



UNIL | Université de Lausanne

Faculté des lettres

**Centre NUCLEUS**



Formation de deux jours

## **Initiation à l'encodage XML-TEI**

Lausanne, février 2025

Sonia Solfrini et Simon Gabay

(Université de Genève)



## Plan du cours 6 : lire un schéma TEI personnalisé

- Qu'est-ce qu'un schéma TEI personnalisé ?
- Quelques exemples
- Comprendre et lire un ODD
- ROMA : un éditeur ODD

# XML-TEI ? Révisions !

Les documents TEI reposent sur la syntaxe du langage XML. Un document TEI est donc dit bien formé s'il respecte la syntaxe de XML, avec des balises ouvrantes et fermantes correctement imbriquées.

Le document TEI est dit valide si les balises qu'il contient se conforment à un schéma qui, en plus des règles syntaxiques de XML, fournit des recommandations sur le vocabulaire à utiliser. Les *Guidelines* de la TEI fournissent le nom et la définition de centaines de balises, ainsi que des règles sur la façon dont elles peuvent être combinées.

**N.B. : La plupart des documents TEI n'a besoin que d'une petite partie de ce qui est fourni !**

N.B.B. : On n'invente pas de balises et elles sont en anglais !

**N.B. : La plupart des documents TEI n'a besoin que d'une petite partie de ce qui est fourni !**



**schéma TEI personnalisé (ODD = One Document Does it all)**

Vu que la TEI doit (théoriquement) apporter une solution à toutes les situations possibles, l'éventail est trop large pour un projet personnel, même ambitieux. Il est donc conseillé de restreindre le schéma TEI All en concevant un modèle adapté à ses besoins.

Il est fondamental de documenter ce modèle pour éviter de se tromper (nous-même et/ou au sein d'une équipe) et pour rendre ses choix compréhensibles et réexploitables par d'autres projets.

## Qu'est-ce qu'un schéma TEI personnalisé ?

```
<?xml version="1.0" encoding="UTF-8"?>
<?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"?>
<?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"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
```


```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://raw.githubusercontent.com/SETAFDH/TEI-SETAF/main/schema/odd-setaf.rng"
  type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="https://raw.githubusercontent.com/SETAFDH/TEI-SETAF/main/schema/odd-setaf.rng"
  type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
```

## Quelques exemples

- Un ODD en cours d'élaboration :

Le schéma TEI du projet FNS SETAF

Lien : <https://github.com/SETAFDH/TEI-SETAF/tree/main/schema>

 odd-setaf.html

 odd-setaf.odd

 odd-setaf.rnc

 odd-setaf.rng

## Quelques exemples

- Un ODD de référence pour les textes de la Renaissance :

Le Manuel d'encodage XML-TEI Renaissance et temps modernes (Imprimés - manuscrits)

Lien : <https://www.bvh.univ-tours.fr/XML-TEI/index.asp>



# Comprendre et lire un ODD

## TEI conformance :

1. Une personnalisation est une restriction de la TEI mais elle reste valide contre TEI All.
2. Elle ne doit pas altérer le sémantisme des éléments.

Si la personnalisation respecte ces deux règles, elle est dite “propre” (*clean modification*), sinon elle est dite “sale” (*unclean modification*). Une modification sale, cependant, peut parfois être nécessaire.



### **Changements propres :**

- \* Suppression d'éléments
- \* Modification d'une liste d'attributs ou de valeurs d'attributs

### **Changements sales :**

- \* Changement du nom d'éléments

### **Changements sales ou propres :**

- \* Ajout de nouveaux éléments (propre si ces nouveaux éléments appartiennent à un autre espace de nom)
- \* Ajout de nouveaux attributs (pas trop sale s'ils ne sont pas autorisés sur un élément mais existent en TEI)

```

<div rend="toc" xml:id="Specifications">
  <head>Specifications</head>
  <schemaSpec ident="myTEI">
    <!-- Checking module "core" -->
    <moduleRef key="core" include="addrLine [...] pubPlace [...] title"/>
    <!-- Checking element "pubPlace" -->
    <elementSpec ident="pubPlace" mode="change">
      <attList>
        <attDef ident="corresp" mode="delete"/>
        <attDef ident="cert" mode="change">
          <valList mode="add" type="closed">
            <valItem ident="high"/>
            <valItem ident="medium"/>
            <valItem ident="low"/>
            <valItem ident="unknown"/>
          </valList>
        </attDef>
      </attList>
    </elementSpec>
  </div>

```

<elementSpec>

# Le <teiHeader> : le <fileDesc> du projet SETAF

## Modèle de base :

```
<imprint>
  <pubPlace ref="geonames:x">Lieu de
publication</pubPlace>
  <respStmt>
    <resp>Imprimeur</resp>
    <persName ref="isni:x">
      <surname>Nom</surname>
      <forename>Prénom</forename>
    </persName>
  </respStmt>
  <date>Date de publication</date>
</imprint>
```

## Exemple :

```
<imprint>
  <pubPlace cert="medium"
ref="geonames:2659496">Neuchâtel</pubPlace>
  <respStmt>
    <resp>Imprimeur</resp>
    <persName role="presumed_printer"
ref="isni:000000007752492X">
      <surname>Vingle</surname>
      <forename>Pierre de</forename>
    </persName>
  </respStmt>
  <date cert="high">1533</date>
</imprint>
```

## <classSpec>

```
<!-- Checking attributes (https://tei-c.org/release/doc/tei-p5-doc/en/html/REF-CLASSES-ATTS.html) -->
  <!-- Checking attributes, letter a -->
    <classSpec ident="att.anchoring" type="atts" mode="delete"/>
    <classSpec ident="att.ascribed.directed" type="atts" mode="delete"/>
  <!-- Checking attributes, letter b -->
    <classSpec ident="att.breaking" type="atts" mode="delete"/>
  <!-- Checking attributes, letter c -->
    <classSpec ident="att.citeStructurePart" type="atts" mode="delete"/>
    <classSpec ident="att.combinable" type="atts" mode="delete"/>
    <classSpec ident="att.cReferencing" type="atts" mode="delete"/>
    <classSpec ident="att.citing" type="atts" mode="change">
      <attList>
        <attDef ident="from" mode="delete"/>
        <attDef ident="tog" mode="delete"/>
      </attList>
    </classSpec>
```

<https://roma.tei-c.org/>

SELECT ODD

UPLOAD ODD

## Select ODD

Choose a preset

TEI All (customize by reducing TEI) ▼

START

# Roma

Roma is an ODD Editor, using the TEI ODD (One Document Does-it-all) format for meta-schema documentation and local encoding guidelines as created by the [Text Encoding Initiative](https://textencodinginitiative.org/).

The old version of Roma can be found at <https://romaantiqua.tei-c.org>

Roma - ODD Customization (TEI with maximal setup)

SETTINGS

START OVER

LANGUAGE (EN)

DOWNLOAD

v 1.1.2

Elements

Attribute Classes

Model Classes

Datatypes

by module

filter it

<input checked="" type="checkbox"/>	<a href="#">ab</a>	(anonymous block) contains any component-level unit of text, acting as a container for phrase or inter level elements analogous to, but without the same constraints as, a paragraph.	Customization as ODD
<input checked="" type="checkbox"/>	<a href="#">abbr</a>	(abbreviation) contains an abbreviation of any sort.	Compiled ODD
<input checked="" type="checkbox"/>	<a href="#">abstract</a>	contains a summary or formal abstract prefixed to an existing source document by the encoder.	RELAX NG schema
<input checked="" type="checkbox"/>	<a href="#">accMat</a>	(accompanying material) contains details of any significant additional material which may be closely associated with the manuscript or object being described, such as non-contemporaneous ...	RELAX NG compact
<input checked="" type="checkbox"/>	<a href="#">acquisition</a>	(acquisition) contains any descriptive or other information concerning the process by which a manuscript or manuscript part or other object entered the holding institution.	W3C schema
<input checked="" type="checkbox"/>	<a href="#">actor</a>	contains the name of an actor appearing within a cast list.	DTD
<input checked="" type="checkbox"/>	<a href="#">add</a>	(addition) contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.	ISO Schematron constraints
<input checked="" type="checkbox"/>	<a href="#">additional</a>	(additional) groups additional information, combining bibliographic information about a manuscript or other object, or surrogate copies of it, with curatorial or administrative information.	Documentation as HTML
			Documentation as TEI Lite
			Documentation as MS Word
			Documentation as LaTeX
			Documentation as PDF