



HAL
open science

TEI: Text Encoding Initiative

Christine Chadier

► **To cite this version:**

Christine Chadier. TEI: Text Encoding Initiative. Les Carnets du LARHRA, 2023, 2023-1. halshs-04466237

HAL Id: halshs-04466237

<https://shs.hal.science/halshs-04466237>

Submitted on 21 Feb 2024

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 - NonCommercial - ShareAlike 4.0 International License

Les Carnets du LARHRA

ISSN : 2648-1782

Éditeur : Laboratoire de recherche historique Rhône-Alpes

2023 | 1 | 2023

Numéro spécial "Les 20 ans du LARHRA"

TEI : TEXT ENCODING INITIATIVE

Christine Chadier

🔗 <https://publications-prairial.fr/larhra/index.php?id=942>

Référence électronique

Christine Chadier, « **TEI : TEXT ENCODING INITIATIVE** », *Les Carnets du LARHRA* [En ligne], 2023 | 1 | 2023, mis en ligne le 18 décembre 2023, consulté le 19 décembre 2023. URL : <https://publications-prairial.fr/larhra/index.php?id=942>

TEI : TEXT ENCODING INITIATIVE

Christine Chadier

TEXTE

```
<div type="Oullins_1603_participants">
  <list>
    <item n="1">Premierement pour l'Eglise de Lyon M<hi rend="sup">r</hi>
      <ref xml:id="a-8" target="#n1603_8">de Brun<ref> Pasteur, les sieurs
      <ref xml:id="a-9" target="#n1603_9">Gras</ref> et <ref xml:id="a-10"
      target="#n1603_10">Huguetan</ref> Anciens.</item>
    <item n="2">Pour les Eglises de <ref xml:id="a-11" target="#n1603_11"
      >Couches</ref> et <ref xml:id="a-12" target="#n1603_12">Autun</ref>
      M<hi rend="sup">r</hi>
      <ref xml:id="a-13" target="#n1603_13">Textor</ref> Pasteur, et M<hi
      rend="sup">r</hi> Lazare Feure Ancien.</item>
    <item n="3">Pour l'Eglise de <ref xml:id="a-14" target="#n1603_14"
      >Vosne</ref> M<hi rend="sup">r</hi>
      <ref xml:id="a-15" target="#n1603_15">Cassegain</ref> Pasteur, et les
      sieurs <ref xml:id="a-16" target="#n1603_16">Fournier</ref> avocat et
      Lavisé marchand, Anciens.</item>
    <item n="4">Pour l'Eglise de <ref xml:id="a-17" target="#n1603_17"
      >Paray</ref>, M<hi rend="sup">r</hi>
      <ref xml:id="a-18" target="#n1603_18">Colinet</ref> Pasteur, et M<hi
      rend="sup">r</hi>
      <ref xml:id="a-19" target="#n1603_19">Pierre Fageous</ref> Ancien. <pb
      xml:id="fol.50vo"/>
    </item>
    <item n="5">Pour l'Eglise d'<ref xml:id="a-20" target="#n1603_20"
      >Issurtille</ref> M<hi rend="sup">r</hi>
      <ref xml:id="a-21" target="#n1603_21">Gaus<ref> Pasteur, sans
      ancien, et seront ceux de ladicte eglise exhortés à s'esvertuer à
      l'advenir.</item>
    <item n="6">Pour l'Eglise d'<ref xml:id="a-22" target="#n1603_22"
      >Arneleduc</ref> M<hi rend="sup">r</hi>
      <ref xml:id="a-23" target="#n1603_23">Heliot</ref> Pasteur et le sieur
      Hugues Robert Ancien.</item>
    ...
    [...]
    <item n="16">Pour les Eglises de Chastillon et <ref xml:id="a-38"
      target="#n1603_38">Aign<ref> Jean Raol ancien n'ayans
      lesdictes Eglises aucun Pasteur. </item>
    <item n="17">Pour l'Eglise de Bourbon Nancis, lesdicts Sieurs Colinet et
      Fageous ayant fait apparoir de procuracion.</item>
  </list>
  <p>Pasteurs 13. </p>
  <p>Anciens 19. </p>
</div>
```



Réalisation Christine Chadier, CC-BY-NC-SA

- 1 Depuis la mise en place du pôle d'histoire numérique (devenu Axe de recherche en histoire numérique) le Laboratoire de recherche historique est résolument tourné vers l'interopérabilité et l'accessibilité des données, une dizaine d'années avant même que le « FAIR¹ » et l'Open Science ne deviennent une recommandation du CNRS. Dans ce cadre, pour les projets d'édition de sources, nous avons cherché un outil qui permette le traitement numérique de l'édition de texte. Nous avons choisi d'utiliser la TEI, Text Encoding Initiative, pour les grandes possibilités de balisage structurel et sémantique qu'elle offre, mais aussi parce qu'elle reste lisible par l'œil humain et, tant pour faciliter la saisie en TEI que le travail sur des corpus de textes qui nécessitent l'utilisation de schémas pour contrôler l'édition, nous avons retenu l'éditeur xml Oxygen, assez

souple pour être indifféremment utilisé par les chercheurs et les ingénieurs mais permettant aussi d'assurer un traitement affiné des données.

- 2 La TEI a été mise en place en 1987 par trois sociétés savantes, l'Association for Computers and the Humanities, l'Association for Computational Linguistics et l'Association for Literary and Linguistic Computing. Depuis 1987, le modèle théorique original s'est adapté à différentes technologies. Le schéma TEI utilisé actuellement est la version P5 de 2007. Ce schéma xml-TEI² est le fruit du travail de comités internationaux chargés notamment de la maintenance et la croissance du schéma, de rédiger la documentation, de développer des outils génériques, d'assurer le support sur des listes de diffusions et de faire connaître le format. La TEI permet de modéliser la structure d'un texte tel qu'il a été conçu ce qui est particulièrement précieux pour l'historien pour lequel la forme du texte peut se révéler aussi significative que son contenu. La TEI fournit un cadre pour penser la nature du texte. Le langage xml-TEI est aujourd'hui un standard en matière d'édition de textes et particulièrement d'édition de sources.
- 3 Pour essayer de résumer un langage aussi structuré et développé que la TEI ³en moins de 5 000 signes, retenons qu'il permet de rassembler dans un seul document :
- 4 les métadonnées dans le TEI Header
 - abstract,
 - idno identifiant standardisé
 - keywords
 - language
 - licence
 - profileDesc (fournit une description détaillée des aspects non bibliographiques du texte, notamment les langues utilisées et leurs variantes, les circonstances de sa production, les collaborateurs et leur statut)
 - publicationStmt (regroupe des informations concernant la publication ou la diffusion d'un texte électronique)
 - revisionDesc
 - titleStmt (regroupe les informations sur le titre d'une œuvre et les personnes ou institutions responsables de son contenu etc.)

- 5 et le texte lui-même (<text>) contenant un <front> (Titre, description du manuscrit, lieu de conservation, etc.) le <body>, structuré en éléments <div>. Chaque <div> peut contenir différents éléments textuels : liste (<list>), paragraphe (<p>), etc. La TEI permet non seulement d'attribuer à chaque élément <div> du <body> un identifiant unique mais aussi de les « typer » (fig. 1).



Fig 1 : Structure en xml TEI des actes du synode de Bourgogne de 1618 qui s'est tenu à Paray-le-Monial (capture d'écran Oxygen, éditeur xml)

- 6 Typer les éléments est particulièrement important lorsque l'on doit traiter un ensemble de textes (recueil d'actes, lettres, etc.), cela permet notamment de cibler les parties du texte sur lesquelles on souhaite effectuer une recherche textuelle (i. e. préambule, liste, signature, etc.) et l'identifiant unique permet de repérer instantanément la portion du texte concerné et ce quel que soit le langage utilisé pour effectuer les recherches. On pourrait par exemple écrire une requête en xslt pour lister des occurrences en

fonction de l'identifiant attribué à une personne ou un lieu (i.e. rechercher les différentes graphies).

- 7 Pour mettre en place un projet d'édition de sources en TEI, il faut un chercheur qui prenne en compte l'intérêt de ce que le numérique peut apporter à sa recherche et un ingénieur pour organiser le travail. C'est cette collaboration étroite qui permet la réussite et la pérennité d'un projet.

NOTES

1 Findable, Accessible, Interoperable, Reusable / Facile à trouver, Accessible, Interopérable, Réutilisable.

2 L'Extensible Markup Language, généralement appelé XML est le métalangage informatique de balisage générique sur lequel s'appuie actuellement la technologie de la TEI.

3 Les *Guidelines*, CONSORTIUM TEI, P5 : *Recommandations pour l'encodage et l'échange de textes électroniques* Version 4.3.0. Last updated on 31st August 2021, revision b4f72b1ff [en ligne : <https://tei-c.org/release/doc/tei-p5-doc/fr/html/index.html>] se sont 23 chapitres et 9 Annexes...

INDEX

Mots-clés

TEI, édition de sources, édition numérique, humanités numériques, données, FAIR

Keywords

TEI, source publishing, digital publishing, digital humanities, data, FAIR

AUTEUR

Christine Chadier