

# Modélisation et structuration des données : structuration XML et XML/TEI

**Descriptif du cours :** Ce cours propose l'étude des concepts, normes et outils logiciels permettant de structurer données et documents, et rendant donc possible une exploitation informatique des documents primaires au-delà de leur signalement et de leur description. Sont étudiés XML et la modélisation XML (DTD, schémas) et la mise en œuvre du format XML/TEI pour l'édition critique des textes.

## Programme

### Éléments déjà acquis et lectures complémentaires

#### -1. Du Texte brut aux balises.

#### 0. Introduction générale à XML et à la modélisation (DTD).

##### *Lectures complémentaires :*

- *TEI Guidelines*, sect. v. « A Gentle Introduction to XML » ;
- Eliotte Rusty HAROLD & W. Scott MEANS, *XML en concentré...*, chap. 1 à 4, p. 3-78.

**I. Introduction à la Text Encoding Initiative (TEI).** A. Histoire, philosophie et principes de la TEI : présentation et historique du projet ; pourquoi et quand l'utiliser ? dans quelle perspective ? Lire les *Guidelines* et construire son modèle ; trouver de la documentation. — B. Premiers éléments d'encodage : fondamentaux et infrastructure générale de la TEI (module `tei`). C. La TEI et les outils.

##### *Lectures complémentaires :*

- *TEI Guidelines*, chap. 1 et 3 ;
- *TEI by Example*, leçons 0 et 1.

**II-III. La structure d'un document TEI :** II. Entête (header) et métadonnées du document. — III. Structuration logique et physique du texte.

##### *Lectures complémentaires :*

- *TEI Guidelines*, chap. 2 et 4 ;
- *TEI by Example*, leçon 2, ainsi que 3 -5.

**Index et entités nommées.** Identification des noms propres, réalisation des index (namesdates). — Bibliographies (cf. `core`) et références (linking).

##### *Lectures complémentaires :*

- *TEI Guidelines*, chap. 13 et 16.

## Deuxième partie du cours

### Les sources primaires et leurs transcriptions

**IV. La transcription des sources ;** représentation de la source, interventions éditoriales, régularisations (transcr).

##### *En complément :*

- lire les *TEI Guidelines*, chap. 11 (et 5) ;
- *TEI by Example*, leçon 6 ;
- terminer l'encodage de la transcription.

**V. La description des manuscrits (msdesc).** Réalisation de notices de manuscrits en TEI.

##### *En complément :*

- lire les *TEI Guidelines*, chap. 10 ;
- terminer l'encodage de la notice de manuscrit.

**VI. L'édition critique et son appareil.** Les différentes perspectives (avec exemples) ; édition à ms. de base, édition critique à appareil de variantes (*textcrit*), génétique textuelle...

**En complément :**

- lire les *TEI Guidelines*, chap. 12 ;
- *TEI by Example*, leçon 7 ;
- terminer l'encodage de l'apparat critique, en vérifier la qualité, et finir l'ensemble de l'encodage du 3<sup>e</sup> document.

**Adapter et mettre en œuvre la TEI**

**VII. Concevoir une personnalisation de la TEI avec ODD.** L'infrastructure de la TEI, la création d'une personnalisation et le langage ODD ; utilisation de Roma. Mise en place et coordination d'un projet d'édition en TEI : définition des pratiques, suivi d'un modèle commun, documentation. — Gérer des projets complexes en TEI.

**En complément :**

- lire les *TEI Guidelines*, chap. 22-23 ;
- *TEI by Example*, leçon 8 ;
- terminer le travail sur le modèle réalisé en cours.

**VIII. Les langages de schéma.** Introduction à RelaxNG et à sa syntaxe. Introduction à Schematron ; pourquoi écrire des règles Schematron dans son ODD ?

**Cas d'étude**

Les séances aborderont successivement deux cas d'étude : tout d'abord, des extraits de manuscrits médiévaux romans, pris sous l'angle de la transcription de manuscrits, de l'écriture de notices, ainsi que de l'établissement d'un appareil critique (séances IV-VI) ; ensuite, l'écriture d'un modèle documenté pour la TEI avec ODD. Ce travail s'appuiera sur des exercices personnels et des séances de travaux dirigés.

**Modalités d'évaluation**

**Devoir final :** travail d'encodage à rendre en février 2018 ; sujets tirés de cas d'encodage concrets et pouvant s'insérer dans des projets d'édition en cours.

**Éléments bibliographiques**

*En complément des références données au cours des présentations.*

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

TEI CONSORTIUM, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>>.

Pour référence, la spécification XML du W3C,

W3C, *Extensible Markup Language (XML) 1.0 (Fifth Edition)*, recommandation du W3C du 26 novembre 2008 (1<sup>re</sup> version, 14 novembre 1996, <http://www.w3.org/TR/WD-xml-961114.html>), dir. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, *et al.*, en ligne : <http://www.w3.org/TR/xml/>.

Enfin, l'article fondateur sur la définition du « texte » (*OHCO*),

RENEAR (A.), MYLONAS (E.), DURAND (D. ), « Refining our notion of what text really is: the problem of overlapping hierarchies », dans *Research in Humanities Computing*, dir. Nancy Ide, Susan Hockey, 1996, Oxford, en ligne : <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>.

**Introductions et tutoriels**

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

HAROLD (Elliott Rusty) & MEANS (W. Scott), *XML en concentré: manuel de référence*, 3<sup>e</sup> 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 :

TEI CONSORTIUM, *Tutorials*, en ligne : <<http://www.tei-c.org/Support/Learn/tutorials.xml>> (consulté en nov. 2015).