



Séance 3

Décrire les documents patrimoniaux

La Text Encoding Initiative

Correction exercice

```
<!ELEMENT e1 (e11 | e12)>
```

L'élément **e1** contient :

SOIT l'élément **e11** une fois

SOIT l'élément **e12** une fois

```
<!ELEMENT e1 (e11 | e12)+>
```

L'élément **e1** contient une ou plusieurs fois :

SOIT l'élément **e11**

SOIT l'élément **e12**

```
<!ELEMENT e1 (e11+ | e12+)>
```

L'élément **e1** contient :

SOIT l'élément **e11** une ou plusieurs fois

SOIT l'élément **e12** une ou plusieurs fois

```
<!ELEMENT e1 (e11 | e12+)>
```

L'élément **e1** contient :

SOIT l'élément **e11** une fois

SOIT l'élément **e12** une ou plusieurs fois

```
<!ELEMENT e1 ((e11 | e12)+ | e13+)>
```

L'élément **e1** contient :

SOIT l'élément **e11 OU e12** une ou plusieurs fois

SOIT l'élément **e13** une ou plusieurs fois

A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter blue, creating a cross-like shape.

Logiciel Oxygen

Raccourcis, boutons et mise en place TEI

Raccourcis Oxygen

Ctrl + E	Cmd + E	Entourer le texte sélectionné avec des balises
Ctrl + G	Cmd + G	Trouver la prochaine occurrence du texte sélectionné
Ctrl + + / -	Cmd + + / -	Zoom avant / arrière
Alt + Shift + D	Ctrl + Alt + D	Découper l'élément en deux en insérant une balise fermante et ouvrante à l'endroit du curseur

Boutons Oxygen



Auto-indentation



Tester la validité



Appliquer un scénario de transformation



Afficher le XPath correspondant à l'endroit du curseur



Visualisation

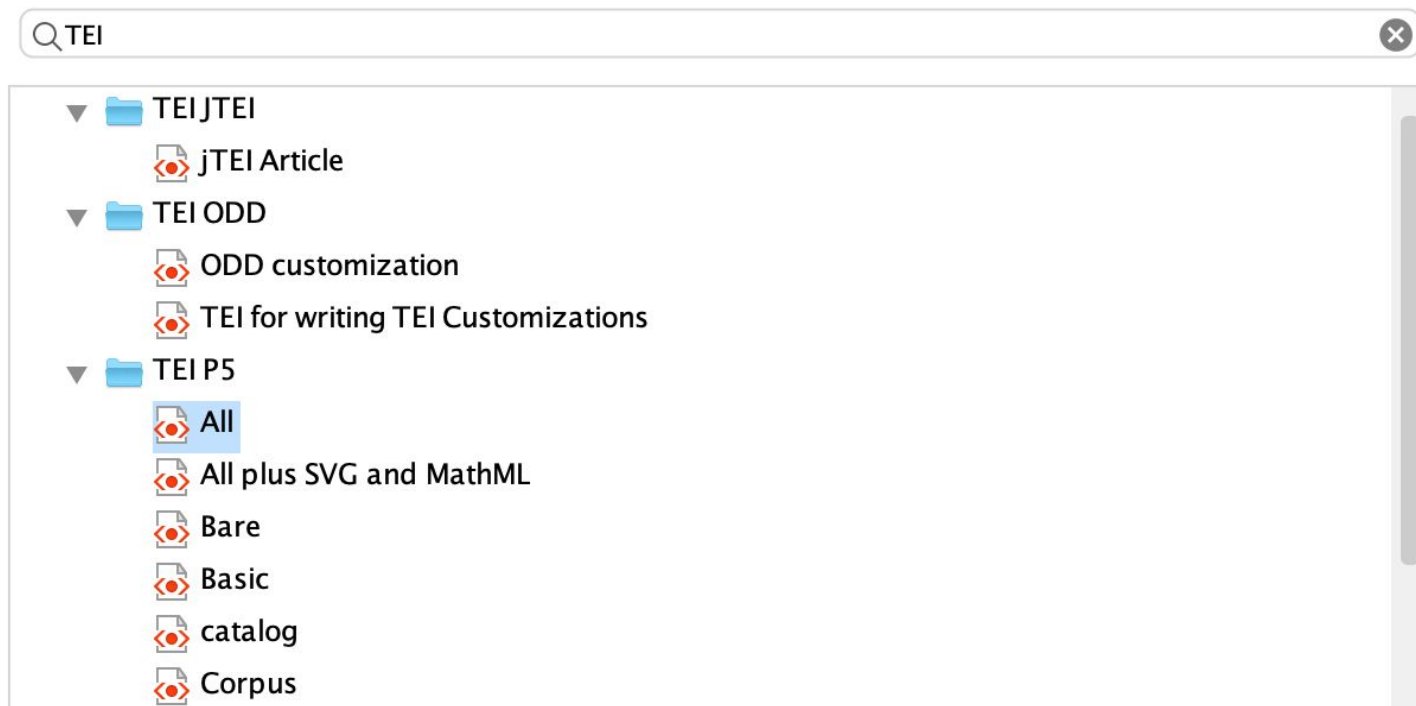
Texte	Encodage brut avec balises
Grille	Imbrication sous forme de tableau
Auteur	Visualisation standard de l'encodage

Texte

Grille

Auteur

Nouveau fichier TEI dans Oxygen



A decorative graphic on the left side of the slide consisting of two overlapping squares. The bottom-left square is a dark blue, and the top-right square is a lighter, medium blue.

Pourquoi la TEI ?

Contexte de création et objectif de la
Text Encoding Initiative

Qu'est-ce que la TEI ?

Lou Burnard et Marjorie Burghart, *Qu'est-ce que la Text Encoding Initiative ?*, 2015

"La TEI a été d'abord développée, il y a plus de trente ans, comme un projet de recherche dans le champ alors émergent du « *Humanities computing* ». L'idée originelle était de proposer un **ensemble de recommandations** sur la façon dont les chercheurs devraient créer des **ressources textuelles « lisibles par ordinateur »**, qui soient **adaptées aux besoins de la recherche** – dans la mesure où un consensus existait sur le sujet –, mais qui soient également extensibles, puisque ces besoins changent et évoluent."

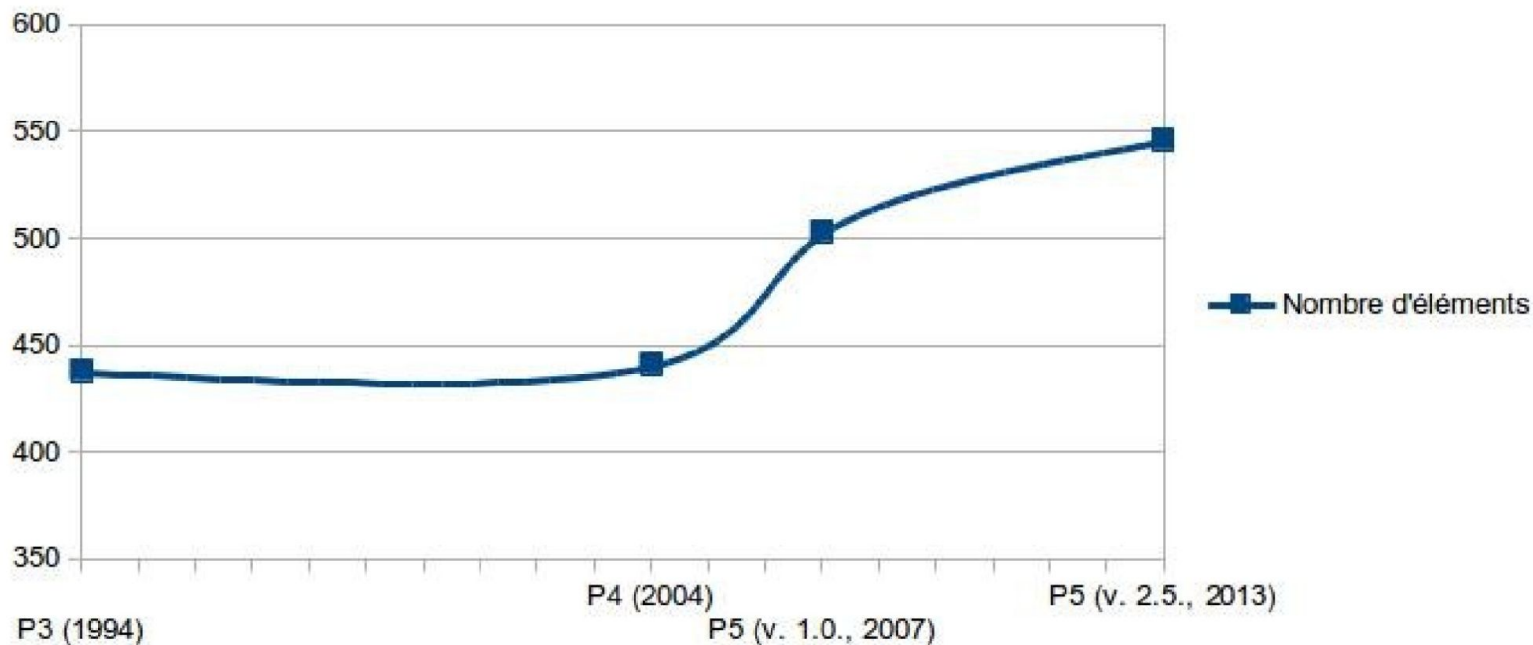
Définition

TEI est un set de balises prédéfini et documenté dans les *TEI guidelines* qui permet de procéder à une description « scientifique » et « sémantique » d'un texte.

Historique

1987	Établissement de la <i>Text Encoding Initiative</i>
1990	<u>TEI P1</u> , dir. Michael Sperberg-McQueen et Lou Burnard
1992-93	<u>TEI P2</u> , expansion
1994	<u>TEI P3</u> , première version complète
2000	Naissance du <u>TEI Consortium</u>
2001-04	<u>TEI P4</u> , introduction du XML
2007	<u>TEI P5</u> , abandon du SGML

Évolution du nombre d'éléments TEI



Principe de Poughkeepsie (1987)

1. Provide a **standard format for data interchange** in humanities research
2. Suggest **principles for the encoding of texts** in the same format
3. Define a **recommended syntax** for the format and a **metalanguage for the description of text-encoding schemes** : 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

TEI > Word

Quel avantage y a-t-il à apprendre le XML ou la TEI quand les traitements de textes existent ?

En XML TEI, **on sépare la question de la présentation de celle du fond et de la structure**. Ce n'est plus la mise en forme que l'on va coder, mais le contenu logique du document.

The slide features two solid blue rectangles: a tall, narrow one on the left and a shorter, wider one at the bottom left.

Penser le texte

La TEI fournit un cadre pour penser la nature du texte : elle constitue une sorte d'**encyclopédie de notions textuelles** associées à des balises et des règles sur la façon dont elles peuvent être combinées.

Points forts de la TEI

Sémantique

Non pas centré sur la mise en page mais sur le contenu intellectuel

Conçu par/pour les chercheurs

Objectif d'aider les chercheurs à créer des ressources textuelles adaptées aux besoins de la recherche

Interopérable

Indépendant de tout environnement logiciel, idéal pour l'échange et la collaboration

Objet d'encodage

Visuel

Aspect physique (codicologie, manques, foliotation, etc.) et aspects paléographiques (abréviations et forme développée, etc.)

Sémantique

Toponymes, anthroponymes, dates, balisage des types de discours, désambiguïsation des références textuelles

Analytique

Correspondance des parties du texte à une grille d'analyse (taxonomie, thésaurus, etc.)

Niveau d'encodage

Close reading

Attention méticuleuse au texte,
description détaillée du
document source

→ Encodage extensif

Distant reading

Considération d'un corpus dans
son ensemble pour en étudier
les généralités

→ Encodage minimal

TEI as data

Le *TEI Consortium*

La communauté TEI est animée par le *TEI consortium*, fondation interdisciplinaire à but non lucratif. Il se compose des unités suivantes :

- [*TEI Board of Directors*](#)
- [*TEI Technical Council*](#)
- [*Membres institutionnels et individuels*](#)
- [*TEI Workgroups*](#)
- [*Special Interest Groups*](#) :
 - [*TEI Manuscripts Special Interest Group*](#)
 - [*Correspondence SIG*](#)

La communauté TEI

- Des listes de diffusion : [TEI-L mailing list](#), [TEI-FR](#)
- Un wiki : [TEIWiki](#)
- Des congrès mensuels/annuels : [TEI Conference](#)
- Une revue : [Journal of the Text Encoding Initiative](#)

A decorative graphic on the left side of the slide consisting of two overlapping squares: a dark blue square at the bottom and a lighter blue square on top of it.

Fonctionnement de la *Text Encoding Initiative*

Comment utiliser la TEI



TEI ≠ schéma



TEI (All) n'est pas un schéma à proprement parler, mais plutôt un **framework**, utile à la conception de son propre schéma. Il est fortement déconseillé d'utiliser un schéma contenant l'intégralité de la TEI.

Implémentation

Un modèle pour de nombreuses applications :

Modèle abstrait

Ensemble des concepts
et les relations entre ces
concepts

Implémentation

Retranscription des
contraintes formelles par le
biais d'un schéma de
validation

Aspect modulaire

Les *Guidelines* de la TEI définissent plusieurs centaines de concepts différents, avec les instructions détaillées sur les éléments qui peuvent être utilisés pour les représenter. La plupart des documents TEI n'a besoin que d'une petite partie de ce qui est fourni.

Les projets d'encodage utilisent bien souvent un **sous-ensemble très restreint de la TEI** : une documentation restreignant l'ensemble utilisé peut être rédigée à l'aide d'outils comme Roma

Tei Lite =

designed to meet

*“90% of the needs of
90% of the TEI user
community”*

Les modules

Les modules sont des ensembles cohérents de balises

Une base commune

Modules obligatoires : pour renseigner les métadonnées principales et couvrir les balises les plus courantes

Des compléments

Modules additionnels qui correspondent à l'encodage d'un type de texte (poésie, dictionnaires, etc.) ou d'un type de données et d'informations (apparat critique, éléments graphiques, etc.)

Modules obligatoires

<u>tei</u>	définition des classes, macros et types de données
<u>textstructure</u>	éléments de base pour structurer un texte de type livre
<u>core</u>	éléments incontournables pour tout type de document
<u>header</u>	en-tête TEI (métadonnées du document)

Utiliser la TEI

Pour utiliser un set de balises TEI, il faut déclarer l'espace de nom TEI (*namespace*) dans l'élément racine du document XML grâce à une adresse URI qui pointe vers une description du set de données.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
```

Exercice

**Convertir en XML TEI
l'encodage de la carte
postale**

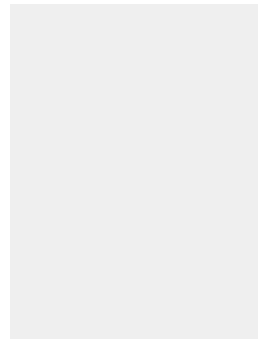
Convertir du XML en XML TEI

Convertir l'encodage de la carte postale en vous appuyant sur les guidelines TEI concernant le module [Core](#) et [Default text structure](#) :

[Addresses](#) / [Division du texte](#)

Exercice

**Convertir en XML TEI
une lettre**



Convertir du XML en XML TEI

Convertir l'encodage de [cette lettre des liaisons dangereuses](#) en vous appuyant sur la [documentation des balises suivantes](#)

```
<opener> <closer> <salute> <dateline> <date>  
<persName> <placeName> <head> <div> <p>
```

Exercice



**Encoder en XML TEI un
poème de Baudelaire**

Encoder un poème

Choisissez un poème des *Fleurs du mal* en vous appuyant sur la documentation TEI pour l'encodage des vers

Exercice



**Encoder en XML TEI une
scène de théâtre**

Encoder une scène de théâtre

Encoder [ce passage du *Misanthrope*](#) en vous appuyant sur la [documentation TEI pour l'encodage du théâtre](#)

```
<castList> <castItem> <speaker> <stage> <head>  
<div> <sp> <l>
```