



HAL
open science

Modélisation des métadonnées en XML-TEI. Pratiques de l'IR Métopes et du Pôle Document numérique de la MRSH de Caen

Marie Bisson, Edith Cannet, Anne Goloubkoff, Stéphane Lecouteux, Émeline Mancel, Subha-Sree Pasupathy, Claire Remoué, Julia Roger

► To cite this version:

Marie Bisson, Edith Cannet, Anne Goloubkoff, Stéphane Lecouteux, Émeline Mancel, et al.. Modélisation des métadonnées en XML-TEI. Pratiques de l'IR Métopes et du Pôle Document numérique de la MRSH de Caen. 2022. hal-03830903v1

HAL Id: hal-03830903

<https://hal.science/hal-03830903v1>

Preprint submitted on 14 Dec 2022 (v1), last revised 5 Oct 2023 (v2)

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 4.0 International License

Modélisation des métadonnées en XML-TEI : pratiques de l'IR Métopes et du Pôle Document numérique de la MRSH de Caen

Marie Bisson, Edith Cannet, Anne Goloubkoff, Stéphane Lecouteux, Émeline Mancel, Subha-Sree Pasupathy, Claire Remoué et Julia Roger

Le groupe de travail « Métadonnées » du pôle Document numérique (PDN) et de l'IR Métopes a été créé en 2020¹. Son objectif visait à modéliser un `<teiHeader>`, à l'échelle de l'ensemble des sources anciennes encodées en XML-TEI relevant du PDN, ainsi qu'aux données textuelles (unité éditoriale de type article et chapitre, volume) prises en charge par la chaîne d'édition structurée Métopes et conformément au schéma d'encodage TEI COMMONS Publishing, partagé avec l'IR OpenEdition.

Le GT a également agrégé le modèle de `<teiHeader>` proposé par le consortium CAHIER pour l'encodage des métadonnées des manuscrits et des imprimés². Il s'est parfois écarté de ce modèle et s'en est justifié, le cas échéant.

Le projet de dépôt dans Nakala³ des sources anciennes traitées au PDN a fourni l'occasion de mener cet effort collectif d'harmonisation.

Ce vadémécum se propose ainsi comme un set « idéal » d'éléments, d'attributs et de valeurs pour tout document XML-TEI du PDN et de Métopes décrivant une source textuelle, assortie d'un travail d'édition au sens large (matérielle et/ou scientifique).

D'abord conçu pour les ingénieurs de Caen, il est ici proposé à une communauté plus large d'encodeurs de `<teiHeader>` de sources anciennes. Nous avons fourni un certain nombre d'exemples afin qu'il soit le plus utile et le plus compréhensible possible.

Rédigé comme un document de travail, il est susceptible d'évoluer encore au gré des nouvelles sources traitées et des objectifs scientifiques qui leur sont liées. Il est accompagné ici d'un fichier type de `<teiHeader>` au format texte.

Ont participé aux travaux du groupe : Marie Bisson (PDN), Édith Cannet (IR Métopes), Anne Goloubkoff (PDN), Stéphane Lecouteux (PDN/Biblissima+), Émeline Mancel (PDN), Subha-Sree Pasupathy (PDN), Claire Remoué (IR Métopes) et Julia Roger (PDN).

Pour plus d'informations, vous pouvez contacter Subha-Sree Pasupathy (PDN) : subha-sree.pasupathy@unicaen.fr et Claire Remoué (IR Métopes) : claire.remoue@unicaen.fr.

Table des matières

Description bibliographique du fichier XML <fileDesc>

- I. Titre et responsabilités <titleStmt>
 - A. Typologie des titres
 - B. Autorités
 - a. Les rôles
 - b. Les noms
 - c. Les institutions de rattachement
 - d. Les identifiants
- II. Mention d'édition <editionStmt>
 - A. Numéro (et date) d'édition scientifique
 - B. Financeur
 - C. Institutions commanditaires et partenaires
 - D. Les chargés d'édition (au sens large)
- III. Informations sur la publication <publicationStmt>
 - A. Responsable de la publication
 - a. Éditeur matériel
 - b. Autre responsable de la publication
 - B. Lieu de publication
 - C. Accessibilité de la ressource
 - D. Identifiant pour le projet / le support de diffusion
 - E. Gestion des données par support de diffusion
 - a. Informations pour le site
 - b. Informations pour un dépôt
 - F. Gestion des métadonnées pour OpenEdition – spécificités Métopes (UE)
- IV. Description de la collection <seriesStmt>
- V. Description des documents sources <sourceDesc>
 - A. Indiquer que le texte est nativement numérique
 - B. Indiquer une référence bibliographique de cartouche de citation
 - C. Indiquer une édition princeps
 - D. Indiquer une édition de sources
 - a. Source manuscrite
 - b. Source imprimée
 - E. Cas particulier de l'édition critique, méthode de la segmentation parallèle

Description de l'encodage <encodingDesc>

- I. Déclaration de balisage <tagsDecl>
- II. Déclaration des pratiques éditoriales <editorialDecl>
- III. Informations d'application <appInfo>
- IV. Déclaration de classification <classDecl>

Description du profil <profileDesc>

- I. Résumé
- II. Création

III. Langues

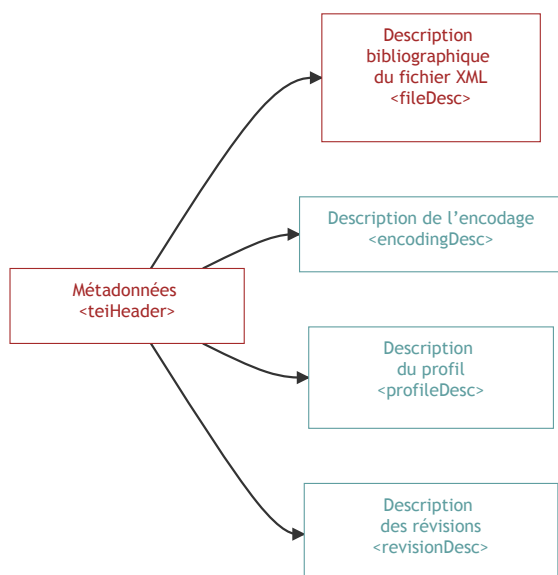
IV. Classification du document

V. Mains annotatrices

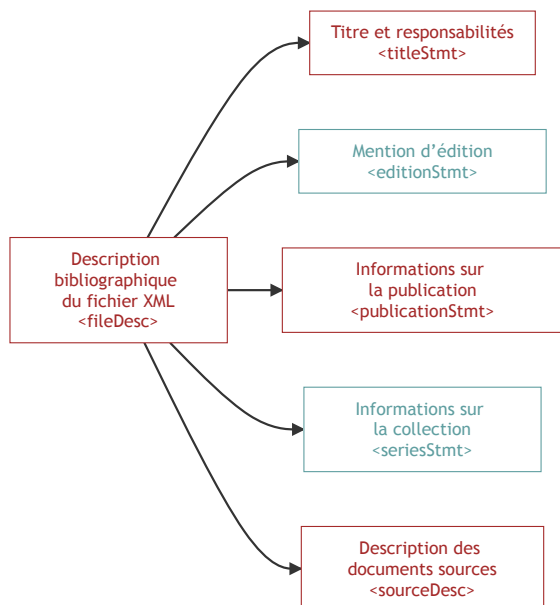
Description des révisions <revisionDesc>

Le module Métadonnées intègre plusieurs sets :

- Description bibliographique du fichier XML lui-même :
 - titre et responsabilités ;
 - mention d'édition ;
 - informations sur la publication ;
 - mention de la collection ;
 - description des sources éditées.
- Description de l'encodage :
 - description de balisage particulier ;
 - description des pratiques éditoriales ;
 - description des applications utilisées ;
 - description des référentiels utilisés.
- Description du profil :
 - résumé ;
 - création ;
 - langue(s) ;
 - classification du texte ;
 - main(s).
- Description des révisions.



Description bibliographique du fichier XML `<fileDesc>`



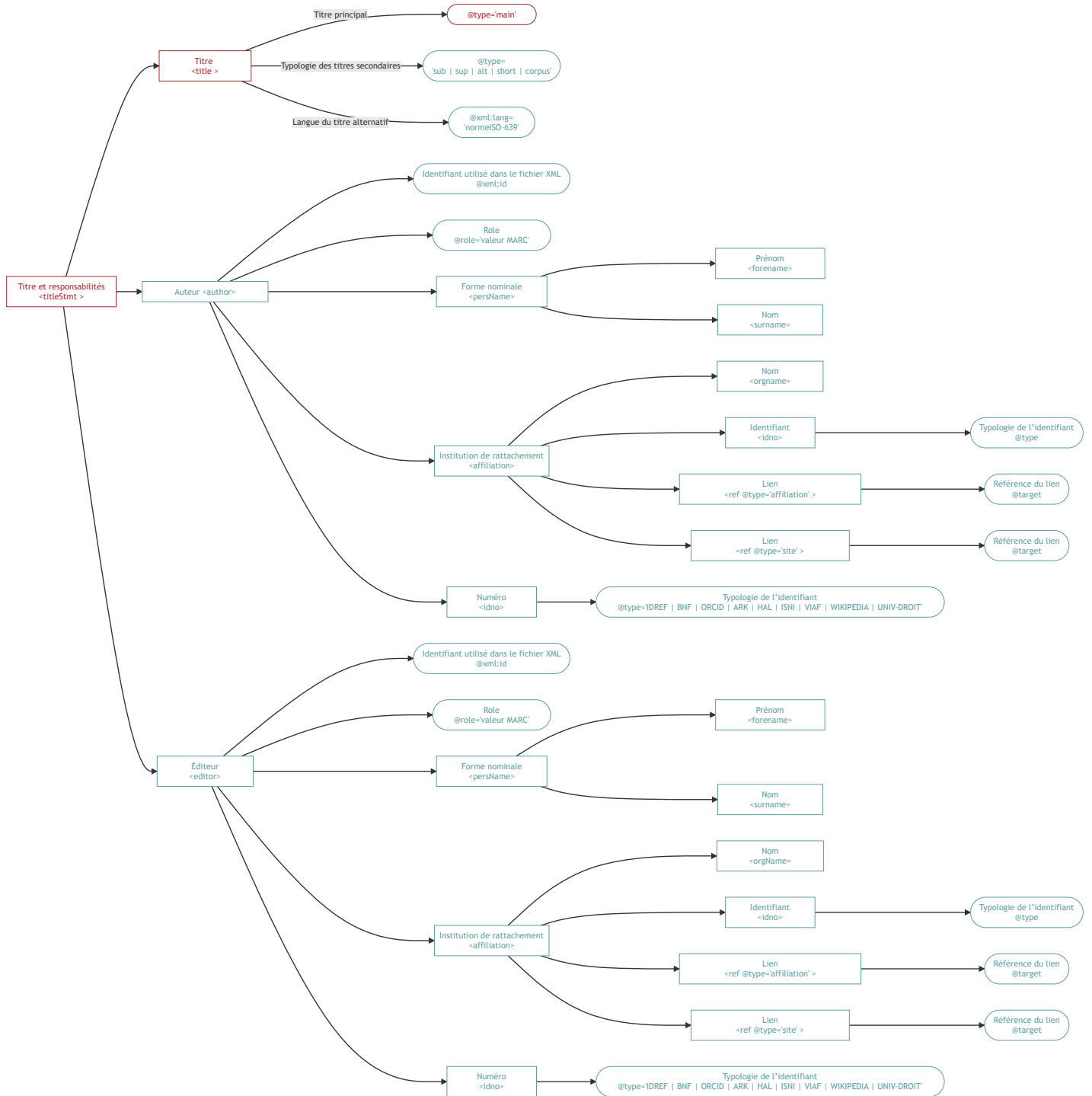
La description bibliographique du fichier s'insère au sein de l'élément `<fileDesc>` et Métopes propose de l'aligner sur le modèle des [FRBR](#) (*Functional Requirements for Bibliographic Records*):

- `<titleStmt>` : niveau *work* (œuvre, expression) ;
- `<editionStmt>` : niveau *manifestation* ;
- `<publicationStmt>` : niveau *item* (format de diffusion).

Cette description est complétée par :

- `<seriesStmt>` : niveau collection ;
- `<sourceDesc>` : documentation de la/des source(s) du document.

I. Titre et responsabilités <titlestmt>



A. Typologie des titres

- Titre principal (unique et obligatoire) : `<title[@type="main"]>` ;
- Sous-titre : `<title[@type="sub"]>` ;
- Sur-titre : `<title[@type="sup"]>` ;
- Titre(s) traduit(s) : `<title[@type="trl"][@xml:lang]>` ;
- Titre court : `<title[@type="short"]>` ;
- Titre(s) de corpus : `<title[@type="corpus"]>` .

Nota Bene

Le titre court est notamment utile pour : les index, le sommaire, le fil d'ariane (pour le [moteur d'affichage MaX](#)), le cartouche de citation, les versions PAO et PDF.

Le titre de corpus est utile pour regrouper des fichiers XML qui appartiennent à un même groupe (collection de sources, ouvrage, tome, etc.). C'est un titre qui dépend d'un éditeur scientifique. L'éditeur scientifique peut affiner sa typologie à l'aide de l'attribut `@subtype` .

```
<titleStmt>
  <title type="main">Titre principal</title>
  <title type="sub">Sous-titre</title>
  <title type="sup">Surtitre</title>
  <title type="trl" xml:lang="en">Translated title</title>
  <title type="short">Titre court</title>
  <title type="corpus">Titre du module</title>
</titleStmt>
```

COMMONS : `<title type="corpus">` n'est pas utilisé dans le schéma Commons Publishing.

B. Autorités

Un attribut `@xml:id` peut être ajouté aux éléments `<author>` et `<editor>` pour différencier des responsabilités au sein d'un texte. La valeur de l'attribut `@xml:id` pourra alors être reprise, précédée de `#` , en valeur des attributs `@resp` ou `@hand` dans le document XML.

a. Les rôles

Faute de référentiel, nous retenons les codes du [MARC](#) (pratique en usage dans Métopes) pour définir les rôles des éléments `<author>` et `<editor>` . Voici la liste des rôles les plus courants :

- auteur : `aut` ;
- éditeur scientifique : `edt` ;
- responsable scientifique : `pdr` ;

- annotateur : `ann` ;
- contributeur : `ctb` ;
- commentateur : `com` ;
- cartographe : `ctg` ;
- dessinateur : `drm` ;
- directeur de publication : `pbd` ;
- illustrateur : `ill` ;
- photographe : `pht` ;
- copiste (ou scribe) : `scr` ;
- traducteur : `trl` ;
- directeur d'opération (archéologie) : `fld` ;
- collectionneur, *collector* (qui peut fonctionner pour catalographe) : `col` ;
- compilateur, *compiler* (qui peut fonctionner pour catalographe) : `com` ;
- expert (pour tous les rôles non parfaitement définis) : `exp` .

b. Les noms

Nous recommandons d'ajouter un élément `<persName>` et à l'intérieur, si possible, les éléments `<forename>` (prénom) et `<surname>` (nom de famille).

```
<author role="aut" xml:id="GC">
  <persName>
    <forename>Grégory</forename>
    <surname>Combalbert</surname>
  </persName>
</author>
```

c. Les institutions de rattachement

L'institution de rattachement d'un auteur ou d'un éditeur se place dans un élément `<affiliation>`. Si une autorité est rattachée à plusieurs institutions, on ajoute autant d'éléments `<affiliation>` que nécessaire. Pour rationaliser l'utilisation de cet élément, on ajoute les éléments enfants suivants :

- `<orgName>` : pour désigner l'institution de rattachement dans sa forme normalisée. L'attribut `@type` permet de préciser, le cas échéant, de quel référentiel (ROR, ISNI, RNSR, idRef, BNF) provient cette forme normalisée.
- `<idno>` : pour renseigner l'identifiant correspondant au référentiel choisi ;
- `<ref @type="site" @target="URL">` : pour pointer vers le site web de l'affiliation ou la page personnelle du chercheur.

```
<affiliation>
  <orgName>Centre Michel de Bouard - Craham - UMR 6273</orgName>
  <ref target="https://www.craham.cnrs.fr" type="site">
    https://www.craham.cnrs.fr
  </ref>
</affiliation>
```

Nota Bene

Il existe deux méthodes pour renseigner l'url d'un site : soit en utilisant simplement l'attribut @ref dans l'élément directement concerné, soit en utilisant l'élément `<ref>` et son attribut `@target`, ce qui permet notamment de faire plusieurs liens.

d. Les identifiants

Les identifiants sont encodés au moyen de l'élément `<idno>`. Une typologie de ces identifiants est proposée pour l'attribut `@type` :

- IDREF (récupération d'une interrogation d'[idref](#)) ;
- BNF ;
- ORCID ;
- ARK ;
- HAL ;
- ISNI ;
- VIAF ;
- WIKIDATA ;
- UNIV-DROIT.

```
<author role="aut" xml:id="GC">
  <persName>
    <forename>Grégory</forename>
    <surname>Combalbert</surname>
  </persName>
  <idno type="BNF">ark:/12148/cb17982062c</idno>
  <idno type="VIAF">213343228</idno>
</author>
```

```

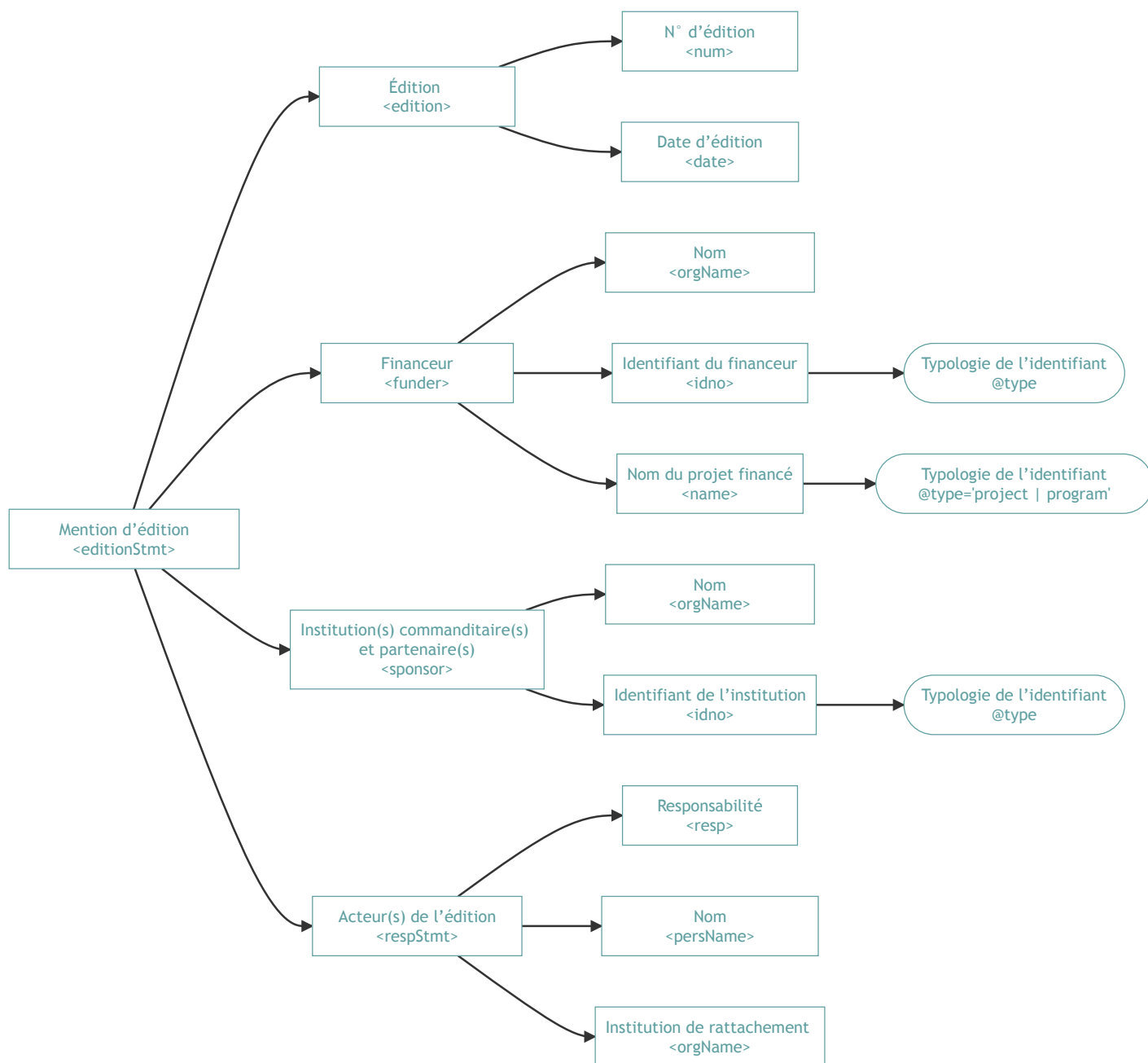
<affiliation>
  <orgName>Centre Michel de Bouard - Craham - UMR 6273</orgName>
  <orgName type="BNF">Centre de recherches archéologiques et historiques
anciennes et médiévales (Caen)</orgName>
  <idno type="BNF">ark:/12148/cb118789712</idno>
  <ref target="https://www.craham.cnrs.fr" type="site">
    https://www.craham.cnrs.fr
  </ref>
</affiliation>

```

COMMONS : les balises `<author>` et `<editor>` peuvent également comprendre la biographie de l'auteur sous la forme `<ref target="#bio01" type="biography"></ref>` qui pointe vers `<div type="biography" xml:id="bio01"><p>[...]</p></div>` dans la balise `<back>`. Ces balises peuvent aussi inclure le mail sous la forme `<email>[...]</email>`.

COMMONS : l'élément `<funder>` qui comprend `<orgName>`, `<idno type="awardnumber">` et `<idno type="funder_registry">`, est dans le `<titleStmt>` alors que le choix a été fait au PDN de mettre cette information dans la balise `<editionStmt>` (voir II.B).

II. Mention d'édition `<editionStmt>`



COMMONS : pour les unités éditoriales, `<editionStmt>` est supprimé.

A. Numéro (et date) d'édition scientifique

Le numéro d'édition scientifique se précise au sein de l'élément `<edition>`. On propose d'y ajouter l'élément `<num>`.

Nota Bene

Cela permet, si cela est toujours nécessaire pour Nakala, d'ajouter la date d'édition scientifique au sein d'un élément `<date>`.

Il faut encore se mettre d'accord sur ce à quoi correspond cette date.

```
<edition>
  <num>2</num>
  <date when="1992-01-06">6 janvier 1992</date>
</edition>
```

Nota Bene

Le format de la date suit la norme ISO 8601 (YYYY-MM-DD).

B. Financier

Le financeur s'ajoute au sein d'un élément `<funder>`, contenant les éléments enfants suivants :

- le nom du financeur au sein d'un élément `<orgName>` ;
- l'identifiant de ce financeur au sein de l'élément `<idno>`, qualifié de l'attribut `@type` ;
- pour les projets de recherche du PDN, nous ajoutons les nom des projet et des programmes dans un élément `<name>` avec un attribut `@type` de valeur `project` ou `program`.

Exemples des Actes des évêques d'Évreux (XI^e siècle-1223) ([e-cartae](#))

```
<funder>
  <orgName ref="https://anr.fr/">Agence nationale de la recherche
  (ANR)</orgName>
  <name type="project">Actépi (2019-2023)</name>
  <idno type="ANR">ANR-18-CE27-0001</idno>
</funder>
```

```
<funder>
  <orgName ref="https://www.normandie.fr/">Région Normandie</orgName>
  <name type="project">Vexicaen (2017-2019)</name>
  <name type="program">CPER (2015-2020)</name>
</funder>
```

C. Institutions commanditaires et partenaires

Les institutions – commanditaires ou partenaires – peuvent être spécifiées au moyen de l'élément

`<sponsor>` . Cela correspond à un soutien scientifique (une caution intellectuelle, pas forcément un apport financier).

On peut préciser comme enfant de cet élément :

- le nom de l'institution au sein d'un élément `<orgName>` , qui permet de distinguer le nom ;
- l'identifiant, ajouté au sein de l'élément `<idno>` et qualifié par l'attribut `@type` .

```
<sponsor>
  <orgName ref="https://www.unicaen.fr/recherche/mrsh/document_numerique">
    Pôle document numérique - MRSB – UAR 3486
  </orgName>
  <orgName type="IDREF">
    Maison de la recherche en sciences humaines (Caen)
  </orgName>
  <idno type="IDREF">033290814</idno>
</sponsor>
```

COMMONS : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

D. Les chargés d'édition (au sens large)

Les différents intervenants dans le processus d'édition se placent au sein d'un élément `<respStmt>` .

On y trouve notamment :

- la (ou les) personne(s) qui a (ont) mis en place l'encodage XML ;
- le(s) chargé(s) d'édition ;
- etc.

L'architecture de l'élément `<respStmt>` est composée de la manière suivante :

- `<resp>` : pour décrire la tâche exercée ;
- `<persName>` : pour spécifier le nom de la personne responsable de la tâche ;
- `<orgName>` : pour spécifier l'institution de rattachement de la personne ou l'institution responsable de la tâche quand celle-ci n'est pas sous la responsabilité d'une personne mais d'une institution.

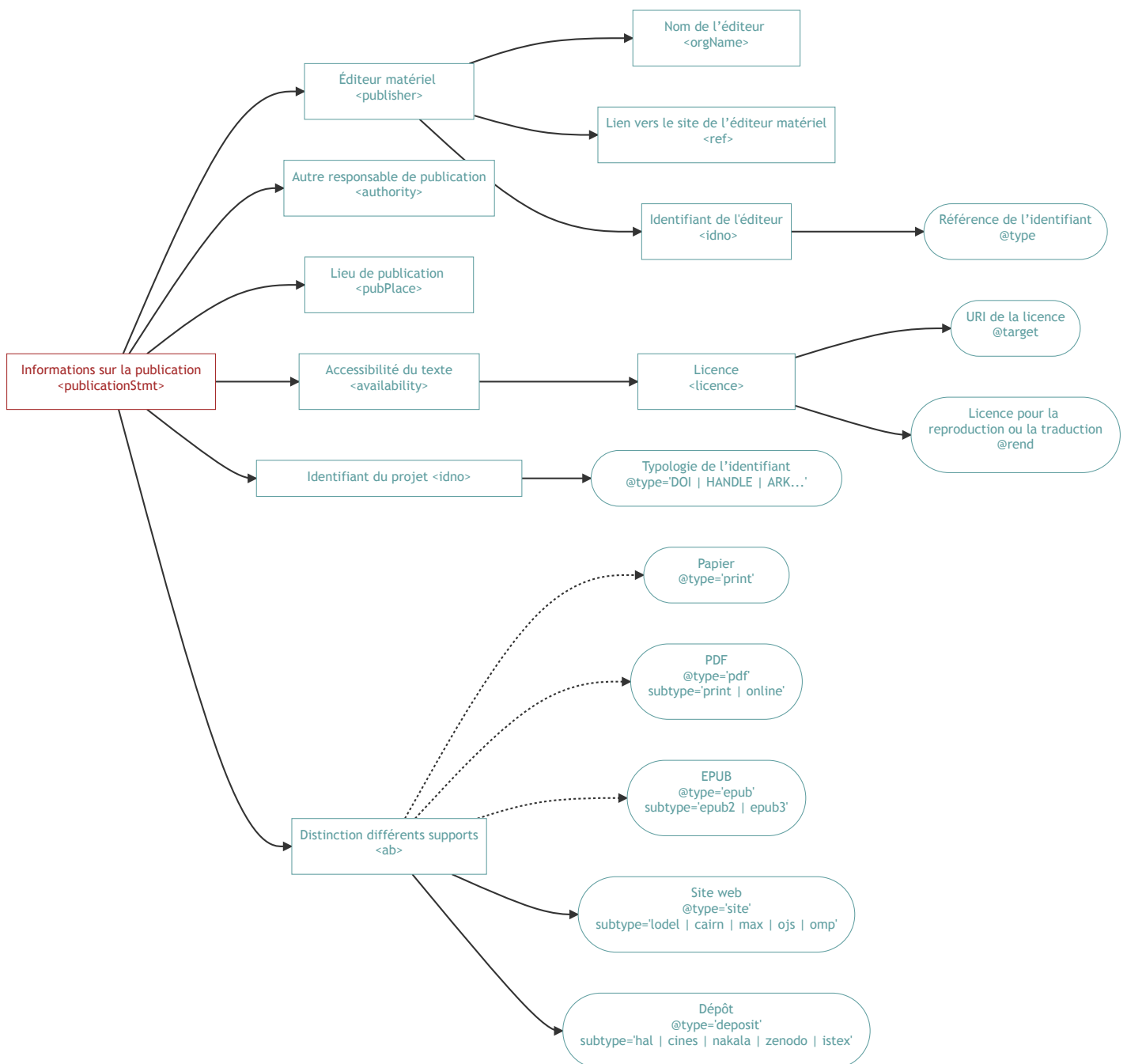
Si on ajoute un `@xml:id` sur l'élément `<respStmt>` , on pourra en convoquer la valeur ailleurs dans le document, notamment au sein de l'attribut `@who` de l'élément consacré à décrire une révision

`<change>` (cf. *infra* [Description des révisions](#)).

```
<respStmt xml:id="GC">
  <resp>Édition critique et encodage XML</resp>
  <persName>Grégory Combalbert</persName>
  <orgName>Centre Michel de Boüard - Craham</orgName>
</respStmt>
```

COMMONS : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

III. Informations sur la publication **<publicationStmt>**



A. Responsable de la publication

a. Éditeur matériel

Les informations concernant l'éditeur matériel sont renseignées au sein de l'élément `<publisher>`. Le nom de l'éditeur matériel est précisé au sein de l'élément `<orgName>`. L'élément `<publisher>` peut également contenir un élément `<ref>` pour préciser l'URL du site de l'éditeur matériel et un élément `<idno>` pour préciser un identifiant.

```
<publisher>
  <orgName type="IDREF">Presses universitaires de Caen</orgName>
  <ref target="https://www.puc-ed.fr"/>
  <idno type="IDREF">194901602</idno>
</publisher>
```

COMMONS : `<publisher>` ne contient pas d'autre élément.

b. Autre responsable de la publication

Pour encoder le responsable de la publication d'un fichier électronique, qui n'est ni éditeur ni distributeur, nous utilisons la balise `<authority>`.

```
<authority>
  Pôle Document numérique
</authority>
```

B. Lieu de publication

Le lieu de publication est renseigné au moyen de l'élément `<pubPlace>`.

```
<pubPlace>Caen</pubPlace>
```

COMMONS : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

C. Accessibilité de la ressource

L'élément `<availability>` est recommandé par CAHIER pour introduire le type de licence.

On ajoute à l'intérieur un élément `<licence>` et on pointe vers l'URI de la licence en valeur de l'attribut `@target` :

- [CC-BY](#) : attribution (licence libre non copyleft) ;

- [CC-BY-SA](#) : attribution + partage dans les mêmes conditions (licence libre copyleft) ;
- [CC-BY-ND](#) : attribution + pas de modification (licence de libre diffusion) ;
- [CC-BY-NC](#) : attribution + pas d'utilisation commerciale (licence de libre diffusion) ;
- [CC-BY-NC-SA](#) : attribution + pas d'utilisation commerciale + partage dans les mêmes conditions (licence de libre diffusion) ;
- [CC-BY-NC-ND](#) : attribution + pas d'utilisation commerciale + pas de modification (licence de libre diffusion).

Il s'agit de la licence choisie par le chercheur pour son fichier XML, avant l'intervention d'un éventuel éditeur matériel.

```
<availability>
  <licence target="https://creativecommons.org/licenses/by/4.0/">
    CC-BY
  </licence>
  <licence rend="translation">
    Licence pour un volume Métopes concernant les droits de traduction.
  </licence>
  <licence rend="reproduction">
    Licence pour un volume Métopes relative aux droits de reproduction de l'œuvre
  </licence>
</availability>
```

COMMONS : `<licence>` peut contenir un élément `<p>` qui permet d'ajouter une description longue des droits d'utilisation.

Nota Bene

Les licences des différents supports de diffusion de ce fichier XML (livre papier, HTML, PDF en ligne, dépôt) sont renseignées **par support** dans un élément `<ab>` (cf. *infra*).

D. Identifiant pour le projet / le support de diffusion

On peut renseigner ici l'identifiant d'un projet au moyen de l'élément `<idno>`. Il faut se poser la question des DOI/HANDLE/ARK pour chaque projet et **s'il est pertinent d'attribuer un identifiant et à quelle étape du fichier XML** ? Concernant les éventuels DOI attribués à des fichiers XML pour des corpus qui sont publiés chez un éditeur matériel, il convient de se mettre d'accord avec ce dernier.

Comme pour les deux points ci-dessus (date et licence), se pose la question du format. On peut donner un identifiant au fichier XML (archivage) mais on en donnera probablement aussi un à chaque support de diffusion.

Exemples

- ARK pour le fichier XML ;
- DOI pour la ressource en ligne ;
- ISBN pour le livre imprimé ;
- ISSN pour la revue ;
- etc.

Nota Bene

Le Guide CAHIER préconise de renseigner ici rétroactivement le DOI délivré par Nakala. Nous préférons renseigner cet identifiant « Nakala » comme une spécificité de support, cf. *infra*.

```
<idno type="ARK | DOI | ISBN | ISSN"></idno>
```

COMMONS : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

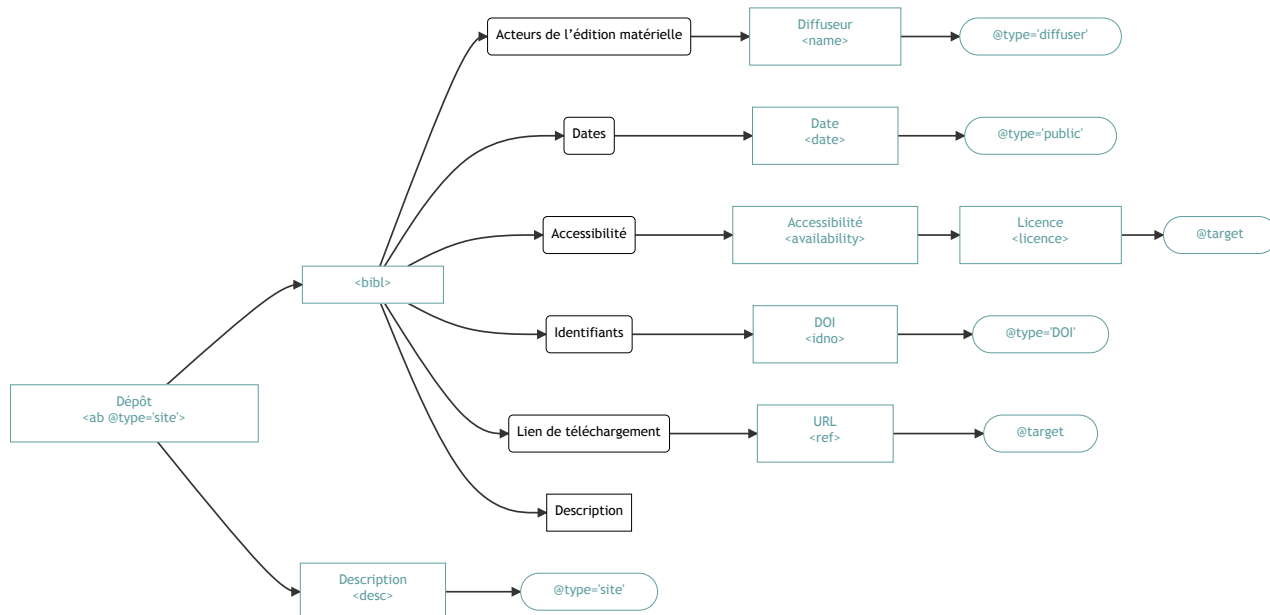
E. Gestion des données par support de diffusion

Afin de spécifier les métadonnées par support de publication, nous utilisons l'élément `<ab>` typé (attribut `@type`) selon la typologie révisée suivante :

- Typologie des `<ab>` :
 - `@type` = `print`, `pdf`, `epub`, `site`, `deposit` ;
 - `@subtype` = `print`, `online` pour le `@type=pdf` ;
 - `@subtype` = `epub2`, `epub3` pour le `@type=epub` ;
 - `@subtype` = `lodel`, `cairn`, `max`, `ojs`, `omp` pour le `@type=site` ;
 - `@subtype` = `hal`, `cines`, `nakala`, `istex` pour le `@type=deposit` .

Nous traitons dans ce document, pour l'instant, uniquement des éléments `<ab>` ayant pour attribut `@type='site'` et `@type='deposit'` .

a. Informations pour le site



Au sein de cet élément, et à l'intérieur de l'élément `<bibl>`, on va pouvoir documenter :

- des noms liés aux différents responsables de l'édition matérielle encodés au moyen de l'élément `<name>` qualifié par l'attribut `@type` (à mettre en lien avec [A. Responsable de la publication](#)) ;
- des mentions de dates encodées au moyen de l'élément `<date>` qualifié par l'attribut `@type` ;
- des liens vers les sites de diffusion `<ref>` ;
- des identifiants encodés au moyen de l'élément `<idno>`, qualifié par l'attribut `@type` (à mettre en lien avec [d. Les identifiant](#)) ;
- des licences encodées au moyen de l'élément `<licence>`, avec l'attribut `@target` (à mettre en lien avec [C. Accessibilité de la ressource](#)). Ces éléments doivent être placés à l'intérieur d'un élément `<availability>` ;

Pour ajouter une description, on peut l'encoder au moyen de l'élément `<desc>`, avec l'attribut `@type`, après l'élément `<bibl>`.

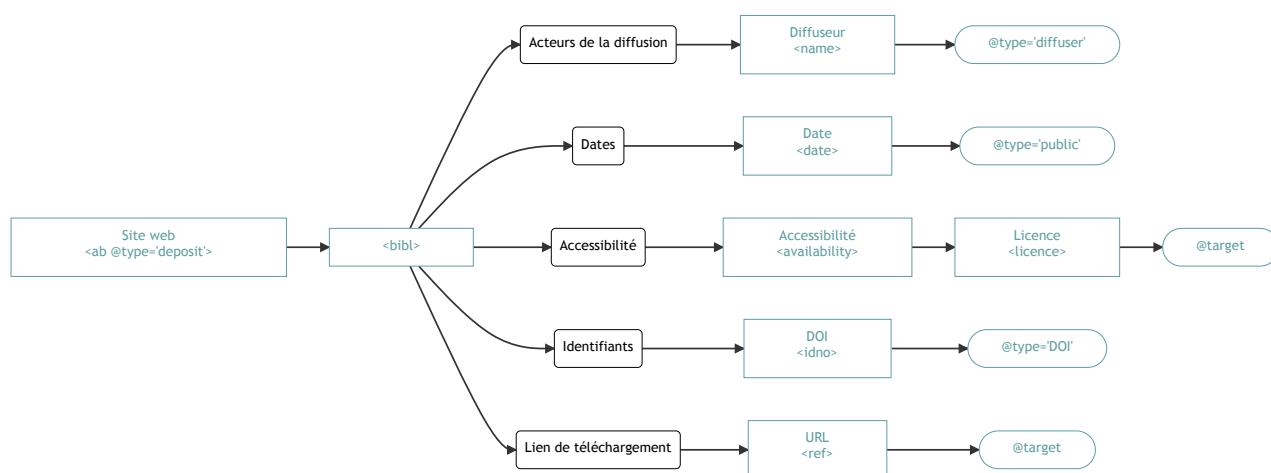
Exemple

```

<ab type="site">
  <bibl>
    <name type="diffuser">Presses universitaires de Caen</name>
    <date type="public">9 décembre 2019</date>
    <ref target="https://www.unicaen.fr/puc/sources/ecartae/evreux/accueil">
      https://www.unicaen.fr/puc/sources/ecartae/evreux/accueil
    </ref>
    <idno type="DOI">[DOI à venir]</idno>
    <availability>
      <licence target="https://creativecommons.org/licenses/by-nc-sa/4.0/">
        Creative Commons - Attribution-NonCommercial-ShareAlike 4.0
        International - CC BY-NC-SA 4.0
      </licence>
    </availability>
  </bibl>
  <desc type="site">Édition en ligne des actes des évêques d'Évreux</desc>
</ab>

```

b. Informations pour un dépôt



Au sein de cet élément, et à l'intérieur d'un élément `<bibl>`, on va pouvoir documenter :

- des noms liés aux différents responsables de l'édition matérielle encodés au moyen de l'élément `<name>` qualifié par l'attribut `@type` (à mettre en lien avec [A. Responsable de la publication](#)) ;
- des mentions de dates encodées au moyen de l'élément `<date>` qualifié par l'attribut `@type` ;
- des liens vers les sites de diffusion `<ref>` ;
- des identifiants encodés au moyen de l'élément `<idno>`, qualifié par l'attribut `@type` (à mettre en lien avec [D. Identifiant](#)) ;

- des licences encodées au moyen de l'élément `<licence>`, avec l'attribut `@target` (à mettre en lien avec [B. Licence](#)). Ces éléments doivent être placés à l'intérieur d'un élément `<availability>`.

Exemple

```
<ab type="deposit" subtype="nakala">
  <bibl>
    <date type="public">01/12/2020</date>
    <ref target="https://nakala.fr">https://nakala.fr</ref>
    <idno type="DOI">[identifiant DOI fourni par nakala]</idno>
    <availability>
      <licence target="https://creativecommons.org/licenses/xxx">
        [Licence relative aux droits du site]
      </licence>
    </availability>
  </bibl>
</ab>
```

COMMONS : les informations générales sont placées dans un bloc `<ab>` avec un attribut `@type` ayant pour valeur `editorial_workflow` :

```
<ab type="editorial_workflow">
  <date type="received">2019-06-01</date>
  <date type="accepted">2020-01-01</date>
</ab>````
```

Puis, pour chaque format, un élément `<ab>` est ajouté contenant toutes les informations nécessaires :

```
````xml
<ab type="print">
 <bibl>
 <publisher> UMR 8504 Géographie-cités</publisher>
 <idno type="documentnumber">108</idno>
 <dim type="pagination">25-36</dim>
 <availability>
 <licence target="https://creativecommons.org/licenses/by-nc-nd/4.0/">
 <p>Creative Commons - Attribution-NonCommercial-NoDerivatives
 4.0 International - CC BY-NC-ND 4.0</p>
 </licence>
 </availability>
 </bibl>
</ab>
<ab type="pdf">
 <bibl>[...]</bibl>
</ab>
```

## F. Gestion des métadonnées pour OpenEdition – spécificités Métopes (UE)

Cartouche de métadonnées pour la publication sur OpenEdition comprenant les informations suivantes :

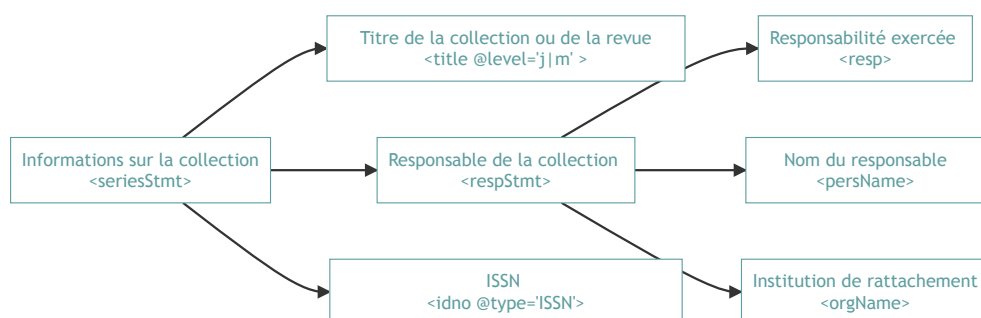
- pagination de la version imprimée ;
- date de publication imprimée ;
- date de publication électronique.

```

<!-- Métadonnées dont a besoin OpenEdition -->
<ab type="print">
 <bibl>
 <dim type="pagination">25-36</dim>
 [...]
 </bibl>
</ab>
<ab @type="online">
 <bibl>
 <distributor>OpenEdition</distributor>
 <date type="publishing">2020-02-01</date>
 <date type="embargoend">2021-02-01</date>
 <idno type="DOI">[DOI]</idno>
 <idno type="URL">[URL]</idno>
 [...]
 </bibl>
</ab>

```

#### IV. Description de la collection `<seriesStmt>`



On utilise `<seriesStmt>` pour l'édition matérielle : collections, revues.

#### Nota Bene

Les corpus et sous-corpus (catégorie de recherche) sont mentionnés *supra* dans `<titleStmt>`.

On peut ajouter la même architecture pour les directeurs de collection ou de revues et pour les chargés d'édition (cf. *supra* `<editionStmt>`).

L'attribut `@level` de l'élément `<title>` permet de distinguer les ouvrages (*monography*) et les revues (*journal*), selon les recommandations de la TEI.

On peut ajouter l'ISSN ici au moyen de l'élément `<idno>` qualifié par l'attribut `@type` de valeur `ISSN`.

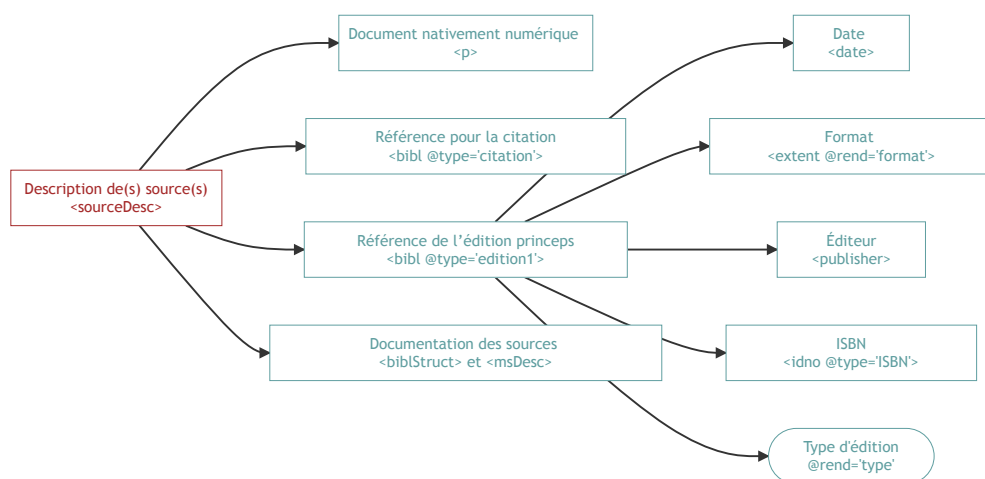
```

<seriesStmt>
 <title level="mlj">[Nom de la collection]</title>
 <respStmt>
 <resp>Directeur de collection</resp>
 <persName>[John Doe]</persName>
 <orgName>[Institution de rattachement de John Doe]</orgName>
 </respStmt>
 <idno type="ISSN">[xxxx-xxxx]</idno>
</seriesStmt>

```

**COMMONS** : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

## V. Description des documents sources <sourceDesc>



**COMMONS** : Dans le cadre d'une unité, <sourceDesc> ne contient qu'une balise <p> qui précise la version du schéma Commons Publishing utilisée.

Nous distinguons 4 types de renseignements possibles au sein de cet élément :

- mentionner que le document est nativement numérique (quand les trois items suivants sont absents) ;
- introduire un cartouche de citation ( <bibl type='citation''> ) ;
- préciser l'édition princeps ( <bibl type='edition1''> ) ;
- documenter les sources d'une édition ( <biblStruct> et <msDesc> ).

Il faut distinguer les productions scientifiques inédites des éditions de sources :

- le fichier XML est natif, ou produit à partir d'un fichier de traitement de texte numérique natif :
  - cas général traité par les outils de Métopes ;
  - cas des introductions à des textes scientifiques ;
  - cas des textes de présentation d'un site web destinés à être des pages HTML statiques (via le moteur d'affichage MaX par exemple) ;



- le fichier XML (natif ou produit à partir d'un fichier de traitement de texte ou autre) est une transcription / édition / annotation / enrichissement d'une ou plusieurs « sources » manuscrites ou imprimées, qu'il faut alors décrire.

## A. Indiquer que le texte est nativement numérique

Une chaîne de caractères est placée au sein d'un élément `<p>`.

```
<sourceDesc>
 <p>Fichier nativement numérique.</p>
</sourceDesc>
```

## B. Indiquer une référence bibliographique de cartouche de citation

Pour préciser la mention de citation du document ou un corpus, on renseigne ici, sous forme libre, un cartouche de citation. L'encodeur utilise l'élément `<bibl>` au sein de l'élément `<sourceDesc>`. L'élément `<bibl>` est typé (attribut `@type`) au moyen de la valeur `citation`.

```
<sourceDesc>
 <bibl type="citation"><hi rend="italique">Actes des évêques d'Évreux
 (<hi rend="small-caps">XI<hi rend="sup">e</hi> siècle-1223)</hi>,
 Grégory Combalbert (éd.), Caen, Presses universitaires de Caen
 (e-Cartae – Actépi), 2019, en ligne :
 https://www.unicaen.fr/puc/sources/ecartae/evreux/accueil</bibl>
</sourceDesc>
```

## C. Indiquer une édition princeps

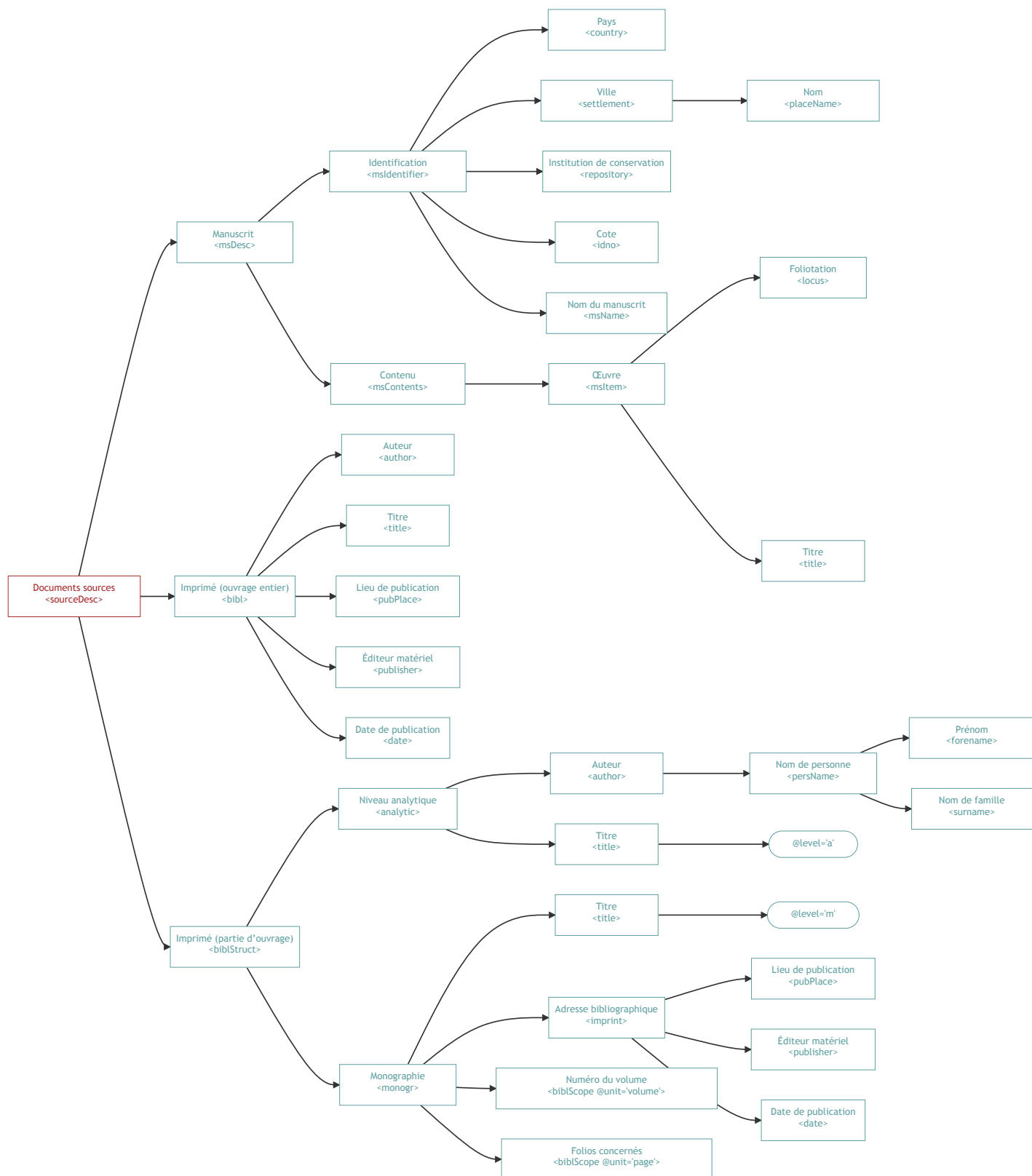
Dans le cas d'une réédition, il peut être utile de mentionner l'édition princeps. Cela peut être fait au moyen de l'élément `<bibl>`, dont l'attribut `@type` prend alors la valeur `edition1`. On peut formater les différents éléments enfants selon la granularité suivante :

- date de l'édition princeps : `<date>` ;
- format : `<extent>` ;
- éditeur matériel : `<publisher>` ;
- ISBN : `<idno>` ;
- Typologie d'édition : `<edition>` (exemple *Édition augmentée, Réédition cf. ONIX*).

```
<sourceDesc>
 <bibl type="edition1">
 <date></date>
 <extent rend="format"></extent>
 <publisher></publisher>
 <idno type="ISBN"></idno>
 <edition rend="type"></edition>
 </bibl>
</sourceDesc>
```

## D. Indiquer une édition de sources

Dans le cas d'une édition de sources, on utilise les éléments `<msDesc>` ou `<biblStruct>` pour décrire précisément la ou les sources utilisées.



## a. Source manuscrite

Lorsque la source est manuscrite, on utilise l'élément `<msDesc>`, qui contient lui-même les éléments `<msIdentifier>` et `<msContents>`.

- La balise `<msIdentifier>` contient les éléments permettant de localiser le manuscrit :

- pays : `<country>` ;
- ville : `<settlement>` ;
- institution de conservation : `<repository>` ;
- cote : `<idno>` ;
- nom du manuscrit `<msName>` .

```
<msDesc xml:id="Avranches_BM42">
 <msIdentifier>
 <country>
 <placeName>France</placeName>
 </country>
 <settlement>
 <placeName>Avranches</placeName>
 </settlement>
 <repository>Bibliothèque municipale, fonds anciens</repository>
 <idno>ms 42</idno>
 <msName>Missel à l'usage du Mont Saint-Michel</msName>
 </msIdentifier>
</msDesc>
```

- L'élément `<msContents>` permet de donner des informations sur une ou des parties du manuscrit étudié. On précise alors dans la balise `<msItem>` la foliation `<locus>` et le titre de la source `<title>` .

```
<msDesc xml:id="Avranches_BM42">
 <msIdentifier>
 <country>
 <placeName>France</placeName>
 </country>
 <settlement>
 <placeName>Avranches</placeName>
 </settlement>
 <repository>Bibliothèque municipale, fonds anciens</repository>
 <idno>ms 42</idno>
 <msName>Missel à l'usage du Mont Saint-Michel</msName>
 </msIdentifier>
 <msContents>
 <msItem>
 <locus>fol. 213 verso</locus>
 <title>Missa de sancto Michael</title>
 </msItem>
 </msContents>
</msDesc>
```

## b. Source imprimée

Lorsque la source est imprimée et qu'elle constitue un ouvrage entier, on peut se contenter d'utiliser l'élément `<bibl>` qui contient les éléments :

- auteur(s) : `<author>` ;
- titre : `<title>` ;
- lieu de publication : `<pubPlace>` ;
- éditeur : `<publisher>` ;
- date : `<date>` ;

```
<bibl xml:id="Epistola_Voetium">
 <author>Descartes René</author>
 <title>Epistola Renati Des Cartes ad celeberrimum virum D.
 Gisbertum Voetium</title>
 <pubPlace>Amsterdam</pubPlace>
 <publisher>L. Elzevirium</publisher>
 <date>1643</date>
</bibl>
```

Cependant, lorsque la source constitue une partie de l'ouvrage imprimé, on utilise l'élément `<biblStruct>`, accompagné d'un attribut `@xml:id` décrivant le type de source (article, chapitre...).

Cet élément `<biblStruct>` contient les éléments `<analytic>` et `<monogr>` :

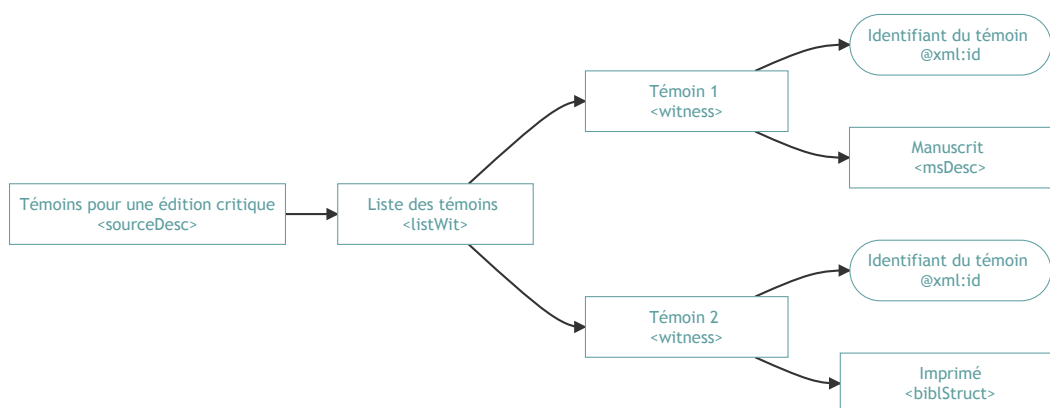
- `<analytic>` comprend les informations sur la source elle-même : titre de la source et auteur ;
- `<monogr>` comprend les éléments sur l'ouvrage dans son entier : titre de l'ouvrage, informations sur la publication, pagination.

## Exemple de l'édition du Pouillé du diocèse de Rouen, qui se trouve aux pp. 228-231 de l'ouvrage *Recueil des historiens des Gaules et de la France*

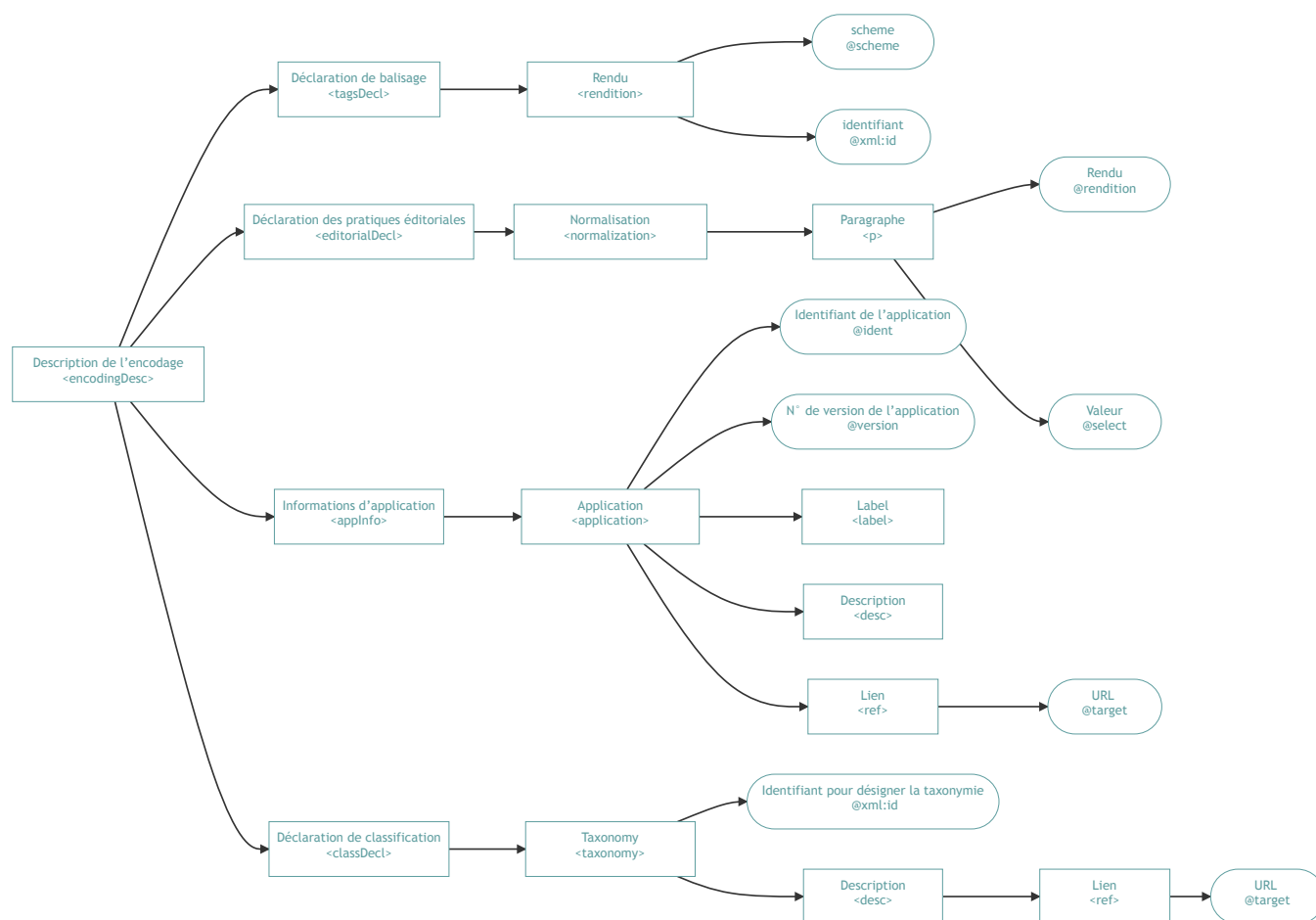
```
<biblStruct type="chapter" xml:id="RHGF_23">
 <analytic>
 <title level="a" type="main">Polyptychum Rotomagensis Diocesis</title>
 <author>
 <persName>
 <forename>Delisle</forename>
 <surname>Léopold</surname>
 </persName>
 </author>
 </analytic>
 <monogr>
 <title level="m">Recueil des historiens des Gaules et de la France</title>
 <imprint>
 <pubPlace>Paris</pubPlace>
 <publisher>H. Welter</publisher>
 <date>1894</date>
 </imprint>
 <biblScope unit="volume">23</biblScope>
 <biblScope unit="page">228-331</biblScope>
 </monogr>
</biblStruct>
```

## E. Cas particulier de l'édition critique, méthode de la segmentation parallèle

Dans le cas d'une édition critique appliquant la méthode de la segmentation parallèle ( `<app>` , `<lem>` , `<rdg>` ), on ajoute dans l'élément `<sourceDesc>` un élément `<listWit>` . À l'intérieur de ce dernier, afin de définir chaque témoin, on utilise l'élément `<witness>` avec un attribut `@xml:id` qui est la référence utilisée en valeur de l'attribut `@wit` dans le corps du texte.



# Description de l'encodage <encodingDesc>



## I. Déclaration de balisage <tagsDecl>

En usage dans Métopes (UE) pour documenter des informations de mise en forme récupérées depuis le fichier de traitement de texte (ex. cellules de tableau, types de numérotation des listes).

```
<encodingDesc>
 <tagsDecl>
 <rendition scheme="css" xml:id="Cell1.A1">
 border:0.5pt solid #000000;
 </rendition>
 </tagsDecl>
</encodingDesc>
```

## II. Déclaration des pratiques éditoriales `<editorialDecl>`

Comme le recommande la TEI, on utilise l'attribut `@method` avec la valeur `markup`, lorsque la normalisation a été décrite en employant un codage.

### Exemple du *Registre de la Société des Amis des Noirs (1788-1790)* ([SAN](#))

```
<editorialDecl>
 <normalization method="markup">
 <p>L'édition est de type diplomatique, effectuée dans un soucis constant de rester fidèle, autant que possible, au manuscrit original : les fautes d'orthographe, les ratures, les ajouts, les substitutions sont saisis tels quels avec ajout de la forme corrigée.</p>
 <p>Les majuscules, les accents et la ponctuation en général ont été rétablis selon les usages actuels.</p>
 <p>les abréviations font l'objet d'un développement, à l'exception des abréviations usuelles.</p>
 <p>Les mentions de chiffres ou nombres sont transcrites telles qu'elles apparaissent dans le manuscrit : en chiffres arabes, romains ou en toutes lettres.</p>
 </normalization>
</editorialDecl>
```

**COMMONS** : Exemple Métopes où cet élément sert à renseigner des consignes de mise en forme des notes, des citations ainsi que des recensions :

```
<editorialDecl>
 <ab>
 <date>[...]</date>
 </ab>
 <normalization rendition="notes">
 <p rendition="note" select="following"></p>
 <p rendition="table" select="following"></p>
 <p rendition="annexe" select="following"></p>
 <p rendition="floatingText" select="following"></p>
 </normalization>
 <normalization rendition="quotes">
 <p rendition="rom" select="romain"></p>
 <p rendition="ital" select="italique"></p>
 </normalization>
 <normalization rendition="recension">
 <p rendition="subtitle" select="false"></p>
 </normalization>
</editorialDecl>
```



### **Nota Bene**

En usage dans Métopes (UE, VOL) : cas particulier pour la gestion de la rétroconversion, l'utilisateur documente les paramètres utilisés pour la version imprimée de façon à ce qu'ils soient conservés pour les versions numériques.

## III. Informations d'application `<appInfo>`

Cet élément et sa descendance sont utilisés pour préciser les outils utilisés et leur version.

### **Nota Bene**

Aligner avec le serveur d'environnements et de modules d'extensions SEME du PDN.

```
<appInfo>
 <application ident="Diplomatique" version="1.0">
 <label>PDN-TEI Édition des sources diplomatiques</label>
 <desc>Environnement d'édition de sources diplomatiques</desc>
 <ref target="http://www.unicaen.fr/recherche/mrsh/document_numerique/
 seme/Diplomatique.zip"/>
 </application>
 <application ident="TEI_indexation" version="1.0">
 <label>PDN-TEI Indexation</label>
 <desc>Environnement XXE d'édition scientifique pour encoder des notices de
 thesauri : personnes, lieux, œuvres et poissons/créatures aquatiques.</desc>
 <ref target="http://www.unicaen.fr/recherche/mrsh/document_numerique/
 seme/TEI_indexation.zip"/>
 </application>
</appInfo>
```

**COMMONS** : Exemple Métopes :

```
<appInfo>
 <application ident="circe" version="1.0">
 <label>Circé</label>
 <desc>Document converti en TEI avec l'application Circé</desc>
 <ref target="#" />
 </application>
 <application ident="Word-tpl" version="1.0">
 <label>Word template</label>
 <desc>Version du modèle de stylage utilisé</desc>
 <ref target="#" />
 </application>
</appInfo>
```

## IV. Déclaration de classification `<classDecl>`

Cet élément sert à déclarer les référentiels externes utilisés pour documenter et indexer le document :

- Bibliographie(s) Zotero.
- Référentiels communautaires :
  - thesauri PDN-Biblissima personnes-lieux-oeuvres ;
  - pactols.
- Notices d'autorités :
  - IDREF ;
  - VIAF.
- Référentiels géographiques :
  - geonames.
- Code ISO langue (ISO 639).

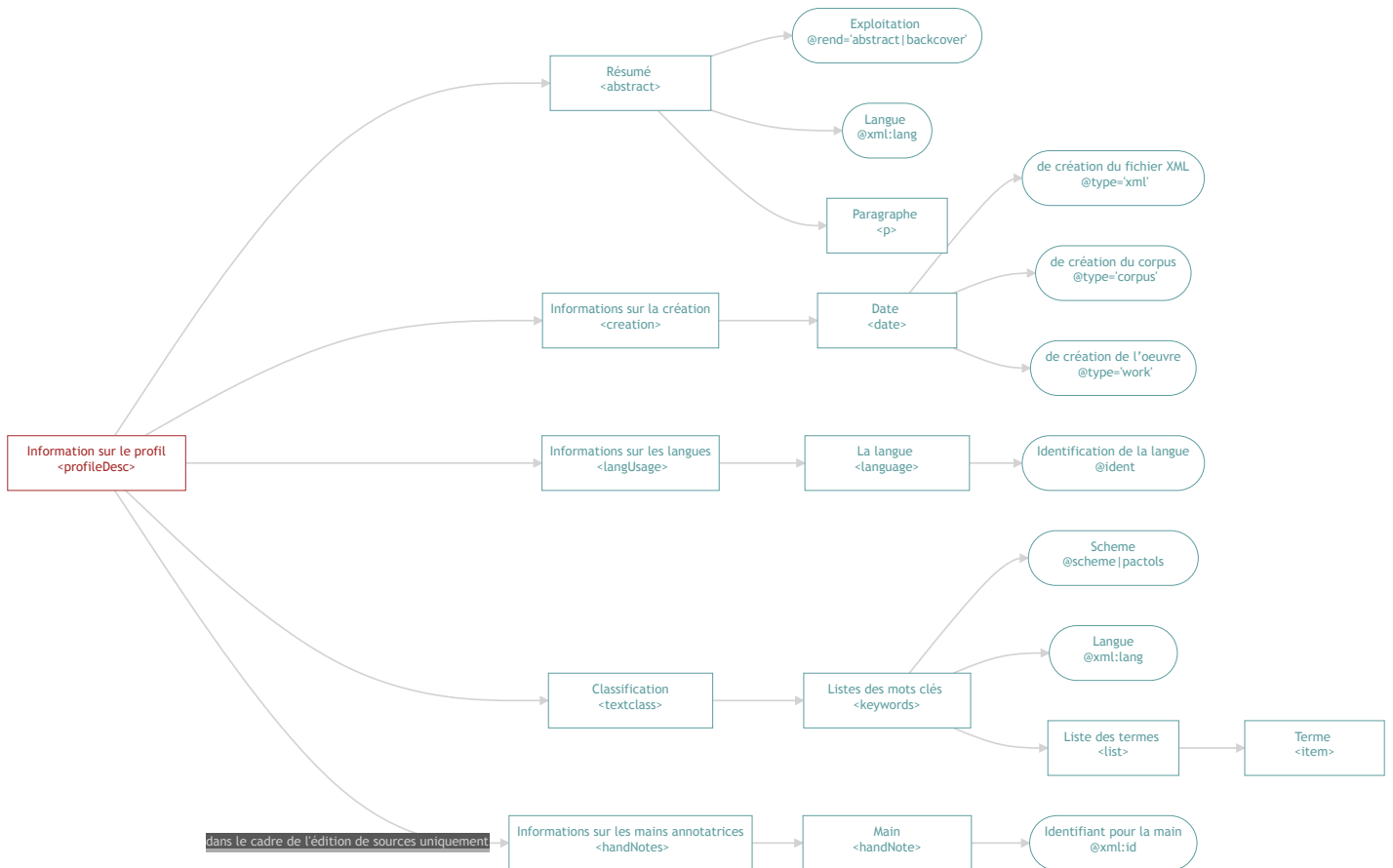
`<classDecl>` (déclaration de classification) : contient une ou plusieurs taxinomies définissant les codes de classification utilisés n'importe où dans le texte.

`<taxonomy>` (taxinomie) : définit une typologie, soit implicitement au moyen d'une référence bibliographique, soit explicitement au moyen d'une taxinomie structurée.

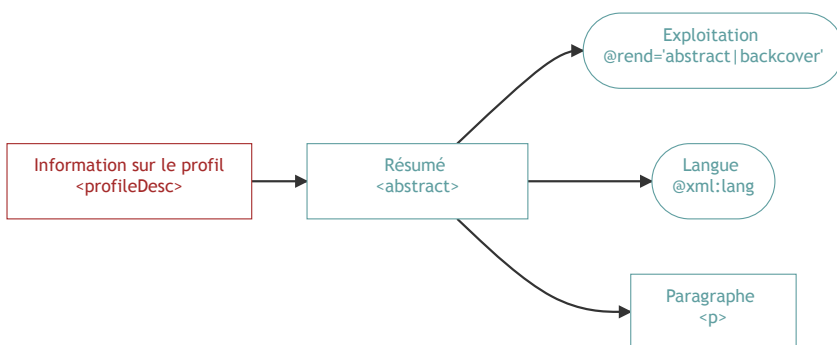
```
<classDecl>
 <taxonomy xml:id="personnes">
 <desc>Base d'autorités personnes (PDN – Biblissima)
 <ref target="https://thesauri.unicaen.fr/autorites/personnes"/>
 </desc>
 </taxonomy>
 <taxonomy xml:id="lieux">
 <desc>Base d'autorités lieux (PDN – Biblissima)
 <ref target="https://thesauri.unicaen.fr/autorites/lieux"/>
 </desc>
 </taxonomy>
 <taxonomy xml:id="zotero">
 <desc>Bibliographie Zotero ; collection
 <ref target="https://www.zotero.org/groups/2272072/e-cartae">
 e-Cartae
 </ref>
 </desc>
 </taxonomy>
 <taxonomy xml:id="geonames">
 <desc>.....<ref/></desc>
 </taxonomy>
</classDecl>
```

**COMMONS** : cet élément n'est pas présent à l'échelle de l'unité éditoriale. Les entrées Pactols sont renseignée dans `<textClass>` présent dans le `<profileDesc>`.

## Description du profil `<profileDesc>`



### I. Résumé



L'élément `<abstract>` appartient au schéma commun même si pour les projets de recherche, c'est plutôt le premier paragraphe du texte qui sera récupéré. Le résumé se place au sein d'un élément `<p>`.

Le résumé est recommandé par le consortium CAHIER pour entrer dans Nakala qui précise : « À défaut d'un résumé fourni par l'équipe, Nakala extrait les premières lignes du texte du document. »

L'usage Métopes est le suivant :

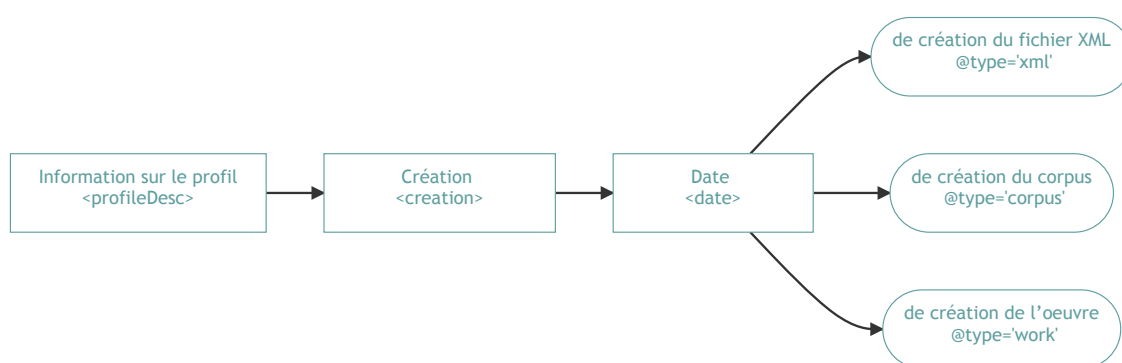
- [UE] : pas de résumé dans les métadonnées mais dans le `<front>` ;
- [VOL] : `<abstract>` `@rend` dont les valeurs sont :
  - `backcover` ;
  - `abstract` .

**Recommandations TEI** : `<abstract>` *This element is intended only for cases where no abstract is available in the original source. Any abstract already present in the source document should be encoded as a [div](#) within the [front](#), as it should for a born-digital document.*

```
<abstract rend="abstract|backcover" xml:lang="ISO-639">
 <p>Résumé rédigé ou à défaut copie du premier paragraphe de texte.</p>
</abstract>
```

**COMMONS** : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

## II. Création



L'élément `<creation>` est utilisé par Métopes pour préciser la date de création du fichier XML ( `@type="xml"` ) lors de sa conversion (XSL) de fichier stylé en fichier XML-TEI.

Il est utilisé pour les projets de recherche ( `@type="corpus"` ) pour préciser une fourchette de dates de constitution d'un corpus.

CAHIER donne la définition suivante :

C'est ici qu'on inscrit la date de création du document ou de l'œuvre source : date de la première représentation, date de la première édition ou de l'édition choisie pour l'encodage, date de l'envoi de la lettre, etc. L'usage des attributs `@notBefore` / `@notAfter` est autorisé. Les dates incomplètes, type "mois-année" ou "année" sont autorisées.

La date de création préconisée par le consortium CAHIER est indiquée au moyen de l'élément `<date>` et de l'attribut `@type="work"`.

```
<creation>
 <date type="xml">
 Date de la création du fichier XML (date d'utilisation de la
 XSL dans le contexte de Métopes, date de création d'un fichier dans le cadre
 de Pluco, etc.).
 </date>
 <date type="corpus">
 Fourchette de dates de création d'un corpus ; usage pour
 la recherche.
 </date>
 <date type="work">
 Date (ou fourchette de dates) pour dater la création d'une œuvre.
 Élément préconisé par Cahier, mais qui ne correspond pas à des usages du PDN.
 </date>
</creation>
```

**COMMONS** : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

### III. Langues



Les éléments `<langUsage>`, `<language>` enchassés permettent de déclarer les langues du document (Code ISO 639).

On distingue plusieurs usages pour :

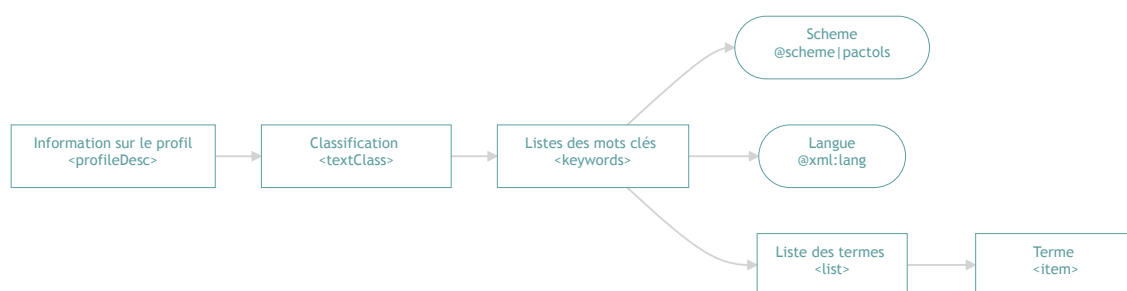
- documenter les langues ;
- déclarer les langues pour centraliser les codes et permettre de pointer vers les identifiants ensuite ;
- configurer les recherches sur une langue.

### **Nota Bene**

**Commons Publishing** : choix de conserver l'encodage de la langue principale du document sur 4 caractères ( `language/@ident` ).

```
<langUsage>
 <language ident="la">Latin</language>
 <language ident="fro">Ancien français</language>
</langUsage>
```

## IV. Classification du document



Il s'agit ici principalement de trouver une place pour les mots-clés.

- Métopes :
  - permet de faire plusieurs listes de mots-clés, triées par langue. Attribut `@xml:lang` sur l'élément `<keywords>` et `@scheme="keywords"` ;
  - possibilité de faire de l'indexation Pactols dans une liste de mots-clés : `textClass/keywords[scheme="pactols"]/list/item` .
- CAHIER : « Les mots clés sont laissés au libre choix des équipes. » Les listes complètement ouvertes sont généralement problématiques car elles sont inutilisables. Si on souhaite typer la source qu'on encode, on peut se référer au thesaurus *Typologie textuelle* produit par le GT Cahier du même nom [4](#).

```

<textClass>
 <keywords scheme="pactols"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:ns="http://www.tei-c.org/ns/1.0">
 <list>
 <item>
 <index indexName="Index">
 <term type="orig">Nécropole</term>
 <index indexName="pactols:Sujets" n="1" rendition="oe"
 source="26678/pcrtNb90Egda4H"
 xml:base="https://ark.frantiq.fr/ark:/">
 <term xml:lang="en">death</term>
 <term xml:lang="nl">dood</term>
 <term xml:lang="fr">mort</term>
 <term xml:lang="it">morte</term>
 <term xml:lang="es">muerte</term>
 <term xml:lang="de">Tod</term>
 <term xml:lang="ar">موت</term>
 <index n="2" rendition="oe" source="26678/crtMK52omryzY">
 <term xml:lang="fr">espace funéraire</term>
 <index n="3" rendition="oe"
 source="26678/pcrtsIm3RuNMGu">
 <term xml:lang="fr">nécropole</term>
 <term xml:lang="it">necropoli</term>
 <term xml:lang="en">necropolis</term>
 <term xml:lang="es">necrópolis</term>
 <term xml:lang="nl">necropool</term>
 <term xml:lang="de">Nekropole</term>
 <term xml:lang="ar">مدينة الموتى</term>
 </index>
 </index>
 </index>
 </item>
 <item>
 [...]
 </item>
 <item>
 [...]
 </item>
 </list>
 </keywords>
 </textClass>

```

## V. Mains annotatrices



Cet élément est assez spécifique. Il ne concerne que les projets de transcription/édition critique de textes, qui étudient plusieurs témoins. Les environnements « Apparat » et « Inventaires anciens » du PDN proposent cet encodage. Mais cela ne concerne ni Métopes, ni probablement CAHIER qui ne le mentionne pas.

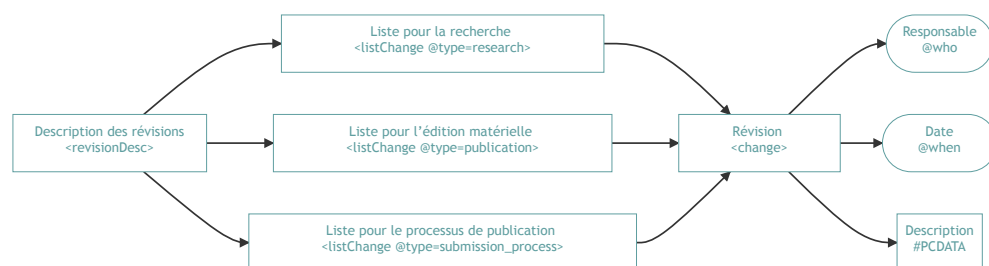
```

<handNotes>
 <handNote xml:id="main1">
 Identification d'une main annotatrice dans un document source dans le
 cas de l'édition scientifique de source.
 </handNote>
</handNotes>

```

**COMMONS** : cet élément n'est pas présent à l'échelle de l'unité éditoriale.

## Description des révisions <revisionDesc>



La description des révisions permet de répertorier l'ensemble des interventions effectuées pour la constitution du document XML. C'est un élément important pour le PDN avec la dernière version de Pluco V2 (plugin de travail collaboratif du PDN) mais il est aussi nécessaire pour le processus éditorial.

Une proposition est de séparer les typologies de révisions dans 3 éléments <listChange> typés :

- encodage recherche : @type='research' ;
- historique éditorial : @type='publication' ;
- processus de publication :  
@type='publication\_history' or 'editorial\_process' or 'submission\_process' .



Exemple : date de soumission du manuscrit par l’auteur, date d’acceptation du manuscrit, etc. Point de vigilance : définir ce qui relève de l’historique du fichier, de ces révisions et des étapes de publication.

```
<revisionDesc>
 <listChange type="research">[...]</listChange>
 <listChange type="publication">[...]</listChange>
 <listChange type="submission_process">[...]</listChange>
</revisionDesc>
```

Ensuite, on peut énumérer autant d’éléments `<change>` que nécessaire.

```
<revisionDesc>
 <listChange>
 <change when="2021-08-18T00:19:42.074+02:00" who="GC">Révision</change>
 </listChange>
</revisionDesc>
```

- `@who` pointe vers l’identifiant déclaré d’une autorité définie dans le `<teiHeader>` ; attribut renseigné automatiquement par Pluco ;
- `@when` : attribut renseigné automatiquement par Pluco ;
- description en plein texte : liste à sélectionner avec le type d’intervention pratiquée, et maintenir la description en plein texte possible.

- 
1. [https://www.unicaen.fr/recherche/mrsh/document\\_numerique](https://www.unicaen.fr/recherche/mrsh/document_numerique) et <http://www.metopes.fr/index.html>. ↩
  2. Idmhand Fatiha, Galleron Ioana. *Version 2 Guide pour la FAIRisation des données des corpus d’auteurs préparé par Fatiha Idmhand et Ioana Galleron [Groupe de travail DataCahier]*. [Rapport de recherche] Huma-Num. 2020. ⟨halshs-03037748⟩. ↩
  3. <https://nakala.fr>. ↩
  4. [https://groupes.renater.fr/wiki/cahier/typologie\\_textuelle](https://groupes.renater.fr/wiki/cahier/typologie_textuelle). ↩