



HAL
open science

XML et espaces de nom

Morgane Pica

► **To cite this version:**

Morgane Pica. XML et espaces de nom. Doctorat. XML et espaces de nom, Bibliothèque du lab. CRISCO EA 4255, France. 2021, pp.44. halshs-03637189

HAL Id: halshs-03637189

<https://shs.hal.science/halshs-03637189>

Submitted on 11 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

XML et espaces de nom

Morgane Pica

Ingénieure d'Études du projet ConDÉ
morgane.pica@unicaen.fr

18/11/2021



XML et espaces de nom

I. XML et espaces de nom

1. *Rappels sur le XML*
2. *Qu'est-ce qu'un espace de nom ?*

II. Schémas de validation

1. *Qu'est-ce qu'un schéma de validation ?*
2. *Les formats de schémas*
3. *Relier un document à son/ses NS*

III. L'espace de nom TEI

1. *Définition*
2. *Structure générale d'un document*
3. *Les métadonnées*
4. *Le texte*
5. *Les informations linguistiques*
6. *Autres informations pertinentes*
7. *Lire la documentation*

I.

XML et espaces de nom

XML et espaces de nom

I. XML et espaces de nom — 1. Rappels sur le XML

Principes généraux du XML

- *eXtensible Markup Language*
- Langage strictement hiérarchique (chaque élément est contenu par d'autres éléments et peut en contenir d'autre, pas de chevauchement d'éléments), modélisable comme arborescence.
- Langage à balises (les débuts et fins d'éléments sont respectivement marqués par des balises ouvrantes et fermantes).
- Permet de stocker les différentes chaînes de caractères formant un fichier dans des éléments qui définissent leur nature.
- Permet donc de récupérer les chaînes de caractères d'une certaine nature.
- Permet un stockage des métadonnées en interne (dans des éléments dédiés uniquement aux m.d.).

XML et espaces de nom

I. XML et espaces de nom — 1. Rappels sur le XML

Vocabulaire

```
<livre>
  <partie numero="1" >
    <chapitre numero="1" >
      <texte>
        Comme texte il y aurait des phrases.
      </texte>
    </chapitre>
    <chapitre numero="2" >
      <texte>
        Mais cela prendrait trop de place.
      </texte>
    </chapitre>
    <chapitre numero="3" >
      <texte>
        Alors je n'en ai pas inventé.
      </texte>
    </chapitre>
  </partie>
  <partie number="2"/>
</livre>
```

L'élément qui contient tous les autres est appelé la racine.

Vocabulaire des relations familiales pour décrire les relations entre éléments :

- ancêtre
- parent
- frère
- enfant
- descendant

XML et espaces de nom

I. XML et espaces de nom — 1. Rappels sur le XML

Court exemple

```
<livre>
  <partie numero="1" >
    <chapitre numero="1" >
      <texte>
        Comme texte il y aurait des phrases.
      </texte>
    </chapitre>
    <chapitre numero="2" >
      <texte>
        Mais cela prendrait trop de place.
      </texte>
    </chapitre>
    <chapitre numero="3" >
      <texte>
        Alors je n'en ai pas inventé.
      </texte>
    </chapitre>
  </partie>
</partie numero="2"/>
</livre>
```

- Racine : <livre>

- <texte> est toujours enfant de <chapitre>.

- <chapitre> peut avoir des frères <chapitre>

- <livre> est un ancêtre de tous les autres élément.

XML et espaces de nom

I. XML et espaces de nom — 1. Rappels sur le XML

Un document "bien formé"

Un document bien formé suit la syntaxe du XML, c'est-à-dire :

1) HIÉRARCHIE STRICTE

= Pas de chevauchement d'éléments : Toujours fermer un élément avant d'ouvrir un autre élément du même niveau.

MAL FORMÉ :

<paragraphe>

<titre>Voici un <phrase>titre.</titre>Puis une autre.</phrase>

</paragraphe>

BIEN FORMÉ :

<paragraphe>

<titre>Voici un titre.</titre><phrase>Puis une autre.</phrase>

XML et espaces de nom

I. XML et espaces de nom — 1. Rappels sur le XML

Un document "bien formé"

Un document bien formé suit la syntaxe du XML, c'est-à-dire :

2) PAS D'ÉLÉMENT NON REFERMÉ

= Fermez toujours les éléments que vous ouvrez.
Si l'élément est vide, utilisez une balise autofermante.

MAL FORMÉ :

<paragraphe>

<titre>Voici un titre.<phrase>Puis une phrase.</phrase>

</paragraphe>

BIEN FORMÉ :

<paragraphe>

<titre>Voici un titre.</titre><phrase>Puis une phrase.</phrase>

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Pourquoi un espace de nom ?

XML ≠ format

- › XML = uniquement syntaxe.
- › Nécessite de définir un vocabulaire et une grammaire spécifiques en plus.
- › Ces vocabulaires/grammaires sont les fameux « espaces de noms ».

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Définition

- Pour pouvoir s'échanger des fichiers compréhensibles les uns par les autres, il ne s'agit pas uniquement de syntaxe mais aussi des éléments et attributs utilisés : lesquels, où, quand et en quel nombre.
- Pour se comprendre et favoriser la pérennité de son travail ainsi que son interopérabilité, on ne réinvente pas chacun son schéma de fichier chacun de son côté.
- Plutôt, on utilise des standards internationaux correspondant à la fois à son champ d'étude, aux objets qu'on souhaite décrire et à ce qu'on souhaite en faire.
- Ces standards s'appellent des **espaces de nom** ("*namespace*" en anglais). Ce sont des vocabulaires précis définis pour un usage précis.

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Interopérabilité ?

- Compatibilité avec d'autres données.
- Données compréhensibles par d'autres personnes et d'autres systèmes informatiques.
- Interrogeable de concert avec d'autres données du même format.
- Données plus pérennes puisque largement comprises et partagées.
- (Attention, la pérennité absolue n'existe pas, on ne peut qu'y tendre.)

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Interopérabilité ?

Ce qui implique :

- de s'assurer qu'il n'existe pas un standard correspondant déjà à nos données,
- de penser aux autres utilisateurs potentiels lorsqu'on met en forme des données.

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Interopérabilité ?

Ce qui implique :

- de choisir des formats standards, bien installés dans la communauté scientifique, et les plus largement compris pour garantir le partage des données,
- de choisir des formats non propriétaires, pour éviter que les logiciels/systèmes associés soient abandonnés et qu'on ne sache plus les lire,
- d'être très attentifs à la documentation officielle d'un format choisi,
- de solliciter la communauté via des lieux d'échange en cas de doute (forums ou listes de diffusions).

XML et espaces de nom

I. XML et espaces de nom — 2. Qu'est-ce qu'un espace de nom ?

Définition

Un espace de nom définit donc :

- › le nom des éléments et attributs,
- › les balises de chaque élément,
- › à quoi ils servent,
- › dans quel élément on peut les placer,
- › pour les éléments, combien on peut en placer dans leur élément parent,
- › ce qu'ils peuvent contenir (éléments, texte, nombres, identifiant, valeurs obligatoires...).

Cette définition se fait dans un schéma de validation.

II.

Schémas de validation

II. Schémas et validation — 1. Qu'est-ce qu'un schéma ?

Définition

- Vous savez à quoi sert un schéma, maintenant voyons ce que c'est.
- Un schéma est d'abord une suite d'informations dans un format numérique dédié.
- Elles peuvent être incluses dans le XML qu'elles documentent.
- La plupart du temps, elles sont dans un fichier à part.

XML et espaces de nom

II. Schémas et validation — 1. Qu'est-ce qu'un schéma ?

Exemple

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!DOCTYPE liste_de_gens [
  <!ELEMENT liste_de_gens (personne)*>
  <!ELEMENT personne (nom, date_de_naissance?, genre?, numero_de_secu?)>
  <!ELEMENT nom (#PCDATA)>
  <!ELEMENT date_de_naissance (#PCDATA)>
  <!ELEMENT genre (#PCDATA | masculin | féminin) "féminin">
  <!ELEMENT numero_de_secu (#PCDATA)>
]>
<liste_de_gens>
  <personne>
    <nom>Fred Bloggs</nom>
    <date_de_naissance>2008-11-27</date_de_naissance>
    <genre>masculin</genre>
  </personne>
</liste_de_gens>
```

XML et espaces de nom

II. Schémas et validation — 1. Qu'est-ce qu'un schéma ?

Exemple

**Déclaration
XML :**

Schéma DTD :

**Exemple
d'application
du schéma :**

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!DOCTYPE liste_de_gens [
  <!ELEMENT liste_de_gens (personne)*>
  <!ELEMENT personne (nom, date_de_naissance?, genre?, numero_de_secu?)>
  <!ELEMENT nom (#PCDATA)>
  <!ELEMENT date_de_naissance (#PCDATA)>
  <!ELEMENT genre (#PCDATA | masculin | féminin) "féminin">
  <!ELEMENT numero_de_secu (#PCDATA)>
]>
<liste_de_gens>
  <personne>
    <nom>Fred Bloggs</nom>
    <date_de_naissance>2008-11-27</date_de_naissance>
    <genre>masculin</genre>
  </personne>
</liste_de_gens>
```

Les formats

Lorsqu'on a un type de fichier complexe ou long à décrire, ou lorsqu'on définit un schéma applicable à plusieurs documents, il vaut mieux écrire un fichier externe. Ce fichier peut être écrit en plusieurs formats :

- **RELAXNG** (<https://relaxng.org/>)
- **ODD** (<https://wiki.tei-c.org/index.php/ODD>)
- **XML Schema** (<https://www.w3.org/XML/Schema>)
→ = « XSD »
- **Doctype** (<https://www.w3.org/QA/2002/04/valid-dtd-list.html>)

Les formats

Lorsqu'on a un type de fichier complexe ou long à décrire, ou lorsqu'on définit un schéma applicable à plusieurs documents, il vaut mieux écrire un fichier externe. Ce fichier peut être écrit en plusieurs formats :

- **RELAXNG** (<https://relaxng.org/>)
- **ODD** (<https://wiki.tei-c.org/index.php/ODD>)
- **XML Schema** (<https://www.w3.org/XML/Schema>)
→ = « XSD »
- **Doctype** (<https://www.w3.org/QA/2002/04/valid-dtd-list.html>)

Lorsqu'il est externe au fichier, on le renseigne sous forme de lien (local ou URL) dans le document.

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Les déclarations initiales

- Tout document XML doit comporter une déclaration XML initiale, avant la balise ouvrante de l'élément racine.

```
<?xml version='1.0' encoding='UTF-8'?>
```

- La déclaration XML peut être accompagnée de déclarations de schémas.

```
<?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"?>
```

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Les balises d'éléments

- Chaque élément XML appartient à un espace de nom. On peut les déclarer dans les balises. Ex :

```
<TEI  
  xmlns="http://www.tei-c.org/ns/1.0"  
  version="5"/>
```

- Attention : même si l'espace de nom se déclare comme un attribut, ce n'est pas un attribut en soi. Ce n'est pas un resserrage de la nature de l'élément. C'est une déclaration du système dont dépend l'élément.
- XMLNS = "*XML Namespace*"

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Les balises d'éléments

- Cette déclaration dans l'élément TEI est celle faite dans la racine d'un document TEI, ce qui signifie qu'il s'agit de l'espace de nom par défaut du document :

```
<TEI  
  xmlns="http://www.tei-c.org/ns/1.0"  
  version="5"/>
```

- Par défaut, les éléments descendant de /TEI héritent de son XMLNS.
- Un autre indice du fait que cet espace de nom est le XMLNS par défaut du document est qu'on ne lui adjoint aucune abréviation.

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Les balises d'éléments

- Lorsqu'on utilise un XMLNS qui n'est pas par défaut dans le document, on lui adjoint un préfixe. Par exemple, ici, où en plus du XMLNS TEI par défaut, on ajoute un XMLNS TXM :

```
<TEI
  xmlns="http://www.tei-c.org/ns/1.0"
  xmlns:txm="http://textometrie.org/1.0"
  version="5"/>
```

- Par défaut, les éléments descendant de /TEI héritent toujours de son XMLNS.
- Sauf si on ajoute un préfixe déclaré, par exemple :

```
<w>
  <txm:form>LA</txm:form>
  <txm:ana type="token-number">1</txm:ana>
  <txm:ana type="#lemma">LE</txm:ana>
  <txm:ana type="#pos">Da|Pp</txm:ana>
</w>
```

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Exemple complet

```
<?xml version="1.1" 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://purl.oclc.org/dsdl/schematron"?>

<TEI
  xmlns="http://www.tei-c.org/ns/1.0"
  xmlns:txm="http://textometrie.org/1.0"
  version="5">

[Corps du fichier]

</TEI>
```

Validation

- La déclaration des XMLNS permet de référencer son schéma.
- Elle permet également de demander à la machine de vérifier que notre fichier est conforme au schéma et nous orienter vers les erreurs.
- Attention, la machine nous montre toujours où l'erreur se manifeste, c'est-à-dire toujours après l'erreur elle-même.

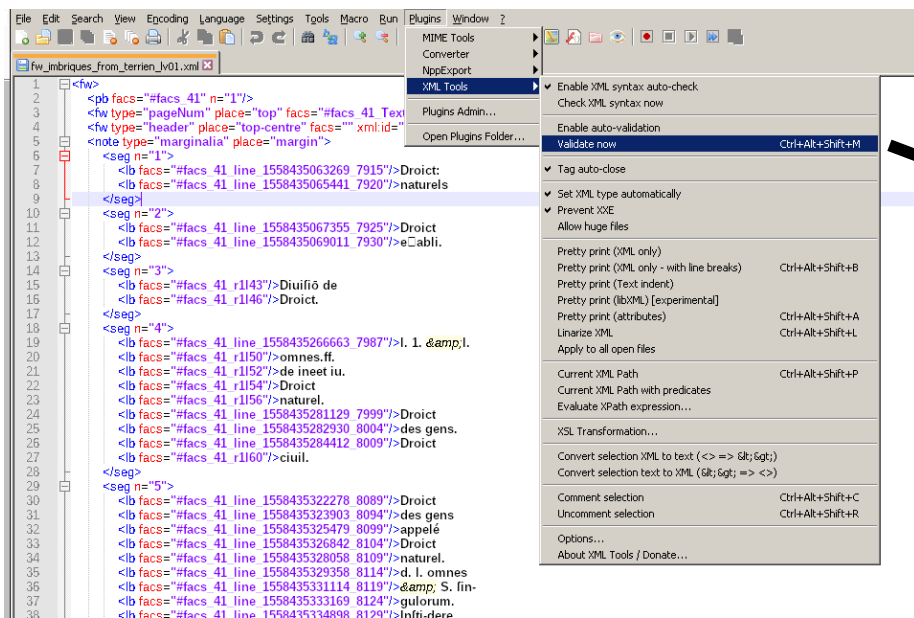
Validation

- La déclaration des XMLNS permet de référencer son schéma.
- Elle permet également de demander à la machine de vérifier que notre fichier est conforme au schéma et nous orienter vers les erreurs.
- Attention, la machine nous montre toujours où l'erreur se manifeste, c'est-à-dire toujours après l'erreur elle-même.

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

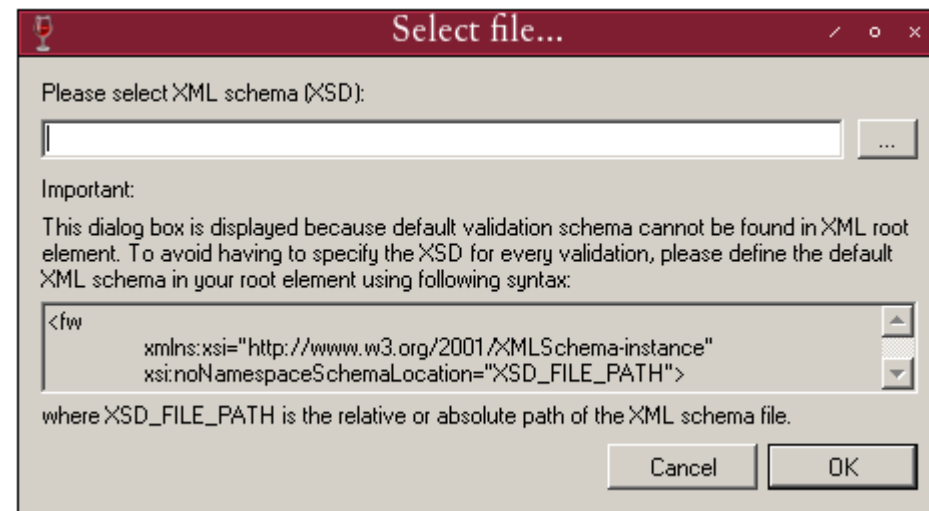
Validation



The screenshot shows the Notepad++ interface with an XML file open. The XML code is as follows:

```
1 <fw>
2 <pb facs="#facs 41" n="1"/>
3 <fw type="pageNum" place="top" facs="#facs 41 Text" />
4 <fw type="header" place="top-centre" facs="" xml:id="
5 <note type="marginalia" place="margin">
6 <seg n="1">
7 <lb facs="#facs 41 line 1558435063269 7915"/>Droict;
8 <lb facs="#facs 41 line 1558435065441 7920"/>naturels
9 </seg>
10 <seg n="2">
11 <lb facs="#facs 41 line 1558435067355 7925"/>Droict
12 <lb facs="#facs 41 line 1558435069011 7930"/>e▯abli.
13 </seg>
14 <seg n="3">
15 <lb facs="#facs 41 r1143"/>Diuifô de
16 <lb facs="#facs 41 r1146"/>Droict.
17 </seg>
18 <seg n="4">
19 <lb facs="#facs 41 line 1558435266663 7987"/>I. I. &amp;l.
20 <lb facs="#facs 41 r1150"/>omnes.ft.
21 <lb facs="#facs 41 r1152"/>de ineet iu.
22 <lb facs="#facs 41 r1154"/>Droict
23 <lb facs="#facs 41 r1156"/>naturel.
24 <lb facs="#facs 41 line 1558435281129 7999"/>Droict
25 <lb facs="#facs 41 line 1558435282930 8004"/>des gens.
26 <lb facs="#facs 41 line 1558435284412 8009"/>Droict
27 <lb facs="#facs 41 r1160"/>ciuil.
28 </seg>
29 <seg n="5">
30 <lb facs="#facs 41 line 1558435322278 8089"/>Droict
31 <lb facs="#facs 41 line 1558435323903 8094"/>des gens
32 <lb facs="#facs 41 line 1558435325479 8099"/>appelé
33 <lb facs="#facs 41 line 1558435326842 8104"/>Droict
34 <lb facs="#facs 41 line 1558435328058 8109"/>naturel.
35 <lb facs="#facs 41 line 1558435329358 8114"/>d. l. omnes
36 <lb facs="#facs 41 line 1558435331114 8119"/>&amp; S. fin-
37 <lb facs="#facs 41 line 1558435333169 8124"/>gulforum.
38 <lb facs="#facs 41 line 1558435334898 8129"/>Infri. dera
```

The XML Tools menu is open, showing options like 'Enable XML syntax auto-check', 'Validate now', and 'Pretty print (XML only)'. An arrow points from the 'Validate now' option to a dialog box.

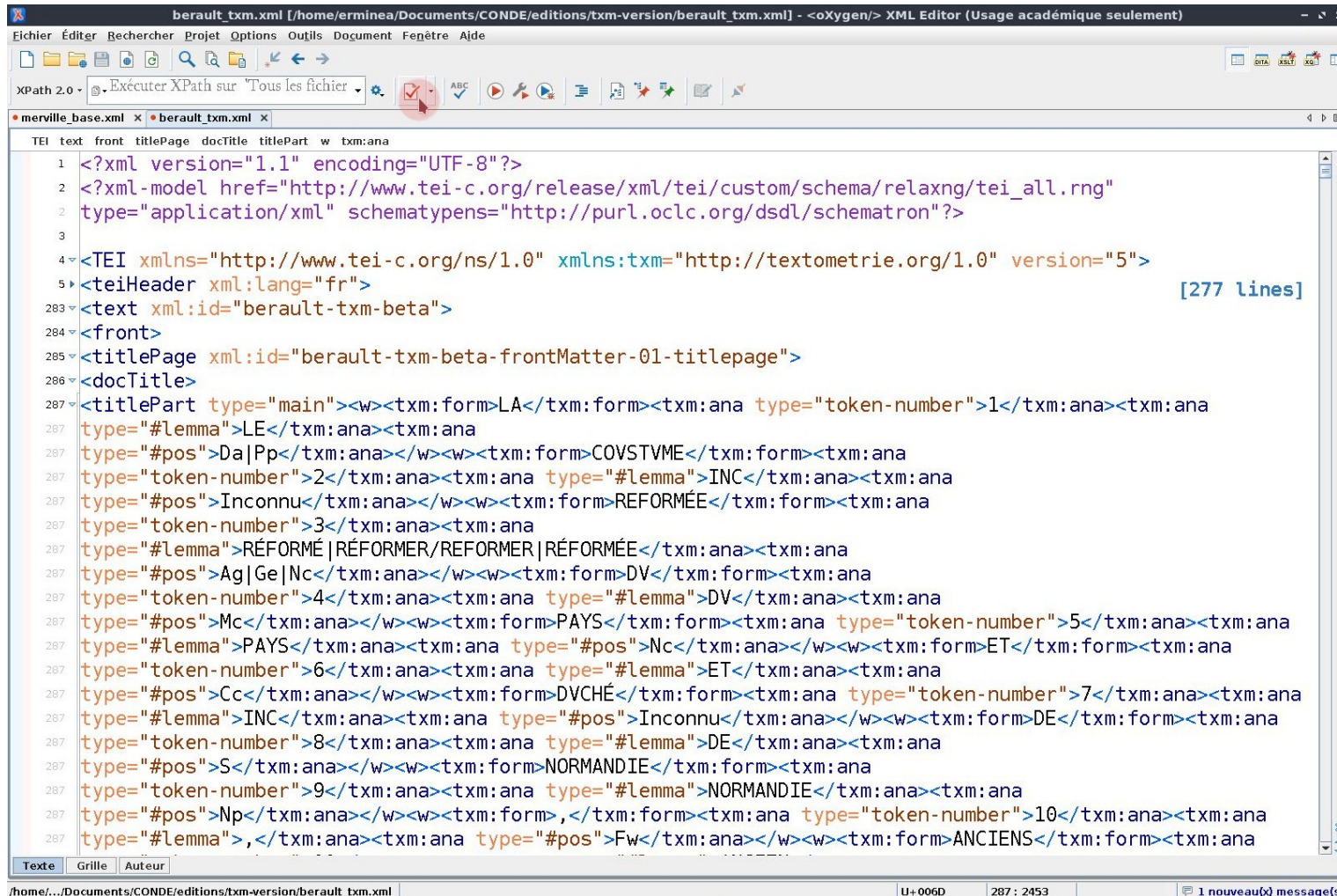


Sur Notepad++.

XML et espaces de nom

II. Schémas et validation — 3. Relier XML et NS

Validation



The screenshot shows the Oxygen XML Editor interface. The title bar indicates the file is 'berault_txm.xml' located at '/home/erminea/Documents/CONDE/editions/txm-version/berault_txm.xml'. The menu bar includes 'Fichier', 'Éditer', 'Rechercher', 'Projet', 'Options', 'Outils', 'Document', and 'Fenêtre Aide'. The toolbar contains various icons for file operations and editing. The main editor area displays XML code for a TEI document. The code starts with a declaration of XML version and encoding, followed by a model href and schema type. The root element is a TEI document with namespaces for TEI and txm. The document contains a title page with a title part that lists several words and their grammatical forms, such as 'LA', 'LE', 'DA', 'INC', 'INCONNU', 'REFORMÉE', 'REFORMÉ', 'REFORMER', 'REFORMÉE', 'DV', 'DVCHÉ', 'DE', 'S', 'NORMANDIE', 'NORMANDIE', 'NP', and 'Fw'. The code is color-coded, with tags in blue and text in black. A status bar at the bottom shows the current line (287) and column (2453), and a message icon indicating '1 nouveau message(s)'. The bottom of the window shows the file path and a status bar with 'U+006D', '287 : 2453', and '1 nouveau message(s)'.

```
<?xml version="1.1" 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://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:txm="http://textometrie.org/1.0" version="5">
<teiHeader xml:lang="fr">
<text xml:id="berault-txm-beta">
<front>
<titlePage xml:id="berault-txm-beta-frontMatter-01-titlepage">
<docTitle>
<titlePart type="main"><w><txm:form>LA</txm:form><txm:ana type="token-number">1</txm:ana><txm:ana
type="#lemma">LE</txm:ana><txm:ana
type="#pos">Da|P</txm:ana></w><w><txm:form>COVSTVME</txm:form><txm:ana
type="token-number">2</txm:ana><txm:ana type="#lemma">INC</txm:ana><txm:ana
type="#pos">Inconnu</txm:ana></w><w><txm:form>REFORMÉE</txm:form><txm:ana
type="token-number">3</txm:ana><txm:ana
type="#lemma">RÉFORMÉ|RÉFORMER/REFORMER|RÉFORMÉE</txm:ana><txm:ana
type="#pos">Ag|Ge|Nc</txm:ana></w><w><txm:form>DV</txm:form><txm:ana
type="token-number">4</txm:ana><txm:ana type="#lemma">DV</txm:ana><txm:ana
type="#pos">Mc</txm:ana></w><w><txm:form>PAYS</txm:form><txm:ana type="token-number">5</txm:ana><txm:ana
type="#lemma">PAYS</txm:ana><txm:ana type="#pos">Nc</txm:ana></w><w><txm:form>ET</txm:form><txm:ana
type="token-number">6</txm:ana><txm:ana type="#lemma">ET</txm:ana><txm:ana
type="#pos">Cc</txm:ana></w><w><txm:form>DVCHÉ</txm:form><txm:ana type="token-number">7</txm:ana><txm:ana
type="#lemma">INC</txm:ana><txm:ana type="#pos">Inconnu</txm:ana></w><w><txm:form>DE</txm:form><txm:ana
type="token-number">8</txm:ana><txm:ana type="#lemma">DE</txm:ana><txm:ana
type="#pos">S</txm:ana></w><w><txm:form>NORMANDIE</txm:form><txm:ana
type="token-number">9</txm:ana><txm:ana type="#lemma">NORMANDIE</txm:ana><txm:ana
type="#pos">Np</txm:ana></w><w><txm:form>,</txm:form><txm:ana type="token-number">10</txm:ana><txm:ana
type="#lemma">,</txm:ana><txm:ana type="#pos">Fw</txm:ana></w><w><txm:form>ANCIENS</txm:form><txm:ana
```

Sur Oxygen.

III.

L'espace de nom XML-TEI

XML et espaces de nom

III. XML-TEI — 1. Définition

Définition

<https://tei-c.org>

« The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. In addition to the Guidelines themselves, the Consortium provides a variety of resources and training events for learning TEI, information on projects using the TEI, a bibliography of TEI-related publications, and software developed for or adapted to the TEI.

« The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world. Members contribute financially to the Consortium and elect representatives to its Council and Board of

III. XML-TEI — 1. Définition

Définition

<https://tei-c.org>

- Consortium formé de savants.
- Développement d'un standard de syntaxe XML pour représenter les textes sous forme numérique.
- Sur le site, un guide du standard TEI, ressources, outils, tutoriels...

- Communauté très active d'encodage XML-TEI, le standard est très largement utilisé par la communauté scientifique et l'édition.

Définition

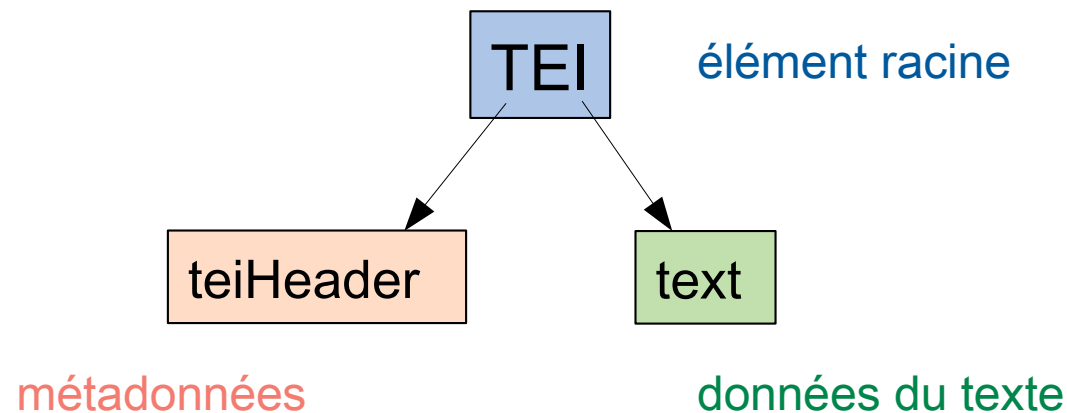
Permet de représenter :

- Métadonnées de la source comme du document numérique.
- Structure logique des textes.
- Support de la source originelle éditée.
- Description des sources premières ou critiques.
- Textes oraux et leurs spécificités.
- Nature morpho-syntaxique des chaînes de caractères.
- Modernisation de graphies anciennes.
- Liaison avec d'autres sources (via identifiants)
- Formules mathématiques.

XML et espaces de nom

III. XML-TEI — 2. Structure générale

Structure générale



<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-TEI.html>
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiHeader.html>
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-text.html>

XML et espaces de nom

III. XML-TEI — 2. Structure générale

Structure générale

```
<?xml version="1.1" 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://purl.oclc.org/dsdl/schematron"?>

<TEI
  xmlns="http://www.tei-c.org/ns/1.0"
  version="5">

  <teiHeader/>

  <text/>

</TEI>
```

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

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

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

XML et espaces de nom

III. XML-TEI — 3. Métadonnées

Métadonnées

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xml:lang="en" xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>
7           Titre du document numérique
8         </title>
9       </titleStmt>
10      <publicationStmt>
11        <p>
12          Modalités de publication et
13          distribution du document numérique.
14        </p>
15      </publicationStmt>
16      <sourceDesc>
17        <p>
18          Description des sources matérielles
19          sur lesquelles est basé le document numérique.
20        </p>
21      </sourceDesc>
22    </fileDesc>
23  </teiHeader>
24  <text>
25    Le corps du texte.
26  </text>
27 </TEI>
```

XML et espaces de nom

III. XML-TEI — 4. Le corps du texte

Le texte

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xml:lang="en" xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader/>
4   <text>
5     <front>
6       Page de titre du livre
7     </front>
8     <body>
9       Le texte lui-même sans paratexte.
10    </body>
11    <back>
12      Appendices suivant le texte.
13    </back>
14  </text>
15 </TEI>
```

XML et espaces de nom

III. XML-TEI — 4. Le corps du texte

Le texte

Une structure à base de divisions typées : les divisions de type livre contiennent leur titre (“head”) et des divisions de type chapitre qui elles-mêmes contiennent des divisions de type article.

```
2379 ▶ <facsimile xml:id="facs_55"> [39 lines]
2419 ◀ <text>
2420 ◀ <body>
2421 ◀ <div type="livre" n="1" xml:id="terrienStruct-lv01" sameAs="terrienLem-lv01">
2422 <pb facs="#facs_41" n="1"/>
2423 ▶ <head facs="#facs_41_TextRegion_1558434724675_7630"> [5 lines]
2429 ◀ <div type="chapitre" n="1" xml:id="terrienStruct-lv01-chp01" sameAs="terrienLem-lv01-chp01">
2430 ▶ <head facs="#facs_41_TextRegion_1560348543743_7540"> [2 lines]
2433 ▶ <div type="article" n="1" subtype="coutume" xml:id="terrienStruct-lv01-chp01-art01" sameAs="terrienLem-lv01-chp01-art01"> [395 lines]
2829 ▶ <div type="article" n="2" subtype="coutume" xml:id="terrienStruct-lv01-chp01-art02" sameAs="terrienLem-lv01-chp01-art02"> [137 lines]
2967 ▶ <div type="article" n="3" subtype="ordonnance" xml:id="terrienStruct-lv01-chp01-art03" sameAs="terrienLem-lv01-chp01-art03"> [17 lines]
2985 ▶ <div type="article" n="4" subtype="coutume" xml:id="terrienStruct-lv01-chp01-art04" sameAs="terrienLem-lv01-chp01-art04"> [31 lines]
3017 </div>
3018 ▶ <div type="chapitre" n="2" xml:id="terrienStruct-lv01-chp02" sameAs="terrienLem-lv01-chp02"> [53 lines]
3072 ▶ <div type="chapitre" n="3" xml:id="terrienStruct-lv01-chp03" sameAs="terrienLem-lv01-chp03"> [147 lines]
3220 ▶ <div type="chapitre" n="4" xml:id="terrienStruct-lv01-chp04" sameAs="terrienLem-lv01-chp04"> [176 lines]
3397 </div>
3398 </body>
3399 </text>
3400 </TEI>
```

XML et espaces de nom

III. XML-TEI — 4. Le corps du texte

Le texte

L'utilisation d'un <quote> (citation) et d'un <p> (paragraphe).

```
788 <div type="section" xml:id="morisse-lighter-beta-001-001-008"
788 corresp="gc_base.xml/#gc-lighter-beta-001-001-008">
789 <head>De terres qui sunt en defens. ¶ viii</head>
790 <note>
791 <quote>
792 <p>
793 ¶ Aucunes bestes sunt qui n ont point de banon. ains doiuent en tous tens estre
793 gardees et les damages que eles font doiuent estre restores si comme chieures qui brostent les borions des arbres et des uignes et porceaus qui foillent les
793 pres et les bles semes et totes bestes fraitieres et malfesans qui tous iors doiuent estre gardees et les damages que eles font doiuent estre restores.
794 </p>
795 </quote>
796 <p> [120 lines]
917 <p> [74 lines]
992 </note>
993 </div>
994 </div>
995 <div type="chapter" xml:id="morisse-lighter-beta-001-003" corresp="gc_base.xml/#gc-lighter-beta-001-003">
996 <head source="gc_base.xml/#gc-lighter-beta-001-003">Des tenemens et des eschaietes et des choses qui appartient.
996 </head>
997 <div type="section" xml:id="morisse-lighter-beta-001-003-003"
997 corresp="gc_base.xml/#gc-lighter-beta-001-003-003">
997 <head source="gc_base.xml/#gc-lighter-beta-001-003-003">De assise. xxii.</head>
```

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

XML et espaces de nom

III. XML-TEI — 5. Les informations linguistiques

Lemmes, POS etc

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.linguistic.html>

```
<s>
  <w pos="AT">The</w>
  <w pos="NN1">victim</w>
  <w pos="GE" join="left">'s</w>
  <w pos="NN2">friends</w>
  <w pos="VVD">told</w>
  <w pos="NN2">police</w>
  <w pos="CST">that</w>
  <w pos="NP1">Kruger</w>
  <w pos="VVD">drove</w>
  <w pos="II">into</w>
  <w pos="AT">the</w>
  <w pos="NN1">quarry</w>
  <w pos="CC">and</w>
  <w pos="RR">never</w>
  <w pos="VVD">surfaced</w>
  <pc pos="." join="left">.</pc>
</s>
```

- “<s> (phrase) contient une division textuelle de type phrase” → <https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-s.html>
- “<w> (mot) représente un mot grammatical (pas nécessairement orthographique)” → <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-w.html>
- “@pos (part of speech) indicates the part of speech assigned to a token (i.e. information on whether it is a noun, adjective, or verb), usually according to some official reference vocabulary (e.g. for German: STTS, for English: CLAWS, for Polish: NKJP, etc.).” → <https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ref-att.linguistic.html>

XML et espaces de nom

III. XML-TEI — 6. Autres informations pertinentes

Entités nommées

Une notice de personne
dans les métadonnées.

```
<person
  xml:id="rheginon"
  sameAs="https://catalogue.bnf.fr/ark:/12148/cb13753355n">

  <persName xml:lang="la">Regino Prumiensis</persName>
  <persName xml:lang="fr">Reginon de Prüm</persName>
  <birth>0840?</birth>
  <death>0915</death>

</person>
```

Références

Dans son ouvrage

`<ref corresp="#hist-francorum">Historiae
Francorum</ref>`,

`<ref corresp="#aimoin">Aimon de Fleury</ref>`

analyse le droit canon qui est repris dans l'arrêt rendu en la Grand-Chambre du 16 janvier 1667.

Merci pour votre attention !

Des questions ?

Bibliographie succincte :

- *BaseX*. Outil collaboratif, depuis 2005. Url : <https://basex.org/>.
- *MaX*. Pôle Document Numérique et Centre de Ressources Technologiques pour les TIC, Université de Caen. Url : http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/max.
- *TEI Guidelines*. Text Encoding Initiative Consortium (éd.) 1994-2020. *Text Encoding Initiative Guidelines*. Url : <https://tei-c.org>.
- *TEI Repository*. Text Encoding Initiative Consortium (éd.) 1994-2020. *Text Encoding Initiative Repository*. Url : <https://github.com/TEIC/TEI>.
- *XML*. World Wide Web Consortium (éd.) 2013-2020. *eXtensible Markup Language, Fifth edition*.
- *XSL*. World Wide Web Consortium (éd.) 2009-2021. *eXtensible Stylesheet Language, version 3.0*. Url : <https://www.w3.org/Style/XSL/>.
- *XPath*. World Wide Web Consortium (éd.) 1999-2020. *XML Path Language*. Url : <https://www.w3.org/TR/xpath/>.