqués. À noter que si l'élément <*tagsDecl*> est employé, il doit contenir un élément <*tagUsage*> pour chaque élément balisé dans l'élément de texte associé.

L'élément <*refsDecl*> est employé pour documenter le fonctionnement d'un éventuel système de référence canonique incorporé dans le codage. Dans sa forme la plus simple, il consiste en une description en texte libre.

Exemple :

```
<refsDecl>
    <p>L'attribut N de chaque DV1 ou DV2 contient
       la référence canonique de chaque sous-division sous
       la forme XX.yyy où XX est le le numéro du tome, en
       nombres romaines, et yyy le numéro de section, en
       chiffres arabes.
</refsDecl>
```

L'élément *<classDecl>* regroupe les définitions pour, ou les sources de, tous les plans de classification descriptifs employés par d'autres parties de l'en-tête. Au moins un de ces plans doit être fourni, codé au moyen des éléments suivants :

<taxonomy>
: définit une typologie employée pour classer des textes soit implicitement, au moyen d'une citation bibliographique, soit explicitement au moyen d'une taxonomie structurée ;

<bibl>
: contient une citation bibliographique à structure informelle, dont les sous-composants peuvent être balisés explicitement ou pas ;

<category>
: regroupe une catégorie descriptive individuelle, éventuellement contenue dans une catégorie de niveau supérieur, à l'intérieur d'une taxonomie définie par l'utilisateur ;

<catDesc>
: décrit une certaine catégorie à l'intérieur d'une taxonomie ou d'une typologie de texte, sous forme d'une description brève en prose.

Dans le cas le plus simple, la taxonomie peut être définie par un référence bibliographique, comme l'exemple suivant :

```
<classDecl>
    <taxonomy id='LCSH'>
         <bibl>Library of Congress Subject Headings
         </bibl>
    </taxonomy>
 </classDecl>
```

Alternativement, ou en outre, le codeur peut définir un plan de classification spécifique, comme dans l'exemple ci-après :

```
<taxonomy id=B>
  <bibl>Brown Corpus</bibl>
  <category id=B.A><catDesc>Press Reportage
     <category id=B.A1><catDesc>Daily</category>
     <category id=B.A2><catDesc>Sunday</category>
     <category id=B.A3><catDesc>National</category>
     <category id=B.A4><catDesc>Provincial</category>
     <category id=B.A5><catDesc>Political</category>
     <category id=B.A6><catDesc>Sports</category>
   ...
  </category>
  <category id=B.D><catDesc>Religion
     <category id=B.D1><catDesc>Books</category>
     <category id=B.D2><catDesc>Periodicals and
                            tracts</category>
  </category>
...
</taxonomy>
```

Le lien entre un texte particulier et une catégorie à l'intérieur d'une telle taxonomie est réalisé au moyen de l'élément <*catRef*> incorporé dans l'élément <*textClass*>, comme décrit plus en détail ci-dessous.

20.3. Description de profil

L'élément <*profileDesc*> permet d'enregistrer dans un cadre unique des informations caractérisant divers aspects descriptifs d'un texte. Il a trois composants facultatifs :

<*creation*>
: contient des informations concernant la création d'un texte ;

<*langUsage*>
: décrit les langues, les sous-langues, les registres, les dialectes etc. représentés à l'intérieur d'un texte ;

<*textClass*>
: contient des informations décrivant la nature ou le sujet d'un texte en termes d'un système standard de classification, d'un thesaurus, etc.

Exemples :

```
<creation>
    <date value='1992-08'>August 1992</date>
    <name type=place>Taos, New Mexico</name>
 </creation>
```

L'élément <*textClass*> classe un texte par référence au(x) système(s) défini(s) par l'élément <*classDecl*>. Il contient un ou plusieurs des éléments suivants :

<*keywords*>
: contient une liste de mots clés ou de phrases clés identifiant le sujet ou la nature d'un texte ; parmi les attributs possibles, citons :

  scheme identifie le vocabulaire contrôlé au sein duquel l'ensemble de mots clés concerné est défini ;

<*classCode*>
: regroupe le code de classification employé pour ce texte dans un système standard de classification ; parmi les attributs possibles, citons :

  scheme identifie le système de classification ou la taxonomie employés ;

<*catRef*>
: précise une ou plusieurs catégories définies à l'intérieur d'une certaine taxonomie ou typologie de texte ; parmi les attributs possibles, citons :

  target identifie les catégories concernées.

L'élément <*keywords*> regroupe une liste de mots clés ou d'expressions clés servant à identifier le sujet ou la nature d'un texte. L'attribut *scheme* relie ceux-ci au système de classification défini dans l'élément <*taxonomy*>.

```
<textClass>
     <keywords scheme=LCSH>
          <list>
          <item>Littérature française -- Histoire et critique --
               Bases de données.</item>
          <item>Littérature française -- Histoire et critique --
                Théorie, etc.</item>
          <item>Langue française -- Style -- Bases de
                données.</item>
          </list>
     </keywords>
 </textClass>
```

20.4. Description de révision

L'élément *<revisionDesc>* fournit un journal de modifications dans lequel chaque modification apportée à un texte peut être enregistrée. Le journal peut être enregistré sous la forme d'une séquence d'éléments *<change>* dont chacun contient :

<date> contient une date dans n'importe quel format ;

<respStmt> fournit une mention de responsabilité indiquant une personne responsable du contenu intellectuel d'un texte, d'une édition, d'un enregistrement ou d'une série, lorsque les éléments spécialisés pour définir les auteurs, les rédacteurs etc. ne suffisent pas ou ne s'appliquent pas ;

<item> contient un composant de liste.

Exemple :

```
<revisionDesc>
     <change><date>6/3/91:</date>
          <respStmt><name>JT</name><resp>ed.</resp>
          </respStmt>
          <item>Modifié le format de fichier</item>
     <change><date>5/25/90:</date>
          <respSmt><name>JT</name><resp>ed.</resp>
          <item>Traité les modids de Jean Sort</item>
 </revisionDesc>
```