

Formation Edition numérique

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

Simon Gabay



**Personnalisier**

# Ma TEI

La TEI doit (théoriquement) apporter une solution à toutes les situations possibles: l'éventail est donc logiquement trop large pour un projet personnel, même ambitieux. Il faut donc

- Restreindre la TEI à ses besoins en concevant un modèle adapté à ses besoins
- Documenter ce modèle pour expliciter ses choix aux autres, afin de les rendre lisibles et réexploitables

# Des exemples de documentation

Les projets proposent leur documentation

- *Bibliothèques virtuelles humanistes*: <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>
- *Ecole des chartes*:  
<http://developpements.enc.sorbonne.fr/diple/schema/>
- *ENRICH*: <http://projects.oucs.ox.ac.uk/ENRICH/>

# Typologie des modification

- Suppression d'éléments
- Changement du nom d'éléments
- Modification d'un modèle de contenu
- Modification d'une liste d'attributs ou de valeurs d'attributs, d'un type de contenu
- Modification de l'appartenance à une classe
- Ajout de nouveaux éléments

# TEI conformance

Deux règles importantes :

1. Une personnalisation est une restriction de la TEI (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 peut cependant être nécessaire.

# Typologie des changements

Changements propres:

- Suppression d'éléments
- Modification d'un modèle de contenu
- Modification d'une liste d'attributs ou de valeurs d'attributs, d'un type de contenu
- Modification de l'appartenance à une classe

Changement sale:

- Changement du nom d'éléments

Changement sale ou propre:

- 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)

# Solutions simples

Les customisations:

- *All* : toute la TEI
- *Drama*: pour le théâtre
- *Corpus*: pour l'encodage de corpus linguistique
- *Lite*: version réduite de la TEI
- *Bare*: version minimale de la TEI

La personnalisation

- *oddByExample*
- *Roma*

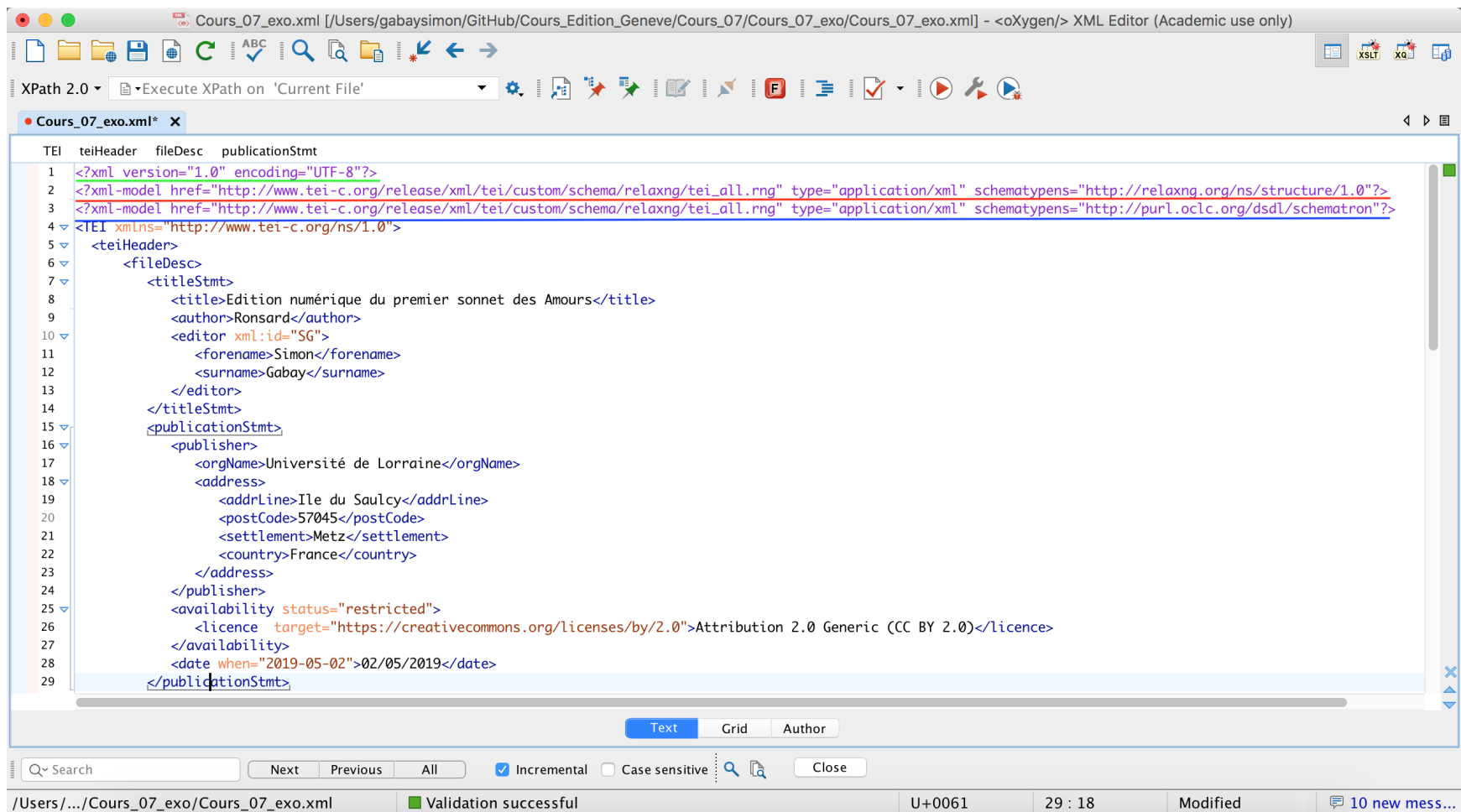


**Un Schema**

# Construction

Les fichiers TEI commencent par trois informations:

- La déclaration XML (Vert)
- Le lien vers les schema (rouge) et les règles schématron (bleu)



```
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">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title>Edition numérique du premier sonnet des Amours</title>
9         <author>Ronsard</author>
10        <editor xml:id="SG">
11          <forename>Simon</forename>
12          <surname>Gabay</surname>
13        </editor>
14      </titleStmt>
15      <publicationStmt>
16        <publisher>
17          <orgName>Université de Lorraine</orgName>
18          <address>
19            <addrLine>Ile du Saulcy</addrLine>
20            <postCode>57045</postCode>
21            <settlement>Metz</settlement>
22            <country>France</country>
23          </address>
24        </publisher>
25        <availability status="restricted">
26          <licence target="https://creativecommons.org/licenses/by/2.0">Attribution 2.0 Generic (CC BY 2.0)</licence>
27        </availability>
28        <date when="2019-05-02">02/05/2019</date>
29      </publicationStmt>
```

## Un nouveau schéma

Notre nouveau schéma est une restriction de la TEI qui reste valide contre *TEI All*. Il existe plusieurs formats:

- XML Schema ( `.xsd` ) [[cf. wikipedia](#)]
- RELAX NG XML ( `.rng` ) [[cf. wikipedia](#)]
- RELAX NG Compact ( `.rnc` )

Le dernier est le plus populaire parmi les utilisateurs de la TEI.

# RELAX NG

*Relax NG* est écrit en XML et est donc à peu près lisible. La règle autorisant cet encodage:

```
<lg>  
  <l>C'est un trou de verdure où chante une rivière</l>  
  <l>Accrochant follement aux herbes des haillons</l>  
</lg>
```

est la suivante (dans une version très minimale)

```
<element name="lg">  
  <oneOrMore>  
    <element name="l">  
      <text/>  
    </element>  
  </oneOrMore>  
</element>
```

# RELAX NG Compact

C'est une version réduite de *RELAX NG* qui n'est pas écrite en XML.

Pour le même texte

```
<lg>
  <l>C'est un trou de verdure où chante une rivière</l>
  <l>Accrochant follement aux herbes des haillons</l>
</lg>
```

La règle est désormais la suivante:

```
element book {
  element page { text }+
}
```

Il est très simple de passer de la version compact à la version XML de *RELAX NG* et inversement.

**ODD**

# Pourquoi l'ODD?

Les règles des schémas sont complexes à écrire et nécessitent une très bonne maîtrise du code. L'ODD sert à contourner ce problème.

- L'*ODD* est un fichier TEI normal, constitué d'un `<teiHeader>` et d'un `<body>`. Si je maîtrise déjà la TEI, je comprends facilement son fonctionnement.
- L'*ODD* permet de générer le schéma, mais aussi la documentation du projet en HTML
- L'*ODD* est accompagné d'une gamme d'outil qui permettent d'effectuer les transformations vers tous les formats nécessaires.

## Un double `<body>`

S'il n'y a bien qu'un seul `<body>` dans l'*ODD*, on peut distinguer deux temps:

1. Une série de `<div>` où j'explique mes choix en rédigeant du texte.
2. Une partie plus technique et plus rigide où je précise la constitution de mes classes d'attributs ( `<specGrp>` ) et surtout des modules pour les éléments ( `<schemaSpec>` ).

```
<body>
  <div>
    <p>J'ai fait comme-ci, comme-ça.</p>
  </div>
  <div>
    <specGrp/>
  </div>
  <div>
    <schemaSpec/>
  </div>
</body>
```



# Les modules

Concentrons-nous sur `schemaSpec` : il s'agit d'autoriser des éléments: comment les sélectionner? La TEI a en fait regroupés chaque élément dans une module. Tous sont amplement commentés dans les *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>

On retrouve le nom du module dans les *guidelines* (et, surtout, après quelques années on les connaît par cœur).

**<lg>**

P5: Recommandations pour l'encodage et l'échange de textes électroniques

Version 4.0.0. Last updated on 13th February 2020, revision ccd19b0ba

---

**<lg>** (groupe de vers) contient un groupe de vers fonctionnant comme une unité formelle, par exemple une strophe, un refrain, un paragraphe en vers, etc. [3.12.1 Core Tags for Verse 3.12 Passages of Verse or Drama 7.2.5 Speech Contents]

Module	core — Elements Available in All TEI Documents
Attributs	<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.divLike</a> (@org, @sample) ( <a href="#">att.metrical</a> (@met, @real, @rhyme)) ( <a href="#">att.fragmentable</a> (@part)) <a href="#">att.typed</a> (@type, @subtype) <a href="#">att.declaring</a> (@decls)
Membre du	<a href="#">macro.paraContent</a> <a href="#">model.divPart</a>
Contenu dans	<b>core:</b> <a href="#">add</a> <a href="#">corr</a> <a href="#">del</a> <a href="#">emph</a> <a href="#">head</a> <a href="#">hi</a> <a href="#">item</a> <a href="#">lg</a> <a href="#">note</a> <a href="#">orig</a> <a href="#">p</a> <a href="#">q</a> <a href="#">quote</a> <a href="#">ref</a> <a href="#">reg</a> <a href="#">said</a> <a href="#">sic</a> <a href="#">sp</a> <a href="#">stage</a> <a href="#">title</a> <a href="#">unclear</a> <b>dictionaries:</b> <a href="#">case</a> <a href="#">colloc</a> <a href="#">def</a> <a href="#">gen</a> <a href="#">gram</a> <a href="#">hyph</a> <a href="#">iType</a> <a href="#">lang</a> <a href="#">lbi</a> <a href="#">mood</a> <a href="#">number</a> <a href="#">orth</a> <a href="#">per</a> <a href="#">pos</a> <a href="#">pron</a> <a href="#">stress</a> <a href="#">subc</a> <a href="#">syll</a> <a href="#">tns</a> <a href="#">usg</a> <b>drama:</b> <a href="#">camera</a> <a href="#">caption</a> <a href="#">castList</a> <a href="#">epilogue</a> <a href="#">performance</a> <a href="#">prologue</a> <a href="#">set</a> <a href="#">sound</a> <a href="#">tech</a> <a href="#">view</a> <b>figures:</b> <a href="#">cell</a> <a href="#">figure</a>

## moduleRef

Pour autoriser un élément, je dois ajouter son module d'abord, puis les éléments du modules que je choisis de conserver (je choisir la logique inverse, et en exclure)

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

Certains modules sont obligatoires, comme `core` ou `tei`.

## elementSpec : les éléments contenus

Je peux ensuite préciser le comportement de mon élément avec `elementSpec`. Je peux préciser les autres éléments qu'il peut contenir

```
<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>
```

## elementSpec : les attributs contenus

Je peux ensuite préciser le comportement de mon élément avec `elementSpec`. Je peux préciser les autres éléments qu'il peut contenir

```
<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>
```

# Les classes

En plus des modules, les éléments sont regroupés par classes. Tous les éléments d'une même classe ont (plus ou moins) le même comportement:

- Ils apparaissent au même endroit
- Ont les mêmes attributs
- ...

On peut les utiliser pour changer tous les éléments d'une classe d'un coup, notamment pour restreindre la liste d'attributs.

On retrouve le nom des classes d'attributs dans les *guidelines* (et, surtout, après quelques années on les connaît par cœur).

**<lg>**

(groupe de vers) contient un groupe de vers fonctionnant comme une unité formelle, par exemple une strophe, un refrain, un paragraphe en vers, etc. [3.12.1 Core Tags for Verse 3.12 Passages of Verse or Drama 7.2.5 Speech Contents]

<b>Module</b>	core — Elements Available in All TEI Documents
<b>Attributs</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.divLike</a> (@org, @sample) ( <a href="#">att.metrical</a> (@met, @real, @rhyme)) ( <a href="#">att.fragmentable</a> (@part)) <a href="#">att.typed</a> (@type, @subtype) <a href="#">att.declaring</a> (@decls)
<b>Membre du</b>	<a href="#">macro paraContent</a> <a href="#">model divPart</a>
<b>Contenu dans</b>	<b>core:</b> <a href="#">add</a> <a href="#">corr</a> <a href="#">del</a> <a href="#">emph</a> <a href="#">head</a> <a href="#">hi</a> <a href="#">item</a> <a href="#">lg</a> <a href="#">note</a> <a href="#">orig</a> <a href="#">p</a> <a href="#">q</a> <a href="#">quote</a> <a href="#">ref</a> <a href="#">reg</a> <a href="#">said</a> <a href="#">sic</a> <a href="#">sp</a> <a href="#">stage</a> <a href="#">title</a> <a href="#">unclear</a> <b>dictionaries:</b> <a href="#">case</a> <a href="#">colloc</a> <a href="#">def</a> <a href="#">gen</a> <a href="#">gram</a> <a href="#">hyph</a> <a href="#">iType</a> <a href="#">lang</a> <a href="#">lbi</a> <a href="#">mood</a> <a href="#">number</a> <a href="#">orth</a> <a href="#">per</a> <a href="#">pos</a> <a href="#">pron</a> <a href="#">stress</a> <a href="#">subc</a> <a href="#">syll</a> <a href="#">tns</a> <a href="#">usg</a> <b>drama:</b> <a href="#">camera</a> <a href="#">caption</a> <a href="#">castList</a> <a href="#">epilogue</a> <a href="#">performance</a> <a href="#">prologue</a> <a href="#">set</a> <a href="#">sound</a> <a href="#">tech</a> <a href="#">view</a> <b>figures:</b> <a href="#">cell</a> <a href="#">figure</a>

## specGrp

Je vais pouvoir modifier les classes 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>
```