

# La TEI c'est quoi?

Lou Burnard

## Qu'est-ce que c'est que la TEI?

- un labyrinthe trop complexe pour les humanistes?
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

## Qu'est-ce que c'est que la TEI?

- **un labyrinthe trop complexe pour les humanistes?**
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

## Qu'est-ce que c'est que la TEI?

- un labyrinthe trop complexe pour les humanistes?
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

## Qu'est-ce que c'est que la TEI?

- un labyrinthe trop complexe pour les humanistes?
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

## Qu'est-ce que c'est que la TEI?

- un labyrinthe trop complexe pour les humanistes?
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

## Qu'est-ce que c'est que la TEI?

- un labyrinthe trop complexe pour les humanistes?
- un système fasciste imposant des normes informatiques à ceux qui n'en ont pas besoin?
- un machin des bibliothécaires numérisés?
- un bibelot inutile qui sort du TAL ?
- une manière de gérer les trucs informatisés vraiment démodée quoi, puisqu'on a tout sur google...

# Les enjeux de la TEI

## "Text Encoding for Interchange"

- faciliter la **création**, l'**échange**, et l'**intégration** des données textuelles informatisées
  - toute sorte de texte
  - toutes les langues
  - toute origine temporelle ou culturelle
- La TEI s'adresse également ...
  - aux débutants, cherchant des solutions bien connues et consensuelles
  - aux experts, cherchant à créer de nouvelles solutions



## Les buts de la TEI

- faire des recommandations qui se basent sur un consensus existant
- préférer les solutions générales à celles spécifiques à une discipline
- en même temps permettant la spécialisation et l'extension

## D'où est sorti la TEI?

- Dans les années 90, c'était un projet de recherche en "digital humanities"..
  - Parainé par trois associations érudites
  - Financé 1990-1994 par NEH, EU LE Programme et SSHRC au Canada
- Influences majeures
  - bibliothèques et archives numérisées
  - ingénierie linguistique
  - édition des sources littéraires ou historiques
- Consortium international établi en 1999 (voir <http://www.tei-c.org/>)

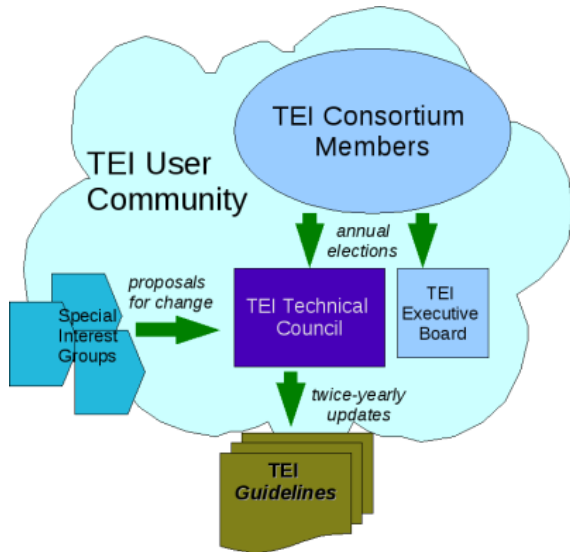
## Qu'est-ce que c'est que la TEI aujourd'hui?

- Un ensemble des *Guidelines* (lignes directrices): peu prescriptives
- représentant un consensus au sujet des distinctions significatives dans un vaste ensemble de matériaux textuels
- qui s'expriment en deux gros volumes de prose et un ensemble de définitions formelles
- ces définitions peuvent être exprimées en plusieurs langues schéma:
  - TEI P1-P3 (1991-1999) : en SGML DTD
  - TEI P4 (2000) : soit en SGML soit en XML DTD
  - TEI P5 (2005-) soit en XML DTD, en W3C Schema, ou en RelaxNG
- derrière cet ensemble se trouve un modèle formel ("conceptual schema") de plus en plus élaboré en système de classes, voire en **ontologie**

et une communauté internationale active...



# Organisation de la TEI



## Il n'y a pas de "TEI.dtd"

- TEI est un système *modulaire*. On s'en sert pour créer un système d'encodage selon ses propres besoins, en sélectionnant des *modules* spécifiques
- Chaque module définit un groupe d'éléments (et leurs attributs)
- on peut sélectionner les éléments voulus, et même en changer des propriétés
- on peut y mélanger des éléments nouveaux, ou bien natifs ou bien d'autres standards

## L'envergure TEI

- Structuration basique des textes continus
- Transcription diplomatique, images, multimédia, annotations...
- Données formelles : dates, noms de lieux ou de personnes...
- Données paratextuelles et "meta"
- Analyses linguistiques à tout niveau
- Documentation de balisage
- Et cetera: voir <http://www.tei-c.org/P5/Guidelines/>

... un encyclopédie du balisage

## L'outillage TEI

Une des raisons fortes pour lesquelles se servir de la TEI est l'existence des outils TEI p.e.

**roma** <http://www.tei-c.org/Roma> permet de construire des schémas TEI

**"Garage"** services web pour conversion TEI/docx/html etc.

**xaira** moteur de recherche xml

**TEI publisher** systeme d'édition web

**Open Text** systeme d'édition de revues.org

Une autre est la possibilité de se servir de n'importe quel outillage XML -- parce que la TEI est 100% standard XML!

# 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 ...



## Création et modification

peut s'effectuer avec

- un éditeur ascii classique (pe emacs)
- un éditeur XML spécialisé (pe Oxygen)
- un système bureautique (pe OpenOffice, Word)

Le Garage transforme entre plusieurs formats bureacratique et XML

# Validation

- devrait se faire en même temps que la modification
- nécessite l'existence d'un schéma
- plusieurs utilités: jing, xmllint, onvdl, xerces...
- 

**Roma** est un outil pour construire facilement des schémas TEI.

# Transformation et visualisation

Opération fondamentale, s'appuyant sur des standards:

- XSLT (Extensible Stylesheet Language) et XPath
- CSS (Cascading Stylesheets)

Fonctionnement intégré dans les navigateurs, ou autres logiciels

## Recherches et analyses

- peuvent bien s'effectuer avec des outils de transformation
- ou bien avec des logiciels spécialisés qui s'y appuient
- XQuery: standard langue de requête, complémentaire de XPath

L'analyse textuelle pose des problèmes intéressants:

- les données non-structurées
- une richesse de métadonnées
- des structures poly-arborescentes

XAIRA est un outil libre permettant l'indexation et recherche des corpus XML

## Quelques exemples (1)

### L'approche resistance minimale...

- Telecharger nos fichiers TEI-XML.  
<http://www.cnrtl.fr/corpus/estrepublikain/>
- Les donnees sont disponibles...  
<http://archimedespalimpsest.net>
- Faire bosser le navigateur: Early America's Digital Archive  
(<http://www.mith2.umd.edu/eada/>)
- traitement de langues: Samyukta Agama  
(<http://buddhistinformatics.ddbc.edu.tw/BZA/>)
- HAL Open Archive (<http://hal.archives-ouvertes.fr>)

## Quelques exemples (2)

- faire lisible l'illisible **Artamène, ou Le Grand Cyrus**  
(<http://www.artamene.org/>)
- combiner l'image et sa transcription: le journal de Robert Graves (<http://www.tapor.uvic.ca:8080/cocoon/graves/index.xml>)
- procédures juridiques du Old Bailey  
<http://www.oldbaileyonline.org>
- archive numérique classique: Swinburne Archive (<http://swinburnearchive.indiana.edu/swinburne/www/swinburne/>)
- ensemble de notices sur les manuscrits (<http://handrit.is>)
- fond patrimoniale des textes du renaissance  
<http://www.bvh.univ-tours.fr/Epistemon/>

## Des références...

Site web du Consortium TEI <http://www.tei-c.org>

Depot sourceforge <http://tei.sf.net>

Quelques références francophones:

- [tei-fr@lists.cru.fr](mailto:tei-fr@lists.cru.fr)
- <http://lespetitescases.net/index102/>
- <http://www.culture.gouv.fr/culture/dglf/riofi/tei.htm>
- [http://artist.inist.fr/article.php3?id\\_article=122](http://artist.inist.fr/article.php3?id_article=122)
- <http://www.mutec-shs.fr/la-tei-en-france-pratiques-et-perspectives>

(et prochainement: <http://meet.tge-adonis.fr>)