

CASTORE



ECOLE DES MINES DE NANTES UNIVERSITÉ DE NANTES



Chaîne de Soumission

Janvier 2006

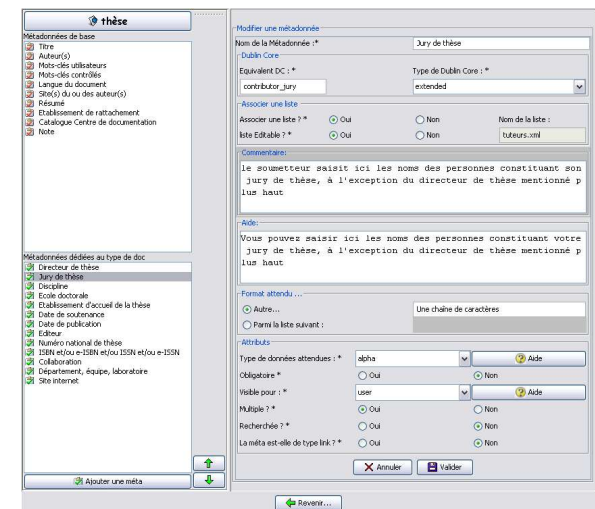


- Un document
 - Word, Sxw, html
 - Convertis en XML
 - Module de retro conversion
- Une notice
 - Classification dans des collections
 - Licence / Diffusion / etc...
 - Applet de gestion des types de documents
- Une chaine de validation
 - Applet de gestion du workflow

types de documents

Gestion des types de documents

1. Classés en catégories
2. Chaîne de validation
 - Séquentielle
 - Par personne ou par groupe
 - Différents niveaux de validation
3. Metadonnées associées
 - Tout est configurable.



The screenshot displays the 'thèse' document type configuration in the Castore system. The interface is split into two main panes.

Left Pane (Metadonnées de base): A tree view showing the hierarchy of metadata fields for a thesis, including:

- Titre
- Auteur(s)
- Mots-clés collaboratifs
- Mots-clés contrôlés
- Langue du document
- Site(s) d'ou des auteur(s)
- Résumé
- Établissement de rattachement
- Catalogue Centre de documentation
- Note

Right Pane (Modifier une métadonnée): Configuration for the 'Jury de thèse' field.

- Nom de la Métadonnée :** Jury de thèse
- Dublin Core :** Dublin Core
- Equivalent DC :** contributor_jury
- Type de Dublin Core :** extended
- Associer une liste :** Oui Non
- Associer une liste ? *** Oui Non
- liste Editable ? *** Oui Non
- Nom de la liste :** tuteurs_viii
- Commentaires :** Le soumetteur saisit ici les noms des personnes constituant son jury de thèse, à l'exception du directeur de thèse mentionné p l'us haut.
- Aide :** Vous pouvez saisir ici les noms des personnes constituant votre jury de thèse, à l'exception du directeur de thèse mentionné p l'us haut.
- Format attendu :** Autre... Parmi la liste suivant : Une chaîne de caractères
- Attributs :** Type de données attendues : alpha
- Obligatoire :** Oui Non
- Visible pour :** user
- Multiple ? :** Oui Non
- Recherchée ? :** Oui Non
- La méta est-elle de type link ? *** Oui Non

Buttons at the bottom: 'Ajouter une méta' (up arrow), 'Revenir...' (down arrow), 'Annuler', and 'Valider'.

1. Les catégories

- `.../init/`
- Gérés en ligne
- `nativeCollections.xml`

```
<collections>  
  <collection name="I-INFORMATIQUE" description="Informatique">  
  ...
```

- `nativeGroups.xml`

```
<groups>  
  <group name="EMN" description="Ecole des Mines de Nantes">  
  ...
```

2. Les chaînes de validation

- `.../workflows/`
- `defaultWorkflow.xml`
`<workflow id="default">`

```
<step number="1">  
  <content>  
    <group name="documentalist" />  
  </content>  
</step>  
<step number="2">  
  <aces>  
    <group name="documentalist" />  
  </aces>  
</step>
```

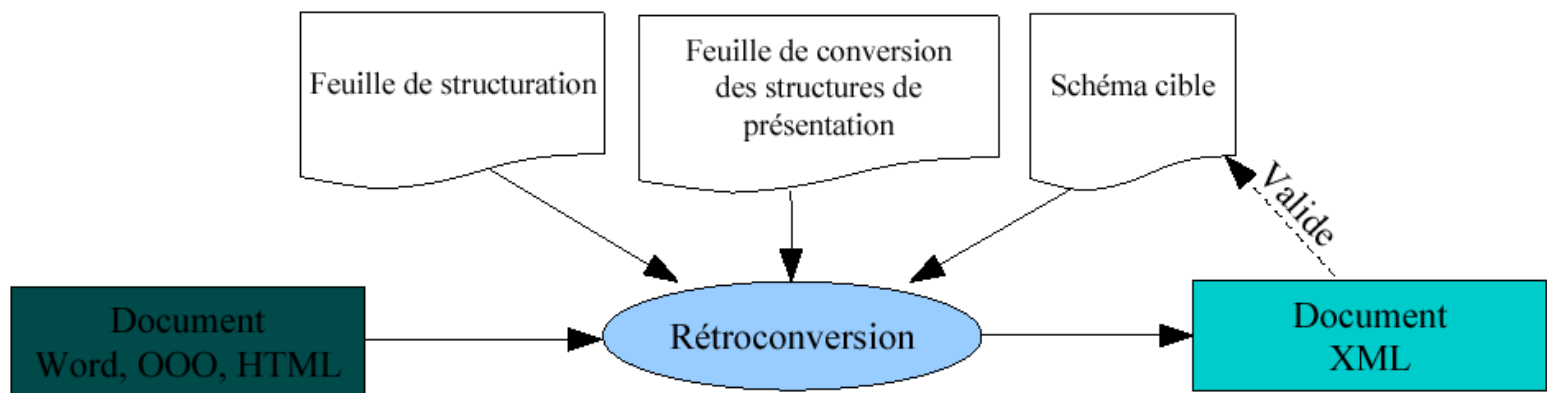
3. Les métadonnées

- Métadonnées
 - .../mapping/
 - Gérées par l'Applet
 - Sauf MD_Base1.xml
métadonnées communes
- Paramètres de l'Applet
 - .../typeDoc/TypeDoc_mapping.xml
+workflow id
- Listes contrôlées :
 - .../listes
 - DTD_listes.xml

```
<list>  
  <element>  
    <value>Castore</value>  
  </element>  
</list>
```

Feuilles de structuration

Retro – niveau 0



- Feuilles de structuration
 - .../s4s/stockOfStructurationFiles/
 - s4s.xml : liste des feuilles publiées
 - emn-bn-conf-acm-s4s .xml .xslt
 - Gérées par l'Applet
- Paramètres de l'Applet
 - .../s4s/structureItemSetup/
 - structure_items.xml
éléments XML du schéma gérés dans l'Applet
- Retro dans .../retro/

Règles de structuration

- Typographie et/ou nom de style -> élément XML

```

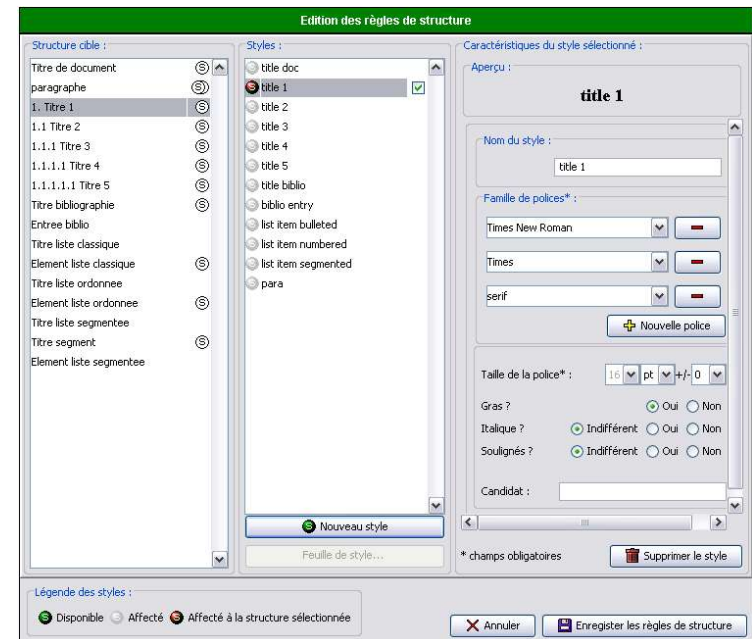
<structure.rule id="r.1">
<criteria scope="unique" id="r.1.c">
<fp>
  <font-family>
    <font-name> Times New Roman </font-name>
    <font-name> Times </font-name>
  </font-family>
  <font-size>
    <value> 16 </value>
    <unit> pixel </unit>
  </font-size>
  <font-weight> normal </font-weight>
</fp>
<source-style-name> title 1 </source-style-name>
</criteria>

<pivot.element> title </pivot.element>
<target.element> sect1 </target.element>

</structure.rule>

```

- Règle de structuration : structure.rule
- Ensemble de propriétés typographiques : fp
- Style nommé : source-style-name
- Format du paragraphe : candidate.value
- Element à créer :
 - Element XML singleton : pivot.element
 - Fragment XML : target.element /pivot.element



- schema-in
- schema-out

```
<mapping>
```

```
  <in element="text list-item"/>
```

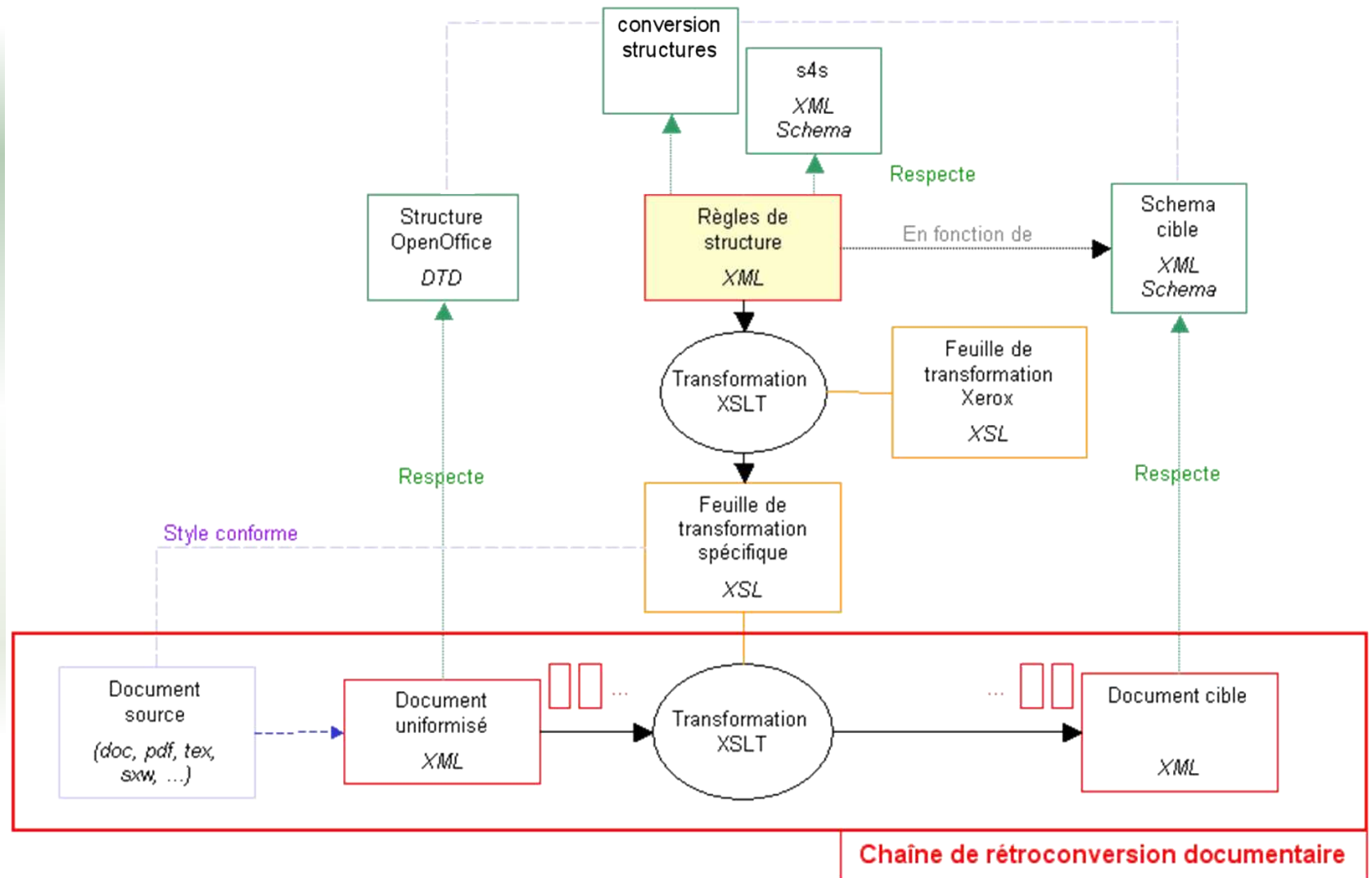
```
  <out>
```

```
    <emn listitem>*</emn listitem>
```

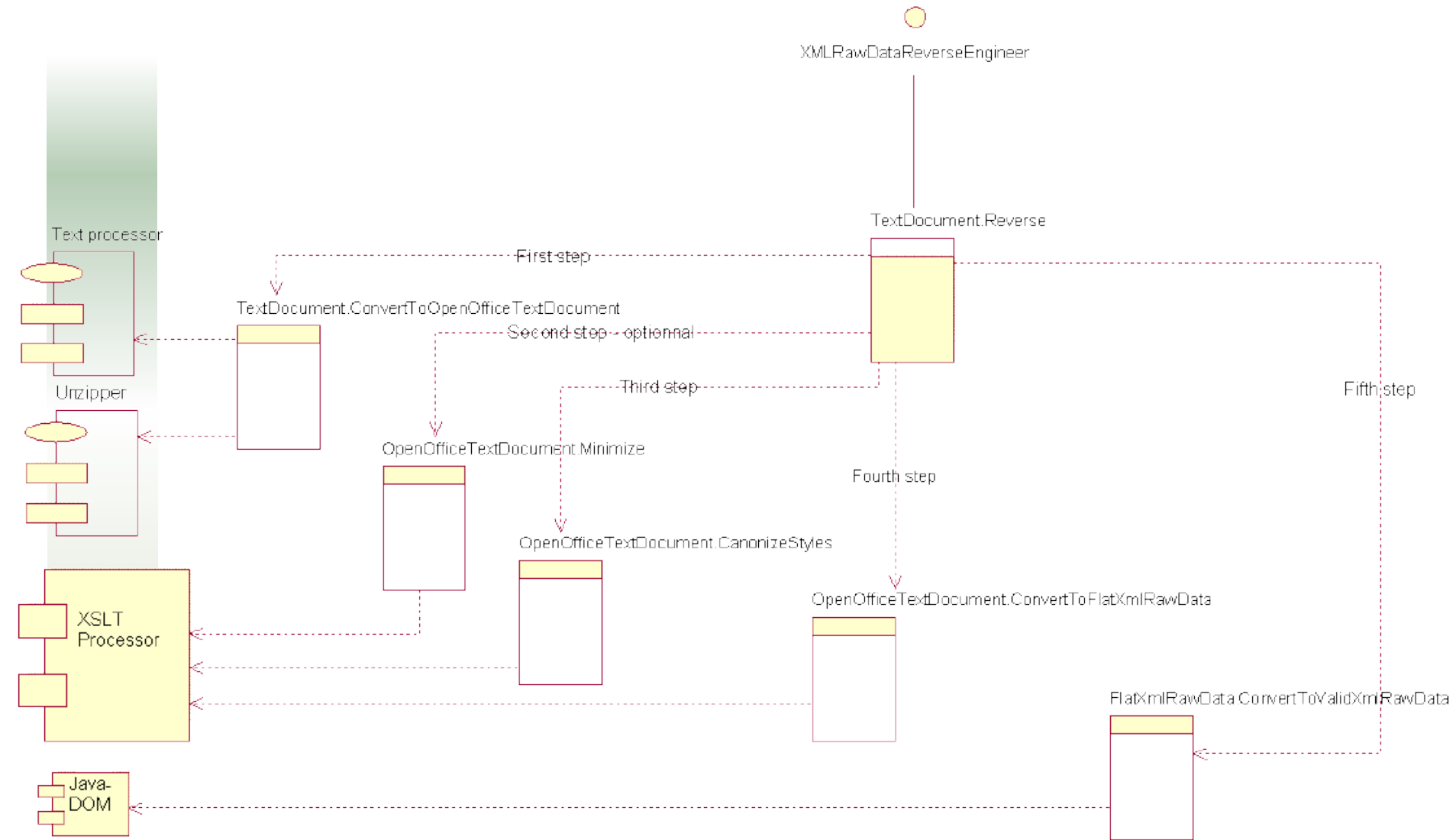
```
  </out>
```

```
</mapping>
```

Retro – niveau 1

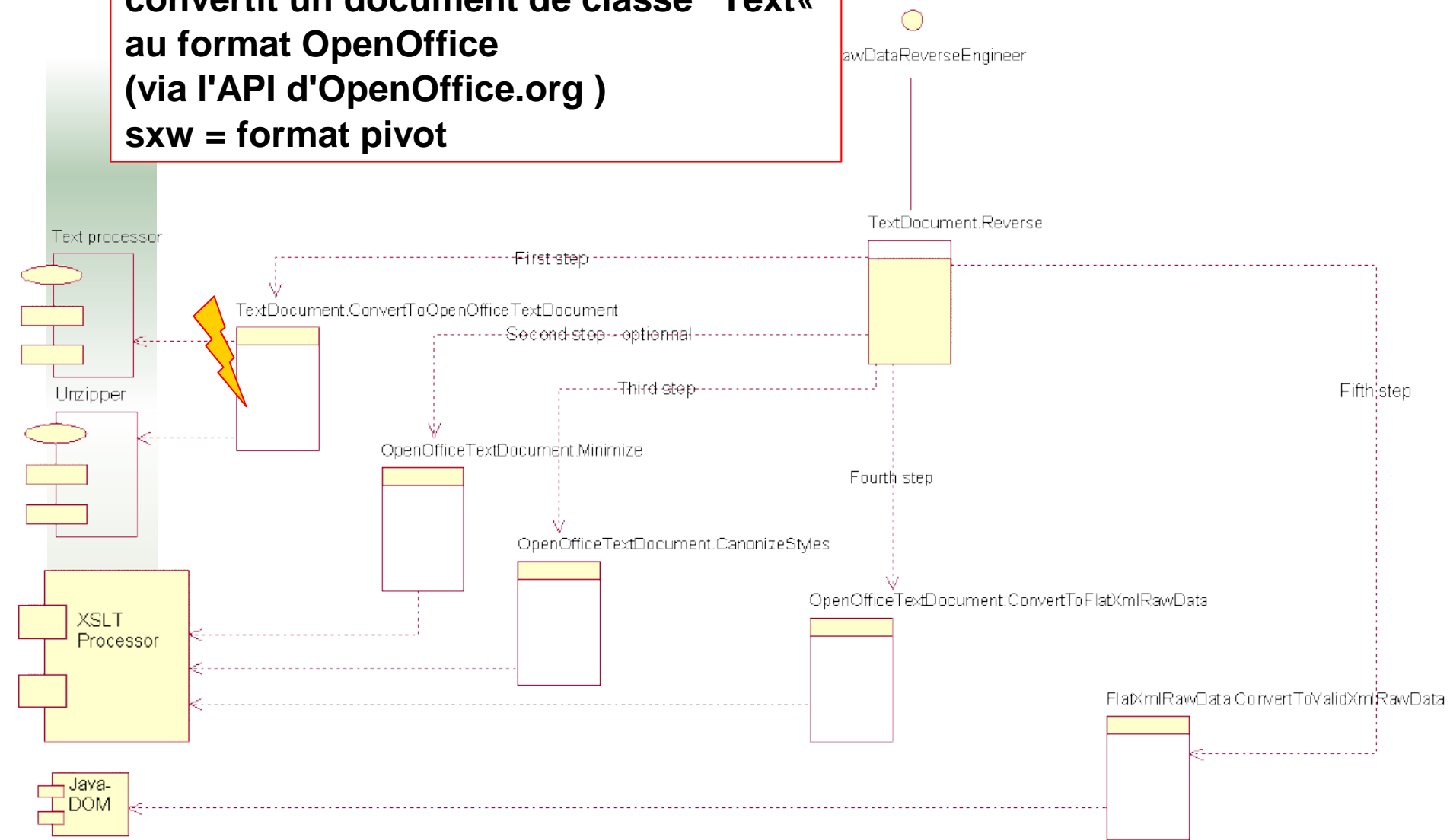


Retro – niveau 2



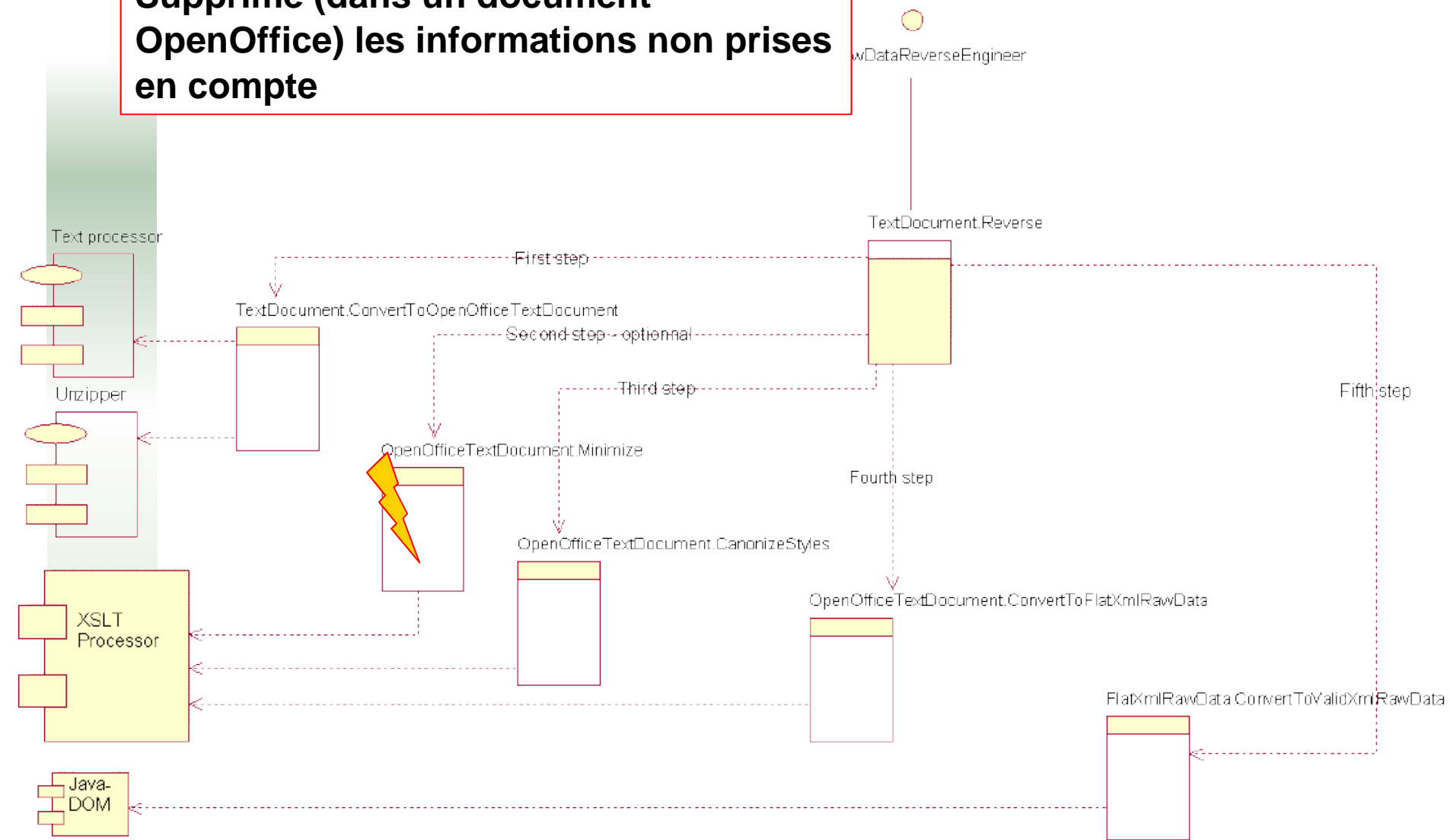
TextDocument.ConvertToOpenOfficeTextDocument

**convertit un document de classe "Text«
au format OpenOffice
(via l'API d'OpenOffice.org)
sxw = format pivot**



OpenOfficeTextDocument.Minimize

Supprime (dans un document OpenOffice) les informations non prises en compte

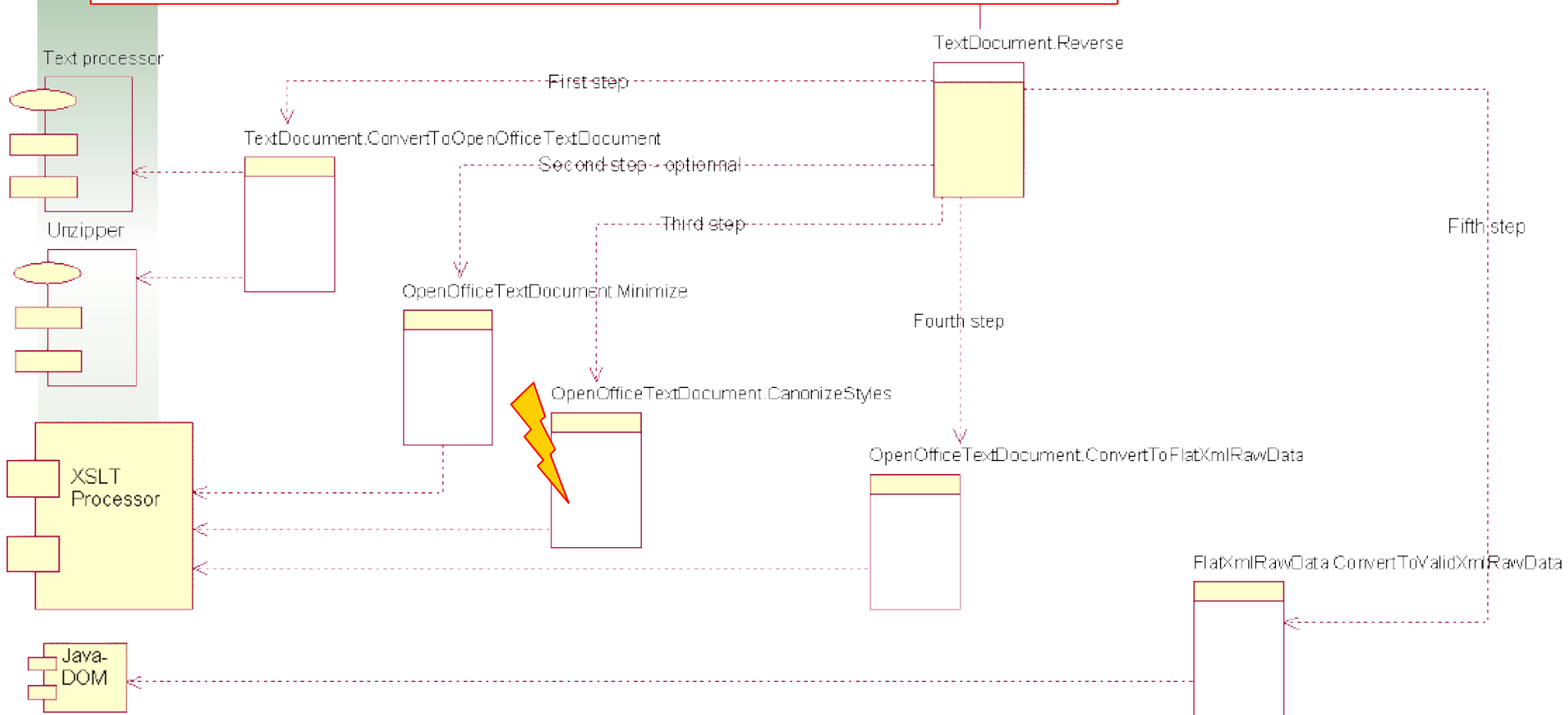


OpenOfficeTextDocument.CanonizeStyles

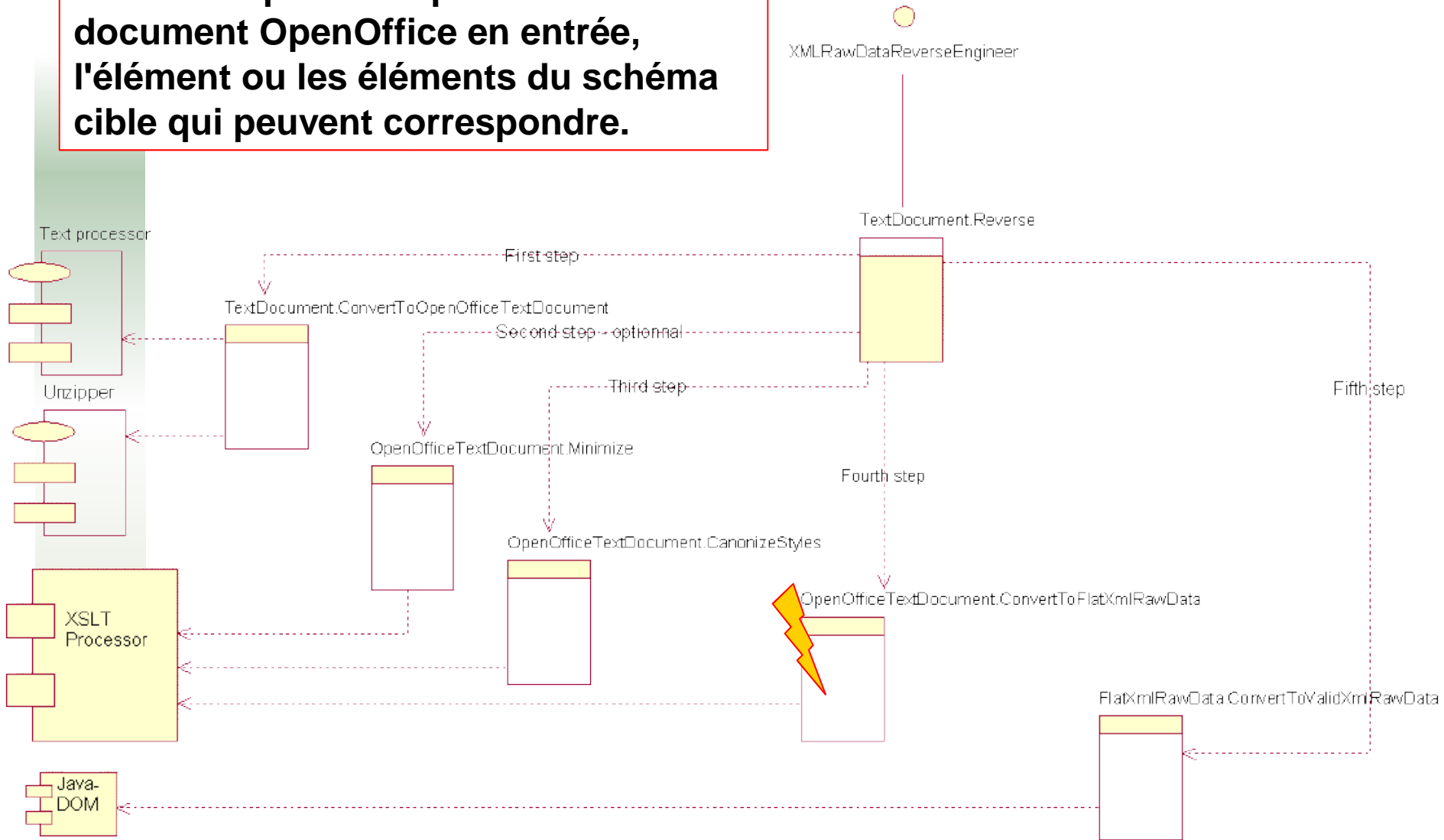
canonise les définitions de styles dans un document OpenOffice.

Un style peut être défini de deux manières :

- 1. Sans être basé sur aucun**
- 2. En le basant sur un style déjà**

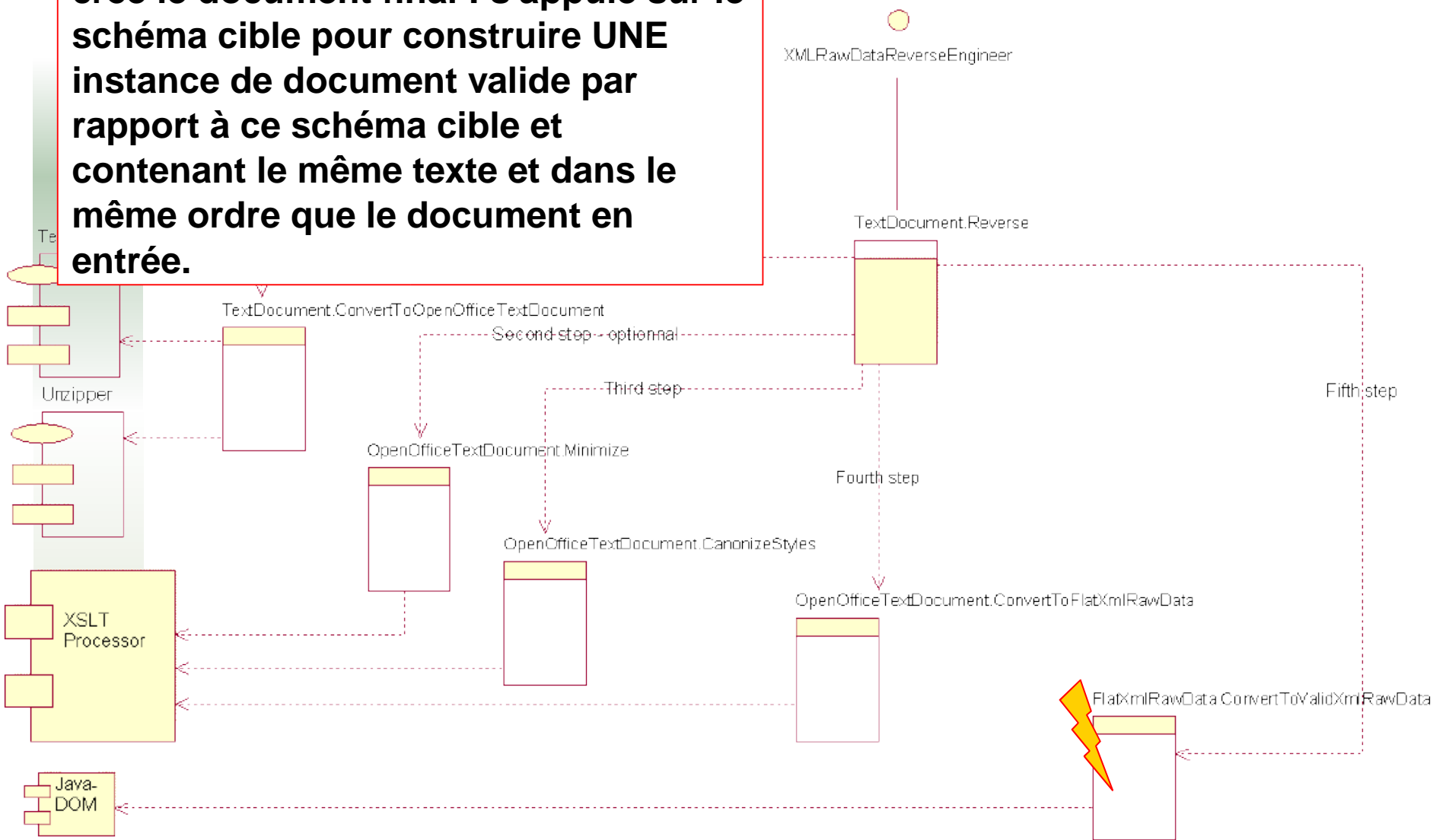


Détermine pour chaque élément du document OpenOffice en entrée, l'élément ou les éléments du schéma cible qui peuvent correspondre.



FlatXmlRawData.ConvertToValidXmlRawData

crée le document final : s'appuie sur le schéma cible pour construire UNE instance de document valide par rapport à ce schéma cible et contenant le même texte et dans le même ordre que le document en entrée.



Fonctionnement de la retro

- Règle 1: il y a deux types possibles de formats de documents en entrée
 - Les formats n'utilisant pas la notion de style nommé mais uniquement les propriétés typographiques :ex.PDF
 - Les formats utilisant la notion de style nommé ex.Word,OpenOffice.

Fonctionnement de la retro

- Règle 2 : traitement par listes triées avec priorités
 - Si aucune règle n'est valide pour un paragraphe, la règle par défaut s'applique : l'élément associé à cette règle est l'unique élément de la liste.
 - Si une seule règle est valide pour un paragraphe, cette règle s'applique : l'élément associé à cette règle et l'élément associé à la règle par défaut composent la liste.
 - Si plusieurs règles sont valides pour un paragraphe : les éléments associés à ces règles et l'élément associé à la règle par défaut composent la liste.

Formats n'utilisant pas de style nommé

Règle		Para	A appliquer	Priorité
Nom de style	Typo	Typo		
S	P	$P' \approx P$	Oui	$2 * N$
S	P	$P' \diamond P$	Règle par défaut	N

Formats utilisant des styles nommés

Règle		Para			A appliquer	Priorité
Nom de style	Typo	Nom de style	Typo du style	Typo surcharge		
S	P	S	\forall	\forall	Oui	5*N
\emptyset	P	S'	$\text{Typo}(S') \sim= P$	\forall	Oui	4*N
S	P	S'	$\text{Typo}(S') \sim= P$	\forall	Oui	3*N
S	P	S'	$\text{Typo}(S')=P'$ $\diamond P$	$P'' / P'+P'' \sim= P$	Oui	2*N
S ou \emptyset	P	S' ou \emptyset	$\text{Typo}(S') \diamond P$	\forall	Règle par défaut	N

Applet de gestion des FS

- Attention à la notion de publication
- Ne jamais supprimer des fichiers s4s