

Université de Neuchâtel

Master en littérature

# TG: Édition de texte (numérique)

## Cours 8: documenter

Élodie Paupe

chaire de philologie classique et d'histoire ancienne

9 novembre 2020

# Pourquoi documenter?

- Au début d'un projet, une étape de modélisation est nécessaire.
- Différents choix éditoriaux sont faits qui vont mener l'éditrice ou l'éditeur à utiliser certains éléments, certains attributs, etc. On dit que l'on "personnalise la TEI".
- Ces choix doivent être documentés pour permettre la bonne lecture des documents TEI, la continuation de projets éditoriaux, la spécificité de certains usages, etc.

## Personnaliser, c'est-à-dire?

Deux priorités:

- personnaliser la TEI = restreindre la TEI globale: votre TEI personnalisée doit être valide du point de vue de la TEI
- conserver le sens des éléments établi par le consortium

On parle de personnalisation "propre" si ces points sont respectés, sinon on la dira "sale".

Modifications propres:

- Supprimer les éléments inutiles
- Modifier un modèle de contenu
- Modifier une liste d'attributs ou de valeurs d'attributs

Propre/sale:

- Ajouter de nouveaux éléments
- Ajouter de nouveaux attributs

# Les personnalisations de la TEI les plus diffusées

Il existe déjà certaines personnalisations de la TEI largement diffusées.

- all: tous les éléments disponibles
- lite: sous-ensemble d'environ cinquante éléments (= les besoins d'env. 90% des utilisatrices et utilisateurs)
- absolutely bare: version minimale
- Epidoc: spécialisation pour les épigraphistes
- ...

## Exemples

- *Bibliothèques virtuelles humanistes*: <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>
- École des chartes: <http://developpements.enc.sorbonne.fr/diple/schema/>
- *Sources du droit suisse*: <https://www.ssrq-sds-fds.ch/wiki/Transkriptionsrichtlinien/fr>

# Rédiger la documentation

La documentation est réalisée à l'aide de langages de définitions de données (*data definition language*):

- une DTD (*Document Type Definition*)
- un schéma:
  - XML Schema ( `.xsd` )
  - RELAX NG XML ( `.rng` )
  - RELAX NG Compact ( `.rnc` )
- un ODD (*One Document Does it all*)

Ces documents fournissent la grammaire et le vocabulaire d'encodage d'un projet.

**ODD:** il fournit l'information nécessaire aux traitements informatiques en même temps que la documentation de cette information destinée à être lue par un être humain, le tout intégré dans un seul document XML. Comme on pouvait s'y attendre, le système TEI lui-même est exprimé en utilisant précisément ce même ensemble d'éléments, mais ici nous nous concentrerons sur son usage pour la création d'une personnalisation de la TEI.

– Burnard 2015: "Personnaliser la TEI", 5



# DTD

```
<div type="test" xml:id="t1">  
  <p>Des choses intéressantes </p>  
  <p>D'autres choses intéressantes chez <persName>Caroline</persName></p>  
</div>
```

- Syntaxe:

```
<!DOCTYPE div [  
<!ELEMENT div (p+)>  
<!ATTLIST div  
    type CDATA #REQUIRED  
    xml:id CDATA #REQUIRED>  
<!ELEMENT p (#PCDATA | persName)*>  
<!ELEMENT persName (#PCDATA)>  

```

Schématiquement, cela donne:

```
<!DOCTYPE élémentracine [  
  <!ELEMENT balise1 (_balises contenues_)>  
  <!ATTLIST balise1  
    attribut1 CDATA (= des données) #REQUIRED (= obligatoire)  
    attribut2_CDATA #IMPLIED (= facultatif)>  

```

Elle peut être interne (au début du document) ou externe (dans un autre fichier ou en ligne):

- Interne:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE div [
  <!ELEMENT div (p+)>
    <!ATTLIST div
      type CDATA #REQUIRED
      xml:id CDATA #REQUIRED>
  <!ELEMENT p (#PCDATA | persName)*>
  <!ELEMENT persName (#PCDATA)>
]>

<div type="test" xml:id="t1">
  <p>Des choses intéressantes </p>
  <p>D'autres choses intéressantes chez <persName>Caroline</persName></p>
</div>
```

- Externe, mais enregistré sur mon ordinateur:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE div SYSTEM "monfichier.dtd">

<div type="test" xml:id="t1">
  <p>Des choses intéressantes </p>
  <p>D'autres choses intéressantes chez <persName>Caroline</persName></p>
</div>
```

# Déclarer les éléments

- Vide: `<!ELEMENT anchor EMPTY>`
- Données textuelles: `<!ELEMENT title (#PCDATA)>`
- Imbrication:
  - élément obligatoire: `<!ELEMENT TEI (teiHeader)>`
  - élément obligatoire répété une ou plusieurs fois: `<!ELEMENT list (item+)>`
  - élément facultatif répété zéro ou plusieurs fois: `<!ELEMENT p (app*)>`
  - élément facultatif répété zéro ou une fois: `<!ELEMENT div (head?)>`
  - élément à choix: `<!ELEMENT sourceDesc (bibl | msDesc)>`
  - plusieurs éléments dans l'ordre: `<!ELEMENT fileDesc (titleStmt, publicationStmt, sourceDesc)>`
  - élément ou texte: `<!ELEMENT p (#PCDATA | persName)*>`

# Déclarer les attributs

```
<!ATTLIST div
  type CDATA #IMPLIED
  xml:id CDATA #REQUIRED
  lang CDATA #FIXED "fr">
```

- facultatif: `#IMPLIED`
- obligatoire: `#REQUIRED`
- fixe: `#FIXED` (dans ce cas, la valeur de l'attribut est défini par défaut et ne peut pas être changée.)

Valeur de l'attribut:

- une chaîne de caractère: `CDATA`
- liste à choix: `<!ATTLIST lg type (distique|tercet|quatrain|sizain) #REQUIRED>`
- valeur unique: `ID`

Exercice 1: DTD interne

Exercice 2: DTD externe



## DTD vs schéma

- DTD peu précise pour la valeurs: on ne peut pas forcer une valeur numérique
- DTD peu sémantique: date? heure? = chiffres (et encore, ce n'est pas sûr...)
- contrôle limité sur les éléments: 0, 1 ou plusieurs occurrences
- DTD pas écrit en XML

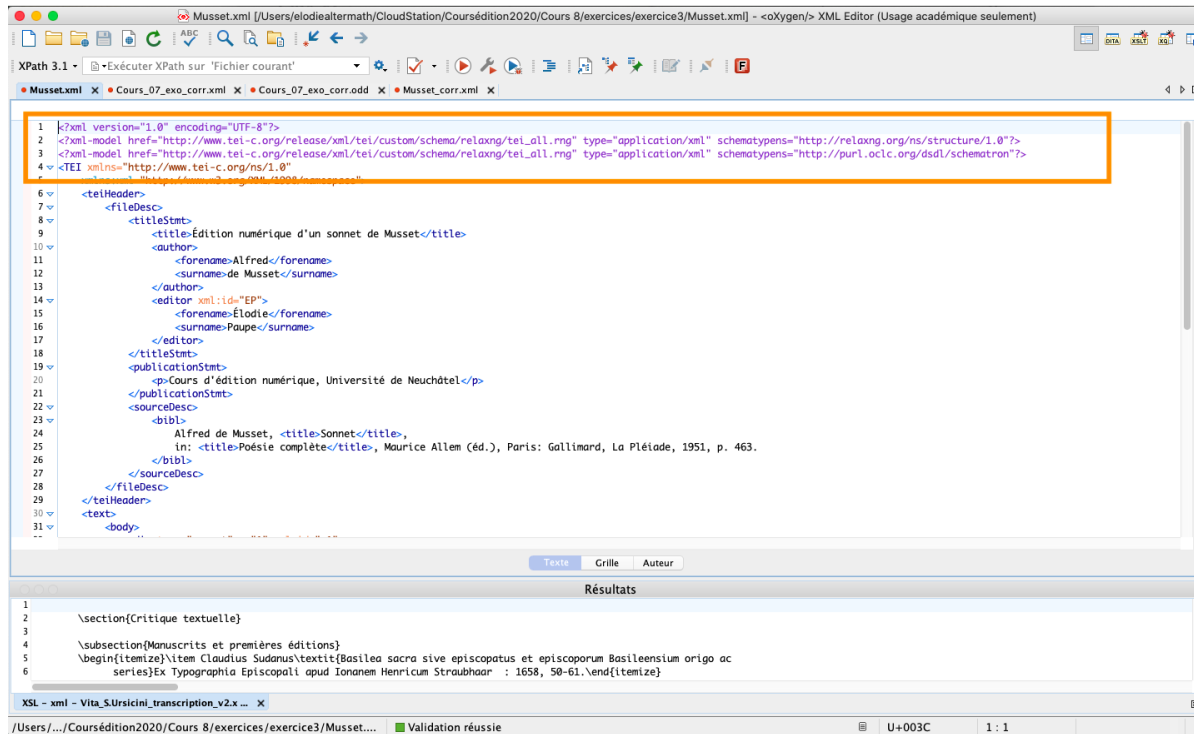
# Schéma

Le schéma est mentionné au début de chaque document TEI:

I. 1: la déclaration XML

I. 2: le lien vers le schéma

I. 3: le lien vers les règles schematron



The screenshot shows the oXygen XML Editor interface. The main window displays an XML document in TEI format. The first four lines of the document are highlighted with an orange box:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
3 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
4 <TEI xmlns="http://www.tei-c.org/ns/1.0">
```

The rest of the document shows the TEI header and body elements, including file description, title, author, editor, publication, and source information.

At the bottom of the editor, there is a 'Résultats' (Results) panel showing the output of an XSL transformation. The results are displayed in a structured format, including section and subsection information.

# RELAX NG

```
<app>
  <lem wit="#AS #P1">religionisque</lem>
  <rdg wit="#Sang #Krusch">relegionisque</rdg>
</app>
```

```
<define name="app">
  <element name="app">
    <element name="lem">
      <attribute name="wit"/>
    </element>
    <oneOrMore>
      <element name="rdg">
        <attribute name="wit"/>
      </element>
    </oneOrMore>
  </element>
</define>
```

Le schéma RELAX NG est écrit en XML.

# RELAX NG Compact

```
<app>  
  <lem wit="#AS #P1">religionisque</lem>  
  <rdg wit="#Sang #Krusch">relegionisque</rdg>  
</app>
```

```
app =  
  element app {  
    element lem {  
      attribute wit { text },  
    },  
    element rdg {  
      attribute wit { text },  
    }  
  }
```

## ***ODD (One Document Does it all)***

Écrire les schémas ci-dessus est complexe. L'ODD a une approche un peu différente qui simplifie la lecture et la maîtrise de la personnalisation de la TEI.

- C'est un fichier TEI.
- Il permet de générer un schéma *et* la documentation en HTML (et d'autres formats).

# Structure d'un *ODD*

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="fr">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title></title>
      </titleStmt>
      <publicationStmt><p></p></publicationStmt>
      <sourceDesc><p></p></sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <div>
        <!-- Où j'explique avec des mots les choix qui sont les miens.
        J'écris des phrases complètes.-->
      </div>
      <div>
        <specGrp/>
        <!-- Où j'annonce les classes d'attributs en XML. -->
      </div>
      <div>
        <schemaSpec/>
        <!-- Où j'annonce les modèles pour les éléments employés en XML. -->
      </div>
    </body>
  </text>
</TEI>
```

TEI	@xmlns	http://www.tei-c.org/ns/1.0					
	@xml:lang	en					
	teiHeader						
	text	body	p	This customization creates a TEI schema with the bare minimum of tags to make a recognizable document. It combines the four basic modules, but removes most of the available elements and changes several attribute classes.			
			div (4 rows)	head	p	specGrp	schemaSpec
				1 Basic text structure	p (6 rows)		
				2 Header structure	p (2 rows)		
				3 Class modifications	p (3 rows)	specGrp	
				4 Formal declaration			schemaSpec

# Modules et <schemaSpec>

- Les éléments nécessaires sont autorisés dans le <schemaSpec> (*schema specification*).
- Les éléments sont réunis en modules:

The following table lists the modules defined by the current release of these Guidelines:

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	<a href="#">17 Simple Analytic Mechanisms</a>
certainty	Certainty and Uncertainty	<a href="#">21 Certainty, Precision, and Responsibility</a>
core	Common Core	<a href="#">3 Elements Available in All TEI Documents</a>
corpus	Metadata for Language Corpora	<a href="#">15 Language Corpora</a>
dictionaries	Print Dictionaries	<a href="#">9 Dictionaries</a>
drama	Performance Texts	<a href="#">7 Performance Texts</a>
figures	Tables, Formulae, Figures	<a href="#">14 Tables, Formulae, Graphics and Notated Music</a>
gaiji	Character and Glyph Documentation	<a href="#">5 Characters, Glyphs, and Writing Modes</a>
header	Common Metadata	<a href="#">2 The TEI Header</a>
iso-fs	Feature Structures	<a href="#">18 Feature Structures</a>
linking	Linking, Segmentation, and Alignment	<a href="#">16 Linking, Segmentation, and Alignment</a>
msdescription	Manuscript Description	<a href="#">10 Manuscript Description</a>
namesdates	Names, Dates, People, and Places	<a href="#">13 Names, Dates, People, and Places</a>
nets	Graphs, Networks, and Trees	<a href="#">19 Graphs, Networks, and Trees</a>
spoken	Transcribed Speech	<a href="#">8 Transcriptions of Speech</a>
tagdocs	Documentation Elements	<a href="#">22 Documentation Elements</a>
tei	TEI Infrastructure	<a href="#">1 The TEI Infrastructure</a>
textcrit	Text Criticism	<a href="#">12 Critical Apparatus</a>
textstructure	Default Text Structure	<a href="#">4 Default Text Structure</a>
transcr	Transcription of Primary Sources	<a href="#">11 Representation of Primary Sources</a>
verse	Verse	<a href="#">6 Verse</a>





## P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.1.0. Last updated on 19th August 2020, revision b414ba550

### <sourceDesc>

[Home](#)  
[C Elements](#)

**<sourceDesc>** (source description) describes the source(s) from which an electronic text was derived or generated, typically a bibliographic description in the case of a digitized text, or a phrase such as "born digital" for a text which has no previous existence. [[2.2.7 The Source Description](#)]

<b>Module</b>	header — The TEI Header
<b>Attributes</b>	<a href="#">att.global</a> (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ( <a href="#">att.global.rendition</a> (@rend, @style, @rendition)) ( <a href="#">att.global.linking</a> (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ( <a href="#">att.global.analytic</a> (@ana)) ( <a href="#">att.global.facs</a> (@facs)) ( <a href="#">att.global.change</a> (@change)) ( <a href="#">att.global.responsibility</a> (@cert, @resp)) ( <a href="#">att.global.source</a> (@source)) <a href="#">att.declarable</a> (@default)
<b>Contained by</b>	<b>header:</b> <a href="#">biblFull</a> <a href="#">fileDesc</a>
<b>May contain</b>	<b>core:</b> <a href="#">bibl</a> <a href="#">biblStruct</a> <a href="#">list</a> <a href="#">listBibl</a> <a href="#">p</a> <b>figures:</b> <a href="#">table</a> <b>header:</b> <a href="#">biblFull</a> <b>linking:</b> <a href="#">ab</a> <b>msdescription:</b> <a href="#">msDesc</a> <b>namesdates:</b> <a href="#">listEvent</a> <a href="#">listNym</a> <a href="#">listObject</a> <a href="#">listOrg</a> <a href="#">listPerson</a> <a href="#">listPlace</a> <a href="#">listRelation</a> <b>spoken:</b> <a href="#">recordingStmt</a> <a href="#">scriptStmt</a> <b>textcrit:</b> <a href="#">listApp</a> <a href="#">listWit</a>

- Les modules `core` (éléments communs à tous les fichiers TEI) ou `tei` sont obligatoires.
- Une fois le module sélectionné ( `@key` ), on ajoute les éléments que l'on souhaite conserver ( `@include` ) ou exclure ( `@except` ):

```
<schemaSpec ident="monProjet" docLang="fr">
  <moduleRef key="header"
    include="teiHeader fileDesc titleStmt publicationStmt sourceDesc"/>
  <moduleRef key="textstructure"
    include="TEI text body div front back"/>
  <moduleRef key="namesdates"
    include="forename surname orgName persName"/>
</schemaSpec>
```

## <elementSpec>

Chaque élément peut être défini pour préciser quels éléments il peut/doit contenir, l'ordre éventuel d'apparition, les occurrences, etc.


```
<elementSpec ident="monElement" mode="change">
  <content>
    <elementRef key="element1"/>
    <elementRef key="element2" minOccurs="1" maxOccurs="unbounded"/>
    <sequence>
      <elementRef key="element3"/>
      <elementRef key="element4"/>
    </sequence>
  </content>
</elementSpec>
```

De même, on peut préciser les attributs associés à un élément et la liste des valeurs que peuvent prendre ceux-ci:

```
<elementSpec ident="monElement" mode="change">
  <attList>
    <attDef ident="att1" mode="delete"/>
    <attDef ident="att2" mode="change" usage="req">
      <datatype>
        <dataRef key="teidata.count"/>
      </datatype>
    </attDef>
    <attDef ident="att3" mode="change" usage="req">
      <valList mode="add" type="closed">
        <valItem ident="valeur1"/>
        <valItem ident="valeur2"/>
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

# Classes d'attributs et <specGrp>

Les attributs sont groupés en classes lorsqu'ils ont des comportements similaires:

 <Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News

P5 Guidelines — English

P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.1.0. Last updated on 19th August 2020, revision b414ba550

<rhyme>

Home  
C Elements

<rhyme> marks the rhyming part of a metrical line. [\[6.5 Rhyme\]](#)

Module	verse — Verse
Attributes	<a href="#">att.global</a> (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ( <a href="#">att.global.rendition</a> (@rend, @style, @rendition)) ( <a href="#">att.global.linking</a> (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ( <a href="#">att.global.analytic</a> (@ana)) ( <a href="#">att.global.facs</a> (@facs)) ( <a href="#">att.global.change</a> (@change)) ( <a href="#">att.global.responsibility</a> (@cert, @resp)) ( <a href="#">att.global.source</a> (@source)) <a href="#">att.typed</a> (@type, @subtype)
	<div><div>@label</div><div>provides a label (usually a single letter) to identify which part of a rhyme scheme this rhyming string instantiates.</div><div><div>Status</div><div>Recommended</div></div><div><div>Datatype</div><div><a href="#">teidata.word</a></div></div><div><div>Note</div><div>Within a particular scope, all <a href="#">rhyme</a> elements with the same value for their @label attribute are assumed to rhyme with each other. The scope is defined by the nearest ancestor element for which the @rhyme attribute has been supplied.</div></div></div>
Member of	<a href="#">model.IPart</a>

On peut modifier les classes d'attributs avec `<specGrp>` :

```
<specGrp xml:id="classmods">
  <classSpec ident="att.global" mode="change"
    module="tei" type="atts">
    <attList>
      <attDef ident="xml:space" mode="delete"/>
      <attDef ident="xml:base" mode="delete"/>
    </attList>
  </classSpec>
  <classSpec ident="att.divLike" mode="delete"
    module="tei" type="atts"/>
</specGrp>
```

## Créer un *ODD* sans encoder

Comme il serait laborieux de devoir écrire les schémas "à la main", il est possible d'utiliser les modèles annoncés au début de la séance. Les schémas peuvent être téléchargés sur [Roma](#).

Pour les projets plus spécifiques, la même plateforme permet de sélectionner les éléments nécessaires et génère automatique l'ODD et le schéma.

# Bibliographie

Burnard, Lou, *Qu'est-ce que la Text Encoding Initiative?* [en ligne], Marseille: OpenEdition Press, 2015. Disponible sur Internet : <http://books.openedition.org/oep/1298>. DOI : <https://doi.org/10.4000/books.oep.1237>.

Ateliers "Philologie numérique", donnés par Simon Gabay et Jean-Baptiste Camp, février 2018.