

# L'outillage TEI

## Workshop « Exploiter les données structurées en XML »

Lou Burnard  
MEET

juin 2011



# La boîte à outils TEI-XML

Pour maîtriser les documents XML, on a besoin de les

- créer, modifier, valider...
- transformer, visualiser...
- rechercher, analyser...
- stocker, gérer, trouver ...



## Où sont les outils TEI-XML?

- La TEI ne vous fournit ni usine à gaz, ni boîte d'outils..
- Dès le début, elle essaie de se distinguer nettement de la production des outils, pour mieux garantir son indépendance
- Les **Guidelines** sont conçus comme expression concrète d'un modèle abstraite des objets -- pour la plupart textuels -- qui sont d'intérêt scientifique dans les SHS...
- Mais c'est réservé aux communautés d'utilisateurs de se décider comment traiter ces objets, et donc de construire les outils pour en profiter.



## OK, mais quand meme...

Les TEI Guidelines s'expriment en XML, comme toute application TEI

- en principe, alors, n'importe quel logiciel conforme au norme XML nous suffirait
- mais il faut le construire!

La TEI a construit, pour ses propres besoins, un ensemble de feuilles de style XSLT, qui permet:

- la génération et la documentation des schémas TEI personnalisés ("ODD")
- la visualisation des documents TEI, à travers les formats actuellement les plus répandus (XHTML, PDF, Word, Open Office, ePUB...)
- une architecture générique pour la transformation entre documents TEI et d'autres formats (Word, OO, Docbook etc.)



# Les Stylesheets de Sebastian Rahtz

(Il faut trouver un nom plus sexy...)

Disponible sous licence ouverte LGPL, ils se trouvent partout:

- emballés avec oXygen, par défaut
- téléchargeables en form de "package" pour Debian/Ubuntu, sur sourceforge
- à l'interieur du web service EGE (on va voir cela)
- soujacent plusieurs applications indépendentes de la TEI



## Pourquoi s'en servir

- ils proposent des solutions aux difficultés que vous n'auriez pas (peut être) encore envisagés
- mieux vaut participer à un projet open source que de réinventer (plusieurs) roues
- ils sont maintenus en tandem avec les Guidelines
- ils sont modulaires, extensibles, bien "architecturés"

## Pourquoi ne pas s'en servir

Ils ont des objectifs génériques, mais avec des limites!

- Ils ne traitent pas tout élément TEI!
- Ils ne traitent pas tous les valeurs possibles des attributs TEI (surtout ceux qui ne sont pas spécifiés par la TEI)

Pour en profiter, il faut:

- savoir se débrouiller en XSLT (un peu)
- ajouter une couche de personnalisation propre à son propre projet (probablement).



# Personnalisation des feuilles de style

(x)HTML Un feuille de style CSS supplémentaire est conseillé pour les détails des blancs, polices etc.

LaTeX Il faut connaitre (evidemment) LaTeX ! mais en partant de cela, la génération d'un PDF de haute qualité est tres efficace (grace aux expertises considérables déjà existant dans l'édition scientifique avec LaTeX)



# Architecture des feuilles de style

Il y a un dossier par format de sortie...

---

docx	transformation vers, ou à partir de, Word OOXML
epub	transformation vers ePub
fo2	transformation vers XSL FO
latex2	transformation vers LaTeX
nlm	transformation à partir de NLM
odds2	transformation à partir de TEI ODD
oo	transformation vers, ou à partir de Open Office / Libre Office

... et des dossiers communs

---

profiles	personalisations
common2	xsl templates utilisés par tout format
tools2	Utilities

# Chaque dossier "format" contient ...

des règles spécialisés, regroupés par module TEI

---

core.xsl	Basic TEI elements
dictionaries.xsl	Dictionaries module
drama.xsl	Drama module
figures.xsl	Figures and tables module
header.xsl	Header module
linking.xsl	Linking module
namesdates.xsl	Names and Dates module
tagdocs.xsl	Processing ODDs
tei-param.xsl	Parameters
tei.xsl	Top-level wrapper
textcrit.xsl	Text critical module
textstructure.xsl	Basic structure
transcr.xsl	Transcription module
verse.xsl	Verse module

## Aspects de personnalisation possible (HTML)

- Layout
- Titres
- Numérotation
- Sélection/génération de texte de sortie
- TOC
- Internationalization
- Invocation de styles CSS ou code javascript
- Tables, figures, graphies
- Stylage caractères ("inline")

# Modes d'emploi

- 1 changement des paramètres proposés en Oxygen
- 2 usage d'un "wrapper" feuille de style
- 3 usage d'un webservice tel que OxGarage

## Usage d'un wrapper (1)

Voici un feuille de style qui invoque la transformation en HTML par défaut:

```
<xsl:stylesheet version="2.0">  
  <xsl:include  
    href="http://www.tei-  
c.org/release/xml/tei/stylesheet/xhtml2/tei.xsl"/>  
</xsl:stylesheet>
```

On s'en sert comme tout autre stylesheet (avec un processeur comme saxon, par exemple)

## Usage d'un wrapper (2)

On peut changer des parametres:

```
<xsl:stylesheet version="2.0">
  <xsl:include
    href="http://www.tei-
c.org/release/xml/tei/stylesheet/xhtml2/tei.xsl"/>
  <xsl:param name="logoFile">../../logo.png</xsl:param>
  <xsl:param name="logoWidth">60</xsl:param>
  <xsl:param name="logoHeight">60</xsl:param>
  <xsl:param name="cssFile">myTEI.css</xsl:param>
  <xsl:param name="pageLayout">CSS</xsl:param>
  <xsl:param name="numberHeadings">>false</xsl:param>
  <xsl:param name="parentWords">Projet MEET</xsl:param>
  <xsl:param name="institution">TGE Adonis</xsl:param>
</xsl:stylesheet>
```

## Usage d'un wrapper (3)

On peut ajouter un template nouveau :

```
<xsl:stylesheet version="2.0">
  <xsl:include
    href="http://www.tei-
c.org/release/xml/tei/stylesheet/xhtml2/tei.xsl"/>
  <xsl:param name="logoFile">../../logo.png</xsl:param>
  <xsl:param name="logoWidth">60</xsl:param>
  <xsl:param name="logoHeight">60</xsl:param>
  <xsl:param name="cssFile">myTEI.css</xsl:param>
  <xsl:param name="pageLayout">CSS</xsl:param>
  <xsl:param name="numberHeadings">>false</xsl:param>
  <xsl:param name="parentWords">Projet MEET</xsl:param>
  <xsl:param name="institution">TGE Adonis</xsl:param>
  <xsl:template match="tei:hi[@rend='upside-down']">
    <span class="upside-down">
      <xsl:apply-templates/>
    </span>
  </xsl:template>
</xsl:stylesheet>
```

# OxGarage

Un outil web qui permet une gamme de transformations à double sens

- OpenOffice XML (ODT) vers TEI (et l'inverse)
- Word 2007 XML (DOCX) vers TEI (et l'inverse)
- Docbook vers TEI
- TEI vers ePub
- TEI P4 vers TEI P5
- ...

Disponible à <http://www.tei-c.org/ege-webclient/> et  
<http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient>





# demo de OxGarage

[OUCS](#) | [Contact](#) | [A to Z](#) | [Help](#) | [Status](#) | [Rules](#) | [Oxford University](#)

Go!



Oxford University Computing Services

Thursday 4. Nov 2010



## OxGarage Document Conversion

Please select the type of the document you want to convert.

Convert from: [?](#)



Presentation documents



Spreadsheet documents



Text documents

# demo de OxGarage

## Convert from: ?



### Presentation documents



### Spreadsheet documents



### Text documents

- Compiled TEI ODD Document
- DocBook Document
- Microsoft Word Document (.doc)
- Microsoft Word Document (.docx)
- ODD Document
- Open Office Text Document (.odt)
- OpenOffice 1.0 Text Document (.sxw)
- Plain Text (.txt)
- Rich Text Format (.rtf)
- TEI P4 XML Document
- TEI P5 XML Document
- WordPerfect Document (.wpd)
- xHTML Document

## Convert to: ?

- Comma-Separated Values (.csv)
- ePub Document
- LaTeX Document
- Microsoft Excel Document (.xls)
- Microsoft Word Document (.doc)
- Microsoft Word Document (.docx)
- National Library of Medicine (NLM) DTD 3.0
- Open Office Spreadsheet (.ods)
- Open Office Text Document (.odt)
- OpenOffice 1.0 Spreadsheet (.sxc)
- OpenOffice 1.0 Text Document (.sxw)
- PDF Document
- Plain Text (.txt)
- Rich Text Format (.rtf)
- Tab-Separated Values (.tsv)
- xHTML Document
- XSL-FO Document

# demo de OxGarage

Applications Places System Thu 4 Nov, 10:19 PM Lou

duBellay.odt (read-only) - Openoffice.org Writer

File Edit View Insert Format Table Tools Window Help

100%

**Extrait du privilege.**


Il est permis par lettre patente  
du Roy nostre sire, à Armeul  
l'Angelier de faire imprimer &  
mettre en vente deux petits li-  
vres intitulés La deffence & Il-  
lustration de la langue Francoyse, & l'autre  
Cinquante Sonnets à la louange de l'Olive,  
l'Anterogive de la vieille & de la jeune amyè,  
& vers Lynques nouvellement composez. Et  
deffence faite à tous libraires & imprimeurs  
d'imprimer ou mettre en vente lesdicts livres,  
fors de ceulx que ledict l'Angelier aura fait  
imprimer, durant le temps & terme de trois  
ans prochains, sur peine de confiscation des  
dictz livres & d'amende arbitraire.

N Buyer

**A MONSEI  
GNEUR LE REVE  
rendissime Cardinal  
du Bellay.  
S.Salut**

*VEU le Personnage,  
que tu joues au Spec-  
tacle de toute l'Europe,  
voys de tout le Mon-  
de en ce grand Thea-  
tre Romain, van sans  
d'affaires, & tu L, que*

home/fo... formatio... lou - Fil... duBellay... OxGarag... Facebook... lou@TH... Inbox for ... Downloads duBellay...



# Matrix of OxGarage conversions

Inputs:	Compiled TEI ODD Document	DocBook Document	Microsoft Word Document (.doc)	Microsoft Word Document (.docx)	ODD Document	Open Office Text Document (.odt)	OpenOffice 1.0 Text Document (.sww)	Plain Text (.txt)	Rich Text Format (.rtf)	TEI P4 XML Document	TEI P5 XML Document	WordPerfect Document (.wpd)	xHTML Document
ODD documentation as TEI Lite	✓				✓								
Comma-Separated Values (.csv)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DTD created from ODD	✓				✓								
ePub Document	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LaTeX Document	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Microsoft Excel Document (.xls)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
xHTML Document	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Microsoft Word Document (.doc)	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓
Microsoft Word Document (.docx)	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
National Library of Medicine (NLM) DTD 3.0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ODD documentation as HTML	✓				✓								
Open Office Spreadsheet (.ods)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Open Office Text Document (.odt)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenOffice 1.0 Spreadsheet (.sxc)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenOffice 1.0 Text Document (.sww)	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
PDF Document	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Plain Text (.txt)	✓			✓	✓	✓	✓		✓	✓	✓	✓	✓
RELAXNG schema created from ODD	✓				✓								
Rich Text Format (.rtf)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tab-Separated Values (.tsv)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
XSL-FO Document	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TEI P5 XML Document		✓	✓	✓			✓	✓	✓	✓		✓	✓
Compiled ODD Document					✓								