Université de Neuchâtel Master en littérature

# TG: Édition de texte (numérique)

### Cours 7: métadonnées

Élodie Paupe chaire de philologie classique et d'histoire ancienne

9 novembre 2020

### Pour commencer

**Métadonnées:** "ensembles de données structurées pour décrire, expliquer, localiser les ressource et en faciliter la recherche, l'usage et la gestion" (Morel-Pair 2007 citant NISO 2014)

• des données exploitables numériquement pour décrire d'autres données

## Six "groupes de fonctions" (Morel-Pair 2007, 3-4)

- 1. améliorer la recherche d'information et la « découverte » des ressources avec des métadonnées descriptives du contenu
- 2. gérer les ressources:
  - i. avec des métadonnées administratives
  - ii. avec des métadonnées instancielles, ou techniques et de structure
- 3. gérer les « archives »
- 4. faciliter le partage de données et leur réutilisation
- 5. participer à la pérennité des ressources numériques avec des *métadonnées de préservation*
- 6. décrire les utilisateurs pour gérer les accès, leur permettre des personnalisations de consultation, analyser les comportements d'usage

## Types de métadonnées et usages (Riley 2017, 7)

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author	Discovery
	Subject	Display Interoperability
	Genre	Interoperating
	Publication date	
Technical metadata	File type	Interoperability
	File size	Digital object management
	Creation date/time	Preservation
	Compression scheme	
Preservation metadata	Checksum	Interoperability
	Preservation event	Digital object management
		Preservation
Rights metadata	Copyright status	Interoperability
	License terms	Digital object management
	Rights holder	
Structural metadata	Sequence	Navigation
	Place in hierarchy	
Markup languages	Paragraph	Navigation
	Heading	Interoperability
	List	
	Name	
	Date	

Les métadonnées sont exploitées depuis longtemps par certains champs disciplinaires ou certaines institutions (bibliothèques, etc.).

Avec l'apparition du web et la diffusion de données numérique, l'établissement des métadonnées est devenu le souci de tous les acteurs de "l'édition numérique".

"ensembles de données **structurées** pour décrire, expliquer, localiser les ressource et en faciliter la recherche, l'usage et la gestion"

- standardisation
- format
- interopérabilité

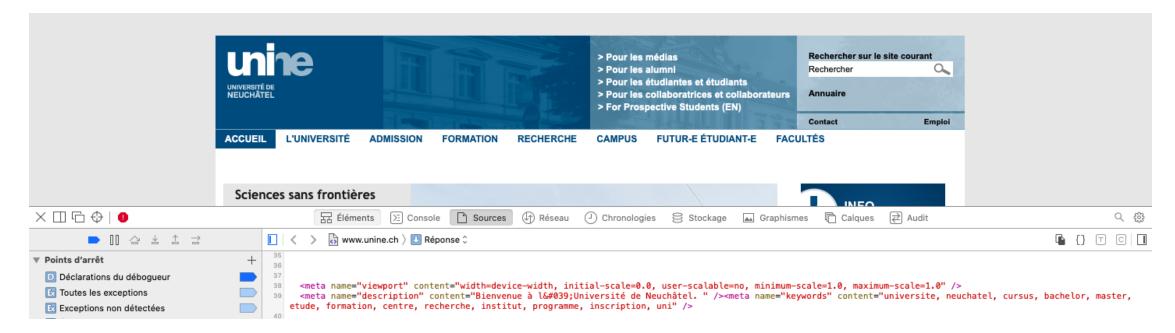
**interopérabilité**: "capacité d'échanger des données entre systèmes multiples disposant de différentes caractéristiques en terme de matériels, logiciels, structures de données et interfaces, et avec le minimum de perte d'information et de fonctionnalités" (NISO 2004)

- → le World Wide Web Consortium travaille en permanence à un web interopérable.
- → la Dublin Core Metadata Initiative travaille à la description généraliste de métadonnées.
- $\rightarrow$  etc.

## Exemples de métadonnées



Format MARC



Format HTML

#### Porrentruy, Bibliothèque cantonale jurassienne, Ms. 34

© Gamper Rudolf / Jurot Romain, Catalogue des manuscrits médiévaux conservés à Porrentruy et dans le canton du Jura, Dietikon-Zürich 1999, p. 115-117.

By courtesy of Urs Graf Verlag, Dietikon. The manuscript description is copyrighted by the publisher.

**Show PDF** 

**Print description** 

Manuscript title: Liber Evangeliorum et Capitulare

Place of origin: Saint-Bertin

Date of origin: 2e tiers du IXe siècle

Alternative name: Evangéliaire de Saint-Ursanne

Support: Parchemin

**Extent:** 414 p.

Format: 270-275 x 170 mm

Foliation: Pagination unilatérale du XIX/XX e s. complétée pour les besoins de la description 1-414.

Collation: Cahiers: IV 16 + (IV-1) 30 + 24 IV 414. Un f. manque entre les p. 28/29 et le tiers inférieur des pages 11/12 a été découpé, sans lacunes textuelle S. P. 113/123 non solidaires à leur base.

Page layout: Réglure à la pointe sèche. Justification: 185 x 100 mm, 26 longues l., 32 dès la p. 399.

Writing and hands: Minuscule caroline pour le texte, capitale carrée pour les incipit des évangiles (p. 30, 137, 210, 316), capitale rustique pour les formules de salutation à la fin des prologues (p. 3 et 7) et onciale pour les titres, sous-titres et les premières l. du texte des évangiles, plusieurs mains.

Decoration: Rubr. Initiales et canons eusébiens peints dans le style iro-franc, de couleur (jaune, vert et rouge), ornés de têtes de serpents et de dragons: p. 31 (Mt), 132 (prol. Mc), 138 (Mc), 199 (prol. Lc), 211 (Lc), 313 (prol. lo), 317 (lo). Quelques esquisses à la plume de figures zoomorphiques dans les marges (voir par exemple, p. 177, 364, 381): homme portant un récipient (p. 184), volutes (p. 365), etc. P. 31, caricature à la pointe sèche du Crucifié, nu avec des pattes d'animal; ce dessin est antérieur à l'écriture, selon l'indice fourni par une perturbation du tracé d'un e à la l. 23 provoquée par le sillon du dessin.

Additions: Dans la marge supérieure de la p. 34, nom inscrit à la pointe sèche par une main insulaire contemporaine de la confection du ms. Eduuardo. Corrections sur grattages.

P. 360 avec une note tironienne signifiant lesus (ihc) au-dessus de venit Bethaniam (lo 12,1) (aimablement déchiffrée par P. Ladner, Fribourg). Inflexions de cantillation, voir Contenu, p. 116-130 et 186-196.

#### Binding:

• de type carolingien, sur ais de bois de 15 mm d'épaisseur, dos long recouvert d'une pièce de peau avec des coiffes débordantes cousues de tissu jaune,

```
▼<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xi="http://www.w3.org/2001/XInclude" xsi:schemaLocation="http://www.tei-
 c.org/ns/1.0 ../xsd/TEI-P5/1.7/tei-p5-e-codices 1.7.xsd" xml:lang="deu" xml:base="http://www.e-codices.unifr.ch/en/description/bcj/0034" version="5.09" xml:id="eCod bcj-0034">
 ▼<teiHeader>
   ▼<fileDesc>
    ▼<titleStmt>
       <title>Liber Evangeliorum et Capitulare</title>
      </titleStmt>
    ▼<editionStmt>
       <edition>Elektronische Version nach TEI P5.031</edition>
      ▼<respStmt>
        ▼<resp>
           Konvertierung nach TEI:
           <persName>Torsten Schaßan</persName>
           <date when="2006-02-17">17.01.2006</date>
         </resp>
        ▼<name>
           e-codices - Virtual Manuscript Library of Switzerland
         </name>
       </respStmt>
      </editionStmt>
    ▼<publicationStmt>
      ▼<publisher>
         e-codices - Virtual Manuscript Library of Switzerland
       </publisher>
      ▼<availability status="restricted" n="copyright">
        ▼<licence>
         ▼
            Mit freundlicher Genehmigung des
            <orgName xml:base="http://www.urs-graf-verlag.com">Urs Graf Verlag, Dietikon</orgName>
             . Das Copyright an der Handschriftenbeschreibung liegt beim Verlag.
           </licence>
       </availability>
      </publicationStmt>
    ▼<sourceDesc>
       <br/>
<br/>
bibl>PDF vorhanden</bibl>
      ▼<msDesc xml:lang="fra" xml:id="bcj-0034">
        ▼<msIdentifier>
           <settlement>Porrentruy</settlement>
           <repository>Bibliothèque cantonale jurassienne/repository>
           <idno>Ms. 34</idno>
           <msName>Evangéliaire de Saint-Ursanne</msName>
         </msIdentifier>
        ▼<head>
           <title>Liber Evangeliorum et Capitulare</title>
           <origPlace>Saint-Bertin</origPlace>
           <origDate notBefore="0833" notAfter="0865">2e tiers du IXe siècle/origDate>
         </head>
        ▼<msContents>
         ▼<msItem>
             <locus from="1" to="7">P. 1-7</locus>
```

# Métadonnées et XML-TEI

2. TEI header

## Structuration générale d'un document TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
     <fileDesc>
        <titleStmt>
           <title>Exercice d'encodage</title>
        </titleStmt>
        <publicationStmt>
           Élodie Paupe, Université de Neuchâtel. Cours d'édition, semestre d'automne 2020
        </publicationStmt>
        <sourceDesc>
           Un poème de Musset.
        </sourceDesc>
     </fileDesc>
  </teiHeader>
 <text>
   <front>
       Introduction, etc.
   </front>
   <body>
       Texte
   </body>
   <back>
       Index, commentaire, apparat, etc.
   </back>
 </text>
</TEI>
```

Au minimum, le <teiHeader> doit contenir une balise <fileDesc> qui contient les métadonnées du fichier informatique lui-même.

#### <fileDesc>

Le <fileDesc> contient au minimum trois balises:

- <titleStmt> (title statement): informations sur le titre du document et la/les personne/s chargée/s l'édition.
- <publicationStmt> (publication statement): informations relatives à la publication et à la diffusion du document
- <sourceDesc> (source description): description de la/des source/s dont dérive la présente édition (par exemple une description bibliographique dans le cas d'une édition numérique d'unt exte préexistant, la liste des manuscrits dans le cas d'une édition critique, dans le cas d'un document original, on le précise également)

#### <titleStmt> VS <sourceDesc>

- Dans le cas de l'édition numérique d'une œuvre existante, on pourrait avoir l'impression que le <titleStmt> et le <sourceDesc> contiennent la même information, mais ce ne'st pas le cas:
  - <titleStmt> est le titre du document numérique, pas celui de la source.
  - <sourceDesc> contient la description de la source qui est éditée, les informations qui s'y trouvent de décrivent pas l'édition numérique, mais son contenu.

#### Métadonnées structurées vs paragraphe

Le <titleStmt> n'admet pas la balise et nous contraint à structurer les informations. Au minimum, il doit contenir un titre:

```
<titleStmt>
    <title></title>
</titleStmt>
```

## ComposantIdu <fileDesc> : <titleStmt>

Mais on peut augmenter la granularité de l'information un jeu de balises:

- <author>
- <editor>
- <respStmt> (statement of responsibility): indication sur la responsabilité intellectuelle
  lorsque le statut d'auteur ou d'éditeur n'est pas suffisamment précis
  - <resp> indique la nature de la responsabilité
  - < <name>
- <principal> (principal researcher): informations sur le chercheur principal
- <funder> (funding body): informations sur l'institution qui finance un projet
- <sponsor> : informations sur une institution qui parraine un projet

#### Préciser un <author>, <editor> OU <name>

On peut encoder le nom, le prénom, le nom d'auteur, etc.

#### cf. 13.2 Names

- <perName> (personal name): un nom qui peut être décomposé en prénom, nom, etc.
- <surname> : nom de famille
- <forename> : prénom
- <rolename> : titre, position hiérarchique, etc.
- <addname> (additional name): surnom, alias, etc.
- etc.

## Balises spécifiques et balises "généralistes"

Les balises qui viennent d'être vues ne sont pas spécifiques au <teiHeader> et pourraient se retrouver ailleurs dans le document.

Ainsi, la balise <title> peut identifier un titre dans un paragraphe ordinaire, mais pas la balise <author> qui est réservée pour les usages bibliographiques (références bibliographiques, de manuscrits, d'édition).

```
 Dans l'exercice, vous allez encoder un poème de <name>Musset</name>
tiré de ses <bibl><title>Poésies complètes</title></bibl> publiées par <rs>Gallimard</rs>.
```

Exercice 1: ici

## Composant II du <fileDesc>: <publicationStmt>

Les informations relatives à la publication du document informatique peuvent être présentées dans un paragraphe ( ).

Cette solution descriptive n'est toutefois pas interrogeable par un ordinateur. Il est plus intéressant de structurer l'information.

#### cf. <publicationStmt>

- <address>
- <date>
- <pubPlace>
- <publisher>
- <authority> (release authority): donne le nom de la personne ou de l'organisation qui met à disposition l'édition (≠ l'éditeur ou le distributeur)
- <availability> : informations sur la disponibilité d'un texte, son utilisation, le droit d'auteur, la licence de partage, etc. (@status propose trois options, mais la balise
   permet de définir plus précisément les critères ou les licences)
- etc.

```
<availability status="restricted">
  Available for academic research purposes only.
</availability>
```

```
<availability status="free">
  In the public domain
  </availability>
```

```
<availability>
  Copyright 1989, Oxford University Press
</availability>
```

## Composant III du <fileDesc> : <sourceDesc>

Dernier composant obligatoire du <fileDesc> , le <sourceDesc> peut

• s'exprimer dans un plus ou moins structuré:

```
<sourceDesc>
  Aucune source: ce document est né numérique
</sourceDesc>
```

• se limiter à une référence bibliographique peut structurée:

```
<sourceDesc>
  <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton Facsimile, 1968)</bibl>
</sourceDesc>
```

• se présenter sous la forme d'une référence bibliographique fortement structurée

```
<sourceDesc>
    <biblStruct xml:id="Nussbaumer1965" type="livre">
        <monogr>
            <author>P. Arnold Nussbaumer</author>
            <title>Die Jura-Heiligen German und Randoald</title>
            <imprint>
                <pubPlace>Laufon</pubPlace>
                <publisher>Verlag der "Nordschweiz"</publisher>
                <date>1965</date>
            </imprint>
        </monogr>
        <note>Nussbaumer traduit le Codex Sangallensis 551.
    </biblStruct>
</sourceDesc>
```

```
<biblStruct xml:id="Bessire1954" type="livre">
    <analytic>
        <author>Pierre-Olivier Bessire</author>
        <title>Actes de saint Germian, abbé du monastère de Moutier-Grandval,
        son martyre et celui de saint Randoald, par le moine Bobolène</title>
    </analytic>
    <monogr>
        <imprint>
            <date>1954</date>
            <biblScope>67-74</biblScope>
        </imprint>
    </monogr>
    <series>
        <title>Société jurassienne d'émulation</title>
    </series>
    <note>Bessire traduit le texte édité par Trouillat,
    autrement dit le manuscrit <ref source="#P1">P1</ref>.</note>
</biblStruct>
```

Dans le <sourceDesc>, on fait donc figurer les documents dont dérive l'édition numérique.

Dans les cas qui nous intéressent, il peut s'agir

- de livres encodés avec <bibl> , <biblFull> ou <biblStruct> (cf. les exemples précédents)
   ou
- de manuscrits encodés avec <msDesc>

Description d'un manuscrit dans une version minale...

#### ... dans une version étendue...

```
<msDesc>
    <msIdentifier>
        <country>Suisse</country>
        <region>Jura</region>
        <settlement>Porrentruy</settlement>
        <repository>Archives de l'ancien Évêché de Bâle</repository>
        <idno>A 55/12-1.1</idno>
        <msName>Vita Sancti Germani/msName>
    </msIdentifier>
   <msContents>
        <msItem>
            <author>Prêtre Bolène
            <title>Vita Sancti Germani</title>
        </msItem>
    </msContents>
   <physDesc>
        <objectDesc>
            <ab>Livret de papier de 16 folios. </ab>
            <ab>Titre, prologue et notes marginales à l'encre rouge. </ab>
        </objectDesc>
   </physDesc>
   <history>
        <ab>Le manuscrit est une copie du XVIIe siècle<origDate notBefore="1600" notAfter="1700"/>
        d'après le répertoire des AAEB et Krusch.</ab>
    </history>
</msDesc>
```

Et si le contenu du manuscrit est plus important?

Exemple sur e-codices avec l'Évangéliaire de Saint-Ursanne, Porrentruy, Bibliothèque cantonale jurassienne, Ms. 34.

https://www.e-codices.unifr.ch/en/description/bcj/0034/

Exercice 2: ici

### <fileDesc> étendu

Il peut également par exemple contenir les balises suivantes:

```
<teiHeader>
    <fileDesc>
        <titleStmt>...</titleStmt>
        <editionStmt>...</editionStmt>
          <publicationStmt>...</publicationStmt>
          <noteStmt>...</noteStmt>
          <sourceDesc>...</sourceDesc>
          </fileDesc>
          </teiHeader>
```

• <editionStmt> permet de donner des informations descriptives sur l'édition (editio princeps, révision, etc.). Son contenu peut être plus ou moins détaillé.

Il permet également de nommer les différents collaborateurs avec leurs attributions spécifiques:

```
<editionStmt>
    <edition>Troisième édition</edition>
    <respStmt>
        <resp>réalisée par</resp>
        <name ref="#EP">Élodie Paupe</name>
        </respStmt>
        </editionStmt>
```

```
<editionStmt>
    <edition>Première version électronique TEI</edition>
    <respStmt>
        <resp>
            Équipe de traduction pour l'année <date when="2020-02">2020</date>
            pour le Groupe de Théâtre Antique composée de
        </resp>
        <name xml:id="CA" type="participante">Camille Aeby</name>
        <name xml:id="NA" type="participante">Noa Aeby</name>
        <name xml:id="MC" type="participant">Maxime Cario</name>
        <name xml:id="MD" type="participant">Marc Duhram</name>
        <name xml:id="PJ" type="participante">Pauline Jacsont</name>
        <name xml:id="SM" type="participante">Séverine Moy</name>
        <name xml:id="IM" type="participante">Ilma Muminovic</name>
        <name xml:id="JRRS" type="participant">Joram Rafael Ribeiro da Silva</name>
        <name xml:id="RR" type="participante">Rania Richard</name>
        <name xml:id="PPS" type="participant">Paul-Philippe Schwab</name>
    </respStmt>
    <respStmt>
        <resp>
            Direction de la traduction et de la publication.
        </resp>
        <name xml:id="EP" type="éditeur">Élodie Paupe</name>
    </respStmt>
</editionStmt>
```

- <publicationStmt> permet de définir qui publie et avec quels droits
- <noteStmt> permet d'introduire des notes complémentaires

## Au-delà du <fileDesc>

```
<teiHeader>
  <fileDesc>...</fileDesc>
  <encodingDesc>...</encodingDesc>
  <profileDesc>...</profileDesc>
  <revisionDesc>...</revisionDesc>
  </teiHeader>
```

#### <encodingDesc>

L' <encodingDesc> permet d'expliquer la relation qui lie la source avec l'édition numérique. On y décrit notamment les pratiques de transcription et d'édition.

Nous l'avons déjà rencontré lorsque nous avons parlé de normalisation de texte et d'apparat critique. Une simple balise peut faire l'affaire.

```
<encodingDesc>
   <editorialDecl>
       <normalization>
           Les graphies majuscules et abréviations de la source ne sont pas conservées.
       </normalization>
       <correction>
           Les corrections sont encodées.
       </correction>
       <hyphenation>
           La césure des mots en fin de ligne ne respecte pas l'original.
       </hyphenation>
       <punctuation>
           La ponctuation n'est pas conservée
       </punctuation>
       <quotation>
           Les citations marquées dans l'original sont encodées.
           Les éventuelles autres marques d'intertextualité du texte sont encodées.
       </guotation>
       <segmentation>
           La segmentation éditoriale de la source est conservée.
               </segmentation>
   </editorialDecl>
   <variantEncoding location="internal" method="parallel-segmentation"/>
</encodingDesc>
```

#### ofileDesc>

Le <profileDesc> (text-profile description) décrit les aspects non-bibliographiques du texte (mots-clés, résumé, langue, liste de personnes mentionnées, processus de création de la source, etc.)

```
fileDesc>
  <creation>
    <date type="type" when="YYYY-MM-DD"/>
 </creation>
  <textClass>
    <keywords scheme="référentiel">
     st>
        <item>Un mot-clé</item>
        <item>Un autre mot-clé</item>
     </list>
    </keywords>
    <languageUsage>
      <language ident="code_ISO">Langue</language>
    </languageUsage>
  </textClass>
</profileDesc>
```

#### <revisionDesc>

Le <revisionDesc> contient les différentes étapes de révision du document numérique.

```
<revisionDesc>
        <change type="creation" resp="#EP">
           <date when="2020-02-20"/>Création du fichier.</change>
        <change type="update" resp="#EP">
           <date when="2020-03-17"/>Mise à jour du texte
           latin.</change>
        <change type="update" resp="#PJ">
           <date when="2020-03-16"/>Apparat critique.</change>
        <change type="update" resp="#RR">
           <date when="2020-04-02"/>Mise à jour du texte latin</change>
        <change type="update" resp="#PJ">
           <date when="2020-04-11"/>Corrections du texte latin.</change>
        <change type="update" resp="#EP">
           <date when="2020-04-15"/>Correction du texte latin, modification de l'apparat critique, ajout de la traduction.
        <change type="update" resp="#PJ">
           <date when="2020-04-22"/>Correction de l'apparat critique.</change>
        <change type="update" resp="#EP">
           <date when="2020-05-22"/>Correction du texte latin, modification de l'apparat critique, ajout de la traduction.
        <change type="update" resp="#EP">
           <date when="2020-05-25"/>Correction du texte latin, modification de l'apparat critique, ajout de la traduction.
        <change type="update" resp="#PJ">
           <date when="2020-06-01"/>Encodage intertextualité.</change>
        <change type="update" resp="#EP">
           <date when="2020-06-07"/>Correction de l'encodage de l'intertextualité, du texte latin, de l'apparat critique, ajout de la traduction.
        <change type="update" resp="#EP">
            <date when="2020-06-30"/>Finalisation de l'édition de base avec numérotation des paragraphes./change>
     </revisionDesc>
```

Exercice 3: ici

Exercice 4: ici

# Bibliographie

Burnard, Lou, *Qu'est-ce que la Text Encoding Initiative?* [en ligne], Marseille: OpenEdition Press, 2015. Disponible sur Internet : http://books.openedition.org/oep/1298. DOI : https://doi.org/10.4000/books.oep.1237.

Mahé, Annaïg, *Métadonnées: mutations et perspectives*, Séminaire Inria IST 2008, https://bbf.enssib.fr/consulter/bbf-2009-01-0094-008.

Morel-Pair, Catherine, "Métadonnées et XML", *Revue des Sciences et Technologies de l'Information - Série ISI: Ingénierie des Systèmes d'Information*, Lavoisier, 2007, 12 (2 Spécial), p.9-39. DOI: 10.3166/isi.12.2.9-39 . hal-01355788

Riley, Jenn, *Understanding Metadata. What is metadata, and what is it for?*, National Information Standards Organisation (NISO), , 2017,

http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.