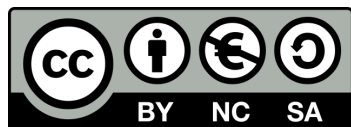


Structuration des données et des documents : balisage XML

Introduction à la TEI

J.B. Camps – jeudi 20 novembre 2015

M2 Technologies numériques appliquées à l'histoire



Ce support de cours est mis à disposition selon les termes de la licence *Creative Commons* Paternité – Pas d'utilisation commerciale – Partage à l'Identique 4.0

Plan de la séance

I. Il était une fois la TEI..

I.1 Historique et philosophie du projet

I.2 Les *Guidelines* et leur structure d'ensemble

I.3 La TEI en France

II. Structure de base et éléments
communs

III. Outils et documentation

les principes de Poughkeepsie (1987)

Proposer des *Guidelines* (recommandations) avec pour objectifs :

1. **provide a standard format for data interchange** in humanities research.
2. **suggest principles for the encoding of texts** in the same format.
3. (a) define a **recommended syntax** for the format, (b) a **metalanguage for the description of text-encoding schemes**, (c) describe the new format and representative existing schemes both in that metalanguage and in prose ;
4. propose **sets of coding conventions suited for various applications**.
5. include a **minimal set of conventions** for encoding new texts in the format.
6. The guidelines are to be drafted by **committees** on text documentation, text representation, text interpretation and analysis, metalanguage definition and description of existing and proposed schemes, coordinated by a **steering committee** of representatives of the principal sponsoring organizations.
7. **Compatibility with existing standards** will be maintained as far as possible.
8. A number of large text archives have agreed in principle to support the guidelines in their function as an interchange format. We encourage funding agencies to support development of tools to facilitate this interchange.
9. Conversion of existing machine-readable texts to the new format involves the translation of their conventions into the syntax of the new format. No requirements will be made for the addition of information not already coded in the texts.

Points de repère chronologiques

- **1987**, établissement de la *Text Encoding Initiative* par trois sociétés savantes : *Association for Computers and the Humanities*, *Association for Literary and Linguistic Computing* et *Association for Computational Linguistics*
- **1990**, **TEI Proposal 1** (TEI P1), *Guidelines for the Encoding and Interchange of Machine-Readable Texts*, dir. Michael Sperberg-McQueen et Lou Burnard,
- **1992-1993**, TEI P2, phase d'expansion
- **1994**, **TEI P3**, considérée comme la première version complète
- **2000**, naissance du **TEI Consortium**,
- **2001-2004**, **TEI P4** :> introduction du **XML** (à côté de SGML)
- **2007-....**, **TEI P5 (v.1.0)** : abandon de SGML, autre phase d'accroissement ; mises à jour bi-annuelles ; 26 juillet 2013, v. 2.5 ; v. 2.9.1, 15 octobre 2015.

Le Consortium de la TEI (2000)

Fondation interdisciplinaire à **but non lucratif**, financée par ses membres

- › TEI Board of Directors
- › TEI (technical) Council
- › Membres institutionnels (70 membres) et individuels
- › TEI Workgroups
- › SIG
- › importante communauté d'utilisateurs (environ 160 projets utilisant TEI recensés par <http://www.tei-c.org/Activities/Projects/>)

Outre les *Guidelines*, organise ou maintient également

- Une liste de diffusion (*TEI-L mailing list* ; NB : il existe également une liste francophone TEI-FR) et un wiki
- des *members meetings* (congrès) annuels, de Pise 2001 à Lyon 2015, la publication d'une revue, la *jTEI* (*Journal of the Text Encoding Initiative*, <http://jte.revues.org/>)

TEI Workgroups

<http://www.tei-c.org/Activities/Workgroups/index.xml>

En activité :

- Text Directionality Workgroup
- TEI Pointer/Stand-off Markup Working Group
- Task Force for TEI Title

Conclus :

- Character Encoding Workgroup (arrêté en 2005)
- Metalanguage Workgroup (2005)
- Stand-Off Markup Workgroup (2004)

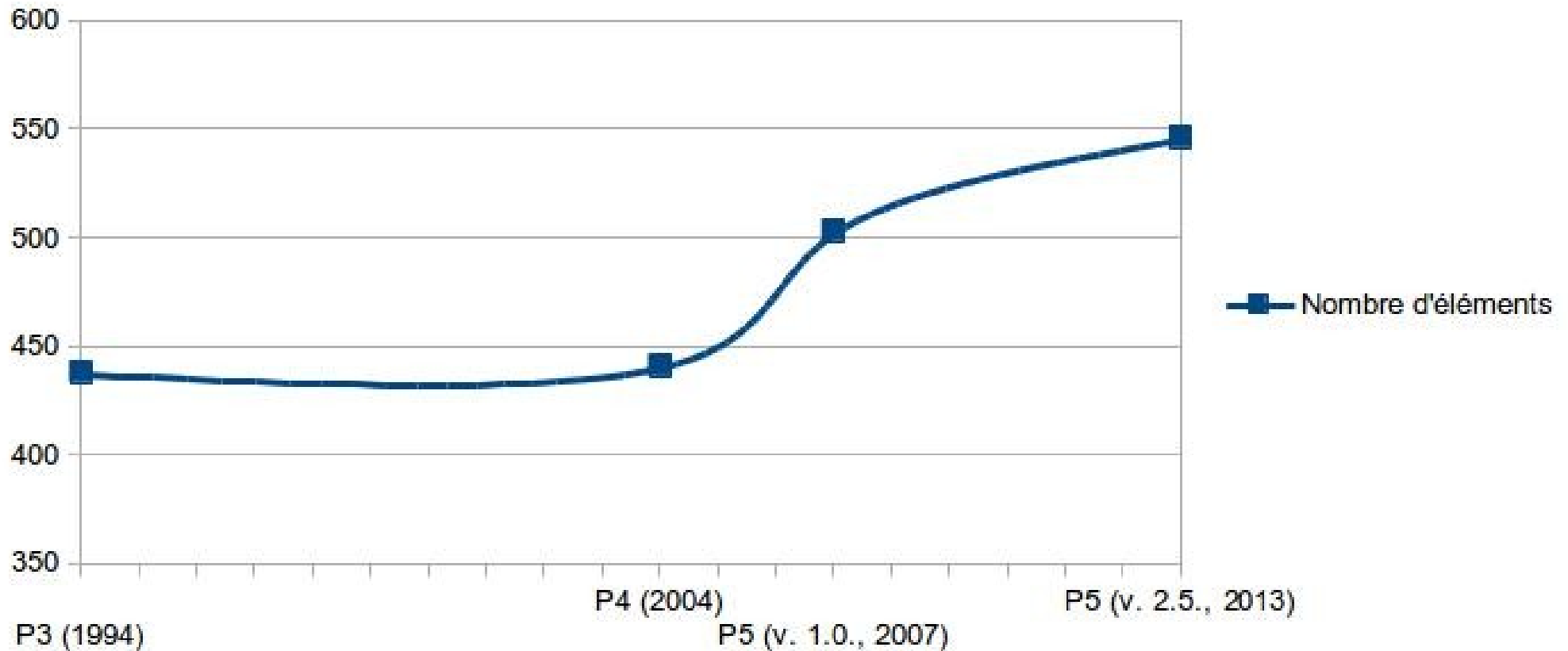
TEI Special Interest Groups

<http://www.tei-c.org/Activities/SIG/>

- **Computer-Mediated Communication**, animateur : Michael Beißwenger
- **Correspondence**, Peter Stadler, Marcel Illetschko, and Sabine Seifert
- **Education**
- **Libraries**, Stefanie Gehrke and Kevin Hawkins
- **Manuscripts**, Dot Porter et Gerrit Brüning
- **Music**, Raffaele Viglianti
- **Ontologies**, Oyvind Eide et Christian-Emil Ore
- **Scholarly Publishing**, Daniel O'Donnell
- **TEI for Linguists**, Piotr Bański et Andreas Witt
- **Text and Graphics**, John Walsh et Martin de la Iglesia
- **Tools**, Serge Heiden

Estimation de l'évolution du nombre d'éléments TEI

(calculée à partir des différentes versions des *Guidelines*)



Plan de la séance

I. Il était une fois la TEI..

I.1 Historique et philosophie du projet

I.2 Les *Guidelines* et leur structure d'ensemble

I.3 La TEI en France

II. Structure de base et éléments
communs

III. Outils et documentation

TEI : un modèle d'ensemble et de nombreuses applications

Distinguer

- **le modèle abstrait de données** (ensemble des concepts et les relations entre ces concepts) et la possibilité d'exprimer des définitions formelles d'une mise en œuvre particulière, un modèle particulier
- **les implémentations techniques du modèle**, (retranscription des contraintes formelles par le biais d'une DTD ou d'un schéma), qui peuvent être
 - générales (correspondre au plus près au modèle des *Guidelines* dans son ensemble)
 - particulières (répondre à des besoins spécifiques, liés à un outil, un projet, etc.)

Les *Guidelines* et leur structure d'ensemble

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>
cf. 1 « The TEI Infrastructure » (définie par le module tei)

P5: Guidelines for Electronic Text Encoding and Interchange

Version 2.5.0. Last updated on 26th July 2013.

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



Front Matter

[Title](#)

- I. [Releases of the TEI Guidelines](#)
- II. [Dedication](#)
- III. [Preface and Acknowledgments](#)
- IV. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

Text Body

- ☐ 1 [The TEI Infrastructure](#)
- ☐ 2 [The TEI Header](#)
- ☐ 3 [Elements Available in All TEI Documents](#)
- ☐ 4 [Default Text Structure](#)
- ☐ 5 [Non-standard Characters and Glyphs](#)
- ☐ 6 [Verse](#)
- ☐ 7 [Performance Texts](#)
- ☐ 8 [Transcriptions of Speech](#)
- ☐ 9 [Dictionaries](#)
- ☐ 10 [Manuscript Description](#)
- ☐ 11 [Representation of Primary Sources](#)
- ☐ 12 [Critical Apparatus](#)
- ☐ 13 [Names, Dates, People, and Places](#)
- ☐ 14 [Tables, Formulae, Graphics and Notated Music](#)
- ☐ 15 [Language Corpora](#)
- ☐ 16 [Linking, Segmentation, and Alignment](#)
- ☐ 17 [Simple Analytic Mechanisms](#)
- ☐ 18 [Feature Structures](#)
- ☐ 19 [Graphs, Networks, and Trees](#)
- ☐ 20 [Non-hierarchical Structures](#)
- ☐ 21 [Certainty, Precision, and Responsibility](#)
- ☐ 22 [Documentation Elements](#)
- ☐ 23 [Using the TEI](#)

TEI sourcecode

- [Using the TEI Sourceforge Repository](#)
- [Sourceforge Subversion Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

La TEI et ses modules

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	17 Simple Analytic Mechanisms
certainty	Certainty and Uncertainty	21 Certainty, Precision, and Responsibility
core	Common Core	3 Elements Available in All TEI Documents
corpus	Metadata for Language Corpora	15 Language Corpora
dictionaries	Print Dictionaries	9 Dictionaries
drama	Performance Texts	7 Performance Texts
figures	Tables, Formulae, Figures	14 Tables, Formulæ, Graphics and Notated Music
gaiji	Character and Glyph Documentation	5 Non-standard Characters and Glyphs
header	Common Metadata	2 The TEI Header
iso-fs	Feature Structures	18 Feature Structures
linking	Linking, Segmentation, and Alignment	16 Linking, Segmentation, and Alignment
msdescription	Manuscript Description	10 Manuscript Description
namesdates	Names, Dates, People, and Places	13 Names, Dates, People, and Places
nets	Graphs, Networks, and Trees	19 Graphs, Networks, and Trees
spoken	Transcribed Speech	8 Transcriptions of Speech
tagdocs	Documentation Elements	22 Documentation Elements
tei	TEI Infrastructure	1 The TEI Infrastructure
textcrit	Text Criticism	12 Critical Apparatus
textstructure	Default Text Structure	4 Default Text Structure
transcr	Transcription of Primary Sources	11 Representation of Primary Sources
verse	Verse	6 Verse

La TEI et ses modules

Modules obligatoires, communs à tous les documents TEI

- **tei** : définition des classes, macros et types de données
- **textstructure** : éléments de base pour structurer un texte de type livre
- **core** : éléments disponibles dans tous les documents TEI
- **header** : en-tête TEI (métadonnées du document)

La TEI et ses modules

Module propres à un type d'objet, une approche, une discipline

- analysis (analyse linguistique);
- **certainty (niveaux de certitude et marquage de la responsabilité) ;**
- **corpus (corpus) ;**
- drama (textes d'art dramatique) ;
- figures (tableaux, figures et formules) ;
- gaiji (caractères non standard et glyphes) ;
- iso-fs (structures de traits) ;
- linking (liens, segmentation, alignements) ;
- **msdescription (description des manuscrits) ;**
- **namesdates (noms, dates, lieux) ;**
- nets (graphes, réseaux, arbres) ;
- tagdocs (documentation) ;
- **textcrit (apparat critique) ;**
- **transcr (transcription des sources primaires) ;**
- **verse (vers).**

La TEI et ses modules

Chacun des modules :

- est documenté par un chapitre des *Guidelines* ;
- définit un certain nombre de composantes du modèle (éléments, classes, macros) ;
- pointe vers des spécifications techniques ;
- correspond à une problématique particulière et peut être inclus ou non dans une personnalisation (à part les 4 obligatoires).

Quelques exemples :

- « 3 Elements Available in All TEI Documents » (et module **core**)
- « 4 Default Text Structure » (et module **textstructure**)
- « 12 Critical Apparatus » (et module **textcrit**)

Lire la documentation d'un élément dans les *Guidelines*

Exemple, l'élément div

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

Les classes

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STEC>

Permettent d'organiser les éléments du modèle. Les éléments héritent des propriétés de la classe à laquelle ils appartiennent, mais aussi de la classe de cette classe (*superclass*), etc.

Deux types de classes: classes d'attributs (att.**) et du modèle (**model.**)**

- même ensemble d'attributs (classes d'attributs, **att.**), par ex.
 - classe **att.naming** fournit des attributs communs (@role,...) aux éléments qui concernent les personnes, lieux, organismes, etc. comme <forename>, <genName>, <nameLink>, <orgName>, <persName>, <roleName>, etc.
 - classe **att.global** (@xml:id, etc.), commune à tous les éléments.
- même emplacement dans le modèle (classes modèle, **model.**)
 - classes **structurelles**/hiérarchiques, qui regroupent des éléments qui ont un même emplacement dans la structure hiérarchique d'un document TEI
ex. : **model.divPart** pour ce qui peut être contenu par une div ; **model.phrase** pour les éléments pouvant apparaître au niveau de la phrase
 - les **classes sémantiques**, qui regroupent les éléments par similarité de sens.
Ex. : classe **model.nameLike** pour les divers types de noms; **model.pLike** pour les éléments de type paragraphe ; **model.rdgLike** pour définir des éléments sémantiquement équivalents à un rdg (une leçon/variante)

Les attributs

- **globaux**, potentiellement applicables à tous les éléments TEI
(att.global) :

xml:id fournit un identifiant unique

n fournit un numéro ou un label, pas nécessairement unique

xml:lang indique la langue du contenu textuel de l'élément (à l'aide du code ISO)

rend donne une indication sur l'apparence, le rendu
style information de style, de présentation

rendition pointe vers description du rendu

xml:base fournit une référence URI

xml:space informations sur la gestion des espaces par les applications

- définis au sein de classes attributives **pour des groupes d'éléments** ;

- définis **au niveau des modules spécifiques**.

xml:lang

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="fr">
  <teiHeader>
    <!-- [...] -->
  </teiHeader>
  <text>
    <body>
      <div>
        <!-- subdivision en français (hérité) -->
      </div>
      <div xml:lang="en">
        <!-- subdivision en anglais (spécifié localement) →
        <p>Italics are used <foreign xml:lang="la">a
priori</foreign> for foreign words.</p>
      </div>
      <div>
        <!-- subdivision en français (hérité) -->
      </div>
      <!-- ... -->
    </body>
  </text>
</TEI>
```

Les macros

Les macros sont des raccourcis qui permettent de **spécifier le contenu permis** (contrainte de contenu) :

- comme contenu d'un élément (*Standard Content Models*)
ex. mélange de PCDATA et de tels et tels sous éléments, répétés tant de fois, ou dans tel ordre
 - comme valeur d'un attribut (*Datatype Macros*)
ex. une chaîne de caractère, ou un nom XML, ou une liste séparée par des espaces, etc.
- Les types utilisés par la TEI s'inspirent en partie des types de XML Schema (xs, aussi utilisés par XSLT), mais en ajoutent d'autres

Les 6 macros de contenu d'élément (*Standard Content Models*) les plus courantes (par ordre décroissant)

macro.phraseSeq (*phrase sequence*) : mélange de PCDATA (de texte) et d'éléments de niveau phrase (0 ou plus et dans n'importe quel ordre)

macro.paraContent (*paragraph content*) contenu d'élément de type paragraphe

macro.specialPara (*'special' paragraph content*) contenu de paragraphes spéciaux, qui peuvent contenir soit du contenu structuré de type sous-éléments, soit du contenu type paragraphe

macro.phraseSeq.limited (*limited phrase sequence*) mélange de texte et de contenu de niveau phrase, *excluant les éléments servant à la transcription de sources*

macro.limitedContent (*paragraph content*) la même chose, mais au niveau paragraphe

macro.xtext (*extended text*) mélange de PCDATA et d'éléments gajji

Macros de valeur d'attributs

(cf. *Guidelines*, 1.4.2 « Datatype Macros » pour la liste complète)

Généralistes

data.word un mot

data.text une chaîne de caractères

data.name un nom XML

data.enumerated une valeur (nom XML) parmi une liste fermée de valeurs possibles

...

Propres au modèle de la TEI

data.certainty degré de certitude (`data.certainty = "high" | "medium" | "low" | "unknown"`)

data.probability probabilité

data.numeric valeur numérique

data.interval intervalle

data.percentage pourcentage

data.count compteur

...

Macros de valeur d'attributs

(cf. *Guidelines*, 1.4.2 « Datatype Macros » pour la liste complète)

Valeurs normalisées

data.duration.w3c durée exprimée selon la spécification W3C

data.duration.iso durée exprimée selon la spécification ISO 8601

...

Valeurs spécialisées

data.namespace espace de nom

data.pattern une expression régulière

data.pointer pointeur

data.version une version de la TEI ou d'Unicode

data.versionNumber un numéro de version

data.xpath une expression XPath

...

Lire la documentation d'un élément dans les *Guidelines*

Exemple, l'élément div

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

Personnaliser et documenter

Il est fortement déconseillé d'utiliser un schéma englobant l'intégralité de la TEI : une phase importante d'un projet est la **conception d'un modèle adapté** aux données et au projet, à l'exploitation des documents

Documenter pour **rendre ces choix lisibles et réexploitables** par un groupe plus large ou d'autres chercheurs

Outils

Roma (<<http://www.tei-c.org/Roma/>>) : application de génération de schémas TEI, permettant d'opérer les choix principaux de modèle et de générer un fichier utilisant la syntaxe **ODD (One Document Does it all)**, fichier TEI utilisant le module tagdocs), ainsi que le schéma demandé au format souhaité.

Opter pour des personnalisations généralistes de la TEI

<http://www.tei-c.org/Guidelines/Customization/>

- **TEI Lite** (90 % des besoins de 90 % des utilisateurs)

Voir Lou Burnard & C. M. Sperberg-McQueen, *TEI Lite: Encoding for Interchange: an introduction to the TEI*, Final revised edition for TEI P5, 2012, en ligne :

<http://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_lite.doc.html> ; NB :

la trad. fr. d'une ancienne version, dépréciée, intitulée *La TEI simplifiée: une introduction au codage des textes électroniques*, trad. François Role et al., se trouve sur

<<http://www.tei-c.org/Vault/P4/Lite/>>. Il existe également des trad. ital. (TEI Lite: introduzione alla codifica dei testi), esp. (TEI U5: Etiquetado para el intercambio: una introducción al TEI), russe (Введение в TEI Lite), ... mais aucune n'est à jour.

- **TEI Tite** (pour la sous-traitance auprès de prestataires privés de numérisation)
- **TEI bare**
- **TEI all**
- **TEI ms**
- **etc.**

Personnalisations pour un projet : quelques exemples

- les *Guidelines* de la California Digital Library
(<http://www.cdlib.org/groups/stwg/index.html>) ;
- documentation de certaines éditions électroniques publiées par l'École des chartes (voir ELEC et développements enc,
<http://developpements.enc.sorbonne.fr/>)
- manuel d'encodage écrit pour le projet BVH du CESR
(<http://www.bvh.univ-tours.fr/XML-TEI/index.asp>) ;
- documentation des schémas TEI utilisés dans le cadre du projet européen Enrich : <http://tei.oucs.ox.ac.uk/ENRICH/>;

Plan de la séance

I. Il était une fois la TEI..

I.1 Historique et philosophie du projet

I.2 Les *Guidelines* et leur structure d'ensemble

I.3 La TEI en France

II. Structure de base et éléments
communs

III. Outils et documentation

La TEI en France : Membres français du consortium

Centres et laboratoires de recherche du CNRS ou des univ. :
ATILF, IRHT, ICAR-Université Lyon 2, Bibliothèques Virtuelles
Humanistes,...

Établissements d'enseignement supérieur

École nationale des Chartes, ...

Autres structures d'enseignement supérieur :

TGE Adonis, Centre pour l'Édition électronique ouverte (Cléo),
Consortium CAHIER (Corpus d'Auteurs pour les Humanités:
Informatisation, Édition, Recherche, TGIR Huma-num),...

La TEI en France : initiatives nationales

- traduction française (inachevée), par un groupe de travail de l'AFNOR, de la documentation de la TEI
- Session publique du TEI Council organisée en France (Lyon, 2009) (<https://listes.cru.fr/wiki/tei-res/public/tei_council_2009_a_lyon>)
- Humanum (ex ADONIS) soutient les projets utilisant TEI ; création d'un nouveau centre de ressources, MEET (Mutualisations d'Expériences pour l'Encodage des Textes, voir <<http://www.huma-num.fr/ressources/presentation-de-laction-transversale-pour-la-mutualisation-dexperiences-pour-lencodage>>)
- Groupe MutEC (Mutualisation pour les éditions critiques et les corpus), <http://www.mutec-shs.fr>

La TEI en France : projets d'édition

Éditions de textes littéraires, anciens et modernes,
avec différentes perspectives (éditions d'imprimés
anciens, éditions à ms. de base, critiques,
génétiques,...)

Éditions de documents historiques

Éditions à visée linguistique ou paléographique

et d'autres usages encore

Éditions d'ouvrages imprimés anciens

Madeleine et Georges de Scudéry, *Artamène ou le Grand Cyrus*,
<http://www.artamene.org/> encodé en utilisant la TEI Lite

```
<body>
  <div0 type="partie" id="page_CYRUS01" n="I">
    <head type="epitre" n="Epitre dédicatoire à Mme la
Duchesse de Longueville placée en tête du 1er volume">
      <pb id="page_epitre_dedicatoire_volume1_1" n="V01-
I001" />
        A MADAME
        <lb/>LA DUCHESSE
        <lb/>DE LONGUEVILLE.
        <lb/>MADAME,
        <lb/>Un des plus Grands et des plus sages Princes de
toute l'Asie, va demander Audience, à une des plus Grandes et des plus
sages Princesses de toute l'Europe. Ce Vainqueur de la moitié du
Monde, qui croit avécque raison que Vostre Altesse seroit digne de le
commander tout entier : vient mettre à vos pieds ses Palmes et ses
Trophées, et advoüer ingenûment, qu'il a moins conquesté de Sceptres
et de Couronnes, que vous ne meritez d'en avoir.
      <pb id="page_epitre_dedicatoire_volume1_2" n="V01-
I002" />Il a sçeu que vous n'avez pas autrefois dénié vostre Glorieuse
protection à un Prince déguisé (...)
    </div0>
  </body>
```


Éditions d'ouvrages imprimés anciens Bibliothèque virtuelle humaniste

http://www.bvh.univ-tours.fr:8080/xtf/view?docId=tei/B410186201_I65/B410186201_I65_tei.xml

Les Bibliothèques Virtuelles Humanistes

Nouvelle recherche | Retour aux résultats

Chercher

Mosaïque | Formats | No

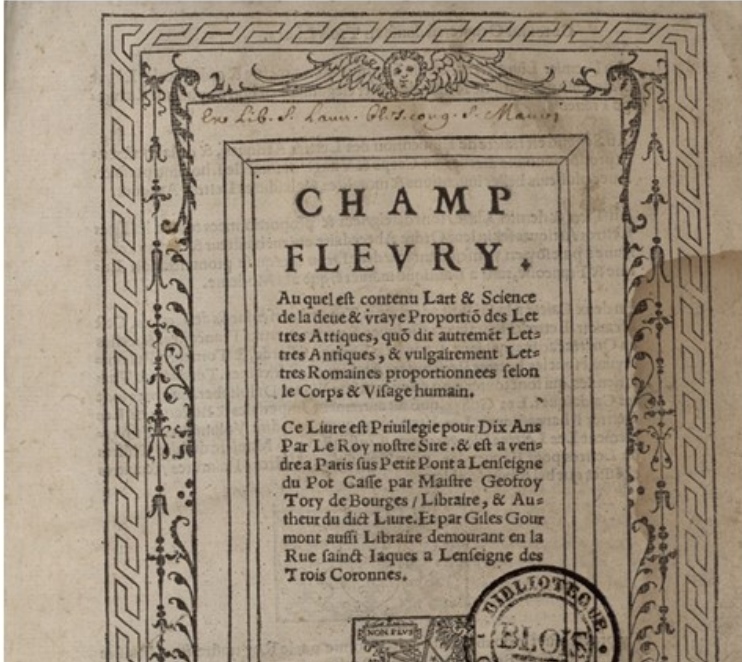
Options de consultation

Tory (Geoffroy)
Champ fleury. Au quel est contenu l'Art & Science de la deue & vraye Proportion des Lettres Attiques, qu'on dit autrement Lettres Antiques, & vulgairement Lettres Romaines proportionnees selon le Corps & Visage humain.
[Détails au survol de la souris]

1529

Ce toutal Oeuvre / est div
PRIVILEGE DU ROY NOST
Geoffroy Tory de Bourges,
LA TABLE.
Table des Dictions Latines
Aux Lecteurs de ce Preser
[Registre]

1. L'ART ET SCIENCE DE LA
2. LE SECOND LIVRE



CHAMP FLEURY.

Au quel est contenu l'Art & Science de la deue & vraye Proportion des Lettres Attiques, qu'on dit autrement Lettres Antiques, & vulgairement Lettres Romaines proportionnees selon le Corps & Visage humain.

Ce Livre est Priuilegie pour Dix Ans Par Le Roy nostre Sire. & est a vendre a Paris sus Petit Pont a Lenefigne du Pot Casse par Maître Geoffroy Tory de Bourges / Libraire, & Auctheur du dict Livre. Et par Giles Gourmont aussi Libraire demourant en la Rue saint Iaques a Lenefigne des Trois Coronnes.

Ce Livre est Priuilegie pour Dix Ans Par Le Roy nostre Sire. & est a vendre



Éditions d'ouvrages imprimés anciens

Bibliothèque virtuelle humaniste

http://www.bvh.univ-tours.fr:8080/xtf/view?docId=tei/B410186201_I65/B410186201_I65_tei.xml

```

<lb/>Ce Livre est Privilegie pour Dix Ans
<lb/>Par Le Roy nostre Sire. & est a ven-
<lb rend="hyphen"/>dre a <pubPlace>Paris</pubPlace> sus <placeName type="rue">Petit
Pont</placeName> a l'Enseigne
  <lb/>du <placeName type="enseigne">Pot Casse</placeName> par <publisher>Maistre
<name>
      <choice>
        <orig>Geofroy</orig>
        <reg>Geoffroy</reg>
      </choice>
      <lb/>Tory</name>
    </publisher> de <placeName type="ville">Bourges</placeName> / Libraire,
& Au-
  <lb rend="hyphen"/>theur du dict Livre. Et par <publisher>
    <name>Giles Gour-
  <lb rend="hyphen"/>mont</name>
  </publisher> aussi Libraire demourant en la
<lb/>
  <placeName type="rue">Rue <name>saint Jaques</name>
</placeName> a l'Enseigne des
<lb/>
  <placeName type="enseigne">Trois Coronnes</placeName>.
</docImprint>
  <lb/>
<imprimatur rend="center">
  <lb/>
  <lb/>PRIVILEGE POUR DIX ANS
</imprimatur>

```

Éditions de textes médiévaux

Parmi de nombreux projets :

- projet Sermones (université de Lyon 2 *et alii* :

<<http://www.sermones.net/>>)

- Les *Chroniques latines du Mont Saint-Michel* (IX^e-XII^e siècle, <http://www.unicaen.fr/services/puc/sources/chroniqueslatines/>)

- l'éd. de la trad. en moyen français par Raoul de Presles de la *Cité de Dieu* de saint Augustin (ATILF)

TEI pour l'annotation linguistique et les traitements de corpus

BFM, éd. de la *Queste del saint Graal* (cf. **normes d'encodage BFM**),...

```
<w type="PRE" xml:id="w106_000001">
<choice>
<me:norm>A</me:norm>
<me:dipl><hi rend="initiale">A</hi></me:dipl>
<me:fac><bfm:lettrine size="15" sizeAct="15" color="black"
decoration="miniature">A</bfm:lettrine></me:fac>
</choice>
</w>
<w type="DETdef" xml:id="w106_000002">
<choice>
<me:norm>la</me:norm>
<me:dipl>la</me:dipl>
<me:fac>la</me:fac>
</choice>
</w>
<w type="NOMcom" xml:id="w106_000003">
<choice>
<me:norm>veille</me:norm>
<me:dipl>ueille</me:dipl>
<me:fac>ueille</me:fac>
</choice>
</w>
<w type="PRE" rend="aggl" xml:id="w106_000004">
<choice>
<me:norm>de</me:norm>
<me:dipl>de</me:dipl>
<me:fac>de</me:fac>
</choice>
```

Éditions de chartes et documents

- IRHT avec la plate-forme TELMA
(<http://www.cn-telma.fr/>) ;
- ENC : Theleme (but pédagogique), ELEC (éditions en ligne), Corpus (mise en ligne de textes numérisés sans révision du texte), Développements (outils et schémas documentés)

```

<text xml:id="morchesne_1-1" n="1.1" type="formule" subtype="article">
  <front>
    <head>Grace a plaidier par procureur</head>
    <index>
      <term type="auth_type" key="act_souv"/>
    </index>
  </front>
  <body>
    <div type="transcription" xml:lang="fro">
      <p>
        <seg function="intitulatio">Charles, par la grace de Dieu roy de France,</seg>
        <seg function="address">a tous ceulz qui ces presentes lettres verront,</seg>
        <seg function="salutatio">salut.</seg>
        <seg function="notification">Savoir faisons</seg>
        <seg function="dispositio">nous de grace especial <seg function="verb">avoir
octroyé</seg> <seg function="beneficiary">a nostre amé et feal clerc, notaire et secretaire
maistre Odart Morchesne, prieur de l'eglise collegial de Saint Aignen en Berry et penancier
d'Orleans</seg>, <note> L'église collégiale, siège de la paroisse de Saint-Aignan (Loir-et-
Cher, ch.-l. cant.), fut fondée, comme le laissent supposer des indices archéologiques, vers
le milieu du XIe siècle. Elle était dite de <q>Saint-Aignan en Berry</q> au Moyen Âge et à
l'époque moderne, car elle relevait du diocèse de Bourges (<bibl><title>Pouillés de la
province de Bourges</title>, t. I, p. 139</bibl>), [...].</note> <seg function="donatio">que
il tant cause de ses benefices [...].</seg></seg>
        <seg function="dating">Donné a Bourges le XXV<hi rend="sup">e</hi> jour de
juing, l'an de grace mil quatre cens vint et cinq, et de nostre regne le tiers.</seg>
      </p>
    </div>
  </body>
</text>

```

Structure générale d'un recueil d'actes

La structure générale d'un recueil d'actes comportera obligatoire un avant-texte avec titre et résumé (<head>, <argument>) afin d'assurer la génération d'un site web. L'élément racine portera en attribut un identifiant de fichier, unique pour l'éditeur.

<TEI>

Attributs	@xml:id
Contenu	text
Usage	Élément racine

<text>

Contenu	group,
Usage	TEI

```

<TEI>
  <teiHeader>...</teiHeader>
  <text>
    <front>
      <head>Cartulaire ...</head>
      <argument>
        <p>Petit texte pour page d'accueil</p>
      </argument>
    </front>
    <group>
      <head>titre court</head>
      <text>acte 1</text>
      <text>acte 2</text>
      <group>
        <head>Possibilité de sous section</head>
        <text>...</text>
      </group>
    </group>
  </text>

```

La TEI en France : TEI Lite

- archivage et publication des thèses, notamment à l'université de Lyon 2 (<<http://theses.univ-lyon2.fr/>>) ;
- structuration et mise en ligne du *Bulletin des bibliothèques de France* (<<http://bbf.enssib.fr/>>);
- format du portail PERSEE (<<http://www.persee.fr/>>) ;
- format de stockage et d'export pour Revues.org
<<http://www.revues.org/>> (logiciel Lodel)

TEI et notices de manuscrits ?

En France, la TEI est encore peu utilisée pour la description des manuscrits dans les institutions patrimoniales (poids pris par EAD dans les bibliothèques françaises). Elle l'est en revanche beaucoup ailleurs dans le monde :

- Portail E-Codices (Suisse) :

<http://www.e-codices.unifr.ch/fr>

- Manuscriptorium (projet Européen) :

<http://www.manuscriptorium.com/>

- Bibliothèque Vaticane, ...

et commence à l'être dans des laboratoires de recherche français (IRHT, ...).

Plan de la séance

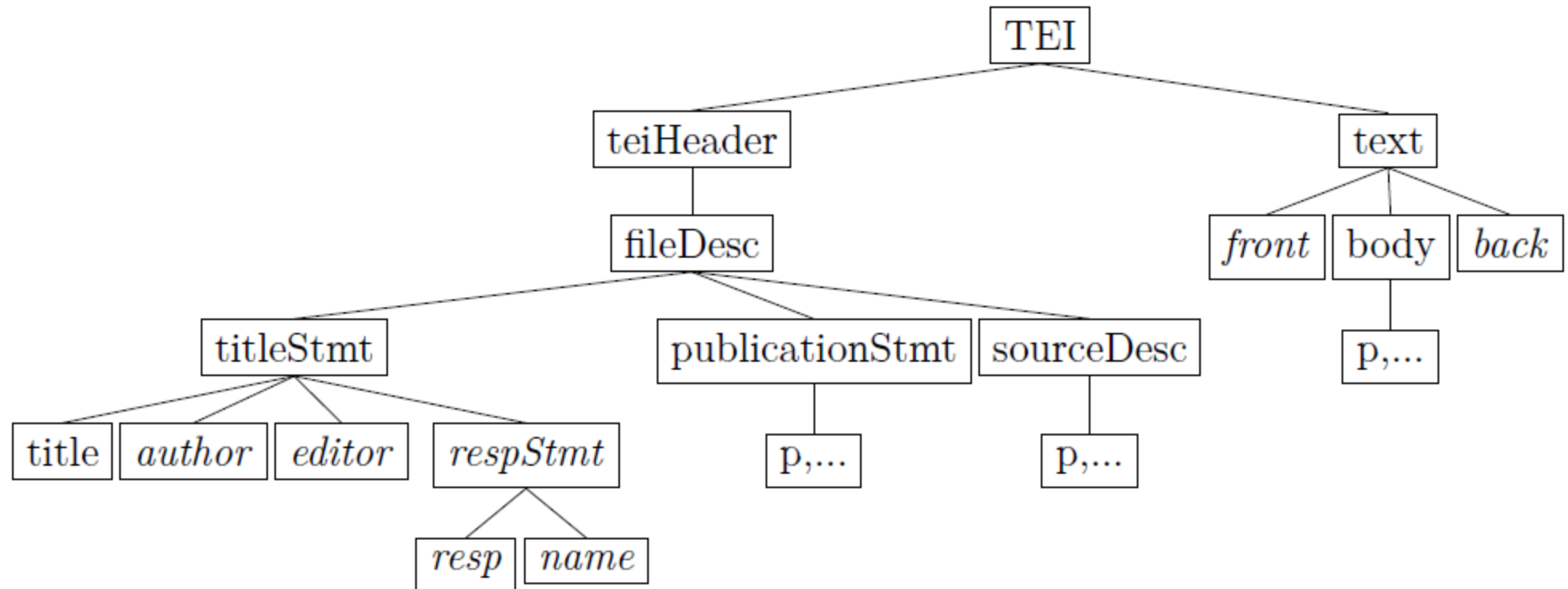
I. Il était une fois la TEI..

II. Structure de base et éléments communs

III. Outils et documentation

Un fichier TEI minimal (cf. oXygen)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```



Quelques éléments courants

- subdivisions du texte : <div>, <p> (paragraphe) ou <ab>, <lg> et <l> pour les vers, <list> pour des listes ; <s> pour un élément de type phrase ; <seg> pour un segment de texte ;
- citations : <q> ou <quote> ;
- mise en valeur du texte : <hi>, <emph> ;
- notes : <note> ;
- mots étrangers dans le contexte : <foreign> ;
- termes et gloses : <term>, <gloss> ;
- dates : <date> ;
- noms de personnes <persName>, de lieux <placeName>, noms propres <name> ...
etc.

T.P.

Proposer un encodage en TEI, utilisant la TEI Lite, du document traité la semaine dernière. Réfléchir à l'encodage des métadonnées.

Plan de la séance

I. Il était une fois la TEI...

II. Structure de base et éléments communs

III. Outils et documentation

III.0 Les technologies plus couramment employées

III.1 Outils de conversion

III.2 Feuilles de style

III.3 Outils de publication

III.4 Moteurs de recherche

III.5 Outils pour l'analyse scientifique

III.6 Outils de modélisation

III.7 Éditeurs et ergonomie

Des technologies couramment employées

- ♦ **CSS** : *Cascading stylesheets*
- ♦ **XSLT** : *eXtensible stylesheet language*, langage de transformation de fichiers XML
- ♦ **XQuery** : *XML Query language* (le SQL du XML), en conjugaison avec des bases de données natives XML (BaseX, eXistDB, Berkeley DB XML d'Oracle...), des serveurs orientés XML (MarkLogic Server)

Outils de conversion I : publication

Un langage de transformation : eXtensible Stylesheet
Language Transformation (**XSLT**)

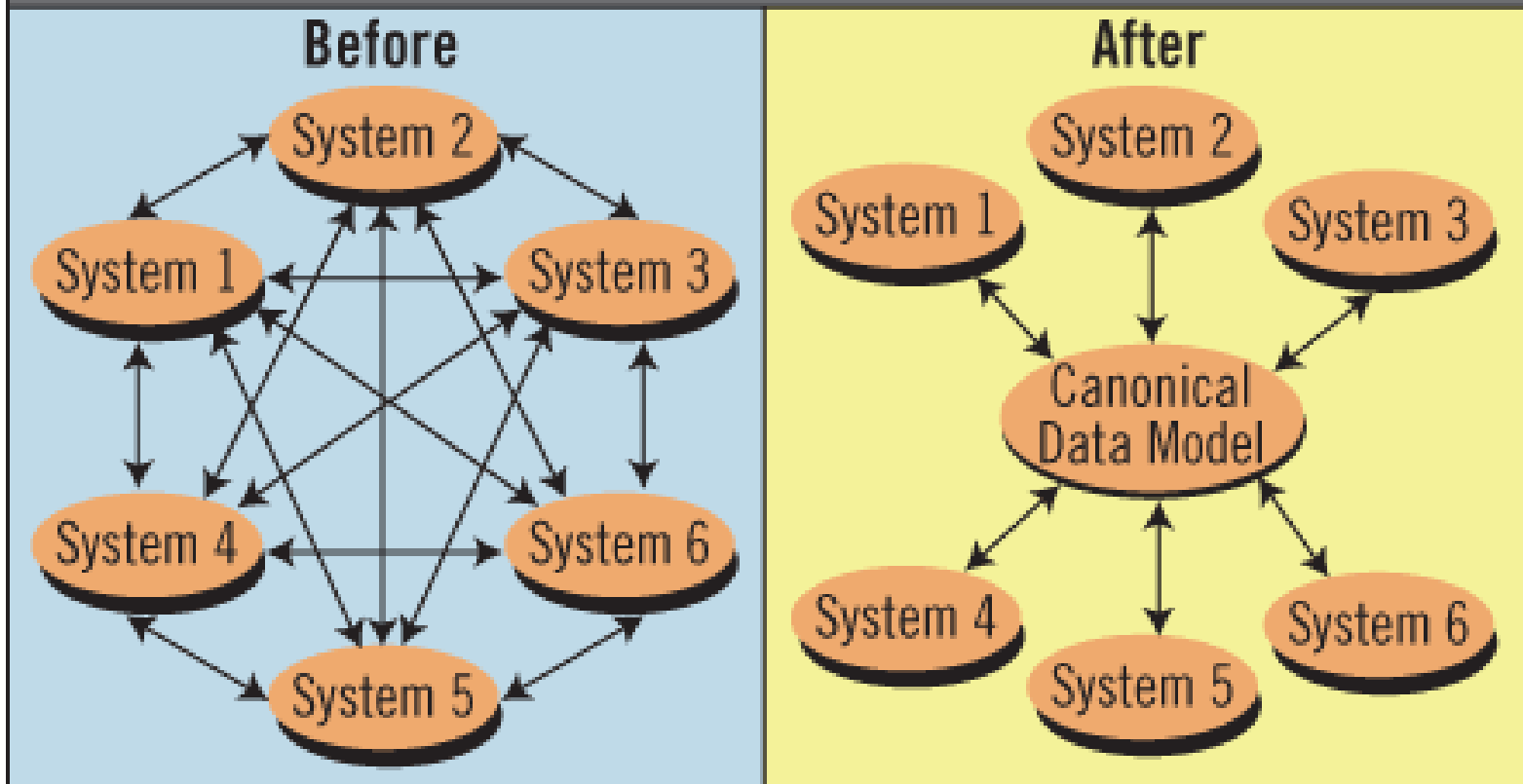
Des outils qui font usage d'XSLT pour des conversions

- XSL du Consortium TEI (dév. S. Rahtz) :
 - version de dév.: <https://github.com/TEIC/Stylesheets>
 - version intégrée au plugin TEI (l'ensemble des feuilles se trouve dans votre dossier Oxygen XML Editor 16/frameworks/tei/xml/tei)
 - documentation, <http://www.tei-c.org/release/doc/tei-xsl/>
- oXgarage (<http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient/>)
- Diple (http://developpements.enc.sorbonne.fr/diple/demo/tei_html)
- Algone (<http://algone.net/>)
- D'autres encore (EPub tools <<http://wiki.tei-c.org/index.php/Epub-tools>>),

Certains de ces outils utilisent la TEI comme format pivot (*Canonical data model*) entre plusieurs formats

Format pivot (*Canonical data model*)

Figure 1: Simplified Example of the Use of a Canonical Data Model Before and After



extr. de Steve Hoberman, « Canonical Data Model », *Information Management Magazine*, en ligne : http://www.information-management.com/issues/2007_50/10001733-1.html (consulté le 14 nov.

Outils de transformation :

II. Métadonnées de document

Problématique spécifique de la transformation des métadonnées de document du teiHeader

- ♦ MarcXML vers teiHeader, Thutmose II
Tentatives de *mappings* tei/rdf, etc.
- ♦ Cf. TEI Libraries SIG, *Best Practices for TEI in Libraries : A guide for mass digitization, automated workflows, and promotion of interoperability with XML using the TEI*, dir. Stefanie Gehrke and Kevin Hawkins, version 3.0, 2011, en ligne : <http://purl.org/TEI/teiinlibraries> (consulté en nov. 2014)

Des feuilles CSS pour mettre en forme de la TEI

- **Comprehensive CSS Stylesheet**

(Tentative de la communauté TEI de développer des conventions CSS exhaustives vis-à-vis des éléments TEI,

http://wiki.tei-c.org/index.php/Comprehensive_CSS_Stylesheet
)

- **CSS TEI d'<oXygen/>**

(dans le dossier d'oXygen, Oxygen XML Editor

`N/frameworks/tei/xml/tei/css`, développées à l'origine par Adrian Edwards et S. Rahtz)

et des centaines d'autres, pour chaque outil, projet...

Publier de la TEI : les CMS

Des CMS dédiés à la TEI ou à l'utilisation de documents XML

- **TEI Boilerplate**
- **Kiln**
- **Lodel** (revues.org)
- **EVT**
- **SDPublisher**
- **eXtensible Text Framework (XTF)**, développé par la California Digital Library

Des CMS généralistes avec des modules adaptés à la TEI

- **Drupal** : CMS généraliste très complet pour lequel des modules TEI existent, tels que **TEI-CHI** (Drupal 7), **XML Content** (Drupal 6, développement arrêté), **DARE** (à venir)
- **Omeka** : CMS spécialement créé pour les usages universitaires ou les institutions patrimoniales, avec le module **TeiDisplay**

Wordpress : module **Anthologize**, qui permet d'exporter du contenu sous une forme dégradée de TEI, mixte avec HTML

Et d'autres outils encore pour le développement TEI :

- **Apache Cocoon** (environnement de développement web utilisé notamment pour les *Guidelines*)
- **SimpleXML** : extension PHP pour manipuler des données XML

Utiliser la TEI dans un environnement dédié à l'analyse scientifique

Plateformes utilisant la TEI

- ◆ TXM
- Text Analysis Portal for Research (TAPoR)
- EHumanities Desktop (<http://www.hucompute.org/ressourcen/ehumanities-desktop>)

Outils de collation et d'affichage parallèle

- ◆ **Juxta**
- ◆ **CollateX**
- ◆ **Versioning Machine**

Outils de recherche dans des corpus

- ◆ PhiloLogic

Outils dédiés à une approche particulière

- ◆ CATMA (Analyse littéraire)
 - ◆ Morphadorner (Annotation morphosyntaxique)
- Cf. Tei Wiki, catégorie , ainsi que le Digital research tools directory (DiRT, <http://dirtdirectory.org/>)
- ◆ WebLicht (étiquetage linguistique), WordHoard, ...

Créer des personnalisations de la TEI

Des outils du consortium pour créer des personnalisations

- PizzaChef, Roma (<http://www.tei-c.org/Roma/>), Byzantium, Vesta, etc.

Des feuilles XSL (disponibles dans le plugin TEI d'<code>oXygen/</code>)

- feuille oddbyexample



Éditer de la TEI : outils et ergonomie

Des éditeurs et outils généralistes

- <oxygen/>, qui propose tout un environnement de développement XML, adapté à la TEI, et propose en outre un mode auteur d'<oxygen/> ; aussi XMLSpy, XMLMind, etc.

Module Ediarum, qui modifie le mode auteur d'<oxygen/> pour le rendre plus *user-friendly* ; module HisTEI pour les historiens et ling.

Éditeurs « WYSIWYG » ou « WYSIWYM » (cf.

<http://wiki.tei-c.org/index.php/Editors>)

- Éditeurs en ligne (*web based*) : FontoXML, Documents ODT stylés

Des outils spécialisés, dédiés à l'édition de texte par exemple, permettant de générer de la TEI

- Classical Text Editor
- ImageMarkupTool (pour annoter des parties d'image en TEI)
- TILE (Text-Image Linking Environment)

Des outils collaboratifs

- FromThePage, pour transcrire des textes dans un environnement de type wiki, et l'exporter ensuite en TEI
- Outils TEI ou non, comme TPen, Elaborate, CWRC-Writer, Textgrid, cf. Elena Spadini, *Outils numériques pour l'édition des textes : transcription et encodage*, mémoire de master 2, École nationale des

Outils spécifiques à une problématique et compatibles avec la TEI

Gestion de références bibliographiques

- ♦ RefDB
- ♦ TEIZoteroTranslator

Outils d'annotation

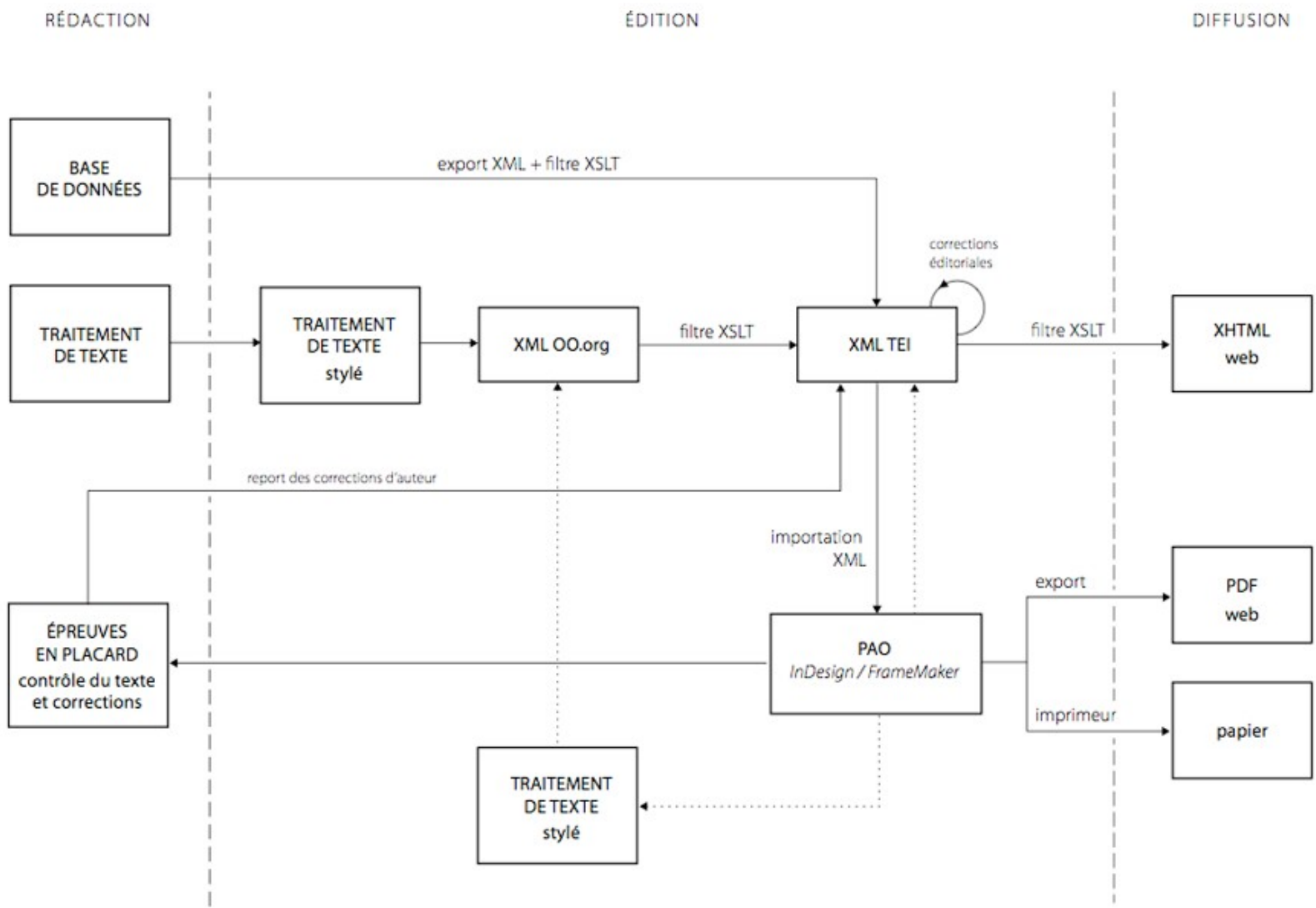
- ♦ Sacodeyl Annotator
- ♦ une variété d'outils, TEI ou non, comme Annotator, Annotation Studio, Annotopia, **Pundit**, Textus (voir J. Daniel, *Les Outils d'annotation et l'édition scientifique de corpus textuels : l'exemple du projet « Les Guides de Paris (les historiens des arts et les corpus numériques) »*, mémoire de master 2, École nationale des chartes, 2014)

Chaîne éditoriale des PU de Caen



http://scd.unicaen.fr/recherche/mrsh/document_numerique/projets/chaine_editoriale

Voir aussi la présentation de Pierre-Yves Buard, disponible en ligne, <http://www.aedres.fr/pdf/XML-CR-19juin2009.pdf>



Bibliographie indicative

En complément des références données au cours de la présentation.

Avant tout et en premier lieu, les *Guidelines* elles-mêmes

TEI CONSORTIUM, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 2014, en ligne :

<<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>> (consulté en novembre 2014).

Introductions et ressources

BURNARD (Lou), *What is the Text Encoding Initiative? : How to add intelligent markup to digital resources*, Marseille, 2014, 114 p. (Encyclopédie numérique), en ligne : <<http://books.openedition.org/oep/426>> (consulté le 12 nov. 2014).

HAROLD (Elliotte Rusty) & MEANS (W. Scott), *XML en concentré: manuel de référence*, 3^e éd., Paris, 2005 (*quelques pages sur la TEI*).

VAN DEN BRANDEN (Ron), TERRAS (Melissa) & VANHOUTTE (Edward), *TEI by Example*, <<http://www.teiByExample.org>> (consulté en nov. 2013) ;

Voir également la liste de tutoriels sur la TEI, en ligne :

<<http://www.tei-c.org/Support/Learn/tutorials.xml>>

Bibliographie: approfondissements

Pour approfondir les questions liées à la TEI et à ses évolutions, il est possible de se référer aux :

- actes des conférences annuelles du Consortium TEI, du colloque annuel *Digital Humanities*, ceux du colloque annuel Balisage ...)
- revues spécialisées en humanités numériques, comme, pour le Moyen Âge, *Digital Medievalist* (<<http://www.digitalmedievalist.org/journal/>>) ; ou la revue *Literary and Linguistic Computing* de l'université d'Oxford (<<http://llc.oxfordjournals.org/>>), ou encore *Digital Humanities Quarterly* (<<http://digitalhumanities.org/dhq/>>)

Pour une approche plus générale :

- *A Companion to Digital Humanities*, dir. Susan Schreibman, Ray Siemens, John Unsworth. Oxford, 2004. <<http://www.digitalhumanities.org/companion/>>.

Pour une histoire des débuts de la TEI :

Ide (Nancy) et Véronis (Jean), *Text encoding initiative: background and context*, Dordrecht, 1995.

Bibliographie: les outils

Pour une présentation générale des outils fonctionnant avec la TEI, se rapporter à :

TEI CONSORTIUM, *TEI Tools*, 2011, en ligne : <<http://www.tei-c.org/Tools/>>
(consulté en nov. 2014)

Pour suivre les activités en ce domaine, le SIG de la TEI consacré aux outils, qui gère notamment :

- la catégorie « Tools » du Wiki de la TEI (plus de 100 p.), qui fournit un recensement des outils principaux :

TEI CONSORTIUM, « Category:Tools », *TEI Wiki*, en ligne, <<http://wiki.tei-c.org/index.php/Category:Tools>> (consulté le 19 nov. 2014).

- une liste de diffusion (cf. TEI Consortium, *TEI Tools SIG*, <<http://www.tei-c.org/Activities/SIG/Tools/>>