

CapiTainS

Thibault Clérice

Décembre 6, 2017

1. La norme

- ▶ CapiTainS a pour objectif de rendre un texte disponible à la citation à travers une série de petites règles que doivent suivre les documents
- ▶ Suite d'outils et une norme
- ▶ Source du cours : <https://github.com/Capitains/DH2016>

2. Les identifiants

urn : cts : froLit	:	jns915 . jns1856 .	ciham-fro1	:	1.1-1.2
namespace		textgroup	work	version	passage
<i>Littérature en ancien français</i>		<i>Wauchier de Denain</i>	<i>Vie de Saint Martin</i>	<i>Edition d'Ariane Pinche au CIHAM</i>	<i>Entre 1.1 et 1.2</i>

Figure 1: Norme basée sur identifiant CTS

3. La structure des dossiers

- ▶ Un dossier corpus avec un sous dossier data
 - ▶ Un dossier avec l'identifiant du textgroup
 - ▶ et un fichier **cts.xml** pour les métadonnées
- ▶ Un dossier avec l'identifiant work
 - ▶ et un fichier **cts.xml** pour les métadonnées
 - ▶ et les fichiers d'éditations, de traductions et de commentaire

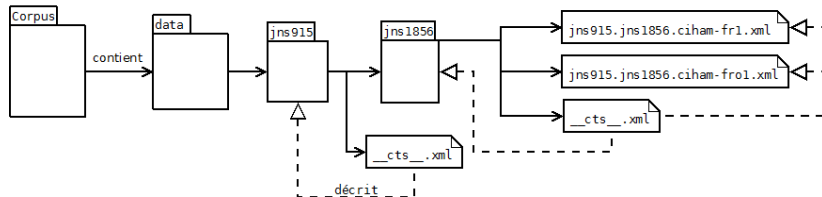


Figure 2: Structure

4. Métadonnées du textgroup

```
<textgroup xmlns="http://chs.harvard.edu/xmlns/cts"  
  urn="urn:cts:froLit:jns915">  
  <groupname xml:lang="fre">Wauchier de Denain</groupname>  
</textgroup>
```

5. Métadonnées du work et de l'édition

https://en.wikipedia.org/wiki/ISO_639:f

```
<work xmlns="http://chs.harvard.edu/xmlns/cts"
  groupUrn="urn:cts:froLit:jns915"
  urn="urn:cts:froLit:jns915.jns1856" xml:lang="fro">
  <title xml:lang="fre">Vie de Saint Martin</title>
  <edition workUrn="urn:cts:froLit:jns915.jns1856" urn="urn:cts:froLit:jns915.jns1856">
    <label xml:lang="fre">Vie de Saint Martin</label>
    <description xml:lang="fre">Édition du CIHAM basé s
  </edition>
</work>
```

5 bis Métadonnées avancées :

```
<textgroup xmlns="http://chs.harvard.edu/xmlns/cts"
  xmlns:cpt="http://purl.org/capitains/ns/1.0#"
  xmlns:saws="http://purl.org/saws/ontology#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:schema="http://schema.org/"
  urn="urn:cts:froLit:jns915">
  <groupname xml:lang="fre">Wauchier de Denain</groupname>
  <cpt:structured-metadata>
    <dc:type>schema:Person</dc:type>
    <schema:birthDate>1200</schema:birthDate>
    <schema:deathDate>1299</schema:deathDate>
    <dc:description>Wauchier de Denain est un écrivain fran
    <dc:identifiant>https://www.wikidata.org/wiki/Q942529</dc:
  </cpt:structured-metadata>
</textgroup>
```

6. Un essai de groupe

Un textgroup pour Perceval -

<http://atilf.atilf.fr/gsouvey/dect/download/Perceval.xml>

7. Tester

- ▶ <https://capitains-validator.herokuapp.com/>
- ▶ Pour ceux-lles avec Python3 sur leur ordinateur (Ne fonctionne pas sous windows) :
- ▶ Installer HookTest : `sudo pip3 install HookTest`
- ▶ Dans le dossier du corpus, ouvrir un terminal et lancer
`hooktest --console table --scheme tei --verbose`
10 .

8. Le fichier texte TEI : l'identifiant

- ▶ On mets l'identifiant du texte dans l'attribut @n de
 - ▶ /TEI/text/body pour de la TEI normale
 - ▶ TEI/text/body/div[@type="edition" or @type="translation" or @type="commentary"] pour de l'epidoc
- ▶ On met sur ce même tag l'attribut xml:lang avec le code en trois caractères

9. Le fichier texte TEI : l'identification des passages

```
<refsDecl n="CTS">
  <cRefPattern
    n="vers" matchPattern="(\w+).(\w+)"
    replacementPattern="#xpath(/tei:TEI/tei:text/tei:body
    <p>Ce pointeur extrait les vers des poèmes</p>
  </cRefPattern>
  <cRefPattern
    n="poeme" matchPattern="(\w+)"
    replacementPattern="#xpath(/tei:TEI/tei:text/tei:body
    <p>Ce pointeur extrait les poèmes</p>
  </cRefPattern>
</refsDecl>
```

Exemple

- ▶ Source :

<http://atilf.atilf.fr/gsouvey/dect/download/Perceval.xml>

<http://www.atilf.fr/dect/>