

# Initiation à l'édition scientifique numérique

Adapter son schéma de modélisation des données à son édition numérique



Séance 2 Édition électronique

Master mention Humanité numériques Rennes 2, 07/12/2021

# Les éditions scientifiques

Zoom sur les éditions “paléographiques” et les éditions critiques

---

# Les éditions scientifiques

Les éditions ont pour but de rendre accessible et compréhensible un texte pour un lectorat donné.

Il existe différents types d'édition scientifique (liste non exhaustive) :

- Les **éditions génétiques** → rendre intelligible la genèse d'une création artistique (cf. [Madame Bovary : l'histoire du texte à travers ses brouillons](#)).
- Les **éditions « paléographiques »** → rendre compte de la matérialité du texte, de sa graphie, de ses signes d'abréviations et de ponctuations, de la disposition des rubriques, de la coupure des lignes et des pages (cf. [Le Didascalicon d'Hugues de Saint-Victor](#)).
- Les **éditions critiques** → établir un texte définitif à partir de différentes versions disponibles, tout en faisant un travail de modernisation et d'annotation du texte et en ajoutant un appareil critique permettant de contextualiser les choix d'édition (cf. [La Queste del saint Graal](#)).

# Les éditions scientifiques

## 1. Les enjeux d'une édition « paléographique »

*Transcription **graphétique**, distinguant précisément chaque variante graphique [...] Transcription **graphémique** [...] respectant la graphie (« spelling ») de chaque mot [...] mais sans tenir compte des éventuelles variations graphiques de ces différentes lettres [...] Transcription **régularisée**, qui normaliserait la graphie des mots, développerait les abréviations [...] On pourrait ajouter Transcription **modernisée** (qui s'apparente, en fait, à une traduction), destinée à la diffusion d'un texte auprès d'un grand public en note : Conceptuellement Robinson et Solopova prévoient l'existence d'un niveau de transcription dit « **graphique** », rendant compte de chaque espace du manuscrit original, Jacques André – Rémi Jimenes, *Transcription et codage des imprimés de la Renaissance. Réflexions pour un inventaire des caractères anciens*, « Document numérique », 16.3 (2013), pp. 113-139 [<https://halshs.archives-ouvertes.fr/halshs-00983575>], p. 117.*

# Les éditions scientifiques

Avant d'encoder, définir les enjeux :

- Identifier et délimiter les phénomènes à signaler ;
- Définir les futurs usages de l'édition ;
- Choisir son protocole d'encodage.

# Les éditions scientifiques

## **Les éléments que l'on veut reproduire :**

- Mise en page ;
- Mise en valeur typographique ;
- Graphie originale du texte.

## **Les éléments que l'on veut apporter :**

- Résolutions des abréviations ;
- Régularisations des graphies ;
- Corrections (si nécessaires).

# Les éditions scientifiques

- **Les solutions TEI**

Cf. le set pour la [structuration des textes](#) dans les TEI Guidelines.

Quelques balises courantes :

- **<lg>** contient un groupe de vers fonctionnant comme une unité formelle, par exemple une strophe, un refrain, un paragraphe en vers ;
- **<l>** contient un seul vers ;
- **<lb>** marque le début d'une nouvelle ligne ;
- **<pb>** marque le début d'un nouveau folio (**@facs** permet de lier l'encodage à une image du folio) ;
- **<cb>** marque le début d'une nouvelle colonne dans une page.

# Les éditions scientifiques

Cf. la [description des sources primaires](#) dans les TEI Guidelines

Quelques balises courantes :

- Développements d'abréviation

*Exemple :* `<choice><abbr></abbr><expn></expn></choice>`

- Régularisations

*Exemple :* `<choice><orig></orig><reg></reg></choice>`

- Corrections

*Exemple :* `<choice><sic></sic><corr></corr></choice>`



# Les éditions scientifiques

Dans une édition numérique, il est conseillé d'utiliser les **caractères UTF-8**. Il existe des [fontes spécialisées](#) dans la représentation des manuscrits (cf. **MUFI**, Medieval Unicode Font Initiative).

Pour les caractères qui n'existent pas dans toutes les fontes, passer par **l'entité hexadécimale XML**.

Astuce : [convertisseur](#) pour les caractères.

# Les éditions scientifiques

- **Traitement des variations graphiques dans un fichier TEI**

Cf. [Le module gaiji](#). *Exemple :*

```
<encodingDesc> <charDecl>
    <char xml:id='q1'>
        <charName>Q SEMICOLON</charName>
        <charProp>
            <localName>entity</localName>
            <value>&#61868;</value>
        </charProp>
    </char>
</charDecl> </encodingDesc>
```

[...] <l>Niule cose non la pouret om<g ref="#q1">que</g> pleier.</l>

# Les éditions scientifiques

Cf. La déclaration d'entités (Nom de fichier vers ppt DTD)

La déclaration se fait dans une DTD interne, soit dans une DTD externe.

*Exemple :*

```
<!DOCTYPE TEI [  
  <!ENTITY pbarre-pre  
    "<choice><abbr>&#58981;</abbr><expn>pre</expn></choice>">  

```

# Les éditions scientifiques

- **Lier son encodage à des zones d'un fac-similé**

Cf. [11.1 Digital Facsimiles](#) dans les TEI Guidelines

- Déclarer un élément <facsimile> dans l'élément racine TEI ;
- Déclarer en élément <surface>, enfant de <facsimile> ;
- Déclarer en élément <graphic>, enfant de <surface>, pour l'image ;
- Déclarer des éléments <zone>, enfants de <surface>, pour chacune des parties de l'image à identifier.

*Exemple :*

```
<facsimile> <surface>  
    <graphic url='adresse_image' width='largeur' height='longueur' />  
    <zone xml:id=" identifiant" ulx="coordonnée " uly="coordonnée"  
lrx="coordonnée" lry="coordonnée"/>  
</surface> </facsimile>
```

# Les éditions scientifiques

Les coordonnées dans <zone> :

Astuce : [TElzon](#) pour les coordonnées.

@	Définition
ulx	la valeur x de l'abscisse du coin supérieur gauche d'un rectangle
uly	la valeur y de l'ordonnée du coin supérieur gauche d'un rectangle
lrx	la valeur x de l'abscisse du coin inférieur droit d'un rectangle
lry	la valeur y de l'ordonnée du coin inférieur droit d'un rectangle

# Les éditions scientifiques

- **Décrire les sources manuscrites**

- EAD (Encoded Archival Description)

EAD est une DTD (un schéma d'utilisation) de XML.

- 2000 : encodage en XML EAD du catalogue général des manuscrits des bibliothèques publiques de France.
    - 2010 : la DeMarch recommandation est publiée. Attention : [DeMArch](#) (Description de MANuscrits ou de fonds d'Archives conservés dans les bibliothèques) est une règle de description indépendante du format informatique.
    - 2012 : deux tiers des manuscrits et documents d'archives conservés par les bibliothèques françaises sont décrits en EAD (« BnF archives et manuscrits », Calames, le sous-domaine « Manuscrits » du Catalogue collectif de France CCFr). Cf. [Inventaire du fonds du Théâtre de Poche-Montparnasse \(1942-2011\)](#).

# Les éditions scientifiques

- Les éléments TEI de description des sources : <msDesc>

Cf. 10.2 [Manuscript Description Element](#) dans les TEI Guidelines et exemple d'un <msDesc> quasi-complet sur [le github d'Arianne Pinche](#).

Naissance du <msDesc> :

- 1999-2001 : projet européen MASTER (*Manuscript Access through Standards for Electronic Records*) ;
- 2001 : publication de la DTD Master, personnalisation de la TEI complétée par l'ajout d'un élément msDescription, et d'autres éléments spécialisés qui représentent un enrichissement des possibilités pour la description des manuscrits.
- 2007-2009 : lancement du projet ENRICH (*European Networking Resources and Information concerning Cultural Heritage*).

# Les éditions scientifiques

## ➤ Que choisir : EAD ou TEI ?

Extrait de Florent Palluault, « Informatiser des descriptions complexes : l'utilisation de l'EAD et de la TEI pour les manuscrits et les livres anciens en France », *IFLA 2012 (Helsinki)*, [en ligne : <http://conference.ifla.org/past/2012/212-palluault-fr.pdf>].

**"La nécessité d'une arborescence est moindre pour ces documents que pour des ensembles d'archives, et la sémantique EAD n'est pas aussi développée que celle de la TEI-P5, utilisée pour les manuscrits médiévaux dans plusieurs pays d'Europe. »**

**"Au-delà des manuscrits et archives, l'EAD est pertinente pour tout ensemble de documents hiérarchisés, quelle que soit leur nature. Elle permet de visualiser aisément la structure de fonds ou de collections composites** comme celles du Département des Arts du Spectacle de la BnF, qui comprennent fréquemment scénarios manuscrits, programmes imprimés, costumes, objets de décor, etc."



# Les éditions scientifiques

## 2. Encoder une édition critique

- Les enjeux scientifiques d'une édition numérique

« L'outil numérique questionne non pas tant le livre, ou même le texte, que nos modes habituels de perception de la littérature. L'outil numérique, parce qu'il questionne notre lecture et le mode même de la lecture, ne ferme pas la porte à la littérature, il ouvre en réalité des fenêtres sur des formes jusqu'ici restées peut-être plus dans l'ombre de l'analyse littéraire. », BUREAU, Bruno, « Quelques réflexions sur la notion de littérature à partir de l'édition numérique de commentateurs anciens », *Interférences. Ars scribendi*, janvier 2012, [En ligne : <http://journals.openedition.org/interferences/186>].

« On ne peut pas connaître l'œuvre sans les documents. De même, on ne peut comprendre les documents sans une compréhension de l'œuvre qu'ils génèrent. On peut en tirer le principe suivant : une édition scientifique doit éclairer autant que possible chaque aspect du texte, aussi bien le texte-comme-œuvre que le texte-comme-document. Les éditions imprimées traditionnelles se sont davantage intéressées au premier. Un avantage évident des éditions numériques est qu'elles peuvent rééquilibrer le balancier en incluant un matériau bien plus riche pour l'étude du texte-comme-document qu'on ne peut y parvenir par le moyen de l'impression. », ROBINSON, Peter, « Where We Are with Electronic Scholarly Editions, and Where We Want to Be », *Jahrbuch für Computerphilologie*, vol. 5 / 5, 2003, p. 126-146.

Pour aller plus loin : Ioana Galleron, Marie-Luce Demonet, Cécile Meynard, Idmhand Fatiha, Elena Pierazzo, et al.. *Les publications numériques de corpus d'auteurs - Guide de travail, grille d'analyse et recommandations* (V1-Novembre 2018). [Rapport de recherche] Huma-Num. 2018, 19 p. [halshs-01932519](https://halshs.archives-ouvertes.fr/halshs-01932519).

# Les éditions scientifiques

- **La critique textuelle**

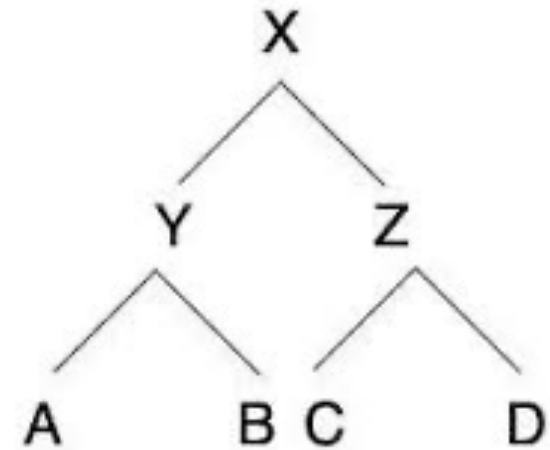
Une édition moderne, afin de transmettre un texte, compare les manuscrits pour **établir le texte original**. Elle rend compte de cette comparaison dans un **apparat critique**.

À partir de la **comparaison des variantes**, on essaie d'établir une **généalogie des textes**, pour pouvoir déterminer quel témoignage (**leçon**) est celui qui a le plus de chance d'être équivalent au texte original (**archétype**).

# Les éditions scientifiques

- Les principes du Stemma

Pour aller plus loin : [CollateX – Software for Collating Textual Sources.](#)



# Les éditions scientifiques

- **Encoder une édition critique**

Définir quel type d'édition, on souhaite faire :

- **Reconstructionniste**, selon la méthode lachmanienne → Appliquée aux textes en langues anciennes. Reconstruction du texte original en examinant l'ensemble de la tradition manuscrite et en comparant les variantes.
- **Conservatrice**, selon la méthode bédieriste → Appliquées aux textes en langue vernaculaire. Dite conservatrice car idée de corriger le moins possible le texte qui semble le meilleur et le plus correct.

Définir quel type d'apparat on veut constituer :

- **Apparat exhaustif** : qui présente l'ensemble des variantes;
- **Apparat sélectif** : qui exclut certains types de variantes jugées moins importantes n'ayant pas d'impact sur le sens (variantes graphiques, séparation des mots, ponctuation).

Pour aller plus loin : Jean-Baptiste Camps, « Copie, authenticité, originalité dans la philologie et son histoire », *Questes* [En ligne], 29 | 2015, mis en ligne le 20 juin 2015, consulté le 30 avril 2021. URL : <http://journals.openedition.org/questes/3535> ; DOI : <https://doi.org/10.4000/questes.3535>.

# Les éditions scientifiques

Avant d'encoder, définir quelle présentation on souhaite :

- **Apparat positif**, qui enregistre et fait figurer explicitement les leçons de chaque témoin accompagnées de son sigle ; le plus lourd, mais le plus clair ;
- **Apparat négatif**, qui ne fait figurer que les témoins dont le texte diffère. Tous les autres, qui ne sont pas signalés, sont supposés porter le texte retenu ;
- **Apparat mixte**, négatif par défaut qui ne devient positif qu'en cas de besoin de clarification.

# Les éditions scientifiques

- Les solutions TEI

Cf. [\*Critical Apparatus\*](#) dans les TEI Guidelines.

Attention : la méthode utilisée dans les exemples suivants suit les prescriptions de la *parallel segmentation method*. Préciser la méthode d'encodage des variantes dans l'<encodingDesc> du header.

Exemple : `<variantEncoding method="parallel-segmentation" location="internal"/>`

Pour aller plus loin : Chap 12.2 [\*Linking the Apparatus to the Text\*](#) dans les TEI Guidelines.

# Les éditions scientifiques

- **Déclaration des sources dans le <sourceDesc> :**
  - **<listWit>** : donne une liste de tous les témoins cités dans un appareil critique
  - **<witness>** : témoin d'une édition

*Exemple :*

<listWit>

<witness xml:id="C1">Paris, BnF, ms 412</witness>

<witness xml:id="C2">Paris, BnF, ms 411</witness>

</listWit>

# Les éditions scientifiques

- **Constitution de l'apparat critique**

- **<app>** contient l'apparat critique, constituée d'un lemme facultatif ou d'au moins une leçon
- **<lem>** contient le lemme
- **<rdg>** contient une variante/leçon du texte

*Exemple :*

```
<app>  
  <lem wit="#El #Hg">Experience</lem>  
  <rdg wit="#La" type="substantive">Experiment</rdg>  
  <rdg wit="#Ra2" type="substantive">Eryment</rdg>  
</app>
```



# Les éditions scientifiques

- **<rdgGrp>** regroupe deux leçons ou plus qui sont perçues comme ayant une relation génétique ou une autre affinité

*Exemple :*

```
<app>
  <lem wit="#fr_wit1">comte</lem>
  <rdgGrp type="orthographic">
    <rdg wit="#fr_wit2">cante</rdg>
    <rdg wit="#fr_wit3">contes</rdg>
  </rdgGrp>
</app>
```

# Les formats de schémas d'encodage

---

# Les formats de schémas d'encodage

La génération d'un schéma d'encodage propre à votre corpus est ce qui va vous permettre de vérifier le respect de vos **choix de structuration** des données lors de la phase d'encodage. Il est aussi un élément de **documentation** pour que des personnes extérieures au projet comprennent comment sont structurées vos données.

# Les formats de schémas d'encodage

Si chaque schéma est unique pour répondre aux enjeux d'une édition, il est important qu'il soit conforme aux règles de la TEI (*TEI-conformant*). Pour cela, respecter les règles suivantes :

1. Le XML doit être bien formé
2. L'encodage proposé doit pouvoir être validé avec un schéma TEI all
3. L'encodage doit être conforme au modèle abstrait de la TEI
4. L'encodage doit faire bon usage de l'espace de nom de la TEI (et des autres espaces de nom si besoin)
5. L'encodage doit être documenté

La notion de **TEI conformance** n'a pas d'autre but que d'aider à la description d'un fichier XML pour faciliter l'échange de données, l'utilisation d'outils ou l'archivage. Elle n'est en rien liée à la qualité scientifique du projet.

---

# Les formats de schémas d'encodage

Un schéma d'encodage en XML-TEI peut prendre différentes formes :

- DTD (Document Type Definition)
- XML Schéma
- Relax NG
- ODD (One Document Does it all)

# Les formats de schémas d'encodage

- **DTD** (Document Type Definition)

- permet la création d'éléments, de sous-éléments, d'attributs, d'entités ;
- pas de documentation du schéma directement dans le fichier ;
- pas de typage précis du contenu des éléments (chaîne de caractères de texte, nombre entier, etc.) ;
- pas de gestion des espaces de nom ;
- pas en XML.

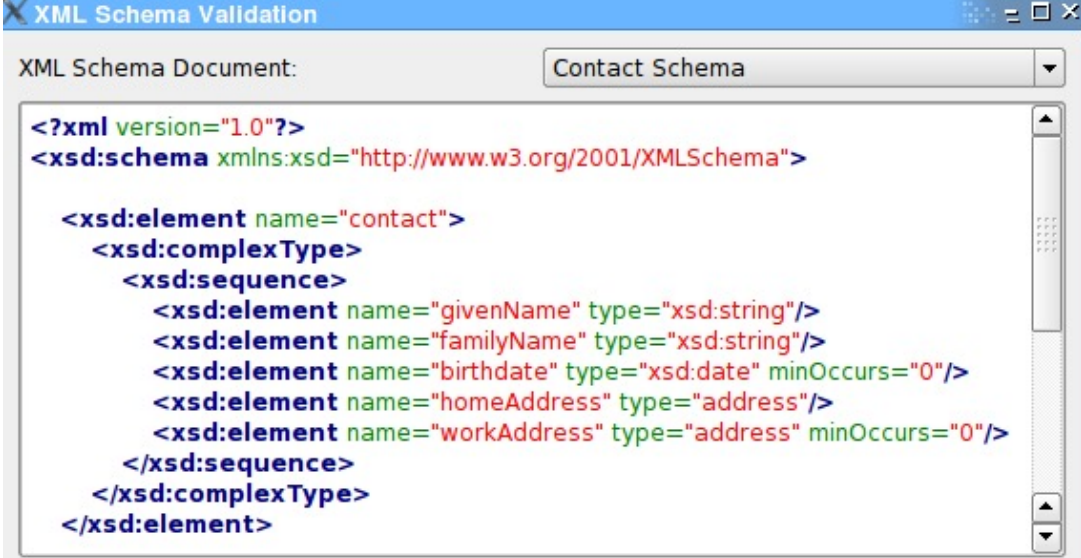
```
15 <!ELEMENT prenom (#PCDATA)>
16 <!ELEMENT entreprise (#PCDATA | nom | ville | pays | adresse)*>
17 <!ELEMENT mention (#PCDATA)>
18 <!ELEMENT experience (#PCDATA | entreprise | debut | duree | intitule)*>
19 <!-- ATTENTION : le type CDATA est implicite -->
20 <!ELEMENT option (#PCDATA)>
21 <!ELEMENT formation (#PCDATA | diplome)*>
22 <!ELEMENT CV (#PCDATA | titre | identite | formation | experiences | autre)*>
23 <!ELEMENT loisir (#PCDATA)>
24 <!ELEMENT nationalite (#PCDATA)>
25 <!ELEMENT etablissement (#PCDATA | nom | ville | pays)*>
26 <!ELEMENT nom (#PCDATA)>
```

Pour plus d'informations sur la structure d'une DTD, cf. cours d'Arianne Pinche, cours TNAH\_XML-TEI, [séance 2](#).

# Les formats de schémas d'encodage

- **XML Schéma**

- permet de typer les données ;
- permet de mettre en place une séquence d'éléments ;
- XML Schema est lui-même un document XML ;
- pas de documentation du schéma directement dans le fichier.



The screenshot shows a window titled "XML Schema Validation". Inside, there's a tab labeled "Contact Schema". The main area displays an XML Schema document in a code editor. The code defines a root element named "contact" which is a complex type containing a sequence of five elements: "givenName" (xsd:string), "familyName" (xsd:string), "birthdate" (xsd:date with minOccurs="0"), "homeAddress" (xsd:string), and "workAddress" (xsd:string with minOccurs="0").

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

  <xsd:element name="contact">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="givenName" type="xsd:string"/>
        <xsd:element name="familyName" type="xsd:string"/>
        <xsd:element name="birthdate" type="xsd:date" minOccurs="0"/>
        <xsd:element name="homeAddress" type="xsd:string"/>
        <xsd:element name="workAddress" type="xsd:string" minOccurs="0"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

</xsd:schema>
```

# Les formats de schémas d'encodage

- **Relax NG**

- permet de typer les données ;
- permet de mettre en place une séquence d'éléments ;
- peut créer des contraintes d'enchaînement très précises grâce à l'intégration du langage [schematron](#) ;
- supporte XML namespaces ;
- écrit en syntaxe XML ;
- compatible avec XML schéma ;
- plus simple que XML schéma ;



# Les formats de schémas d'encodage

- peut s'appuyer sur les modules et les macros de la TEI ;
- pas de documentation du schéma directement dans le fichier.

```
<define name="att.dimensions.attribute.unit">
  <optional>
    <attribute name="unit">
      <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">names the unit used for the measurement
Suggested values include: 1] cm (centimetres); 2] mm (millimetres); 3] in (inches); 4] line; 5] char (characters)</a:documentation>
      <choice>
        <value>cm</value>
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">(centimetres) </a:documentation>
        <value>mm</value>
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">(millimetres) </a:documentation>
        <value>in</value>
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">(inches) </a:documentation>
        <value>line</value>
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">lines of text</a:documentation>
        <value>char</value>
        <a:documentation xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">(characters) characters of text</a:documentation>
        <data type="token">
          <param name="pattern">{\p{L}}{\p{N}}{\p{P}}{\p{S}}+</param>
        </data>
      </choice>
    </attribute>
  </optional>
</define>
```

# Les formats de schémas d'encodage

- **ODD** (*One Document Does it all*)

ODD signifie « *One Document Does it all* ». C'est un fichier *TEI-XML-conformant* qui permet de spécifier un encodage TEI pour personnaliser un usage de la TEI P5.

- documentation intégrée au fichier ;
- possibilité de définir précisément le nombre d'occurrences d'un élément, ainsi que des séquences d'éléments ;
- possibilité de typer les données d'un élément ;
- supporte des espaces de noms ;
- peut d'appuyer sur le schéma TEI all et sa structure en modules et macros ;
- intégralement en syntaxe XML ;
- autorise l'insertion en son sein d'autres langages de schéma comme RelaxNG ou schematron pour aller encore plus loin dans la finesse des spécifications.

# Les formats de schémas d'encodage

À partir d'un fichier ODD, on peut générer le schéma RelaxNG qu'on liera à son projet XML, ainsi qu'une documentation HTML (ou PDF, .docx, .odt etc.) à l'image des teiGuidelines.

Pour aller plus loin, cf. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html> ;  
<http://www.tei-c.org/guidelines/customization/getting-started-with-p5-odds/>.

# Les formats de schémas d'encodage

Nous allons voir ici des solutions pour créer son ODD, car il constitue un type de schéma qui présente de nombreuses possibilités d'affinement des spécifications, tout en permettant d'écrire une documentation détaillée.

On peut créer de différentes manières une ODD :

- On peut la rédiger à partir de rien ce qui peut être très chronophage. Pour cela, on peut s'aider de l'outil en ligne [Roma](#) ;
- ou bien on peut la créer à partir d'un ou de plusieurs fichiers XML (**oddbysample**).

Les ODD ainsi produites s'appuient sur la structure du schéma TEI all et de ses modules (cf. séance 1).

N. B. : L'ensemble des manipulations ci-dessous sont dans les différentes vidéos de démonstration de Roma. Reproduisez les manipulations sur votre ordinateur afin de vous familiariser avec l'outil.

N.B. : Pour les scénarios de transformation, votre installation de Oxygen n'a peut-être pas tous les add-on de la TEI, cf. tutoriel à la fin de la séance 2 pour l'installation et la mise à jour des add-ons liés à TEI.

# Les formats de schémas d'encodage

- Structurer son ODD avec [Roma](#) :
  - À l'aide des [modules TEI](#) ;
  - À l'aide des [classes TEI](#) ;
  - Au niveau d'un élément ;
  - Au niveau d'un attribut.

Cf. tutoriel : <https://tei-c.org/guidelines/customization/customizing-the-tei-with-roma/>.

# Les formats de schémas d'encodage

- Structurer son ODD avec **Oddbyexample**

Mise à jour des add-on dans Oxygen :

- Options/Préférences/Association de type de document
  - Tout désactiver
  - Bien vérifier que la ligne "options globales" est cochée

Add-ons - aide/installer les nouveaux add-ons - Ajouter <https://www.tei-c.org/release/oxygen/updateSite.oxygen> - Appliquer et accepter

- Aide/Gérer les Add-ons/Installer
  - Tout activer et tout appliquer
- Options/Préférences/Association de type de document
  - Tout activer, appliquer, accepter

# Les formats de schémas d'encodage

Mise en place du scénario :

- Configurer un scénario de transformation (CTRL+MAJ+C ou menu Document/Transformation/Configurer...)
- Créer un nouveau scénario : XML transformation with XSLT
- Renseigner le chemin de l'XSL
  - `${frameworks}/tei/xml/tei/stylesheet/tools/oddbyexample.xsl;`
- Sélectionner processeur Saxon 9.xX
  - Options avancées, template (-it) : main
  - Paramètres : corpus `${cfdu}` (i.e. répertoire courant)
- Configurer la sortie (onglet Sortie) : définir un nom et un emplacement pour la future ODD.

Cf. tutoriel : <http://teic.github.io/TCW/howtoGenerate.html>.

Personnaliser son  
schéma d'encodage :  
ODD

---



# Personnaliser son schéma d'encodage

La personnalisation d'un encodage peut passer par :

- L'ajout d'éléments
- La suppression d'éléments
- Le changement d'éléments
- L'assignation d'attributs et de valeurs d'attributs à un élément

Ces modifications peuvent soit être conformes à la TEI, en piochant dans les modules ce qui est pertinent pour l'encodage, sans supprimer d'éléments obligatoires ; soit en être non-conformes à la TEI, par invention d'un élément, attribut, etc.

# Personnaliser son schéma d'encodage

- **Supprimer un élément**

- Suppression simple : @mode="delete" (*exemple* : `<elementSpec ident="camera" mode="delete"/>`)
- Suppression d'un **élément** ou d'une **classe** au sein d'un module : @except (*exemple* : `<moduleRef key="drama" except="camera"/>`)
- Non-insertion dans un **module** ou une **classe** : @include (*exemple* : `<moduleRef key="drama" include="castList role move"/>`)

N.B. : Si l'on veut respecter la TEI, attention à ne pas supprimer d'enfants obligatoires. Par exemple, dans les métadonnées, un <fileDesc> doit contenir au moins un <titleStmt> et un <sourceDesc>.

# Personnaliser son schéma d'encodage

---

- **Modifier** un élément

Il suffit d'appliquer l'`@mode=change` :

- sur les classes (`<classSpec>`)
- sur les éléments (`<elementSpec>`)
- sur les attributs (`<attDef>`) ou sur les valeurs d'attributs (`<valList>`)

```
<elementSpec ident="dimensions" mode="change">
  <attList>
    <attDef ident="type" mode="change">
      <valList mode="add" type="closed">
        <valItem ident="leaf"/>
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

# Personnaliser son schéma d'encodage

- **Modalités d'apparition** d'un élément

On place la règle définissant une séquence d'éléments dans un élément `<content>`.

L'**@preserveOrder** peut avoir comme valeur « true » ou « false », pour savoir si l'ordre de déclaration des éléments est significatif.

Les **@minOccurs** et **maxOccurs**, dont la valeur peut être « 0 », « 1 » ou « unbounded », permet de définir le nombre d'apparition d'un élément au sein d'une séquence.

```
<elementSpec ident="msIdentifieur" mode="change">
  <!-- Définition d'un enchaînement d'éléments
  à l'intérieur de l'élément <elementSpec> : la séquence
  comprend au minimum l'élément <idno>. -->
  <content>
    <sequence preserveOrder="true">
      <elementRef key="settlement" minOccurs="0" maxOccurs="1"/>
      <elementRef key="repository" minOccurs="0" maxOccurs="1"/>
      <elementRef key="idno" minOccurs="1" maxOccurs="1"/>
      <elementRef key="altIdentifieur" minOccurs="0" maxOccurs="1"/>
    </sequence>
  </content>
```

# Personnaliser son schéma d'encodage

- **Modalités d'apparition** d'un élément

**<alternate>** permet de définir plusieurs séquences possibles au sein d'un élément.

*Exemple ci-contre* : Définition de séquences d'éléments au sein de l'élément **<choice>**.

N.B. : Pour autoriser du texte comme contenu, on peut ajouter dans la séquence : **<textNode/>**.

```
<alternate>
  <sequence>
    <elementRef key="sic"/>
    <elementRef key="corr"/>
  </sequence>
  <sequence>
    <elementRef key="orig"/>
    <elementRef key="reg"/>
  </sequence>
  <sequence>
    <elementRef key="abbr"/>
    <elementRef key="expa"/>
  </sequence>
</alternate>
```

# Personnaliser son schéma d'encodage

- **Typier les données**

Utilisation de **<dataType>** et de son enfant **<dataRef>** avec l'attribut @key si on pointe vers un type de données défini par la TEI.

Cf. Liste des types de données en TEI :  
<https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/REF-MACROS.html>.

```
<attDef ident="n" mode="change">  
  <datatype>  
    <dataRef key="teidata.count"/>  
  </datatype>  
</attDef>
```

# Personnaliser son schéma d'encodage

- Contraindre des valeurs et enchaînements en fonction du contexte : le langage **schematron**

La contrainte est introduite par un élément **<constraintSpec>**. Le langage utilisé est déclaré dans l'attribut **scheme="schematron"**, la règle est nommée à l'aide de l'attribut **ident**.

La règle est contenue dans une balise **<constraint>**.

N.B. : Attention à bien déclarer le nom de domaine *schematron* dans le préambule sous la forme xmlns:s=<http://purl.oclc.org/dsdl/schematron>.

# Personnaliser son schéma d'encodage

- Contraindre des valeurs et enchaînements en fonction du contexte : le langage schematron

Pour définir les contraintes :

- **<s:assert>** pour vérifier l'existence de la contrainte rédigée
- **<s:rule>** pour ajouter un contexte à l'application de la règle

N.B. : Pour définir ces règles, il faut savoir manipuler le langage [XPath](#), servant à cibler différentes parties d'un document XML, et les [expressions régulières](#) pour capter des chaînes de caractères.

```
<constraintSpec ident="fromTo" scheme="schematron">
  <constraint>
    <s:rule context="tei:app[@type='structure']">
      <s:assert test="@from and @to">
        The beginning and the endpoint of the
        lemma have to be identify/
      </s:assert>
    </s:rule>
  </constraint>
</constraintSpec>
```



# Personnaliser son schéma d'encodage

- **Documenter son ODD**

L'ODD comporte :

- Un élément racine <TEI>
- Un <teiHeader>
- Un élément <text>

Le <body> de l'élément <text> peut lui-même être avec des <div> de différents niveaux (<div1>, <div2>, ...) afin de structurer la documentation de son édition.

# Personnaliser son schéma d'encodage

- **Documenter son ODD**

Les spécifications XML sont englobées dans un **<schemaSpec>** à l'intérieur duquel on peut retrouver les éléments suivants :

- **<moduleSpec>** (structure, contenu et fonctions d'un module)
- **<moduleRef>** (incorporation d'un module dans un schéma)
- **<elementSpec>** (structure, contenu et emploi d'un élément)
- **<classSpec>** (informations de référence pour une classe d'éléments TEI)

# Personnaliser son schéma d'encodage

- Documenter son ODD

Pour rédiger la documentation, on peut utiliser les balises :

- **<specList>** avec **<specDesc>** comme élément enfant, pour faire des listes de spécification
- **egXML** et **@xmlns="http://www.tei-c.org/ns/Examples"** pour insérer des exemples en XML dans sa documentation

```
<div5>
  <head>Les métadonnées descriptives</head>
  <p>Le <gi>sourceDesc</gi> répertorie tous les lieux de conservation d'où
    proviennent les manuscrits du corpus. Chaque lieu est encodé dans une
    balise <gi>msDesc</gi> qui contient : <specList>
      <specDesc key="msIdentifier"/>
      <specDesc key="repository"/>
    </specList></p>
  <egXML xmlns="http://www.tei-c.org/ns/Examples">
    <msDesc>
      <msIdentifier>
        <repository>Biblioteca civica de Vérone</repository>
      </msIdentifier>
    </msDesc>
    <msDesc>
      <msIdentifier>
        <repository>Università degli studi de Padoue</repository>
      </msIdentifier>
    </msDesc>
  </egXML>
```

# Personnaliser son schéma d'encodage

- **Documenter son ODD**

Balises à connaître pour rédiger sa documentation :

- **<att>** pour indiquer le nom d'un attribut
- **<gi>** pour indiquer le nom d'un élément
- **<tag>** pour indiquer le nom d'une balise
- **<val>** pour indiquer la valeur d'un attribut