

Structuration des données et des documents : balisage XML

Jean-Baptiste Camps

▶ **To cite this version:**

Jean-Baptiste Camps. Structuration des données et des documents : balisage XML. Master. Structuration des données et des documents : balisage XML, Paris, France. 2018. cel-01706530

HAL Id: cel-01706530
https://shs.hal.science/cel-01706530

Submitted on 12 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Jean-Baptiste Camps (jean-baptiste.camps@enc.sorbonne.fr) – 2015-2016 – École nationale des chartes

Structuration des données et des documents : balisage XML

Descriptif du cours : ce module, qui est programmé au milieu du semestre, élargit le champ déjà abordé dans le module relatif aux métadonnées, pour proposer 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. Ce module se compose de cours magistraux étroitement associés à de nombreuses heures de travaux dirigés. Des exercices sont à faire entre chaque cours.

Programme

Cours

0. ***Introduction générale à XML.***

En complément :

- *TEI Guidelines*, sect. v. « A Gentle Introduction to XML » ;
- Elliotte 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 TEI ; pourquoi et quand utiliser la TEI, dans quelle perspective (exploitation, etc.) ; lire les Guidelines et construire son modèle, outils ; trouver de la documentation. — B. Premiers éléments d'encodage : fondamentaux et infrastructure générale de la TEI (module `tei`). C. La TEI et les outils.

En complément :

- lire les *TEI Guidelines*, chap. 1 et 3 ;
- *TEI by Example*, leçons 0 et 1 ;
- terminer l'encodage du document abordé en cours.

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

En complément :

- lire les *TEI Guidelines*, chap. 2 et 4 ;
- *TEI by Example*, leçon 2, ainsi que 3 -5 ;
- terminer l'encodage du `<teiHeader/>` du document abordé en cours.

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

Sur ce sujet, pour lequel le temps risque de manquer durant les cours, il est nécessaire de se former individuellement, en ayant recours à

- lire les *TEI Guidelines*, chap. 13 et 16 ;
- Exercice à faire :
  - créer une bibliogr. encodée en TEI (par exemple, via un export de Zotero) et l'intégrer à un document ;
  - encoder les noms de personne et de lieu du document en cours de traitement ;
  - terminer l'encodage de la lettre.

IV-V. Les sources primaires et leur transcription. Transcriptions et éditions ; interventions éditoriales (`transcr`). — La description des manuscrits (`msdesc`).

En complément :

- lire les *TEI Guidelines*, chap. 10-11 (et 5) ;
- *TEI by Example*, leçon 6 ;
- terminer l'encodage de la notice de manuscrit et de la transcription.

V. L'édition critique et son apparat. Les différentes perspectives (avec exemples) ; édition à ms. de base, édition critique à apparat 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 3e document.

VI. Adapter et mettre en œuvre la Tei. Schémas, DTD, choix des modules et la syntaxe ODD (`tagdocs`) ; 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.

Cas d'étude

Les séances aborderont successivement trois cas d'étude : un document factice servant à illustrer quelques principes généraux (séances 0-I) ; des correspondances contemporaines (séances II-III), des extraits de manuscrits médiévaux romans (séances IV-VI). Ce travail s'appuyera sur des exercices personnels et des séances de travaux dirigés.

Modalités d'évaluation

Devoir final : travail d'encodage à rendre en mars 2016 ; 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 (1re 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/oep/426> (consulté le 12 nov. 2014).

Harold (Elliotte Rusty) & Means (W. Scott), *XML en concentré: manuel de référence*, 3e 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).