

header, core, textstructure

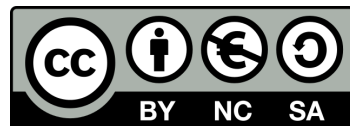
Guidelines, chap. 2-4

Structuration des données et des documents : balisage XML

La TEI : métadonnées du document et éléments de structuration

J.B. Camps – vendredi 4 décembre 2015

M2 Technologies numériques appliquées à l'histoire



Ce support de cours est mis à disposition selon les termes de la licence *Creative Commons* Paternité – Pas d'utilisation commerciale – Partage à l'Identique 4.0

Plan de la séance

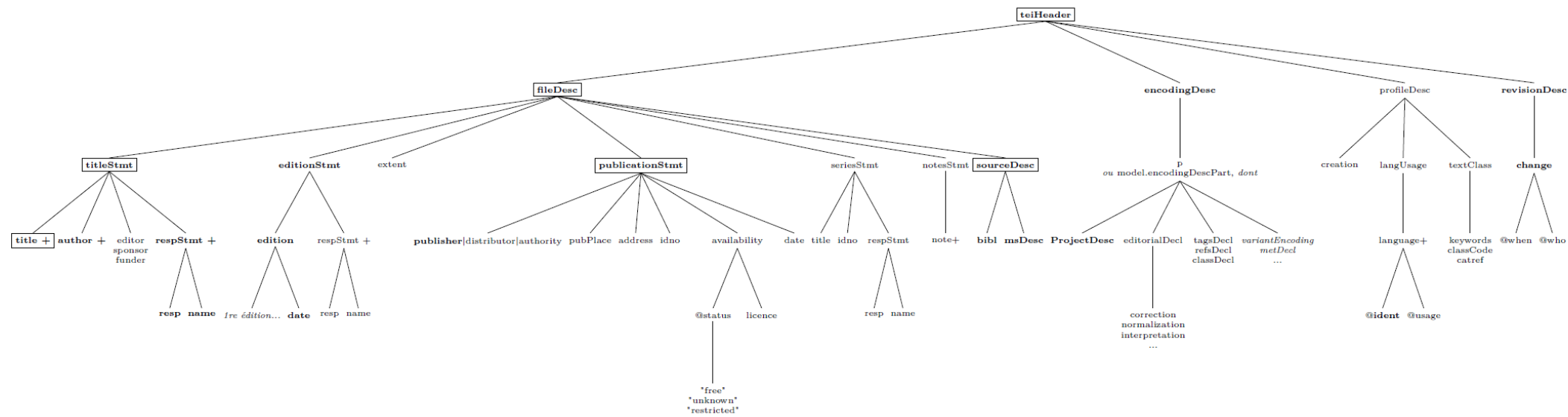
I. Les métadonnées du document : l'entête teiHeader

II. Divisions et structuration du document

teiHeader : modules concernés et statut

- module **header** (module obligatoire)
 - éléments ajoutés au <teiHeader> par d'**autres modules (optionnels)**, pour des emplois spécialisés, par ex. :
 - <variantEncoding> (module textcrit)
 - <msDesc> (module msdescription)
 - etc.
- N.B. :
- *N'a pas vocation à intégrer tous les éléments possibles*
 - *À calibrer/normaliser en fonction du projet et des besoins*

le teiHeader : arborescence générale



N.B. sur le nommage des éléments et leur niveau de structuration

xStmt : appelle un contenu structuré en sous-éléments (type « base de données »)

xDecl : une série de déclarations, associant structuration en sous-éléments et description en prose

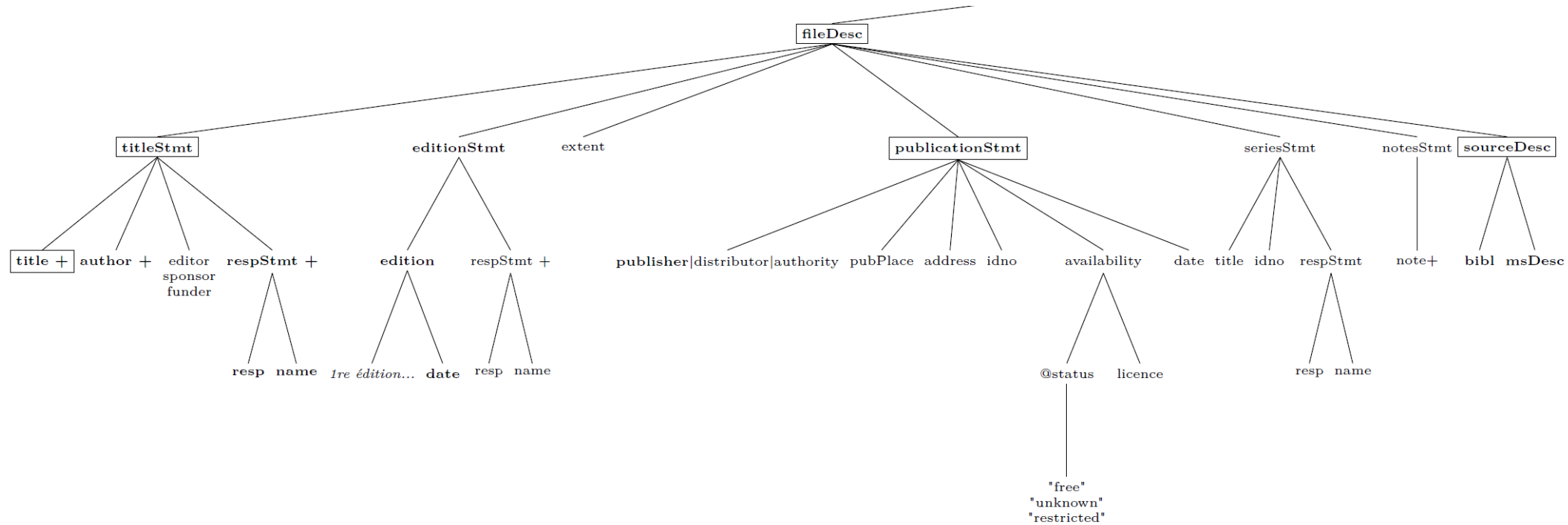
xDesc : description, en prose courante, pouvant être structurée en sous-éléments

Classes modèles du `teiHeader`

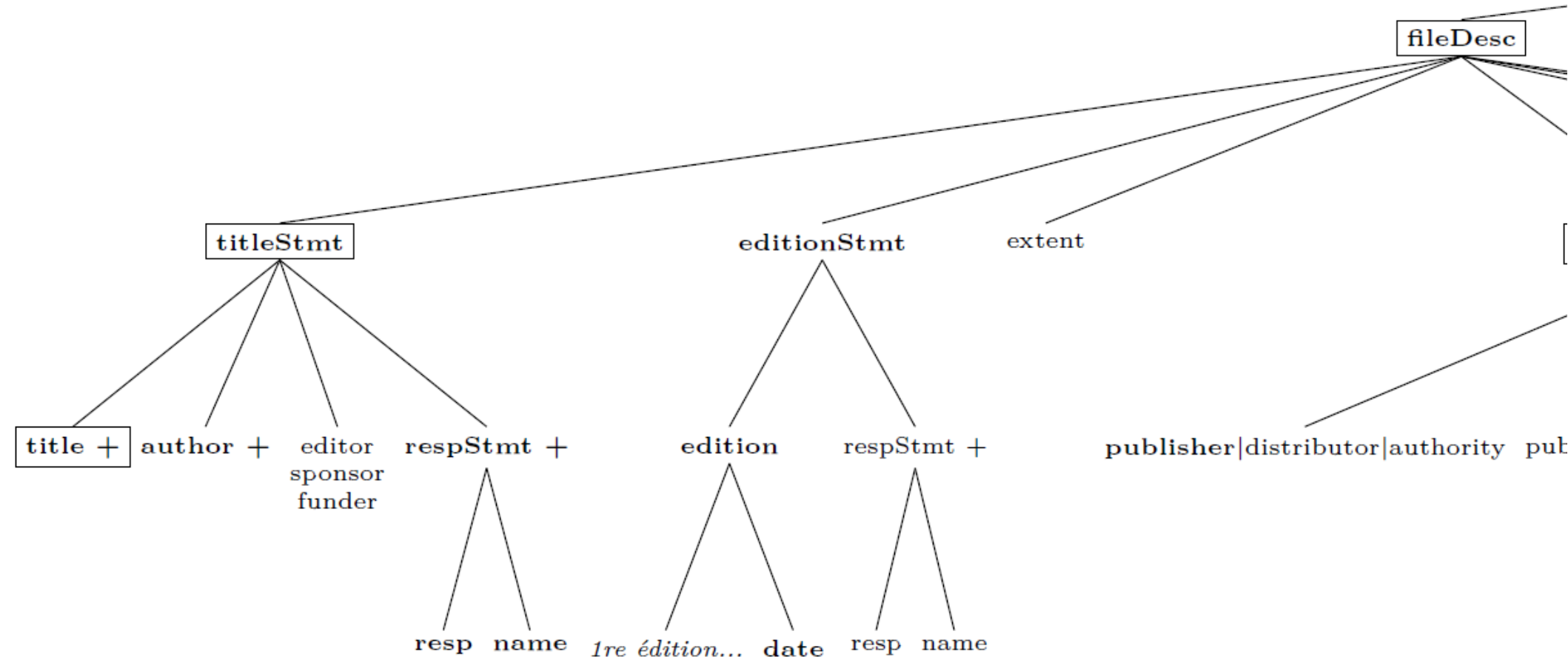
Voir 2.1.3 « Model classes in the TEI Header »

- **model.applicationLike** groups elements used to record application-specific information about a document in its header.
- **model.availabilityPart** groups elements such as licences and paragraphs of text which may appear as part of an availability statement
- **model.catDescPart** groups component elements of the TEI header Category Description.
- **model.editorialDeclPart** groups elements which may be used inside `editorialDecl` and appear multiple times.
- **model.encodingDescPart** groups elements which may be used inside `encodingDesc` and appear multiple times.
- **model.profileDescPart** groups elements which may be used inside `profileDesc` and appear multiple times.
- **model.teiHeaderPart** groups high level elements which may appear more than once in a TEI header.
- **model.sourceDescPart** groups elements which may be used inside `sourceDesc` and appear multiple times.
- **model.textDescPart** groups elements used to categorize a text for example in terms of its situational parameters.

le teiHeader : arborescence fileDesc



le teiHeader : arborescence fileDesc (1-3)




```
<teiHeader>
```

```
<fileDesc>
```

```
<titleStmt>
```

```
<title>Le recueil de prières de Philippe de Mézières</title>
```

```
<author>Giraud, Cédric (ATER à l'École nationale des Chartes)</author>
```

```
<respStmt>
```

```
<resp>document encodé en XML/TEI P5 par</resp>
```

```
<name>Monnin, Alexandre (doctorant à Paris I, vacataire de  
recherche à l'École nationale des Chartes)</name>
```

```
</respStmt> <respStmt>
```

```
<resp>sous la direction de</resp>
```

```
<name>Clavaud, Florence (directrice des nouvelles technologies et  
de l'informatique à l'École nationale des Chartes)</name>
```

```
</respStmt>
```

```
</titleStmt>
```

```
<publicationStmt>
```

```
<idno type="numero">96</idno>
```

```
<publisher>École nationale des Chartes</publisher>
```

```
<date when-iso="2008-11">novembre 2008</date>
```

```
</publicationStmt>
```

```
<sourceDesc>
```

```
<p>Paris, Bibliothèque Mazarine, 516, 2r<hi rend="exposant">o</hi></p>
```

```
</sourceDesc>
```

```
</fileDesc>
```

```
<profileDesc>
```

```
<langUsage>
```

```
<language ident="lat"/>
```

```
</langUsage>
```

```
</profileDesc>
```

```
</teiHeader>
```

[source : dossier Thélème n° 96, en ligne à : <http://theleme.enc.sorbonne.fr/dossiers/notice96.php>]
J.B. Camps – La TEI : métadonnées et éléments de structuration



École
nationale
des
chartes

Artamene... fileDesc/titleStmt et fileDesc/editionStmt

```
<titleStmt>  
  <title>Artamène ou le Grand Cyrus</title>  
  <author>Scudéry, Madeleine (1608-1701)</author>  
  <respStmt>  
    <resp>numérisation par </resp>  
    <name>Claude Bourqui et Alexandre Gefen</name>  
  </respStmt>  
</titleStmt>  
<editionStmt>  
  <edition>  
    <date>1656</date>  
  </edition>  
</editionStmt>
```

Queste... fileDesc/titleStmt

```
<titleStmt>
  <title type="normal">Queste del saint Graal</title>
  <title type="source">Manuscrit : Lyon, Bibliothèque municipale,
Palais des arts 77</title>
  <title type="reference">qgraal_cm</title>
  <title type="gmd">Édition électronique</title>
  <author>anonyme</author>
  <editor role="editor">Christiane Marchello-Nizia</editor>
  <editor role="associate_editor">Alexei Lavrentiev</editor>
  <respStmt>
    <resp>Etablissement du texte</resp>
    <name type="person" xml:id="cmn">Christiane Marchello-
Nizia</name>
  </respStmt>
  <respStmt>
    <resp>Relecture du manuscrit, encodage XML-TEI, mise en
forme</resp>
    <name type="person" xml:id="al">Alexei Lavrentiev</name>
  </respStmt>
  <respStmt>
    <resp>Étiquetage morphosyntaxique</resp>
    <name>Christiane Marchello-Nizia</name>
    <name type="person" xml:id="sp">Sophie Prévost</name>
  </respStmt>
</titleStmt>
```

BVH ... fileDesc/titleStmt

```
<titleStmt>
  <title>
    <hi rend="it">Champ fleury</hi>, Paris, 1529</title>
  <respStmt>
    <name xml:id="MLD">Marie-Luce Demonet</name>
    <resp>
      <date>2008</date>Responsable BVH CESR, relecture</resp>
  </respStmt>
  <respStmt>
    <persName xml:id="DL" resp="transcription">Delphine Lancellata et Angela
Axworthy</persName>
    <resp>
      <date>1998</date>Transcription</resp>
  </respStmt>
  <respStmt>
    <name xml:id="AA">Angela Axworthy</name>
    <resp>
      <date>1999</date>Révision</resp>
  </respStmt>
  <respStmt>
    <name xml:id="LB">Laetitia Bontemps</name>
    <resp>
      <date>2008</date>Révision, contrôle de l'encodage</resp>
  </respStmt>
  <respStmt>
    <name xml:id="VLR">Vanessa Le Rolle</name>
    <resp>
      <date>2008</date>encodage</resp>
  </respStmt>
  <respStmt>
```

BVH ... fileDesc/editionStmt et fileDesc/extent

```
<editionStmt>
  <edition n="1" xml:base="http://www.bvh.univ-
tours.fr:8080/xtf/view?
docId=tei/B410186201_I65/B410186201_I65_tei.xml;query=;brand=default">
  <date type="mise_en_ligne" when="2010-07-20"
n="1">20/07/2010</date>
  <date type="mise_a_jour" when="2013-07-18"
n="1">18/07/2013</date>
  </edition>
  <respStmt>
    <name xml:id="TU">Toshinori Uetani</name>
    <resp>Révision</resp>
  </respStmt>
</editionStmt>
<extent>801 Ko</extent>
```

Queste... fileDesc/editionStmt et fileDesc/extent

```
<editionStmt>
```

```
  <p>Les principes de la présente édition sont décrits dans  
  l'<ref
```

```
target="http://textometrie.risc.cnrs.fr/txm/images/qgraal_intro_2011-  
06.pdf">Introduction</ref>.</p>
```

```
  <p>Les attributs 'ed' de certaines balises font référence aux  
  éditions ou transcriptions suivantes :
```

```
    <list>
```

```
      <item><bibl xml:id="Pauphilet1923">La Queste del Saint  
      Graal, roman du XIIIe siècle, édité par Albert  
      Pauphilet, 1923. Paris, Champion.</bibl> (numéros de  
      page)</item>
```

```
      <item><bibl xml:id="norm">Transcription normalisée</bibl>  
      (sauts de ligne insérés après les mots coupés)</item>
```

```
      <item><bibl xml:id="fac">Transcription fac-  
      similaire</bibl> (sauts de ligne coupant des mots)</item>
```

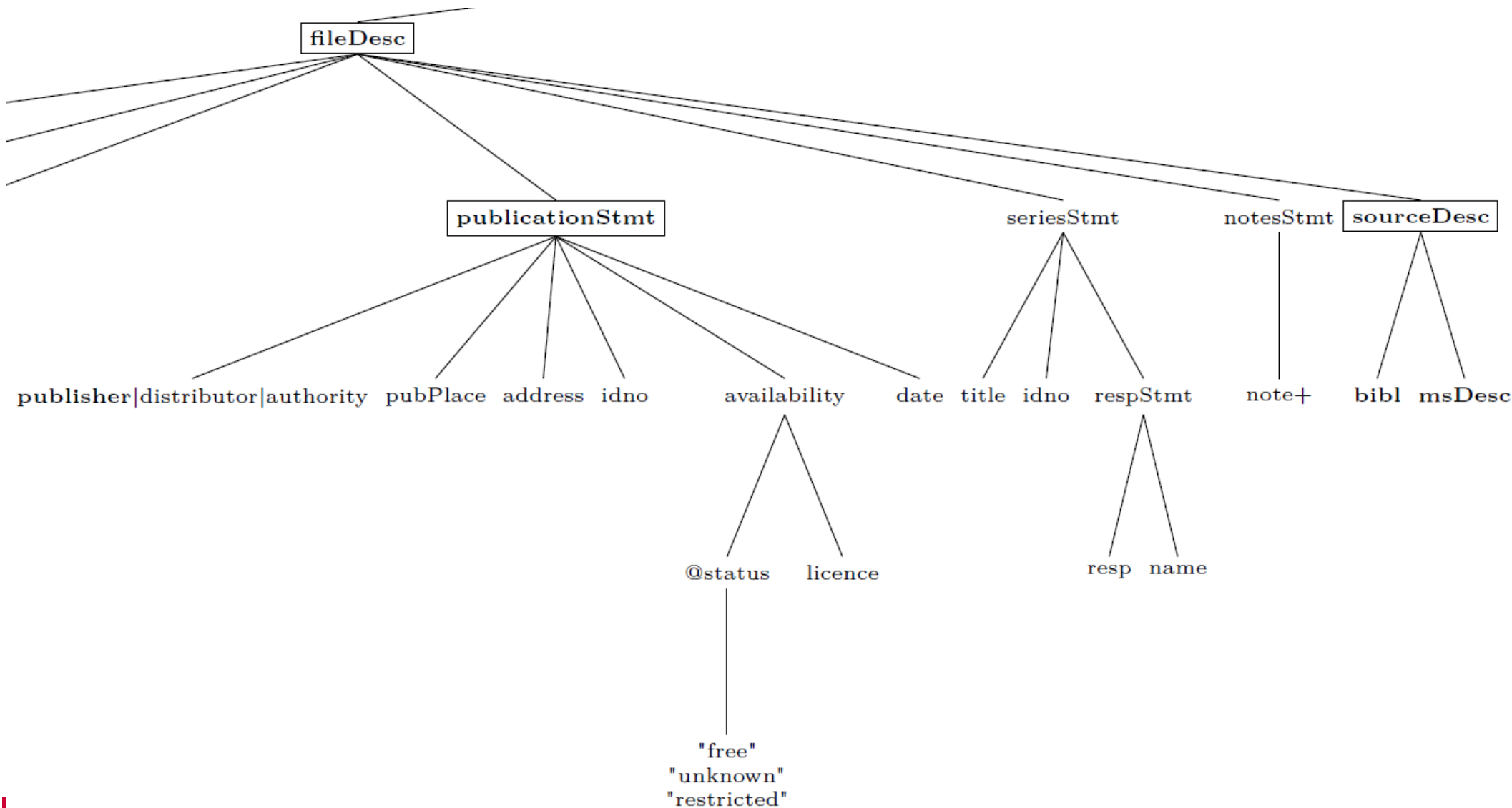
```
    </list></p>
```

```
</editionStmt>
```

```
<extent>- Taille approximative du fichier encodé en TEI non  
compressé : xxx octets - Nombre
```

```
d'occurrences-mots : xxx - Numéros de pages saisies : de 160-recto à  
224-verso </extent>
```

le teiHeader : arborescence fileDesc (4-7)



ENC, dossier Thélème n° 96

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <!--[...] -->
    </titleStmt>
    <publicationStmt>
      <idno type="numero">96</idno>
      <publisher>École nationale des Chartes</publisher>
      <date when-iso="2008-11">novembre 2008</date>
    </publicationStmt>
    <sourceDesc>
      <p>Paris, Bibliothèque Mazarine, 516, 2r<hi rend="exposant">o</hi></p>
    </sourceDesc>
  </fileDesc>
  <profileDesc>
    <langUsage>
      <language ident="lat"/>
    </langUsage>
  </profileDesc>
</teiHeader>
[source : dossier Thélème n° 96, en ligne à : <
http://theleme.enc.sorbonne.fr/dossiers/notice96.php>]
```


ENC... Ducange

```
<publicationStmt>
  <date>2009</date>
  <idno>http://ducange.enc.sorbonne.fr/</idno>
  <publisher>École nationale des chartes</publisher>
  <address>
```

```
  <addrLine>19, rue de la Sorbonne</addrLine>
  <addrLine>75005 Paris.</addrLine>
  <addrLine>tél. : +33 (0)1 55 42 75 00</addrLine>
  <addrLine>http://enc.sorbonne.fr/</addrLine>
  <addrLine>recherche@enc.sorbonne.fr</addrLine>
```

```
</address>
```

```
<availability status="restricted">
```

<p> L'École nationale des chartes met à disposition cette ressource électronique structurée, protégée par le code de la propriété intellectuelle sur les bases de données (L341-1), selon les termes de la licence Creative Commons : « Paternité - Pas d'Utilisation Commerciale - Pas de Modification ; 2.0 France ». Cette licence est disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/2.0/fr/> (...) </p>

```
</availability>
```

```
</publicationStmt>
```

```
<seriesStmt>
```

```
  <title>Éditions en ligne de l'École des chartes</title>
```

```
  <idno type="URI">http://elec.enc.sorbonne.fr/</idno>
```

```
  <idno type="vol">18</idno>
```

```
</seriesStmt>
```

```
<sourceDesc>
```

```
  <bibl><author>Charles du Fresne, sieur du Cange</author>, <foreign>et al.</foreign>, <title>Glossarium mediae et infimae latinitatis</title>, éd. augm., Niort : École nationale des chartes, 1883-1887, 10 vol. in-quarto.</bibl>
```

```
</sourceDesc>
```

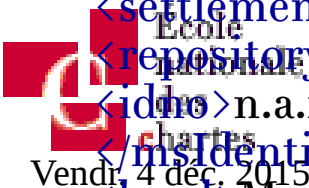


Queste... fileDesc/publicationStmt

```
<publicationStmt>
  <distributor>
    <name>Projet BFM, UMR 5191 ICAR, CNRS/ENS LSH</name>
    <address>
      <addrLine>15 parvis René-Descartes</addrLine>
      <addrLine>B.P. 7000 </addrLine>
      <addrLine>69342 Lyon Cedex 07</addrLine>
      <addrLine>France</addrLine>
      <addrLine>Tél : 04 37 37 63 10</addrLine>
      <addrLine>Fax : 04 37 37 62 65</addrLine>
      <addrLine>Mèl : bfm@ens-lsh.fr</addrLine>
      <addrLine>http://bfm.ens-lsh.fr/</addrLine>
    </address>
  </distributor>
  <availability status="restricted">
    <p>(c) 2008, Equipe BFM, CNRS/ENS-LSH. <hi>Conditions
d'utilisation</hi> :
    Cette édition est mise à disposition sous un <ref
target="http://creativecommons.org/licenses/by-sa/2.0/fr/">contrat Creative
Commons</ref>.
    Il s'agit d'une version non définitive.
  </p>
</availability>
</publicationStmt>
```

Queste... fileDesc/sourceDesc

```
<sourceDesc>
<msDesc>
<msIdentifier xml:id="ms__K">
<country>France</country>
<settlement>Lyon</settlement>
<repository>Bibliothèque municipale</repository>
<idno>Palais des arts 77</idno>
</msIdentifier>
<head>Manuscrit de base</head>
<msPart>
<altIdentifier>
<idno>lyonbm__pa77</idno>
</altIdentifier>
<history>
<origin> Date du manuscrit : <date when="1250-01-01" notBefore="1200-01-01"
notAfter="1299-01-01">13e</date>
</origin>
</history>
</msPart>
</msDesc>
<msDesc>
<msIdentifier xml:id="ms__Z">
<country>France</country>
<settlement>Paris</settlement>
<repository>BnF</repository>
<idno>n.a.fr. 1119</idno>
</msIdentifier>
<head>Manuscrit complémentaire</head>
<msPart>
```



Vendredi, 4 décembre 2015

```

<sourceDesc>
  <msDesc>
    <msIdentifier xml:id="I65">
      <country>France</country>
      <settlement>Blois</settlement>
      <repository key="B410186201">Bibliothèque municipale</repository>
      <idno type="cote">I65</idno>
    </msIdentifier>
    <msContents>
      <p>
        <biblStruct>
          <monogr>
            <author role="auteur">
              <persName key="pers1">
                <surname>Tory</surname>
                <forename>Geoffroy</forename>
              </persName>
            </author>
            <title type="titre_long">Champ fleury. Au quel est contenu l'Art & Science de la deue & vraye Proportion
des Lettres Attiques, qu'on dit autrement Lettres Antiques, & vulgairement Lettres Romaines proportionnees selon le Corps & Visage
humain.</title>
            <title type="titre_court">Champ fleury</title>
            <imprint>
              <pubPlace>Paris</pubPlace>
              <publisher role="libraire">
                <persName key="pers12">
                  <surname>Tory</surname>
                  <forename>Geoffroy</forename>
                </persName>
              </publisher>
              <publisher role="libraire">
                <persName key="pers12">
                  <surname>Gourmont</surname>
                  <forename>Gilles de</forename>
                </persName>
              </publisher>
              <date>1529</date>
            </imprint>
          </monogr>
        </biblStruct>
      </p>
    </msContents>
    <physDesc>
      <objectDesc>
        <supportDesc>
          <collation>
            <p>[8], I-LXXX f. (sig. A<sup>8</sup> B-N<sup>6</sup> O<sup>8</sup>)</p>
            <dimensions>
              <dim type="format">2°</dim>
              <dim unit="ligne" rend="large"/>
            </dimensions>
          </collation>
        </supportDesc>
      </objectDesc>
      <additions>
        <p>Le f. 3v de l'exemplaire de Blois présente une correction manuscrite, probablement de l'époque, dans la dernière note
de la dernière page est corrigé en "Mesiere".</p>
        <p>L'exemplaire du British Museum, cote 60.e.14, présente une correction identique.</p>
        <p>Voir le fac-similé publié par J.W. Jolliffe, S.R. Publishers Ltd, Johnson Reprint Corporation, Mouton éditeur, 1970.</p>
      </additions>
    </physDesc>
  </msDesc>
</sourceDesc>

```

BVH ... fileDesc/sourceDesc

(à voir directement dans le fichier, cf. bibliogr.)



Vendredi 4 décembre 2015

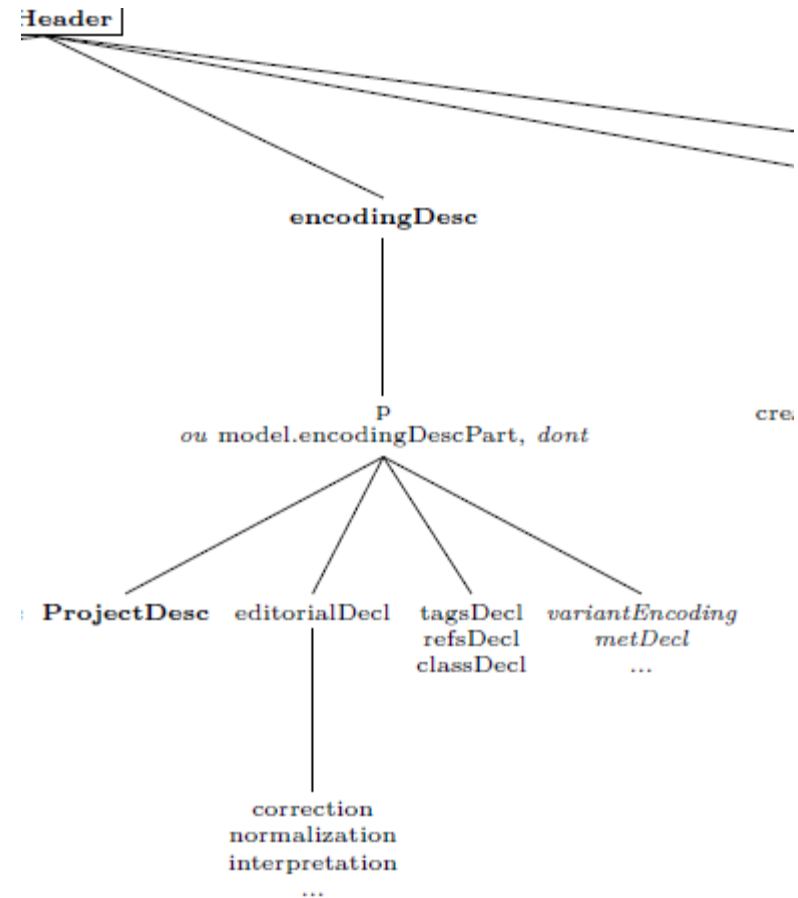
J.B. Camps – La TEI : métadonnées et

éléments de structuration

fileDesc/sourceDesc, cas particulier : fichier TEI dérivé d'un autre fichier TEI

```
<sourceDesc>
  <biblFull>
    <!-- Copie complète du fileDesc du
fichier d'origine -->
    <titleStmt>
      <title/>
      <!-- [...] -->
    </titleStmt>
    <!-- [...] -->
  </biblFull>
</sourceDesc>
```

le teiHeader : arborescence encodingDesc



editorialDecl et les membres de la classe modèle model.editorialDeclPart

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.editorialDeclPart.html>

Une série d'éléments spécialisés

correction hyphenation interpretation normalization
punctuation quotation segmentation stdVals

Des attributs spécialisés avec des listes fermées de valeurs prédéfinies

correction (@status ['high', 'medium', 'low', 'unknown'],
@method ['silent', 'markup']), **hyphenation** (@eol ['all', 'some',
'hard', 'none'])

etc.

encodingDesc/editorialDecl/correction
ex. de jeu de valeur d'attributs

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-correction.html>

ENC... Ducange

```
<encodingDesc>
  <projectDesc>
    <p>[Description scientifique du projet]</p>
  </projectDesc>
  <samplingDecl>
    <p>Lettre </p>
  </samplingDecl>
  <editorialDecl>
    <p>L'attitude générale à l'égard du texte imprimé a été d'établir le texte le
plus utile au lecteur, ainsi la fidélité n'a pas été jusqu'à conserver la mémoire des
coquilles qui ont été corrigées. Cependant, toute information retrouvée a été balisée,
ainsi l'histoire éditoriale de l'ouvrage inscrite par des puces ou des crochets a été
formalisée. Les caractères qui peuvent être générés automatiquement n'ont pas été
conservés.</p>
  </editorialDecl>
</encodingDesc>
<revisionDesc>
  <change when="2009-04-01" who="FG">TEI Council, présentation du schéma</change>
  <change when="2008-11" who="FG">Publication 0-Z</change>
  <change when="2008-01" who="FG">Publication O, P, Q</change>
  <change when="2007-06" who="FG">Cahier des charges Jouve</change>
  <change when="2007-03" who="FG">Echantillon numérisé par Dataland (200
p.)</change>
</revisionDesc>
```

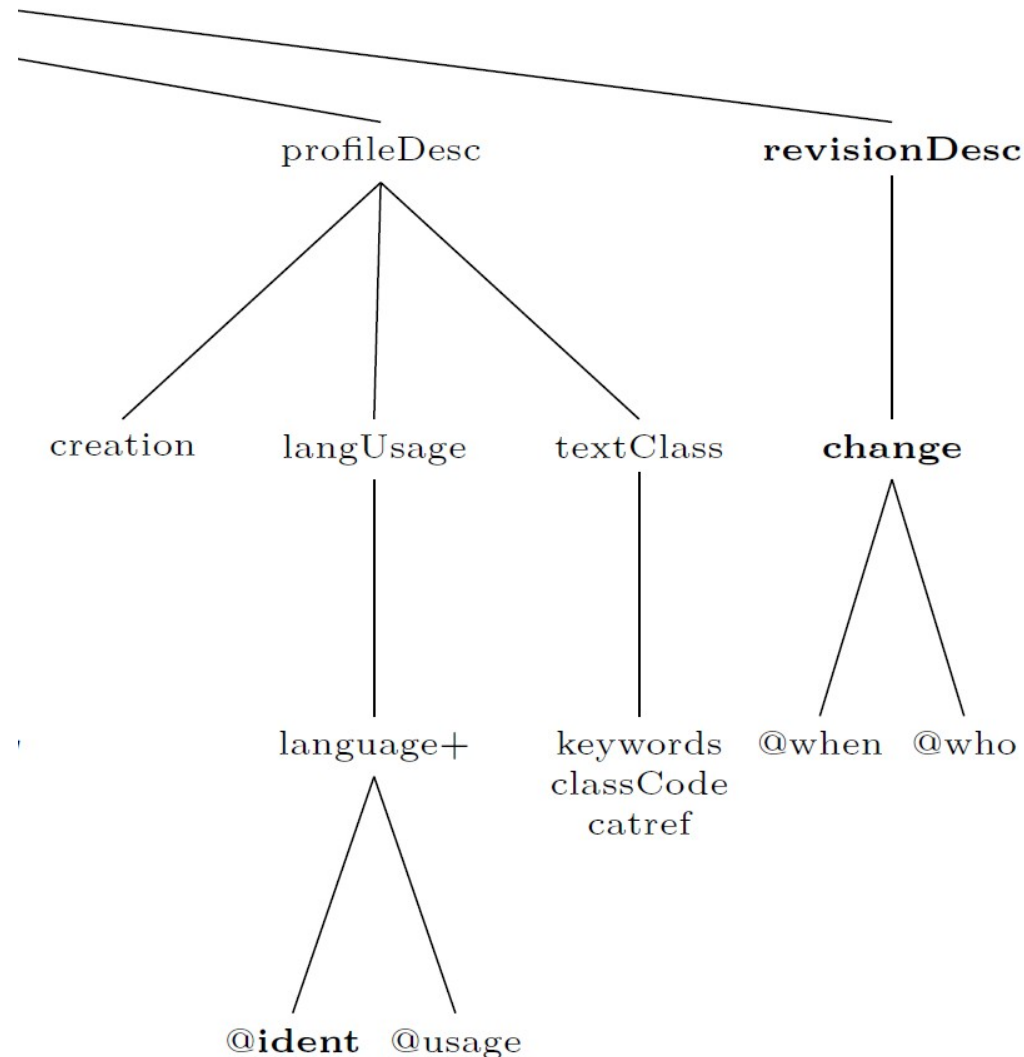
Queste...

```
<encodingDesc>
  <projectDesc>
    <p>Projet : BFM - Graal Équipe : Diachronie et bases textuelles d'ancien et de
moyen français Laboratoire : UMR5191 ICAR Institution : CNRS / ENS LSH, Lyon
    <address>
      [...]
    </address>
  </p>
</projectDesc>
<editorialDecl>
  <p>See BFM-MSS Encoding Guidelines at
http://bfm.ens-lsh.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf. See also BFM XML-TEI
Encoding
Guidelines at http://bfm.ens-lsh.fr/IMG/pdf/Manuel_Encodage_TEI.pdf and BFM XML-TEI
Encoding Instructions at http://bfm.ens-lsh.fr/IMG/pdf/Consignes_BFM.pdf</p>
</editorialDecl>
<classDecl>
  <taxonomy xml:id="forme">
    <category xml:id="forme_vers">
      <catDesc>vers</catDesc>
    </category>
    <category xml:id="forme_prose">
      <catDesc>prose</catDesc>
    </category>
    <category xml:id="forme_mixte">
      <catDesc>mixte</catDesc>
    </category>
  </taxonomy>
</classDecl>
</encodingDesc>
```

BVH : encodingDesc

```
<encodingDesc>
  <p>Les langues sont encodées selon la norme iso 639-3. Pour les dialectes qui n'ont pas encore de
code iso, un code non standard est appliqué sur le modèle "oci-gascon".
  Pour l'encodage des langues artificielles et du latin macaronique, nous avons conservé le code
639-2 "art".</p>
  <projectDesc>
    <p>BVH : programme Bibliothèques Virtuelles Humanistes</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p>Les corrections sont encodées. <lb/>Quelques corrections suivent la réédition de 1549,
en ligne sur le site de la <ref
      target="http://iris.lib.virginia.edu/speccol/gordon/gordonimages/Typ1549_T67/">Gordon
collection</ref>, cote Typ1549_T67.</p>
    </correction>
    <hyphenation>
      <p>La division des mots en fin de ligne respecte l'original.</p>
    </hyphenation>
    <normalization>
      <p>La transcription suit l'exemplaire de la Bibliothèque municipale de Blois, cote I 65.
<lb/>Les apostrophes et les majuscules sont rétablies selon l'usage moderne.
      <lb/>f.42r: la lettre fantastique inversée qui correspond à un "L" n'a pas été transcrite mais
encodée comme un glyphe.
      <lb/>f.50v: Les caractères inversés "D" et "Q" n'ont pas été transcrits mais encodés comme des
glyphes.
      <lb/>f. 74v et f. 75: Le caractère unicode 5.2 "A76E" n'étant pas encore affiché correctement par
les navigateurs, il est mis en commentaire et remplacé par "9".
    </p>
  </normalization>
</editorialDecl>
```

le teiHeader : arborescences profileDesc et revisionDesc



profileDesc et les éléments de model.profileDescPart


<abstract/> résumé que l'encodeur veut apposer au document

<creation/>

<langUsage/> (*language usage*)

<textClass/> (*text classification*) (qui peut utiliser un système formalisé de mots-clés, une ontologie, qui doit alors être décrite auparavant dans encodingDesc/classDecl

<calendarDesc/> (description des calendriers)

<correspDesc>  (*correspondence description*), qui permet de décrire les différentes actions d'une correspondance (envoi, réception, ...) et son contexte.

(...)

<langUsage> et <language>

(ex. tiré de la BVH)

```
<langUsage>
  <language ident="frm">[moyen] Français</language>
  <language ident="lat">Latin</language>
  <language ident="ell">Grec</language>
  <language ident="ita">Italien</language>
  <language ident="ara">Arabe</language>
  <language ident="oar">Ancien araméen</language>
  <language ident="art">Langue artificielle</language>
  <language ident="art">Latin macaronique</language>
  <language ident="pcd">Picard</language>
  <language ident="oci-gascon">Gascon</language>
  <language ident="fra-lorrain">Lorrain</language>
</langUsage>
```

Les codes linguistiques (@ident) doivent théoriquement (ce n'est pas systématiquement le cas dans cet exemple*) être conformes au protocole défini par l'*Internet Engineering Task Force* (BCP 47), composés des codes linguistiques ISO 639 et des codes nationaux ISO 3166, cf.

http://en.wikipedia.org/wiki/IETF_language_tag et

<http://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>

Pour les deux derniers cas ; en la matière, la pratique de la BVH, est que « pour les dialectes qui n'ont pas encore de code iso, un code non standard est appliqué sur le modèle "oci-gascon" ».

<langUsage> et <language> (autre exemple)

```
<langUsage>  
  <language ident="fr-CA" usage="99">français du  
  Canada</language>  
  <language ident="en-CA" usage="1">anglais (quelques  
  mots)</language>  
</langUsage>
```

textClass :

contient une série de mots-clés utilisés pour décrire le texte, et qui peuvent correspondre à une taxonomie décrite dans

encodingDesc/classDecl

```
<encodingDesc>
  <!-- [...] -->
  <classDecl>
    <taxonomy xml:id="Rameau">
      <bibl>Répertoire d'autorité-matière encyclopédique et alphabétique unifié
        (RAMEAU) de la <orgName>Bibliothèque nationale de France</orgName>. <ptr
target="http://rameau.bnf.fr/" />
      </bibl>
    </taxonomy>
  </classDecl>
</encodingDesc>
<profileDesc>
  <!-- [...] -->
  <textClass>
    <keywords scheme="#Rameau">
      <term>Text Encoding Initiative (langage de balisage) – Problèmes et
exercices</term>
      <term>École nationale des chartes (Paris) – Thèses et écrits académiques</term>
    </keywords>
  </textClass>
</profileDesc>
```


Éléments ajoutés par le module corpus

textDesc : *(description de texte) fournit la description d'un texte sous l'angle du contexte situationnel. i.e. la situation dans laquelle prend place une communication (canal ou medium, constitution, dérivation, domaine concerné, , degré de factualité/fictionnalité, intention, degré de préparation, degré d'interaction, fonction...)*

particDesc : *(description des participants) décrit les locuteurs, voix ou autres participants identifiables d'une interaction linguistique.*

éléments permettant de fournir des informations détaillées sur les personnes impliquées dans le document (auteurs, locuteurs, éditeurs, ...), jusqu'à un degré très fin d'un point de vue prosopographique. Peut inclure une `<listPerson>`.

settingDesc : *(setting description) describes the setting or settings within which a language interaction takes place, or other places otherwise referred to in a text, edition, or metadata /*

(description du contexte) décrit le(s) contexte(s) dans lesquels se situe une interaction linguistique, soit sous la forme d'une description en prose, soit sous celle d'une série d'éléments décrivant le contexte. Peut inclure une `<listPlace>`.

profileDesc (*Queste...*)

```
<profileDesc>
  <creation>
    <name type="author">anonyme</name>
    <title>Queste del saint Graal</title>
    <date type="compo" when="1220-01-01" n="13">vers 1220</date>
    <date type="compo_periode">ancien</date>
    <date type="compo_sous_siecle" n="13_1">début</date>
    <region type="dialecte_auteur">non_defini</region>
  </creation>
  <langUsage>
    <language ident="fr" usage="100">Le texte est entièrement écrit en vers
en français médiéval</language>
  </langUsage>
  <textDesc n="roman">
    <channel mode="w">manuscript</channel>
    <constitution type="single"/>
    <derivation type="original"/>
    <domain type="litteraire"/>
    <factuality type="inapplicable"/>
    <interaction type="none"/>
    <preparedness type="prepared"/>
    <purpose type="narrative"/>
  </textDesc>
  <textClass>
    <catRef target="#forme_prose"/>
  </textClass>
</profileDesc>
```

revisionDesc

Une séquence d'éléments `change`, *par ordre antéchronologique*, un pour chaque révision du fichier, qui peuvent être groupés en `list` ou `listChange`.

Attributs `@when`, `@who` (ou, comme alternative, les sous-éléments `name` et `date`)

revisionDesc

ex. tiré de la *Queste*...

```
<revisionDesc>
  <change when="2012-10-20" who="#a1">Corrections ponctuelles</change>
  <change when="2011-12-20" who="#a1">Corrections ponctuelles</change>
  <change when="2011-06-03" who="#a1">Corrections ponctuelles, traitement des
segmentations</change>
  <change when="2010-12-22" who="#a1">Augmentation et mise à jour de la transcription
multi-facette (syntaxe compacte) ; intégration des corrections signalées
auparavant</change>
  <change when="2010-10-27" who="#a1">Correction des balises supplied sur plusieurs
phrases</change>
  <change when="2010-06-17" who="#a1">Ajout de # aux attributs resp</change>
  <change when="2010-05-21" who="#a1">Corrections ponctuelles : remplacement
étiquette ADVgen par ADVneg pour "non", points autour des chiffres, étiquette de "jel"
; mise en minuscule de "saint" dans le titre</change>
  <change when="2010-01-25" who="#a1">Corrections ponctuelles</change>
  <change when="2009-07-14" who="#a1">Intégration des corrections, conversion des
étiquettes en Cattex2009</change>
  <change when="2009-04-04" who="#cmn">Révision du texte</change>
  <change when="2009-02-10" who="#a1">Intégration des étiquettes
morphosyntaxiques</change>
  <change when="2009-02-06" who="#cmn #sp">Révision de l'étiquetage
morphosyntaxiques</change>
  <change when="2008-07-07" who="#a1">Corrections ponctuelles</change>
  <change when="2008-05-28" who="#a1">Préparation à la tokenisation, corrections
ponctuelles</change>
  <change when="2008-04-07" who="#a1">Ajout d'une licence CC</change>
  <change when="2008-02-21" who="#a1">Éléments de structuration
du doc Word en XML-TEI et premier
```



teiHeader, autres formats de métadonnées et catalogues de bibliothèques

teiHeader/fileDesc conçu en s'appuyant sur la norme ISBD (tous les éléments en -stmt, cf. TEI Consortium, « **2.1.2 Types of Content in the TEI Header** »),

Il est donc possible d'envisager des mises en correspondance des données (ou *data mappings*) et des conversions **depuis** des formats bibliographiques vers teiHeader, ou depuis teiHeader, **vers** d'autres formats de métadonnées bibliographiques (MARCXML, DublinCore, Mods, Mets...) et les technologies du web sémantique (RDF,...)

Sur le sujet, voir le TEI Consortium, « 2.7 Note for Library Cataloguers », dans *TEI Guidelines...*, en ligne :

<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD8>>, ainsi que TEI Libraries SIG, « 4.1. The TEI Header », dans *Best Practices for TEI in Libraries...*

Norme ISBD (IFLA) (1/3)

Dernière version publiée en 2011 (*ISBD intégré*), voir la trad. fr. par la BnF :
http://www.bnf.fr/documents/isbd_trad_francais.pdf

(tentative partielle de mise en correspondance entre les champs de
l'International standard bibliographic description et le `teiHeader`)

0. Zone de la forme du contenu et du type de médiation => texte (visuel),
électronique

1. Zone du titre et de la mention de responsabilité => `titleStmt`

1.1 Titre propre => `title`

1.2 Titre parallèle => `title`

1.3 Complément du titre => `title`

1.4 Mention de responsabilité => `author`, `editor`, `respStmt`

2. Zone de l'édition => `editionStmt`

2.1 Mention d'édition => `edition`

2.3 Mention de responsabilité relative à l'édition => `respStmt` (ou éléments
spécialisés)

3. Zone particulière à certains types de ressources => pas de correspondance
unique

Norme ISBD (IFLA) (2/3)

4. Zone de la publication, production, distribution, etc. => publicationStmt

4.1 Lieu de publication, production et/ou distribution => pubPlace

4.2 Nom de l'éditeur, producteur et/ou distributeur => publisher,
distributor, authority

4.3 Date de publication, production et/ou distribution => date

4.4 Lieu d'impression, de fabrication ou de gravure => ? (ne s'applique pas,
ou sourceDesc)

4.5 Nom de l'imprimeur ou du fabricant => ? (ne s'applique pas, ou
sourceDesc)

4.6 Date d'impression ou de fabrication => ? (ne s'applique pas, ou
sourceDesc)

5 Zone de la description matérielle => extent (si l'on veut) pour le fichier élect.,
sinon, sourceDesc, et bibl ou msDesc

Norme ISBD (IFLA) (3/3)

6. Zone de la collection et de la monographie en plusieurs parties => seriesStmt

6.1 Titre propre de collection, sous - collection ou monographie en plusieurs parties. => title

6.3 Complément du titre de collection ou monographie en plusieurs parties => title

6.4 Mention de responsabilité relative à la collection ou à la monographie en plusieurs parties => editor, respStmt

6.5 Numéro international normalisé de la collection ou sous – collection => idno

6.6 Numérotation à l'intérieur de la collection, sous-collection ou monographie en plusieurs parties => idno

7. Zone des notes => notesStmt

8. Zone de l'identificateur et des modalités d'acquisition => ne s'applique pas

Conversion vers ou depuis le teiHeader

Vers teiHeader

Depuis	Vers	Outil(s) existant(s)
MARXML	teiHeader	Thutmose II (feuilles XSLT)

Depuis teiHeader (NB : certains outils de gestion bibl. permettent d'importer et/ou d'exporter dans un certain nombre de ces formats, par ex. *Zotero* ou *RefDB*)

Depuis	Vers*	Outil(s) existant(s)
teiHeader	MARC	Voir le (vieux - 1998) mapping réalisé par l'ALA Voir aussi les travaux en cours du SIG Libraries
teiHeader	MODS	TEI2MODS (XSLT)
teiHeader	Dublin Core	XSLT du Consortium TEI Nombreux mappings, dont celui d' Algone/Obvil

* Ces formats sont classés, peu ou prou, par ordre de complexité décroissante.

J.B. Camps – La TEI : métadonnées et éléments de structuration

Le Dublin Core

Éléments pour une mise en correspondance des données

Se rapporter également à la bibliographie pour plus d'éléments à ce sujet.

Dublin Core	Corresp. possibles en TEI
dc:title	fileDesc/titleStmt/title
dc:creator	fileDesc/titleStmt/author, editor, respStmt
dc:subject	profileDesc/textClass/keywords/term
dc:description	encodingDesc ou profileDesc/abstract
dc:publisher	fileDesc/publicationStmt/publisher <i>etc.</i>
dc:contributor	fileDesc/titleStmt/editor, respStmt
dc:date	fileDesc/publicationStmt/date
dc:type	"text"
dc:format	"text/xml"
dc:identifiant	fileDesc/publicationStmt/idno
dc:source	fileDesc/sourceDesc
dc:language	@xml:lang ou profileDesc/langUsage/
dc:relation	fileDesc/seriesStmt/
dc:coverage	profileDesc/textDesc
dc:rights	fileDesc/publicationStmt/availability



Plan de la séance

I. Les métadonnées du document :
l'entête `teiHeader`

**II. Divisions et structuration du
document**

Types de structuration et modules concernés

1. Structuration logique

i. Éléments généraux et différents niveaux

ii. Forme du discours

α. prose

β. vers

iii. Cas et genres particuliers

α. Dictionnaires

β. Théâtre

γ. Épistolaire/Correspondance

2. Structuration physique

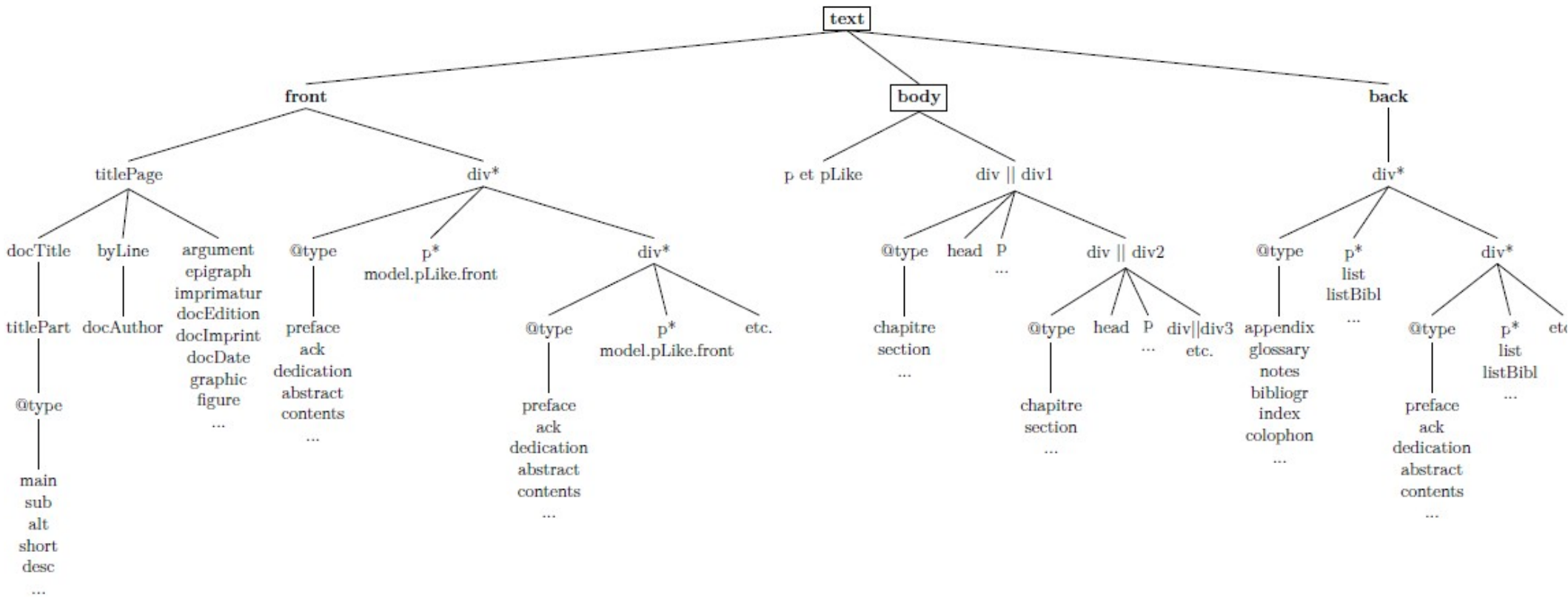
Structuration logique : Structuration générale du (corps du) document

Modules et éléments concernés :

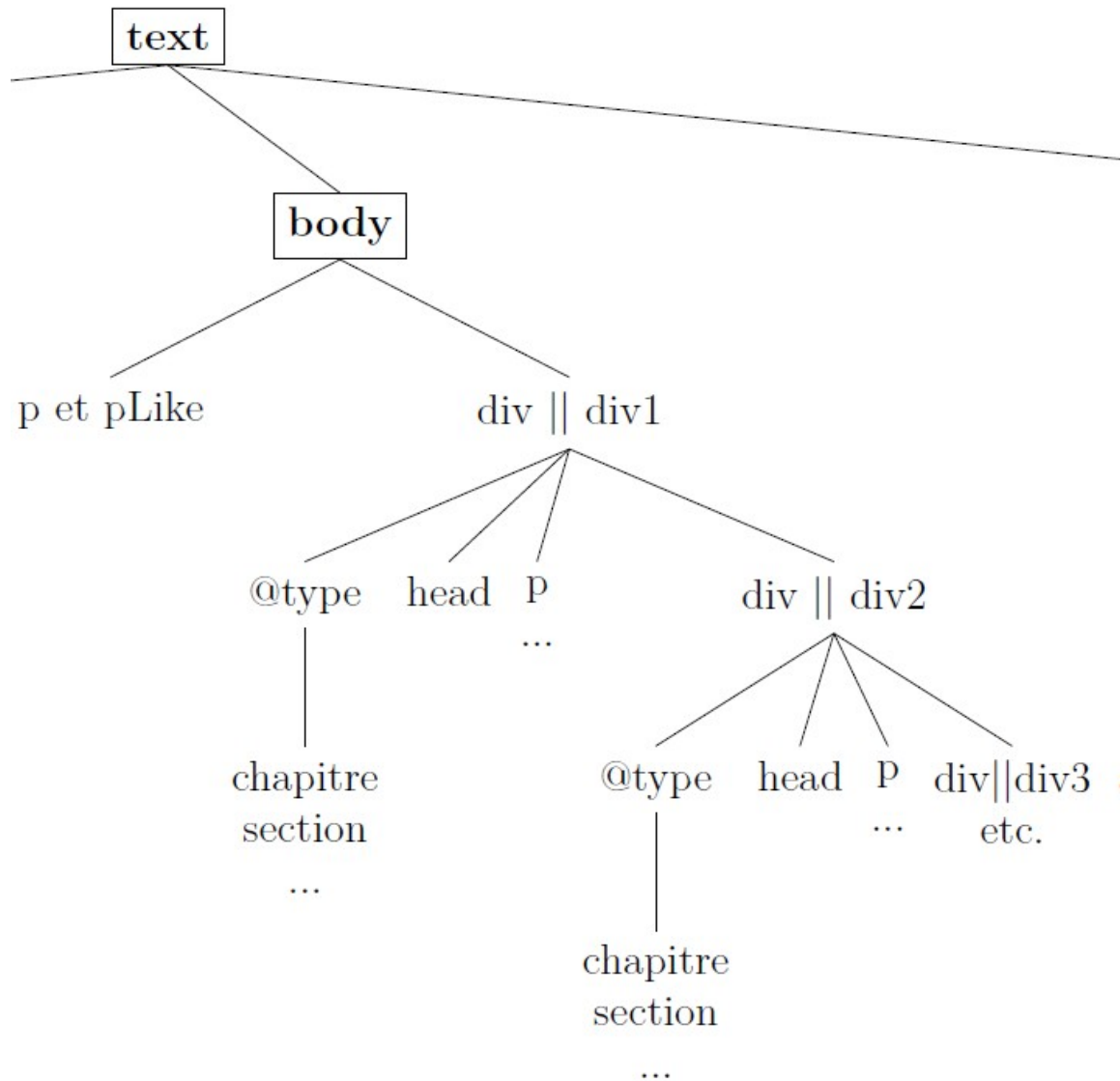
Module textstructure [module obligatoire]:
Default text structure

Elements defined: TEI argument back body
byline closer dateline div div1 div2 div3
div4 div5 div6 div7 docAuthor docDate
docEdition docImprint docTitle epigraph
floatingText front group imprimatur opener
postscript salute signed text titlePage
titlePart trailer

Arborescence générale text (niveau document)



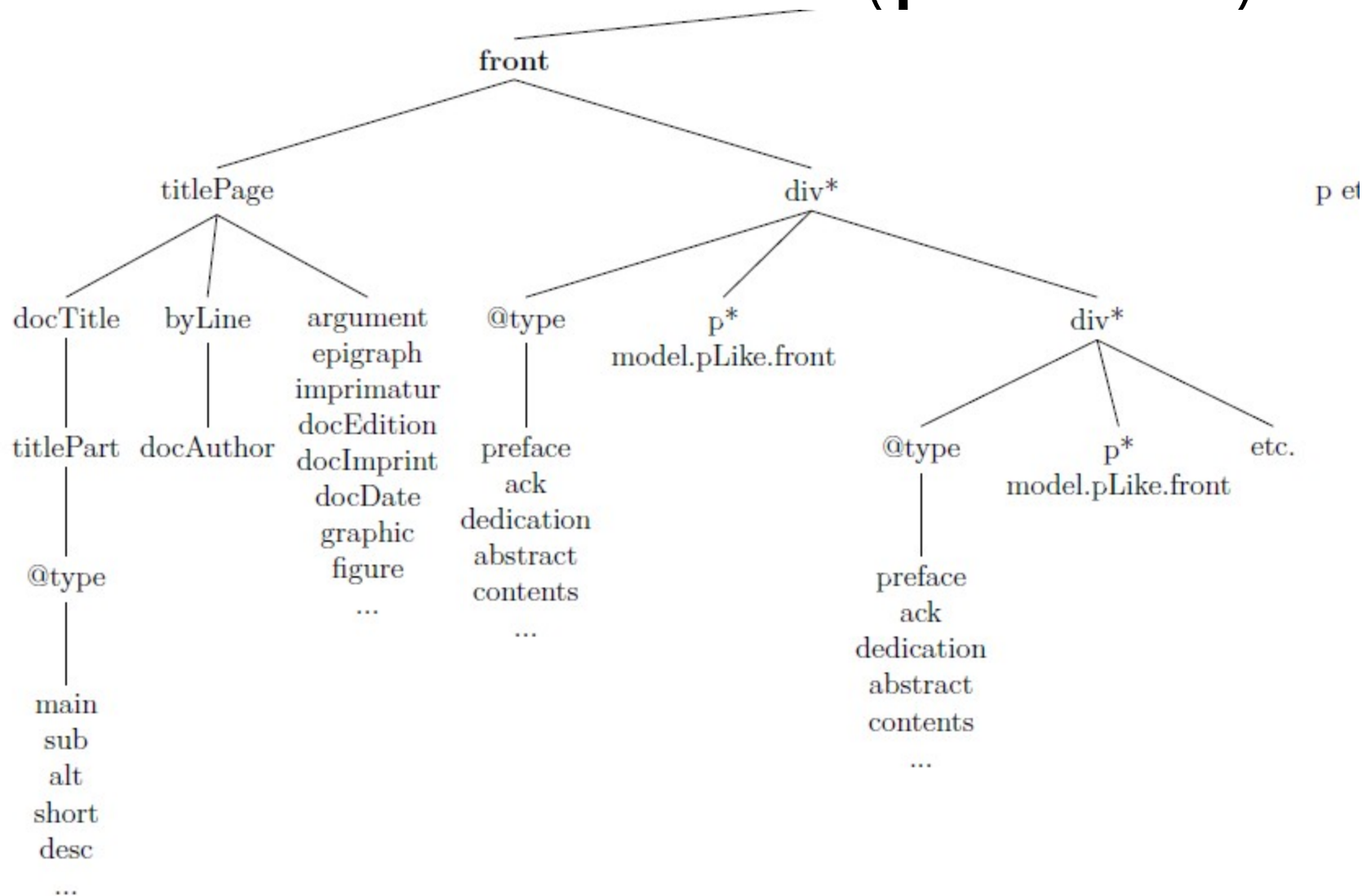
Arborescence **body** (très partielle)



Exemple de *div* imbriqués

```
<div type="niveau1" n="1">
  <head>1. Titre de premier
niveau : Quelques travaux pratiques</head>
  <p>Un premier paragraphe,
contenant du texte qui se situe à ce premier
niveau.</p>
  <div type="niveau2" n="1.1">
    <head>1.1. Titre de second
niveau : Un premier document XML</head>
    <p>Mon premier paragraphe de
ce second niveau [...]</p>
  </div>
</div>
```


Arborescence front (partielle)

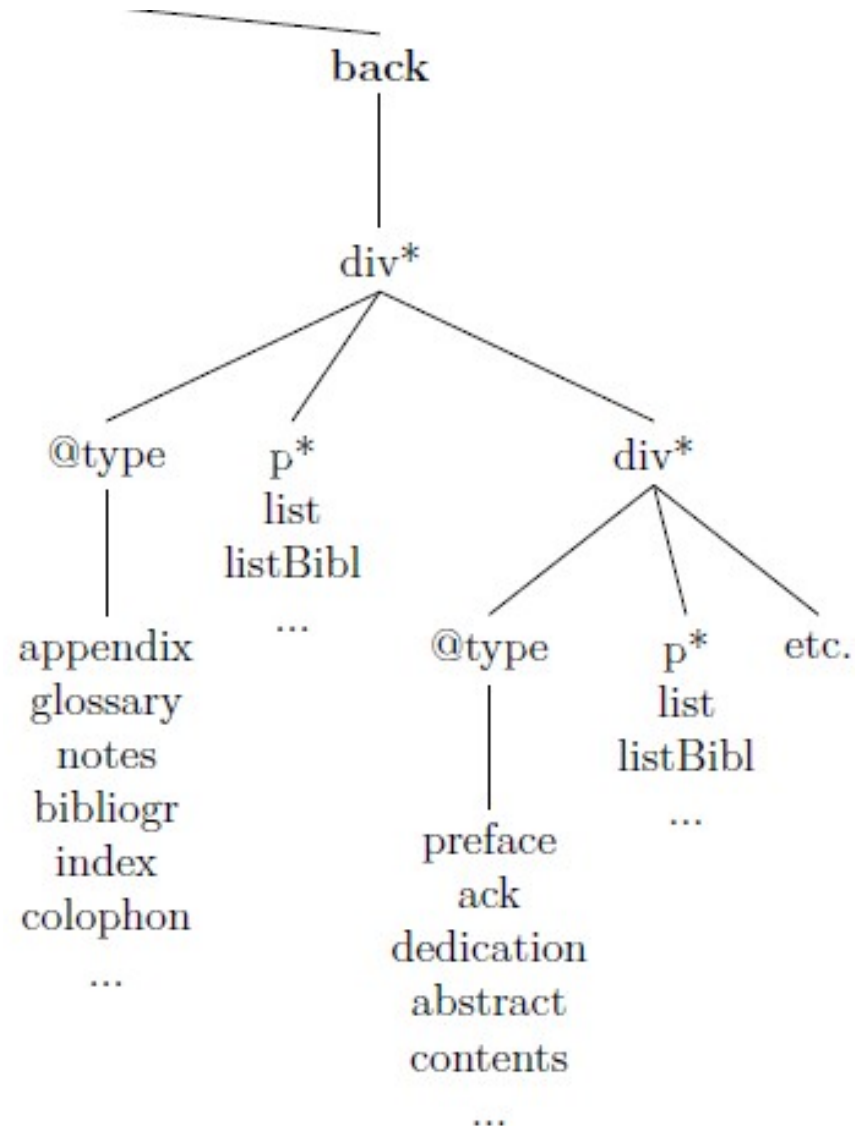


cf. les [model.frontPart](#), [model.titlepagePart](#) et [model.pLike.front](#)

ex. d'élément titlePage (BVH)

```
<titlePage>
  <figure>
    <graphic url="B410186201_I65.jpg" xml:base="http://www.bvh.univ-
tours.fr/Consult/index.asp?numfiche=649"/>
    <figDesc>Bibliothèques Virtuelles Humanistes - Bibliothèque municipale de
Blois</figDesc>
  </figure>
  <docTitle>
    <titlePart rend="center">
      <lb/>
      <hi rend="large">CHAMP
<lb/>FLEURY.</hi>
    </titlePart>
    <titlePart>
      <lb/>Au quel est contenu l'Art & Science
<lb/>de la deue & vraye Proportion des Let
<lb/>tres Attiques, qu'on dit autrement Let-
<lb rend="hyphen"/>tres Antiques, & vulgairement Let-
<lb rend="hyphen"/>tres Romaines proportionnees selon
<lb/>le Corps & Visage humain.
    </titlePart>
  </docTitle>
  <docImprint>
    <lb/>
    <lb/>Ce Livre est Privilegie pour Dix Ans
<lb/>Par Le Roy nostre Sire. & est a ven-
<lb rend="hyphen"/>dre a <pubPlace>Paris</pubPlace> sus <placeName type="rue">Petit
Pont</placeName> a l'Enseigne
<lb/>du <placeName type="enseigne">Pot Casse</placeName> par <publisher>Maistre <name>
  <choice>
    <orig>Geoffroy</orig>
    <reg>Geoffroy</reg>
  </choice>
  <orig>J.B. Camps</orig>
  <reg>J.B. Camps</reg>
  </choice>
  </docImprint>
</titlePage>
```

Arborescence **back** (partielle)



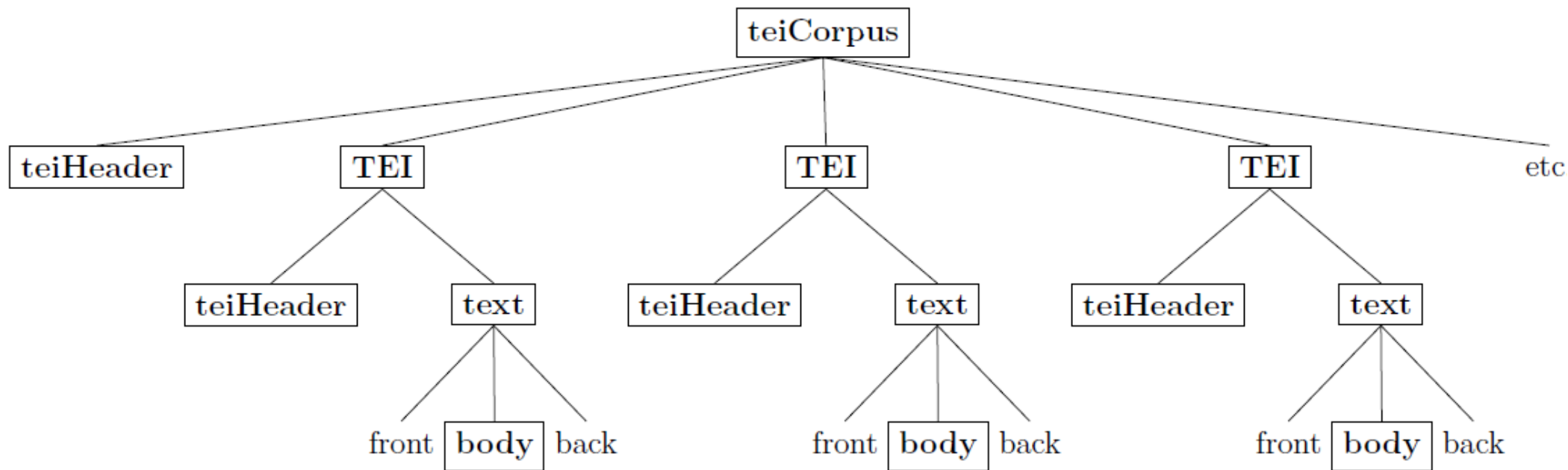
Cas particuliers des collections de textes et des textes composites

teiCorpus (module core)

group (module textstructure)

Voir aussi la présentation dans le chapitre 15
(« Language Corpora »),
15.1 « **Varieties of Composite Text** »

Niveau du corpus de textes : teiCorpus



teiCorpus pour une coll. de textes

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0" xml:lang="fr">
  <teiHeader>
    <!-- en-tête du corpus -->
  </teiHeader>
  <TEI>
    <teiHeader>
      <!-- en-tête du premier texte -->
    </teiHeader>
    <text>
      <!-- contenu du premier texte -->
    </text>
  </TEI>
  <TEI>
    <teiHeader>
      <!-- en-tête du deuxième texte -->
    </teiHeader>
    <text>
      <!-- contenu du deuxième texte -->
    </text>
  </TEI>
  <!-- Éléments TEI supplémentaires-->
</teiCorpus>
```

Niveau des textes composites : group

TEI Guidelines, « 4. Default Text Structure »

« contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc. »

(<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>)

« 4.3.1 Grouped Texts »

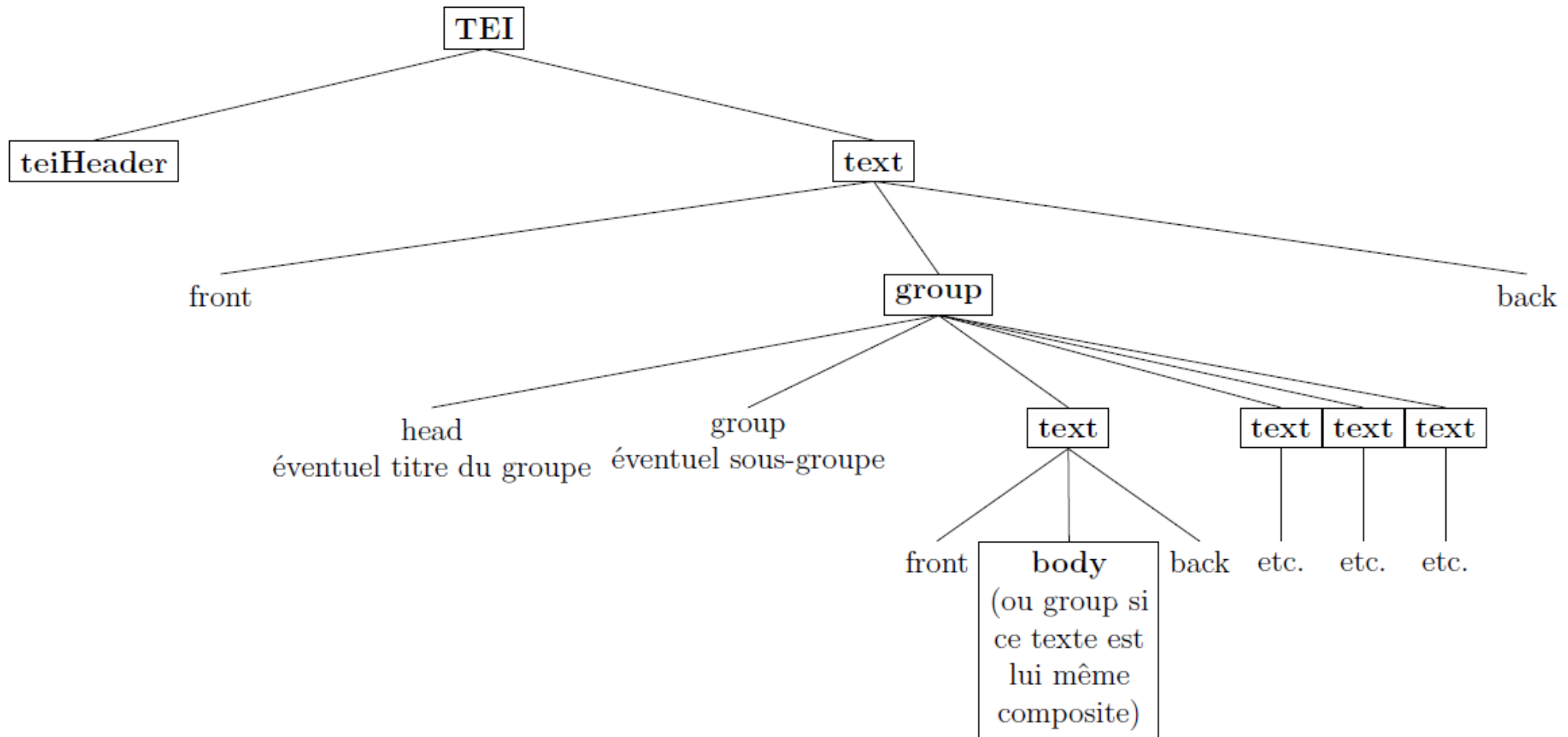
« Examples of composite texts which should be represented using the group element include anthologies and other collections. The presence of common front matter referring to the whole collection, possibly in addition to front matter relating to each individual text, is a good indication that a given text might usefully be encoded in this way; this structure may be found useful in other circumstances too. »

(<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSGRP>)

« 15 Language Corpora »

« The teiCorpus element is intended for the encoding of language corpora, though it may also be useful in encoding newspapers, electronic anthologies, and other disparate collections of material. The individual samples in the corpus are encoded as separate TEI elements, and the entire corpus is enclosed in a teiCorpus element. [...] If it is essential to reflect a single permanent organization of a corpus into sub- and sub-sub-corpora, then the corpus or the high-level subcorpora may be encoded as composite texts, using the group element [...] The group element is provided to simplify the encoding of collections, anthologies, and cyclic works; as noted above, the group element can also be used to record the potentially complex internal structure of language corpora. »

Niveau des textes composites : group



Niveau des textes composites : group

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ..... -->
  </teiHeader>
  <text>
    <front>
      <!-- front de l'ensemble du groupement de textes -->
    </front>
    <group>
      <text>
        <front>
          <!-- front du premier texte, si besoin -->
        </front>
        <body>
          <!-- corps du premier texte -->
        </body>
        <back>
          <!-- back du premier texte -->
        </back>
      </text>
      <text>
        <body>
          <!-- corps du second texte -->
        </body>
      </text>
    </group>
    <back>
      <!-- back du groupement de textes -->
    </back>
  </text>
</TEI>
```

```

<text>
  <front>
    <docTitle><titlePart>Anthologie de poèmes
divers</titlePart></docTitle>
  </front>
  <group>
    <group>
      <head>Recueil des années 1880-1889</head>
      <text><front><head>Odes et poèmes à mon
chat</head></front><body><!--[...]--></body></text>
      <text><front><head>Sonnets</head></front><body><!--[...]--
></body></text>
    </group>
    <group>
      <head>Poèmes épiques</head>
      <text><front><head>La Nouvelle Iliade</head></front><body><!--
[...]--></body></text>
      <text><front><head>La Nouvelle Odyssée</head></front><body><!--
[...]--></body></text>
    </group>
  </group>
</text>

```

ENC Morchesne...

```
<group>
  <head>Le formulaire d'Odart de Morchesne</head>
  <group xml:id="morchesne_1" n="1">
    <pb n="fol. 1"/>
    <head>[Graces et debitis]<note>Omis dans le manuscrit, le
titre est ici reporté d'après la table.</note></head>
    <text xml:id="morchesne_1-1" n="1.1" type="formule"
subtype="article">
      <front>
        <head>Grace a plaidier par procureur</head>
        <index>
          <term type="auth_type" key="act_souv"/>
        </index>
      </front>
      <body>
        <div type="transcription" xml:lang="fro">
          <p>
            <seg function="intitulatio">Charles, par la grace de
Dieu roy de France,</seg>
            (...)
          </p>
        </div>
      </body>
    </text>
```

Structuration de la prose (module core et quelques autres)

- **<p>** : paragraphes
 - **<seg>** pour des subdivisions internes au paragraphe, de même qu'un certain nombre d'éléments du module analysis (seg, s, ...)
- **<list>** contenant des **<item>**
- **<quote>** pour les citations simples, **<q>** pour ce qui est entre guillemets
- **<cit>** avec **<quote>** et **<bibl>** pour les citations avec réf.
- **<table>** pour des tableaux
- **<figure>** pour des images

Structuration de la poésie

- **<lg>** (line group) pour les groupes et sous-groupes de vers (poèmes, strophes, etc.) avec un @type
 - **<l>** pour chacun des vers

Structuration du théâtre

- **<sp>** (speech) pour du texte parlé, dont
 - **<speaker>** pour celui qui parle
 - **<stage>** pour les didascalies
 - **<p>** ou **<lg>** le cas échéant pour ce qui est dit

Théâtre et vers, ex.

```
<sp>
  <speaker>Don Rodrigue</speaker>
  <lg>
    <l>Oui ; tout autre que moi</l>
    <l>Au seul bruit de ton nom pourrait trembler
d'effroi.</l>
    <l>Les palmes dont je vois ta tête si couverte</l>
    <l>Semblent porter écrit le destin de ma perte.</l>
    <l>J'attaque en téméraire un bras toujours
vainqueur;</l>
    <l>Mais j'aurai trop de force, ayant assez de cœur.</l>
    <l>À qui venge son père il n'est rien d'impossible.</l>
    <l>Ton bras est vaincu, mais non pas invincible.</l>
  </lg>
</sp>
```

Structuration matérielle

- `<pb/>` pour les changements de page (avec @n pour la page qui commence)
- `<cb/>` pour un changement de colonne
- `<lb/>` pour les changements de ligne

Bibliographie

En complément des références données au cours de la présentation, et des Guidelines.

1. Tutoriels et exercices

Ron Van den Branden, Melissa Terras & Edward Vanhoutte, *TEI by Example*,

<<http://www.teibyexample.org>> (consulté en nov. 2013) ;

Particulièrement les tutoriels, exercices, exemples et QCM des leçons

« 1: Common Structure and Elements »

« 2: The TEI Header »

« 3: Prose »

« 4: Poetry »

« 5: Drama »

Liste de tutoriels sur la TEI, en ligne : <<http://www.tei-c.org/Support/Learn/tutorials.xml>>

2. Texte de référence

TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*,

<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>>, particulièrement

« 1 The TEI Infrastructure »

« 2 The TEI Header »

« 3 Elements Available in All TEI Documents »

« 4 Default Text Structure »

3. Sources des exemples

Queste del saint Graal, éd. numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77), par Christiane Marchello-Nizia, avec la collab. d'Alexei Lavrentiev, v. 0.8.1, janv. 2013, en ligne : <http://portal.textometrie.org/txm/>.

Le formulaire d'Odart de Morchesne, dir. Olivier Guyotjeannin et Serge Lusignan, 2011 (Éditions en ligne de l'École des chartes, 10), en ligne : <http://elec.enc.sorbonne.fr/morchesne/> (sources : <http://elec.enc.sorbonne.fr/morchesne/src/morchesne.xml>).

Fin du XIVe siècle. Le recueil de prières de Philippe de Mézières, éd. Cédric Giraud *et al.*, 2008 (Thélème, 96), en ligne : <http://theleme.enc.sorbonne.fr/dossiers/notice96.php> (sans les sources).

DU CANGE, et al., *Glossarium mediæ et infimæ latinitatis*. Niort : L. Favre, 1883-1887, éd. sous la dir. d'Anita Guerreau-Jalabert, 2008 (Éditions en ligne de l'École des chartes, 18), en ligne : <http://ducange.enc.sorbonne.fr/>

'*Champ fleury*', Paris, 1529, éd. sous la dir. de Marie-Luce Demonet, 1^{re} éd. (mise à jour en) 2013, Tours, Bibliothèques Virtuelles Humanistes, en ligne : http://xtf.bvh.univ-tours.fr/xtf/view?docId=tei/B410186201_I65/B410186201_I65_tei.xml

SCUDÉRY (Madeleine de), *Artamène ou le Grand Cyrus*, éd. Claude Bourqui et Alexandre Gefen, Neuchâtel, 2003, en ligne : <http://www.artamene.org/>.

4. **teiHeader** et métadonnées bibliographiques

4.1 **TEI et bibliothèques : généralités**

TEI CONSORTIUM « 2.7 Note for Library Cataloguers », dans *TEI Guidelines...*, en ligne : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD8>.

TEI LIBRARIES SIG, *Best Practices for TEI in Libraries : A guide for mass digitization, automated workflows, and promotion of interoperability with XML using the TEI*, ver. 3.0, 2011, en ligne : <http://purl.org/TEI/teiinlibraries>, part. « 4.1. The TEI Header ».

Se reporter également à la section « Tools » du TEI-Wiki.

En ce qui concerne ISBD, voir : FÉDÉRATION INTERNATIONALE DES ASSOCIATIONS DE BIBLIOTHÉCAIRES ET DES BIBLIOTHÈQUES, ISBD : Description bibliographique internationale normalisée, Paris, BnF, 2013, en ligne : http://www.bnf.fr/documents/isbd_trad_francais.pdf.

4.2 **Dublin Core**

Se reporter aux mappings proposés par :

ALGONE/OBVIL, *Schéma Teibook*, « TEI vers Dublin Core » (voir aussi Unimarc|Onix vers TEI), en ligne : http://obvil-dev.paris-sorbonne.fr/developpements/teibook/#output_mapping (consulté en déc. 2015).

CALIFORNIA DIGITAL LIBRARY, *CDL Encoding Guidelines for the TEI Header*, éd. Jared Campbell & Brian Tingle, University of California, 2006, en ligne : http://www.cdlib.org/groups/stwg/META_BPG.html#d52e344.

On trouve d'autres exemples et initiatives, comme :

YOUNGLOVE (April), « TEI Header and Dublin Core Mapping », School of Library and Information Management Emporia University, 2007, en ligne : <http://www.april.secac.com/website/YoungloveTEI.pdf>.

Pour les éléments du Dublin Core, <http://dublincore.org/documents/dces/>.