



HAL
open science

TEI: de l'image au texte : Décrire son corpus grâce aux métadonnées

Charles Riondet

► **To cite this version:**

Charles Riondet. TEI: de l'image au texte : Décrire son corpus grâce aux métadonnées. TEI: de l'image au texte, Feb 2018, Le Mans, France. pp.1-69. hal-01708839

HAL Id: hal-01708839

<https://inria.hal.science/hal-01708839v1>

Submitted on 14 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.

TEI: de l'image au texte.

Session 1: Décrire son corpus grâce aux métadonnées

Charles Riondet

13 février 2018

Inria

charles.riondet@inria.fr

Programme de la session

- Présentation théorique
- Pause
- Travail d'encodage
- Discussion

Que sont les métadonnées ?

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le <teiHeader>

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le <teiHeader>

Étymologie :

méta (préfixe d'autoréférence) et *donnée* (Élément d'une réalité)

Propositions de définition :

Données sur les données

Description d'une entité tangible

Pour rendre les choses plus claires

À une chose (une entité, un objet) correspondent un certain nombre d'informations qui permettent de l'identifier, de la décrire, de la gérer, etc.

Ex. :

La carte d'identité contient les métadonnées d'une personne.

Pour rendre les choses plus claires



Fausse carte d'identité de Pierre Georges, dit *Colonel Fabien*, 1943.
Coll. Musée de la Résistance nationale

- **Identification et description.**
Renseignements sur la substance de l'entité décrite :
Nom, Prénoms, Date de naissance, etc.
- **Provenance et contexte.**
Renseignement sur la vie de l'entité décrite :
Profession, Adresse

Métadonnée : un terme récent

Invention du terme avec le web :

Fournir des informations sur les sites et les pages web aux moteurs de recherche.

```
<html>
  <head>
    <meta name="description" content="Une page web sur ..." />
    <meta name="keywords" content="mots,clés,keys" />
    <meta name="author" content="John Doe" />
  </head>
  <body>
    <p>Bienvenue sur la page web dédiée aux clés et aux mots</p>
  </body>
</html>
```

Balises <meta> d'une page web (HTML)

Mais une idée ancienne

On produit des métadonnées lorsqu'on catalogue, indexe, décrit, classe, etc. : Les bibliothécaires et les archivistes sont des grands producteurs de métadonnées depuis longtemps :

Il y a 2000 ans, dans les rayons de la bibliothèque de la Villa des Papyrus à Herculaneum, des petites étiquettes étaient fixées sur chaque document pour connaître leur auteur, leur titre, leur sujet.

À partir du moment où on collecte des données (numériques ou non) pour un but précis, il est primordial de réfléchir à la façon de les décrire.

Quelles informations ?

Les métadonnées permettent de décrire des données.

Elles donnent des informations sur :

- *Le contenu intellectuel* : titre, résumé, mots-clés, type ;
- *Le contexte de production* : date, créateur, procédés de création ;
- *Les caractéristiques techniques* : formats, taille, organisation ;
- *Les propriétés et droits d'usage* : détenteurs des droits, conditions d'accès, conditions de partage, conditions d'usage.

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le <teiHeader>

Trouver de l'information

Découvrir des sources sur lesquelles travailler

Consulter la bibliographie sur son sujet de recherches



(BnF Catalogue général) Espace

Tout Recherche avancée AUTEURS A-Z SUJETS A-Z PÉRIODIQUES

Accueil Univers jeunesse Liste de notices Notice bibliographique

Notice de collection éditoriale

Notices Au format public

Titre clé : L'Histoire de la vie

Titre(s) : L'Histoire de la vie [Texte imprimé]

Publication : [Paris] : Galliard, 1997-

Description matérielle : ill. en noir et en coul., couv. ill. en coul., 19 cm

Note(s) : Numérotation commune à la collection : "Mes premières découvertes" et à ses sous-collections. - Commence avec le n° 135 de la collection principale

Périodicité : Collection

Titre(s) en liaison :
- Est une sous-collection de : [Mes premières découvertes \(Paris\)](#) - ISSN 1147-4831

Indice(s) Dewey : 000.83

Numéros : ISSN 1285-9825 = L'Histoire de la vie
ISSN-L 1285-9825

Notice n° : FRBNF36131295



Aller dans un centre d'archives, consulter plusieurs cartons, prendre des notes => *création de métadonnées*

- Qui a fait ce dépouillement ?
- Où ?
- Quand ?
- Combien de cartons ?
- Combien de photos prises ? Avec quelle organisation ?
- Quelles notes prises ? Dans quel fichier ? Sous quel format ?
-

Les métadonnées rendent plus facile, pour leur créateurs et pour des tiers, la recherche d'informations, la réutilisation et l'identification des données. Cela vous permet de :

- retrouver et **réutiliser** facilement **vos** propres données ;
- découvrir, évaluer, et réutiliser les **données produites par d'autres** ;
- faciliter la **préservation** des données numériques, alors que les logiciels et les formats ne cessent d'évoluer au fil du temps.

- Donner une description bibliographique approfondie et détaillée dans un format normalisé permet l'**échange de données** ;
- **Lier le document numérique à l'original** et inversement
- Lier différentes versions d'un document
- **Gérer les droits** d'accès (droits d'auteur, confidentialité) et d'usage (droits d'impression, de reproduction, de modification...);
- Préserver les **informations techniques** nécessaires à la lecture des fichiers ;
- Garantir l'**intégrité des fichiers** et le suivi de leurs éventuelles **modifications**.

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le <teiHeader>

Description d'une œuvre, ou plutôt d'une de ses expressions

Pas un objet physique mais une *création intellectuelle* qui possède pour attributs (entre autres) :

- un titre ;
- un auteur ;
- une forme (un texte, un film, un tableau) ;
- une date ;

Description de l'objet qui en résulte

L'expression d'une œuvre se manifeste par un objet (une *manifestation*) qui a pour attributs :

- des informations sur l'édition de l'objet ;
- des informations physiques sur cet objet ;
- ...



En contexte académique, une édition critique a souvent pour base un *exemplaire* d'une *manifestation* d'une œuvre.

- Un exemplaire de la 1^{re} édition ;
- Un exemplaire annoté ;
- Le manuscrit.



D'autres éléments de description sont nécessaires, qui se réfèrent spécifiquement à cet *objet* :

- Lieu de conservation, collection, cote, etc. :
- État matériel ;
- Historique de la conservation.

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus
- La transcription
- La correction
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.)
- L'annotation : Ce qu'on appelle en philologie l'appareil critique

- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription
- La correction
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.)
- L'annotation : Ce qu'on appelle en philologie l'appareil critique

- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription [Qui transcrit et comment ?]
- La correction
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.)
- L'annotation : Ce qu'on appelle en philologie l'appareil critique

- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription [Qui transcrit et comment ?]
- La correction [Qui corrige ? Quel niveau de correction ?]
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.)
- L'annotation : Ce qu'on appelle en philologie l'appareil critique

- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription [Qui transcrit et comment ?]
- La correction [Qui corrige ? Quel niveau de correction ?]
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.) [Qui encode ?]
- L'annotation : Ce qu'on appelle en philologie l'appareil critique

- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou tout autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription [Qui transcrit et comment ?]
- La correction [Qui corrige ? Quel niveau de correction ?]
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.) [Qui encode ?]
- L'annotation : Ce qu'on appelle en philologie l'appareil critique [Qui annote ?] [Liens vers d'autres textes, ou des sources d'informations : dictionnaires, bases de données biographiques, ...]
- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche)

La plupart des étapes du travail éditorial nécessitent l'utilisation (parfois la création) de métadonnées.

- La constitution du corpus [Quels textes ? Quels formats d'images ?]
- La transcription [Qui transcrit et comment ?]
- La correction [Qui corrige ? Quel niveau de correction ?]
- La structuration logique du texte (entête, paragraphe, liste, etc.) et présentation matérielle (changement de ligne, soulignement, espaces, etc.) [Qui encode ?]
- L'annotation : Ce qu'on appelle en philologie l'appareil critique [Qui annoté ?] [Liens vers d'autres textes, ou des sources d'informations : dictionnaires, bases de données biographiques, ...]
- L'indexation : Liste alphabétique des entités intéressantes citées dans le texte (personnes, organisations, lieux, ou toute autre information utile pour le travail de recherche) [Qui indexe ?] [Liens vers sources d'informations : vocabulaires contrôlés, ontologies, dictionnaires]

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le <teiHeader>

Deux modes de fonctionnement principaux

Les métadonnées peuvent être enregistrées dans des documents spécifiques.



Les fameux fichiers des bibliothèques

(BnF Catalogue général) Espace

Tout Recherche avancée AUTEURS A-Z SUJETS A-Z PÉRIODIQUES

Accueil Univers jeunesse Liste de notices Notice bibliographique

Notice de collection éditoriale

Notice Au format public

Titre cité : L'Histoire de la vie

Titre(s) : L'Histoire de la vie [Texte imprimé]

Publication : [Paris] : Gallimard, 1997.

Description matérielle : ill. en noir et en coul., couv. ill. en coul., 19 cm

Note(s) : Numérotation commune à la collection : "Mes premières découvertes" et à ses sous-collections. - Commence avec le n° 135 de la collection principale

Périodicité : Collection

Titre(s) en liaison :
- Est une sous-collection de : [Mes premières découvertes \(Paris\)](#) = ISSN 1147-4831

Indice(s) Dewey : 000.83

Numéros : ISSN 1285-9826 = L'Histoire de la vie
ISSN-L 1285-9826

Notice n° : FRBNF36131295

Les catalogues en ligne

Deux modes de fonctionnement principaux

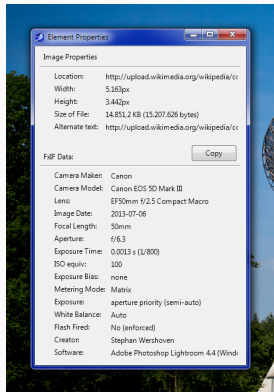
Ou bien elles peuvent être contenues dans le fichier qu'elles décrivent.

```
<html>
<head>
  <meta name="description" content="Une page web sur ..." />
  <meta name="keywords" content="mots, clés, keys" />
  <meta name="author" content="John Doe" />
</head>
<body>
  <p>Bienvenue sur la page web dédiée aux clés et aux mots</p>
</body>
</html>
```

Les balises <meta> des pages web

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" type="researchScenario" xml:lang="fr">
  <teiHeader>
    <fileDesc>
      <titleStm>
        <title>Description of the scenario `Corpus modelling in T</title>
      </titleStm>
      <author>
        <persName>Klaus Illmayer</persName>
        <affiliation>OEAW-ACDH</affiliation>
      </author>
      <author>
        <persName>Susanne Haaf</persName>
        <affiliation>BBAW</affiliation>
      </author>
      <author>
        <persName>Piotr Bański</persName>
        <affiliation>IDS Mannheim, CLARIN</affiliation>
      </author>
    </fileDesc>
    <publicationStm>
      <authority>Parthenos ...</authority>
      <availability>
        <licence target="http://creativecommons.org/licenses/by/4.0/">
          <p>The Creative Commons Attribution 4.0 Unported (CC BY) license applies to this document.</p>
        </licence>
      </availability>
    </publicationStm>
  </teiHeader>
```

Le TEI Header



Métadonnées EXIF pour les photographies numériques

Quelques standards utilisés en SHS

| ISO | DC | MARC et dérivés | EAD | TEI |
|--|--|--|---|--|
| International Organization for Standardization | Dublin Core | Machine Readable Cataloging | Encoded Archival Description | Text Encoding Initiative |
| Standard international de règles génériques | Structuration générique des métadonnées d'objets physiques ou numériques | Métadonnées bibliographiques principales | Métadonnées bibliographiques des manuscrits et fonds d'archives | Données et métadonnées textuelles |
| Utilisations multiples | Format d'échange | Bibliothèques | Archives et bibliothèques | Communauté SHS, édition, bibliothèques |

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

Encodage des métadonnées en TEI : le `<teiHeader>`

Recommandations pour la création et la gestion sous forme numérique de tout type de données créées et utilisées par les chercheurs en sciences humaines :

Manuscrits, documents d'archives, inscriptions anciennes, ...

3 principes fondamentaux

- privilégier le sens avant le format
- s'assurer de l'indépendance vis à vis des logiciels
- s'appuyer sur une communauté forte

Les documents TEI utilisent un langage d'encodage formel très largement utilisé appelé XML ou "extensible markup language".

Un document XML

- Une séquence de caractères lisibles par un humain, sans addition de codes spéciaux ou données binaires.
- Commence toujours par une instruction indiquant que ce qui suit est un document XML conforme à la version du standard XML indiquée (dans ce cas, version 1.0)

```
<?xml version="1.0"?>
<doc xmlns="http://example.org/?
  namespace">
  <p n="1">Ceci est un
    paragraphe.</p>
  <p n="2">Ce paragraphe
    mentionne <placeName>
      Bristol</placeName>.</p>
</doc>
```

- Élément => Tout composant d'un document XML identifié par le balisage
<p> => élément paragraphe
- Attribut => Information supplémentaire relative à un élément
<p n="1"> => l'attribut n attribue un numéro d'ordre à ce paragraphe

XML : un langage de balisage

On introduit un balisage dans un document pour **l'organiser** et **l'annoter** en vue d'un traitement automatisé.

Par exemple, si les paragraphes sont clairement marqués, alors un logiciel de mise en forme pourra les mettre en page correctement.

Pour ce faire, il faut **contrôler** la façon dont les balises sont introduites dans le document et l'endroit où elles apparaissent :

Utilisation d'un **schéma XML** => *un lexique et une grammaire pour les documents XML valides.*

Un schéma spécifie :

- Un ensemble de noms d'éléments,
- Les noms et le type de données de tous les attributs qui leur sont associés,
- Les règles relatives aux contextes dans lesquels ils peuvent apparaître.

La **TEI** est un schéma : fournit le nom et la définition de centaines de balises, en même temps que des règles sur la façon dont elles peuvent être combinées.

- Un document TEI est le substitut numérique d'une source physique.
- On favorise la sémantique (le fond) plutôt que la mise en page (la forme)
- La publication requiert une étape de transformation (XSLT ; ePub, pdf, HTML, etc.)
- Pour structurer le document :
 - Macro-structure : front - body - back
 - Meso-structure : divisions, paragraphes, etc.
 - Micro-structure : mécanismes d'annotations : Dates, noms, notes, references, expressions étrangères, etc.

Un outil essentiel : les TEI guidelines

www.tei-c.org/Guidelines/P5/index.xml

- Dictionnaire des balises et attributs
- En plusieurs langues
- Nombreux exemples d'utilisation

Un document TEI minimal

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

Que sont les métadonnées ?

Pourquoi est-ce important pour les SHS ?

Métadonnées d'un texte, d'une transcription, d'une édition : quelles différences ?

Modèles et formats de métadonnées

La TEI en quelques slides

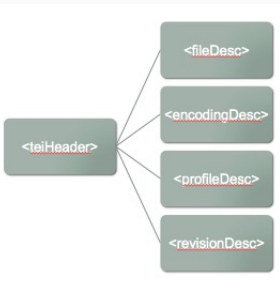
Encodage des métadonnées en TEI : le <teiHeader>

Chaque document TEI valide comprend un *header*, habituellement suivi par un texte.

Le *header* est essentiel pour :

- Le contrôle bibliographique et l'identification
- La documentation sur la ressource et son traitement

Éléments du header



<fileDesc> La page du titre du document [obligatoire]

<encodingDesc> Contraintes et choix éditoriaux
[optionnel]

<profileDesc> Caractéristiques générales du contenu
[optionnel]

<revisionDesc> Histoire numérique du document
[optionnel]

Description bibliographique de l'ensemble du document (du fichier électronique XML-TEI)

Les descripteurs principaux comprennent :

<**titleStmt**> Titre et responsabilités associés au document

<**publicationStmt**> Détails concernant la publication du document,
Par exemple, la licence

<**sourceDesc**> Information sur le(s) source(s) du document (un des principaux éléments).

Un aperçu des autres composants du header

<encodingDesc> = Encoding Description. Caractères, éléments utilisés, références géographiques, choix éditoriaux (normalisation, corrections, traitement des fins de ligne, interventions éditoriales, etc.)

<profileDesc> = Text-profile Description. La description des aspects non bibliographiques du texte (circonstances de la production de la source, langue, sujet, etc.)
Index des termes, résumé, langues du document, participants (qui parle, qui écrit, qui annote)

<revisionDesc> = Revision description. Historique des modifications

```
<revisionDesc status="embargoed">  
  <change when="1991-11-11" who="#LB">deleted  
    chapter 10</change>  
</revisionDesc>
```

Le TEI header : <fileDesc>

Description bibliographique de l'ensemble du document (du fichier électronique XML-TEI)

Les descripteurs principaux comprennent :

<**titleStmt**> Titre et responsabilités associés au document

<**publicationStmt**> Détails concernant la publication du document,
Par exemple, la licence

<**sourceDesc**> Information sur le(s) source(s) du document (un des principaux éléments).

fileDesc est le seul élément obligatoire du Header

<fileDesc> : exemple

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Thomas Paine: Common sense, a machine-readable
        transcript</title>
      <respStmt>
        <resp>compiled by</resp>
        <name>Jon K Adams</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <distributor>Oxford Text Archive</distributor>
    </publicationStmt>
    <sourceDesc>
      <bibl>The complete writings of Thomas Paine, collected
        and edited by Phillip S. Foner. New York, Citadel
        Press 1945</bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

éléments du <fileDesc> : le <titleStmt>

titleStmt (Title statement) contient un élément obligatoire

<**title**> titre du fichier électronique

Et des éléments facultatifs

<**author**> Il s'agit ici de l'auteur du fichier xml.

<**respStmt**> autres responsabilités du fichier

<*resp*> : type de la responsabilité intellectuelle
(transcription, annotation, édition, etc.)

<*name*>

<*date*>

<**editor**>, <**sponsor**>, <**funder**>, <**principal**>, etc.

<fileDesc> : exemple

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Thomas Paine: Common sense, a machine-readable
        transcript</title>
      <respStmt>
        <resp>compiled by</resp>
        <name>Jon K Adams</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <distributor>Oxford Text Archive</distributor>
    </publicationStmt>
    <sourceDesc>
      <bibl>The complete writings of Thomas Paine, collected
        and edited by Phillip S. Foner. New York, Citadel
        Press 1945</bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Publication, diffusion du texte électronique (responsables, droits, etc.)

- Description non structurée dans un <p></p>
- Structuration avec 3 éléments :
 - <**authority**> Les institutions responsables du travail d'encodage, d'édition, etc.
 - <**availability**> *Important* : les droits liés au travail d'édition.

<publicationStmt> : exemple

```
<publicationStmt>
  <distributor>ATILF (Analyse et Traitement Informatique de la
    Langue Française)</distributor>
  <idno type="FRANTEXT">L434</idno>
  <address>
    <addrLine>44, avenue de la Libération</addrLine>
    <addrLine>BP 30687</addrLine>
    <addrLine>54063 Nancy Cedex</addrLine>
    <addrLine>FRANCE</addrLine>
  </address>
  <availability status="free">
    <p>Dans un cadre de recherche ou d'enseignement</p>
  </availability>
</publicationStmt>
```


éléments du <fileDesc> : le <sourceDesc>

Élément obligatoire, Description bibliographique de la source (=> *Métadonnées descriptives*)

- Description non structurée dans un <p></p>
- Structurée de différentes manières, plus ou moins complète
 - <bibl> Faiblement structuré
 - <biblFull> Référence bibliographique structurée (pour tout type de ressource, sur le modèle du <header>)
 - <biblStruct> Référence bibliographique structurée (plus spécifiquement pour un texte imprimé)
 - <listBibl> Liste de références bibliographiques
 - <msDesc> Description d'un manuscrit

<sourceDesc> : exemple

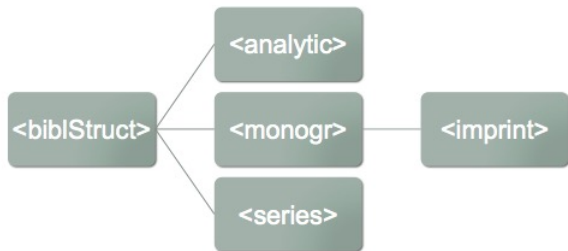
```
<sourceDesc>
  <biblStruct xml:lang="fr">
    <monogr>
      <author>Eugène Sue</author>
      <title>Martin, l'enfant trouvé</title>
      <title type="sub">Mémoires d'un valet de chambre</title>
      <imprint>
        <pubPlace>Bruxelles et Leipzig</pubPlace>
        <publisher>C. Muquardt</publisher>
        <date when="1846">1846</date>
      </imprint>
    </monogr>
  </biblStruct>
</sourceDesc>
```

Documentation approfondie

- Références bibliographiques : <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/CO.html#COBI>
- Description de manuscrits : <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/MS.html>

<biblStruct>

Élément le plus communément utilisé pour décrire une source imprimée.



<biblStruct> : exemple

```
<sourceDesc>
  <biblStruct xml:lang="fr">
    <monogr>
      <author>Eugène Sue</author>
      <title>Martin, l'enfant trouvé</title>
      <title type="sub">Mémoires d'un valet de chambre</title>
      <imprint>
        <pubPlace>Bruxelles et Leipzig</pubPlace>
        <publisher>C. Muquardt</publisher>
        <date when="1846">1846</date>
      </imprint>
    </monogr>
  </biblStruct>
</sourceDesc>
```

<msDesc> : Description de manuscrits

Description fine de l'exemplaire imprimé ancien ou manuscrit :
description physique de la source, histoire de la source, conservation,
reproductions, etc.

Éléments principaux

<msIdentifier> obligatoire Identifiant et localisation de l'exemplaire

<msContents> Description du contenu

<physDesc> Description physique du document

<history> Histoire du document, depuis sa création jusqu'à
l'acquisition par l'institution conservatrice

<additional> bibliographie, des indications sur ses reproductions, ou
des informations sur sa conservation et sur sa gestion, etc.

<msPart> Décrit les parties d'un manuscrit composite

<msDesc>

Identification du manuscrit avec <msIdentifier>

Obligatoire - un seul par <msDesc>

Peut contenir

<institution> Institution conservatrice

<repository> Fonds - collection

<idno> cote

<msName> Autre nom du manuscrit / surnom. *Ex : Manuscrit Voynich*

<country> & <settlement> pays et ville de conservation

<msDesc>

Description du contenu du manuscrit avec <msContents>

Non structuré, avec <p>

```
<msContents>  
  <p>Atlas of the world from Western Europe and Africa to  
    Indochina, containing 27 maps and 26 tables</p>  
</msContents>
```

Structuré, avec <msItem>

```
<msContents>  
  <msItem>  
    <author>Geoffrey Chaucer</author>  
    <title>The Canterbury Tales</title>  
  </msItem>  
</msContents>
```

<msDesc>

Description du contenu du manuscrit avec <msContents> et <msItem>

<msItem>, fils de <msContents> => Informations sur le contenu intellectuel du manuscrit

<author> Auteur du manuscrit,

<respStmt> (mention de responsabilité) comme pour <fileDesc>
indique une responsabilité quant au manuscrit

<title> titre

<textLang> langues du texte

+ Divers éléments plus spécifiques liées à : <rubric>, <incipit>, <explicit>, <finalRubric>, <colophon>, etc.

<msDesc>

Description physique du manuscrit avec <physDesc>

Description physique du document.

Peut contenir :

<objectDesc> Description de l'objet physique

<supportDesc> Description du support

<condition> État matériel

<additions> Ajouts, annotations marginales

<bindingDesc> Description de la reliure

...

<msDesc>

Description physique du manuscrit avec <physDesc>

```
<physDesc>
  <objectDesc form="codex">
    <supportDesc material="perg">
      <support>Parchment.</support>
      <extent>i + 55 leaves
      <dimensions scope="all" type="leaf" unit="inch">
        <height>7 1/4</height><width>5 3/8</width>
      </dimensions>
    </extent>
  </supportDesc>
  <layoutDesc>
    <layout columns="2">In double columns.</layout>
  </layoutDesc>
</objectDesc>
<handDesc><p>Written in more than one hand.</p></handDesc>
<decoDesc><p>With a few coloured capitals.</p></decoDesc>
</physDesc>
```

<msDesc>

Historique de conservation du manuscrit avec <history>

Historique complet du manuscrit, de sa création jusqu'à son arrivée dans la collection qui le contient aujourd'hui

Peut contenir

<origin>

<provenance> Informations sur un épisode précis de l'histoire du manuscrit

<acquisition> Modalités et circonstances de l'entrée du manuscrit ou de la partie du manuscrit dans l'institution qui le détient

<msDesc>

Historique de conservation du manuscrit avec <history>

```
<history>
  <origin>
    <p>Written in <origPlace>Durham</origPlace> during <origDate
      notBefore="1125" notAfter="1175">the mid-twelfth century<
      /origDate>.</p>
  </origin>
  (...)
  <provenance>
    <p>Given to <name type="person">W. Olleyf</name> by <name type
      ="person">William Ebchester, Prior (1446-56)</name> and
      later belonged to <name type="person">Henry Dalton</name>
      , Prior of Holy Island (<name type="place">Lindisfarne</
      name>) according to inscriptions on ff. 4v and 5.</p>
  </provenance>
  <acquisition>
    <p>Presented to <name type="org">Trinity College</name> in <
      date>1738</date> by <name type="person">Thomas Gale</name
      > and his son <name type="person">Roger</name>.</p>
  </acquisition>
</history>
```

Conclusion

Le <teiHeader> est un outil pour produire des métadonnées très diverses et de très grande qualité :

http://www.e-codices.unifr.ch/xml/tei_published/mhv-1346_Bernasconi.xml

Points à retenir pour éviter l'éparpillement et le vertige :

- Lire la ?\$#@! de documentation,
- Une description minimale MAIS structurée vaut mieux qu'une grande quantité d'informations en désordre,
- La description s'affine au fur et à mesure qu'on travaille sur ses données,
- Connaître les besoins (Organisation du travail, citation, publication = différents types de métadonnées) et les contraintes institutionnelles éventuelles (travail dans le cadre d'un projet notamment),
- Idéal : Documenter sa façon de produire ses métadonnées.

On essaie ?