

Text editing in the digital age

Digital editing and XML TEI

Ariane Pinche ¹², Matthias Gille Levenson ³²

¹CNRS, ²CIHAM, ³ÉNS Lyon

9 avril 2024

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

Scholarly Editing in the Digital Age : Structuring and Annotation of Textual Data in XML-TEI

Digital editing allows for recording numerous pieces of information about the text which will be :

- Preserved
- Parsable
- Queryable
- Reusable

The editorial object can be multiplied :

- Multiple visualizations
- Corpora with linguistic annotations
- Databases of text variations

What is a Digital Edition ?

There are different levels of digital editions :

- Provision of a structured text
- Provision of a structured and enriched text
- Scientific and/or critical editions with multiple layers of information

These are the most common types of texts :

- Project Perseus, example : **Amphitryon de Plaute**
- Labex OBVIL, example : **Mercure Galant**

Their purpose is to make the text accessible and comprehensible (addition of footnotes, indexes, glossaries) to the reader, but also to provide information on how the text works (its transmission, variants, links to sources).

- Documentary or palaeographic editions, for example : **le Didascalicon d'Hugues de Saint-Victor**
- Genetic editions, for example : **Madame Bovary : the history of the text through its drafts**,
- Critical editions, example : *Li Seint Confessor* by Wauchier de Denain

And more...

Today, XML TEI is no longer only used for text editing purposes, but also as a format for ensuring the longevity of textual information. The founders of TEI are considering the development of a minimal and generic XML TEI encoding scheme ("comity") to facilitate corpus analysis and enable what is known as "distant reading."

Lou BURNARD, Christof SCHÖCH et Carolin ODEBRECHT. « In search of comity : TEI for distant reading ». fr. In : *Journal of the Text Encoding Initiative Issue 14* (mars 2021). Number : Issue 14 Publisher : Text Encoding Initiative Consortium. ISSN : 2162-5603. DOI : 10.4000/jtei.3500. URL : <https://journals.openedition.org/jtei/3500> (visité le 02/01/2022)

And more...

The Gallic(orpor)a project also integrates the creation of TEI XML files to conserve its corpora built from ATR.

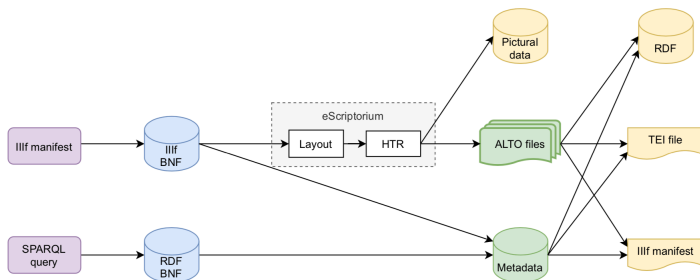


Figure – Data creation protocol for the Gallic(orpor)a project, Kelly CHRISTENSEN, Ariane PINCHE et Simon GABAY. « Gallic(orpor)a : Traitement des sources textuelles en diachronie longue de Gallica ». In : *DataLab de la BnF*. Paris, France, juin 2022. URL : <https://hal.archives-ouvertes.fr/hal-03716534>

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

XML is a pure data format, very simple and well-documented, designed for the **description** of textual documents. XML does not have a predefined set of tags.

```
<note>
```

```
<to>Tove</to>
```

```
<from>Jani</from>
```

```
<heading>Reminder</heading>
```

```
<body>Don't forget me this weekend !</body>
```

```
</note>
```

XML : An International Standard

Since 1996, XML has been a free and well-documented language. XML is also a **standard language** adhering to the recommendations of the **W3C** (World Wide Web Consortium), facilitating :

- readability by machines or human eyes ;
- data exchange ;
- migration to other platforms, software, or formats.

XML and its Associated Languages

XML is part of a comprehensive environment and has associated languages that allow parsing, transforming, and/or querying XML files.

- Xpath is a language for parsing/navigating XML documents :
TEI/text/body/lg/l
- XSLT is a transformation language for XML documents
- Xquery is a language for querying XML documents, which then form a sort of database.

XML : A Bit of History...

XML was born in 1996 under the auspices of the W3C (World Wide Web Consortium).

- SGML (1970), Standard Generalized Markup Language ;
 - ▶ HTML, HyperText Markup Language : displays data, notably on the Web ;
 - ▶ XML, eXtensible Markup Language : contains and structures textual data.

XML : Basic Principles

- Data is included in the XML document as strings delimited by markup describing them.
- The basic unit that includes data and markup is called an element.
Example : `<elementName>characterString</elementName>`
- Elements can be empty : `<element>text</element>` or `<emptyElement/>`
- XML elements follow a strict principle of nesting hierarchy.
- Child elements inherit properties from parent elements.

XML : Basic Principles

- XML attributes can be multiplied as necessary
- The same attribute cannot be added twice to an element.
- Multiple values can be placed in an attribute.

```
<MisEnValeur rendering='red italic' position='centerPage'> text  
</MisEnValeur>
```


Some important rules :

- 1 Every opening tag must have a corresponding closing tag.
- 2 Elements must be nested, but **they cannot overlap**.
- 3 There should only be one root element.
- 4 An element should not have two attributes with the same name.

An encoding that adheres to these XML principles is called **well-formed**.

Well-Formed or Not ?

<paragraphe>du texte</paragraphe>

<paragraphe><article>du</article><nom>texte</nom></paragraphe>

<paragraphe><article>du <nom></article>texte</nom></paragraphe>

<paragraphe type="texte">du texte</paragraphe>

<paragraphe type=texte>du texte</paragraphe>

<paragraphe type="texte">du texte<paragraphe/>

<paragraphe type="texte">du texte<nomPersonnage>nom de personnage</paragraphe>

<paragraphe type="texte">du texte</Paragraphe>

<segment type="texte" type="nombre">du texte</paragraphe>

- Alongside **well-formedness**, there is another important concept which is called **validity**.
- A document is **valid** if it conforms to the rules set out in a document called **schema**.
- Those rules are designed by the human and can be of multiple shape :
 - ▶ what should be the datatype of a given attribute
 - ▶ which attribute is mandatory for a given element
 - ▶ what can contain an element

Schemas are important tools to check the **consistency** of a document or of a set of documents. It is what allows **interoperability**.

TEI : An Introduction

- What is TEI ?
 - ▶ TEI is a set of tags for describing textual sources. It comprises over 550 elements and is constantly evolving.
- What are the advantages of TEI ?
 - ▶ XML TEI provides a common vocabulary for tags.
 - ▶ XML TEI focuses on the meaning of the text rather than its appearance.
 - ▶ XML TEI is independent of any particular software environment.
 - ▶ XML TEI was designed by the scholarly community, which is also responsible for its continuous development.
 - ▶ XML TEI is well documented.

Lou BURNARD. *What is the Text Encoding Initiative? How to add intelligent markup to digital resources.* fr. OpenEdition Press, 2014. URL : <http://books.openedition.org/oep/426> (visit  le 15/04/2014)

"The TEI was initially developed over thirty years ago as a research project in the then-emerging field of 'Humanities computing.' The original idea was to provide a set of recommendations on how researchers should create 'computer-readable' textual resources that were suitable for research purposes - to the extent that a consensus existed on the subject - but were also extensible, as these needs change and evolve."

Some Key Dates

- 1987 : Establishment of the *Text Encoding Initiative* ;
- 1990 : TEI P1 (proposal 1), edited by Michael Sperberg-McQueen and Lou Burnard ;
- 1994 : TEI P3, first complete version ;
- 2000 : Foundation of the TEI Consortium ;
- 2001-2004 : TEI P4, introduction of XML ;
- 2007-... : TEI P5, abandonment of SGML.

The TEI community

The TEI community is driven by the **TEI Consortium**, an interdisciplinary non-profit foundation. It is composed of the following units :

- TEI Board of Directors ;
- TEI Technical Council ;
- Membres institutionnels et individuels ;
- TEI Workgroups, par exemple :
 - ▶ TEI Manuscripts Special Interest Group ;
 - ▶ Correspondence SIG ;

The TEI Community

The community can exchange and meet through :

- An international mailing list : TEI-L mailing list ;
- A French-speaking list : TEI-FR and a wiki ;
- Members meetings ;
- Annual conferences : TEI Conference ;
- A journal : Journal of the Text Encoding Initiative ;
- Guidelines ("recommendations") that document each element in detail.

TEI is a predefined set of tags documented in the TEI guidelines, allowing for a "scientific" and "semantic" description of a text.

TEI tags form a framework useful for designing one's own encoding. **It is strongly discouraged to use the entire TEI for a document.** It is necessary to design a data model that is as simple as possible and adapted to one's project and research question.

For further information : [#DARIAH TEACH. Why Do We Encode? Mai 2019. URL : \[https://www.youtube.com/watch?v=R0ncI_rr1z4\]\(https://www.youtube.com/watch?v=R0ncI_rr1z4\) \(visité le 28/03/2024\)](#)

The importance of the documentation

- TEI is **too flexible** to be used without specifying the uses of a particular project
- What's more, nobody uses the more than 500 elements at our disposal, and it's neither useful nor advisable to do so.
- It is therefore essential to **restrict** the number of elements available in a project, and to **document** both the elements used and the way in which we encode the text.
- This is called **documentation** ; what's nice is that the TEI mixes documentation with schema generation : the **ODD** (One Document Does it All) : BURNARD 2019

- There are subsets of TEI that provide access to a reduced set of tags.
- There are also customizations of TEI organized around an active scholarly community : Epidoc.

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

Every TEI document has at least two parts :

- A header, represented by the `<teiHeader>` element containing metadata describing the document ;
- The text itself, represented by a `<text>` element which can be subdivided into subunits.

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- métadonnées décrivant un texte -->
  </teiHeader>
  <text>
    <!-- texte -->
  </text>
</TEI>
```

Structure of the `teiHeader`

The minimal **teiHeader** includes at least a *fileDesc* element containing the following three sections :

- a **titleStmt** element : information identifying the document itself
- a **publicationStmt** element : information about how it is distributed or published
- a **sourceDesc** element : indications about its origins

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Titre de l'œuvre</title>
    </titleStmt>
    <publicationStmt>
      <p>Informations sur la publication de l'œuvre.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Informations sur la source du texte</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Structure of the Text Body

The **text** element can contain the following elements :

- a **front** element for preliminary text (preface, foreword, etc.)
- a **body** element for the main body of text
- a **back** element for appendices, afterword, etc.

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- métadonnées décrivant un texte -->
  </teiHeader>
  <text>
    <front/>
    <body/>
    <back/>
  </text>
</TEI>
```

Main Structural Elements of the Text

Textual information can be structured using the following elements :

- **div** for major divisions or subdivisions of the text
- **p** for paragraphs
- **lg** for stanzas
- **l** for lines
- **sp** for speeches
- These elements can be customized with the *type* and *n* attributes.

```
<text>
  <body>
    <div type="book" n="1">
      <div type="chapter" n="1">
        <p>Some text here.</p>
      </div>
    </div>
  </body>
</text>
```


Structure in XML TEI the poem by Paul Verlaine, *Mon rêve Familier* Link to the text :

https://fr.wikipedia.org/wiki/Mon_rêve_familier

- Open Oxygen Editor
- Open a new *XML TEI* *all* file
- Fill in a minimal `teiHeader`
- Copy Verlaine's text into the **body**
- Structure the poem using the elements seen previously

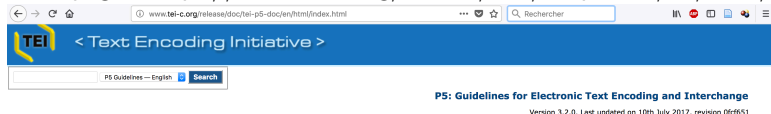
The guidelines aim to :

- Provide a standard format
- Facilitate the exchange of texts
- Offer sets of encoding conventions tailored to different applications

TEI Guidelines, How-to

How to read the guidelines ?

Homepage : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>



[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[한국어\]](#) [\[中文\]](#)



Front Matter

[Title](#)

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

Text Body

- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Characters, Glyphs, and Writing Modes](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Dictionaries](#)
- 10 [Manuscript Description](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulae, Graphics and Notated Music](#)
- 15 [Language Corpora](#)
- 16 [Linking, Segmentation, and Alignment](#)
- 17 [Simple Analytic Mechanisms](#)
- 18 [Feature Structures](#)
- 19 [Graphs, Networks, and Trees](#)
- 20 [Non-hierarchical Structures](#)
- 21 [Certainty, Precision, and Responsibility](#)
- 22 [Documentation Elements](#)
- 23 [Using the TEI](#)

TEI sourcecode

- [Getting and Using the TEI Sources](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

How to read element specifications ?

www.tei-c.org/release/doc/tei-p5-doc/tei/html/ref-msIdentifier.html

Rechercher

Version 3.2.0. Last updated on 10th July 2017, revision 06cf651

<msIdentifier>

Accueil
C Éléments

<msIdentifier> (identifiant du manuscrit) Contient les informations requises pour identifier le manuscrit en cours de description. [10.4 The Manuscript Identifier]

Module	msdescription — Manuscript Description	L'élément est documenté dans le module msdescription (10 Manuscript Description)
Attributs	<div><div>att.global</div><div>@xml:id, @n, @xml:lang, @xml:base, @xml:space (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @include)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) (att.global.source (@source))</div><div>attribute class</div></div>	
Membre du	model.bibliPart	model class
Contenu dans	<div>core: bibli</div> <div>msdescription: msDesc msFrag msPart</div> <div>éléments regroupés par modules</div>	
Peut contenir	<div>header: idno</div> <div>msdescription: altIdentifier collection institution msName repository</div> <div>namesdates: bloc country district geoName placeName region settlement</div>	
Exemple	<pre><msIdentifier> <country>France</country> <settlement>Paris</settlement> <repository xml:lang="fr">Bibliothèque nationale de France. Réserve des livres rares</repository> <idno>B-73</idno> <!-- dans le cas des recueils : cote uniquement sans les sous-cotes --> <altIdentifier> <idno>B-121</idno> <note>Cote de la bibliothèque royale au XVIIIe siècle (inscrite à l'encre, sur la doublure de table).</note> </altIdentifier> <altIdentifier> <idno>Double de B. 274. A (Réserve)</idno> <note>Cote inscrite face à la page de titre, en remplacement de la cote "1541", barrée</note> </altIdentifier> </msIdentifier></pre> <div>Toute la liste bibliograph</div>	

Structuring an Index

In XML TEI, we can mark up place or person names and even begin to structure an index of characters or locations as we edit a text.

Named entities are marked using the *persName* or *placeName* tags.

```
<text n="edition">
  <body n="urn:cts:froLit:jns915.jns2117.ciham-fro1">
    <div n="1"
      type="chapter"
      facs="http://gallica.bnf.fr/ark:/12148/btv1b84259980/f237, http://gallica.bnf.fr/ark:/12148/btv1b84259980/f238"
      corresp="urn:cts:latinLit:stoa0270.stoa003:1.1, urn:cts:latinLit:stoa0270.stoa003:1.2">
      <head>Ci coumence<lb/>li dialogues que <persName ref="#postumien">postumiens</persName> <persName ref="#gallus">Gal
      <persName ref="#severus">seuerus</persName> le
      rate<lb/> les oi <app>
        <lem>ler</lem>
        <rdg wit="#C2 #C3" type="ajout">parler et qi i fu</rdg>
      </app>
    </head>
    <div n="1" type="section">
      <p>
        <pb n="114" facs="http://gallica.bnf.fr/ark:/12148/btv1b84259980/f237"/>
        <hi rend="rubricated orig">Ci coumence<lb/>li dialogues que <persName ref="#postumien">postumiens</persName>
        de <persName ref="#martin">seint martin</persName> des
        moines <placeName ref="#egypte" full="yes">egy<lb/>pte</placeName> si <persName ref="#severus">seuerus</persName>
        oi ler<hi>
          <lb/>
          <hi rend="decorated-initial">U</hi>N ior auint<lb/>qeie
        <persName ref="#gallus">gallus</persName>
          <lb/>
          <cb n="b"/>
        Et mes chiers compains qi deciples Fu<lb/>seint martin estions ensamble laou
        nos<lb/>liions depluiseurs choses<persName ref="#postumien">Postumiens</persName>.i.<lb/>miens tres chiers amisuint
        <placeName ref="#orient" full="yes">orit</placeName>
        <lb/>deuers la terre <placeName full="yes">egypte</placeName>.ouil auoit<seg type="number">iii</seg>
      </p>
    </div>
  </body>
</text>
```

Structuring an Index

To ensure all occurrences of an entity point to a single entry, it must be declared either in the *profileDesc* element of the *teiHeader* or in a *standOff* element in a list of people or places.

```
<listPerson>
  <person xml:id="agaridut">
    <persName>Agariduz</persName>
    <note type="biographical">Honnête homme, témoin de l'adoucissement des mœurs cruelles
d'Avicien.</note>
  </person>
  <person xml:id="agnes">
    <persName>Agnies</persName>
    <death when="0300">4<ex>e</ex> siècle</death>
    <note type="biographical" source="http://data.bnf.fr/11958468/agnes/">Martyre
romaine.</note>
  </person>
  <person xml:id="antoine">
    <persName>Antoines</persName>
    <birth when="0251">251</birth>
    <death when="0356">356</death>
    <note type="biographical" source="http://data.bnf.fr/11888967/antoine/">
      <surname type="complex" full="yes">Antoine le grand, Antoine d'Égypte ou Antoine
l'Ermite</surname>, considéré comme le fondateur du l'érémisme chrétien.</note>
  </person>
  <person xml:id="arborius">
    <persName>Arborius</persName>
    <note type="biographical">Magnus Arborius, élu comte des largesses sacrées en 379, puis
préfet de Rome en 380, il aurait été l'un des protecteurs de <persName ref="#saintMartin">saint Martin</persName>.</note>
  </person>
  <person xml:id="arpage">
    <persName>Arpage</persName>
    <note type="biographical">Prêtre.</note>
  </person>
```

Documentation :

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

Putting into Practice : Structuring an Index

Linking the index to occurrences using a set of identifiers and pointers.

- The identifier is unique, declared in the list of entities in the index with the *xml:id* attribute
- ex : `<person xml:id="agaridut">`
- The pointer is added to occurrences of the entity in the text with the *ref* attribute
- ex : `<persName ref="#agaridut">`
- the pointer value is the id string preceded by a hash.

Putting into Practice : Structuring an Index

Encode the first two paragraphs of "Notre Dame de Paris", book II, chapter 1, "From Charybdis to Scylla" (beginning of chapter to "La cohue admirait").

- Structure the text into books, chapters, and paragraphs
- Mark up place names
- Create an index

Link to the text : https://fr.wikisource.org/wiki/Notre-Dame_de_Paris/Livre_deuxi%C3%A8me#II._La_place_de_Gr.C3.A8ve

Putting into Practice : Structuring a Theater Play

Convert a excerpt from "The Misanthrope" by Molière into XML TEI, using the documentation and the following tags : `<castList>`, `<castItem>`, `<speaker>`, `<stage>`, `<head>`, `<div>`, `<sp>`, and `<l>`.
Link to the text :

https://fr.wikisource.org/wiki/Le_Misanthrope/Édition_Louandre,_1910
Documentation in the guidelines :

<https://tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

Structuring a "Paleographic" Encoding

Among the many possibilities offered by TEI, we can produce an **imitative encoding** of a manuscript, for example, marking abbreviations while providing their expansion, noting a faulty spelling and suggesting a correction. These alternative proposals can be rendered throughout the text using the tag *choice*.

Quelques balises courantes

- Développements d'abréviation : `<choice><abbr></abbr>
<expan></expan></choice>` ;
- Régularisations : `<choice><orig></orig><reg></reg>
</choice>` ;
- Corrections : `<choice><sic></sic><corr></corr></choice>` .

Documentation sur la représentation des sources primaires :
<http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/PH.html>

Structuring a Critical Apparatus

XML TEI also allows for the creation of critical apparatuses. There are several methods for encoding apparatuses, with the simplest being the *parallel segmentation method*, which involves adding the apparatus directly to the text of the base manuscript.

```
<app>
  <lem wit="#E1 #Hg">Experience</lem>
  <rdg wit="#La" type="substantive">Experiment</rdg>
  <rdg wit="#Ra2" type="substantive">Eryment</rdg>
</app>
```

Documentation on representing apparatuses :

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPLK>

Structuring a Critical Apparatus

To link the variants to the witnesses, it will be necessary to declare the list of witnesses in the *teiHeader* beforehand.

Déclaration des sources dans le `<sourceDesc>` :

- `<listWit>` : donne une liste de tous les témoins cités dans un apparat critique;
- `<witness>` : témoin d'une édition.

```
<listWit>
  <witness xml:id="C1">Paris, BnF, ms 412</witness>
  <witness xml:id="C2">Paris, BnF, ms 411</witness>
  <witness xml:id="C3">London, British Library,
    royal 20 D. VI.</witness>
</listWit>
```

Structuring Linguistic Data

Digital editions can be enriched with numerous pieces of information that cannot be added to a paper edition. Among these possibilities is the addition of linguistic annotations, which will subsequently allow for the study of the language of the text.

```
<w xml:id="t78"  
  n="78"  
  lemma="cel"  
  pos="PR0dem"  
  msd="NOMB.=p|GENRE=m|CAS=r">ceus</w>  
<w xml:id="t79"  
  n="79"  
  lemma="venir"  
  pos="VERcjc"  
  msd="MODE=ind|TEMPS=psp|PERS.=3|NOMB.=s">vint</w>  
<w xml:id="t80" n="80" lemma="por2" pos="PRE">por</w>  
<w xml:id="t81"  
  n="81"  
  lemma="nostre"  
  pos="DETpos"  
  msd="PERS.=1|NOMB.=s|GENRE=m|CAS=r">nostre</w>  
<w xml:id="t82"  
  n="82"  
  lemma="seigneur"  
  pos="NOMcom"  
  msd="NOMB.=s|GENRE=m|CAS=r">Signor</w>  
<w xml:id="t83"  
  n="83"  
  lemma="öir"  
  pos="VERinf"  
  msd="NOMB.=x|GENRE=x|CAS=x">oïr</w>
```

Documentation on linguistic annotation :

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/A1.html#A1LALW>

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

- *TEI Publisher*. URL :
<https://teipublisher.com/exist/apps/tei-publisher-home/index.html> (visité le 28/03/2024)
- Roberto Rosselli Del TURCO et al. « Edition Visualization Technology : A Simple Tool to Visualize TEI-based Digital Editions ». en. In : *Journal of the Text Encoding Initiative* Issue 8 (déc. 2014). Number : Issue 8 Publisher : Text Encoding Initiative Consortium. ISSN : 2162-5603. DOI : 10.4000/jtei.1077. URL :
<https://journals.openedition.org/jtei/1077> (visité le 28/03/2024)
- *TEIC/CETELcean*. original-date : 2016-05-03T17 :29 :24Z. Fév. 2024. URL : <https://github.com/TEIC/CETELcean> (visité le 28/03/2024)
- *TEI Critical Apparatus Toolbox*. URL :
<http://teicat.huma-num.fr/> (visité le 28/03/2024)

What to do with the TEI-encoded documents?

- Produce a digital edition
- Use it as a database to retrieve information
- Perform complex queries on the structure
- use it as a base for other tasks (NLP ? tokenization ? linguistic annotations ?)

Table of Contents

- 1 General Principles
- 2 What is XML TEI ?
 - XML
 - TEI
- 3 Let's practice
 - TEI Guidelines
 - Structuring an Index
- 4 Some Examples of Common Encoding
- 5 A few ideas for visualising your TEI
- 6 References

Références I

- [1] #DARIAH TEACH. *Why Do We Encode ?* Mai 2019. URL : https://www.youtube.com/watch?v=R0ncI_rriz4 (visité le 28/03/2024).
- [2] Lou BURNARD. « What Is TEI Conformance, and Why Should You Care ? » In : *Journal of the Text Encoding Initiative* 12 (2019). ISSN : 2162-5603. URL : <http://journals.openedition.org/jtei/1777> (visité le 03/07/2019).
- [3] Lou BURNARD. *What is the Text Encoding Initiative ? How to add intelligent markup to digital resources*. fr. OpenEdition Press, 2014. URL : <http://books.openedition.org/oep/426> (visité le 15/04/2014).
- [4] Lou BURNARD, Christof SCHÖCH et Carolin ODEBRECHT. « In search of comity : TEI for distant reading ». fr. In : *Journal of the Text Encoding Initiative* Issue 14 (mars 2021). Number : Issue 14 Publisher : Text Encoding Initiative Consortium. ISSN : 2162-5603. DOI : 10.4000/jtei.3500. URL : <https://journals.openedition.org/jtei/3500> (visité le 02/01/2022).
- [5] Kelly CHRISTENSEN, Ariane PINCHE et Simon GABAY. « Gallic(orpor)a : Traitement des sources textuelles en diachronie longue de Gallica ». In : *DataLab de la BnF*. Paris, France, juin 2022. URL : <https://hal.archives-ouvertes.fr/hal-03716534>.
- [6] *TEI Critical Apparatus Toolbox*. URL : <http://teicat.huma-num.fr/> (visité le 28/03/2024).
- [7] *TEI Publisher*. URL : <https://teipublisher.com/exist/apps/tei-publisher-home/index.html> (visité le 28/03/2024).
- [8] *TEIC/CETEIcean*. original-date : 2016-05-03T17 :29 :24Z. Fév. 2024. URL : <https://github.com/TEIC/CETEIcean> (visité le 28/03/2024).
- [9] Roberto Rosselli Del TURCO, Giancarlo BUOMPRISCO, Chiara Di PIETRO, Julia KENNY, Raffaele MASOTTI et Jacopo PUGLIESE. « Edition Visualization Technology : A Simple Tool to Visualize TEI-based Digital Editions ». en. In : *Journal of the Text Encoding Initiative* Issue 8 (déc. 2014). Number : Issue 8 Publisher : Text Encoding Initiative Consortium. ISSN : 2162-5603. DOI : 10.4000/jtei.1077. URL : <https://journals.openedition.org/jtei/1077> (visité le 28/03/2024).