

Une carte routière des TEI Guidelines

Les Cinq Etapes du chemin de l'Eveil TEI-ODD

- ① Modélisation : quels sont les buts et les objets de votre projet?
- ② Orientation : est-ce que vos objets sont déjà décrits par la TEI?
- ③ Déclaration : comment créer une specification TEI-conforme?
- ④ Documentation : comment documenter vos pratiques TEI ?
- ⑤ Validation : comment valider vos efforts ?



Modelisation



Ceci n'est pas une pipe.

La modelisation de vos données est un préalable essentiel. Que vous vous serviez de UML, de RDBMS, de SKOS, ou quoique ce soit d'autre, si vous n'avez pas un modèle explicite des choses que vous souhaitez gérer, vous rencontrerez de grandes difficultés.



Comment s'orienter?

Les Guidelines, imprimés ou en ligne, se divisent en deux grandes parties :

- une séquence de 23 chapitres, dont la plupart correspondent à un module
- des listes alphabétiques de définitions exhaustives de classes (146), d'éléments (521), d'attributs (), de macros, de types de données, etc.

Comment savoir quel élément (etc) choisir pour telle ou telle entité identifiée dans votre analyse préalable ? Comment savoir que vous avez besoin d'une 'licorne'?

La triste vérité...

- Désolé, mais il n'y a aucune méthode scientifique ; aucun raccourci pour cela...
- Il faut étudier les exemplaires et les définitions pour savoir si cet élément TEI si prometteur s'applique en effet à votre cas
- comme c'est le cas pour d'autres types de mapping...

Pour chacun(e) des entités/concepts identifiés dans votre modèle, il faut décider :

- quel objet TEI lui correspond parfaitement? (s'il existe)
- quel objet TEI lui ressemble et quelles petites modifications seraient nécessaires pour qu'il lui corresponde ?
- quelle lacune de la TEI votre analyse vous permettra de corriger (oui, elles existent les licornes !)

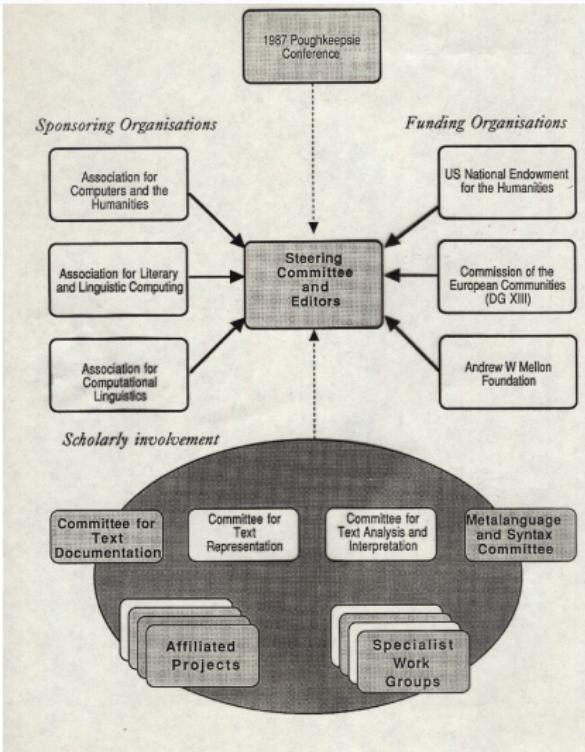


Phases de la TEI

- 1988 - 1990 Développement cycle 1 : Production de TEI P1
(consultation avec une cinquantaine d'experts mondiaux)
- 1990 - 1992 Développement cycle 2 : production des fascicules TEI P2 (en consultation avec plusieurs groupes de travail... un ensemble de quelques centaines d'experts)
- 1993 - 1994 Intégration des fascicules P2 comme TEI P3 : la version "finale"
- 1995 - 1999 Promotion et prise en main (pas financée !)
- 2000 **Etablissement du Consortium TEI**
- 2001 - 2003 Conversion de P3 en XML (TEI P4), lancement d'une révision complète qui apparaîtra comme TEI P5
- 2003 - ? TEI P5 sur sourceforge ; des révisions régulières jusqu'à présent (on est à la version 2.6.0)



La TEI jadis

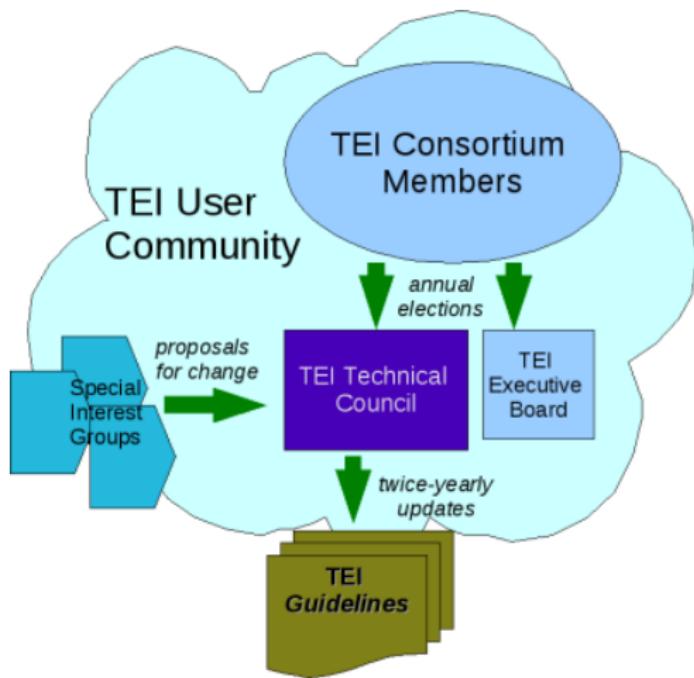


Qu'est-ce que la TEI aujourd'hui ?

- Consortium international établi en 2000 (voir <http://www.tei-c.org/>)
- Un ensemble de *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 utilisant différents langages de 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, organisé en système de classes, voire en **ontologie**



Organisation de la TEI



La TEI facilite un balisage 'intelligent'

La TEI s'applique à l'encodage des...

- composants structuraux et fonctionnels d'un texte
- transcriptions diplomatiques des sources historiques, des images, des annotations
- liens, correspondances, alignements
- données et entités : par exemple de temps, personnes, lieux ou événements
- annotations peritextuelles et métatextuelles (correction, suppression, ajouts)
- analyses linguistiques
- métadonnées de plusieurs types
- ... et définitions formelles de schéma XML !

On est vraiment gâté par ce choix



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

Les TEI Guidelines

Un tour d'horizon très rapide du contenu des TEI *Guidelines*...

Dans ses 1 400 pages imprimées, vous trouverez :

un lexique de 555 éléments
des règles d'usage et des contraintes

- exprimées en langue naturelle
- exprimées en langages formels

beaucoup de discussions notamment de plusieurs exemples
d'utilisation



La liste des chapitres (1)

À part les pièces liminaires, les *Guidelines* contiennent

- 1. The TEI Infrastructure (tei)
- 2. The TEI Header (header)
- 3. Elements Available in All TEI Documents (core)
- 4. Default Text Structure (textstructure)
- 5. Representation of Non-standard Characters and Glyphs (gaiji)
- 6. Verse (verse)
- 7. Performance Texts (drama)
- 8. Transcriptions of Speech (spoken)
- 9. Dictionaries (dictionaries)
- 10. Manuscript Description (msdescription)
- 11. Representation of Primary Sources (transcr)
- 12. Critical Apparatus (textcrit)

...



La liste des chapitres (2)

...

- 13. Names, Dates, People, and Places (namesdates)
- 14. Tables, Formulæ, and Graphics (figures)
- 15. Language Corpora (corpus)
- 16. Linking, Segmentation, and Alignment (linking)
- 17. Simple Analytic Mechanisms (analysis)
- 18. Feature Structures (declarefsiso-fs)
- 19. Graphs, Networks, and Trees (nets)
- 20. Non-hierarchical Structures
- 21. Certainty and Responsibility (certainty)
- 22. Documentation Elements (tagdocs)
- 23. Using the TEI



De plus, en annexes ...

- une bibliographie
- définition de référence pour tout
 - élément, plus ses attributs locaux
 - classe d'attribut, avec les attributs qu'elle fournit
 - macro (patron)
 - datatype

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

Niveaux de validation

Un document TEI-XML doit:

- ① respecter les règles syntaxiques d'XML;
- ② être valide par rapport à un schéma quelconque;
- ③ respecter la sémantique définie de chaque élément TEI utilisé

Un schéma exprime d'une manière formelle et validable par logiciel une partie importante de ces règles

Un TEI ODD nous permet de construire un schéma adapté à nos besoins



Comment effectuer notre choix de balises?

Nous aurons besoins de plusieurs choses :

- un schéma formel (en langage informatique tel que DTD, RELAXNG, W3C Schema, Schematron) qui peut contrôler :
 - quelles balises sont disponibles ?
 - dans quels contextes ?
 - avec quels attributs ?
 - avec quelles valeurs ?
 - en respectant quelles contraintes ?
- une documentation pour expliquer nos principes éditoriaux, nos principes de choix de balises, etc. aux utilisateurs/developpeurs :
 - en plusieurs langues naturelles
 - en plusieurs formats bureautiques (PDF, Word, HTML, epub...)
- des outils informatiques pour transformer et gérer nos données XML

Propositions de la TEI

Ayant elle-même ces mêmes besoins, la TEI vous propose :

ODD Un vocabulaire XML pour définir les vocabulaires XML
Roma

Un logiciel pour traîter (ou générer) des documents exprimés dans ce vocabulaire, pour en faire ressortir

- des schémas RELAXNG, DTD, etc.
- des manuels "mode d'emploi" en HTML, PDF etc.

OxGarage

Un logiciel pour effectuer des transformations documentaires, par ex

- de DOCX en TEI, et l'inverse
- de HTML en TEI, et l'inverse

Stylesheets Un ensemble de feuilles de style XSLT très générique pour la conversion des documents XML TEI; utilisé par les deux logiciels déjà cités, et par oXygen

