



# Modelling humanities data with TEI-XML

SCHOLARLY EDITING AND MANUSCRIPT CATALOGUING IN THE DIGITAL AGE

Dr Katarzyna Anna Kapitan  
16 October 2024

# Course Materials

► **GitHub Repo:**

[https://github.com/KAKDH/ENC\\_TNAH\\_2024/](https://github.com/KAKDH/ENC_TNAH_2024/)

# DTD Recap

- You need at least two files, one xml file (e.g. AM\_30\_fol.xml) and one dtd file (e.g. myDTD.dtd).
  - To associate your DTD file with your XML file you need to add the following close to the beginning of the XML file:  
`<!DOCTYPE XML SYSTEM "myDTD.dtd">`
- \* Remember that both files must be saved in the same folder for this declaration to work.

A DTD file  
correctly  
associated with  
the XML file

The screenshot shows a TEI-XML editor interface. At the top, there's a toolbar with various icons. Below it, a status bar displays 'XPath 2.0' and 'Execute XPath on 'Current File''. The main window has tabs for 'AM\_30\_fol.dtd.dtd' and 'AM\_30\_fol.xml'. The XML code is displayed in a syntax-highlighted text area:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Note this is a mock-up msDesc file for the "Modelling humanities data" course at the University of Amsterdam.
The file is based on a file created by Katarzyna Anna Kapitan and N. K. Kavuz.
which in turn is based on the description of AM 30 fol. by Katarzyna Anna Kapitan.
This file can be freely used under the CC-BY 4.0 license. --&gt;
<!-- This file is used for teaching introductory concepts in XML and XSLT.
&lt;!DOCTYPE XML SYSTEM "AM_30_fol_dtd.dtd"&gt;
&lt;XML&gt;
&lt;fileDesc&gt;
    &lt;titleStmt&gt;
        &lt;title&gt;Basic Description of AM 30 fol.&lt;/title&gt;
    &lt;/titleStmt&gt;
    &lt;publicationStmt&gt;
        &lt;publisher&gt;Katarzyna Anna Kapitan and N. K. Kavuz&lt;/publisher&gt;
        &lt;p&gt;This is an open access file licensed under a Creative Commons International license.&lt;/p&gt;
    &lt;/publicationStmt&gt;
&lt;/fileDesc&gt;</pre>

Annotations include a green arrow pointing from the 'Validation successful' message to the validation icon in the status bar, and a green circle highlighting the validation icon in the toolbar.



Validation successful



Text Grid



/Users/.../AM_30_fol.xml Validation successful



ABC ABC



35 : 30 9 new mess...


```

# DTD Recap

- ▶ **Elements need to be defined only once**
  - ▶ `<!ELEMENT titleStmt (title)>`
    - ▶ definition of "titleStmt" which has "title" as content (the title element is defined elsewhere)
  - ▶ `<!ELEMENT msItem (title, author)>`
    - ▶ definition of "msItem" which has "title" as one of the contents (the title element is defined elsewhere)
  - ▶ `<!ELEMENT title (#PCDATA)>`
    - ▶ definition of "title", which takes PCDATA as content.

# DTD Recap

- ▶ Attributes are defined by ATTLIST following the syntax:
  - ▶ <!ATTLIST msItem n CDATA #REQUIRED>

# DTD Recap

- ▶ What is wrong with this definition?
  - ▶ Note that our file won't validate
- ▶ `<!ELEMENT msContents (summary, msItem)>`

# DTD Recap

- ▶ **<!ELEMENT msContents (summary, msItem)>**
- =>
- ▶ **<!ELEMENT msContents (summary, msItem+)>**

# DTD Recap

- ▶ What is potentially wrong with this definition?
  - ▶ note that our file will validate
- ▶ `<!ELEMENT publicationStmt (publisher, p)>`

# DTD Recap

- ▶ **<!ELEMENT publicationStmt (publisher, p)>**
- =>
- ▶ **<!ELEMENT publicationStmt (publisher, p+)>**

# DTD Recap

- ▶ What is potentially wrong with these definitions?
  - ▶ note that our file will validate
- ▶ `<!ELEMENT msItem (title, author)>`

# DTD Recap

- ▶ **<!ELEMENT msItem (title, author)>**
- =>
- ▶ **<!ELEMENT msItem (title+, author?)>**

# Text Encoding Initiative

```
<?xml version="1.0" encoding="UTF-8"?>
<workshop name="XML_workshop">
  <instructors>
    <name>
      <firstName>Katarzyna</firstName>
      <lastName>Kapitan</lastName>
    </name>
  </instructors>
  <participants>
    <name>
      <firstName>John</firstName>
      <lastName>Doe</lastName>
    </name>
    <name>
      <firstName>Anna</firstName>
      <lastName>Smith</lastName>
    </name>
    [...]
  </participants>
</workshop>
```

**Workshop/Class/Course**  
**Instructors/Teachers:** KAK  
**Participants/Students:** JD, AS, [...]



```
<?xml version="1.0" encoding="UTF-8"?>
<class title="XML_class">
  <teachers>
    <person>
      <firstName>Katarzyna</firstName>
      <lastName>Kapitan</lastName>
    </person>
  </teachers>
  <students>
    <person>
      <firstName>John</firstName>
      <lastName>Doe</lastName>
    </person>
    <person>
      <firstName>Anna</firstName>
      <lastName>Smith</lastName>
    </person>
    [...]
  </students>
</class>
```

# Need for a lingua franca: TEI



**XML**

**TEI**

**Concepts**



**Syntax**

```
<element>
  <element attribute="value">
    content
  </element>
</element>
```

**Language:  
vocabulary and grammar**

```
<p>
<note type="foot">
<head>
```

© 2007 Syd Bauman, Julia Flanders, and the Women Writers Project. Creative Commons Attribution-ShareAlike 3.0 (Unported) license.

Kapitan, *Modelling humanities data with TEI-XML*

# The TEI Consortium

- ▶ The TEI is an **international** and **interdisciplinary** standards project
- ▶ Established in **1987** to develop, maintain and promote hardware- and software-independent methods for encoding humanities data in electronic form.
- ▶ Since **1994**, the TEI Consortium has been issuing Guidelines which specify encoding methods for machine-readable texts.



Text  
Encoding  
Initiative

# The TEI Consortium

- ▶ Community (Membership )
- ▶ Training & Outreach
- ▶ Annual Conference
- ▶ Journal
- ▶ TEI Guidelines
- ▶ TEI Infrastructure

# TEI: Guidelines for Electronic Text Encoding and Interchange

The screenshot shows the homepage of the TEI Guidelines website. The header features the TEI logo and a navigation menu. The main title is "TEI: Guidelines for Electronic Text Encoding and Interchange". Below it is the subtitle "P5 Version 4.8.0. Last updated on 2nd September 2024, revision 4e6e41b0b". Language links for English, Deutsch, Español, Italiano, Français, 日本語, 한국어, and 中文 are provided. There are also icons for PDF, EPUB, and XML formats. The page is divided into two main sections: "Front Matter" and "Text Body". The "Front Matter" section includes links for Title, Acknowledgments, and Sets. The "Text Body" section is a numbered list of 10 chapters: 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. Computer-mediated Communication, and 10. Dictionaries.

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

**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](#)

**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 [Computer-mediated Communication](#)
- + 10 [Dictionaries](#)

Kapitan, Modelling humanities data with TEI-XML

# TEI Infrastructure

- ▶ Templates: <https://github.com/TEIC/TEI/blob/dev/P5/Exemplars/>
- ▶ Stylesheets: <https://github.com/TEIC/Stylesheets>
- ▶ Tools:
  - ▶ TEI Garage: <https://teigarage.tei-c.org>
  - ▶ Roma (<https://roma.tei-c.org>)

# Why to use TEI?

- ▶ International standard
- ▶ Adopted by many institutions in Europe and the USA
- ▶ Interoperable
- ▶ Machine readable
- ▶ Easy to transport/transform/reuse
- ▶ Easy to analyse

# Modelling humanities data with the TEI Guidelines



Text  
Encoding  
Initiative

The TEI Guidelines allow us to model our data so that they are:

- Sustainable
- Sharable
- Analytically rich

```

<front>
<!--<titlePage>書名頁</titlePage>-->
<divGen type="toc"/>
<div>
  <head>序</head>
  <p> ... </p>
</div>
</front>

```

```

<ab>
  <lb n="1"/> بـسـم اللـه الرـحـيم
  <lb n="2"/> هـذـا مـعـنـى لـفـاظـيـنـا
  <lb n="3"/> عـلـى أـيـوب مـوـلـي أـيـانـبـنـعـاصـمـ
  <lb n="4"/> فـعـلـي أـبـو دـارـه يـدـفـع إـلـي مـقـارـه سـلـمـون وـأـتـنـاسـ
  <lb n="5"/> فـإـن لـم يـجـدـهـم فـعـلـي أـبـو دـارـه مـا كـانـ عـلـيـهـمـ مـنـ شـيـ
  <lb n="6"/> شـهـدـ عـلـى ذـكـرـ أـيـوب مـوـلـي أـيـانـبـنـعـاصـمـ
  <lb n="7"/> وـعـطـاـبـنـ تـوـابـةـ الـلـخـمـيـ وـسـعـيدـ مـوـلـيـ أـبـيـ إـبـرـاهـيمـ
  <lb n="8"/> وـإـبـرـاهـيمـبـنـ سـالـمـ الـبـخـتـيـ وـسـالـمـبـنـ إـسـحـاقـ الـبـخـتـيـ وـعـبـيـدـبـنـ مـجـنـادـ
</ab>

```

```

<front>
  <divGen type="toc"/>
  <div>
    <head>Préface</head>
    <p> ... </p>
  </div>
</front>

```

```

<front>
<!--<titlePage>...</titlePage>-->
<divGen type="toc"/>
<div>
  <head>Preface</head>
  <p> ... </p>
</div>
</front>

```

Source: <https://tei-c.org/release/doc/tei-p5-doc/fr/html/ref-div.html>

Source: <https://arabic-tei-workshop.github.io/index.html>

# What to use TEI for?

- ▶ Manuscript descriptions
- ▶ Transcriptions of texts
- ▶ Editions of texts (documentary, genetic, critical)
- ▶ Annotated corpora (linguistic, spatial, prosopography)
- ▶ Dictionaries
- ▶ etc.

# Projects Using the TEI

The following is a list of projects that use the TEI encoding scheme. If you would like to add your project to the list or have corrections or updates, please fill in this [form](#).

- [African American Women Writers of the 19th Century](#)
- [African Languages Lexicon Project \(ALLEX\)](#)
- [Album interactif de paléographie médiévale](#)
- [Alex Catalogue of Electronic Texts](#)

[American Memory from the Library of Congress](#)

## Who uses TEI?

<https://tei-c.org/activities/projects/>

# TEI for Text Editing

Kapitan, Modelling humanities data with TEI-XML

# PERSEUS DIGITAL LIBRARY



PERSEUS DIGITAL LIBRARY

GREGORY R. CRANE, EDITOR-IN-CHIEF  
TUFTS UNIVERSITY

[Home](#) [Collections/Texts](#) [Perseus Catalog](#) [Research](#) [Grants](#) [Open Source](#) [About](#) [Help](#)

## Welcome to Perseus 4.0, also known as the Perseus Hopper.

Read more on the [Perseus version history](#).  
New to Perseus? Click [here](#) for a short tutorial.



### Perseus News and Updates

- Please visit the [Perseus Updates](#) blog for news on project activities, research, and initiatives. We invite you to contact us via [email to the Perseus webmaster](#) if you have any comments, questions, or concerns.
- April 24, 2019: Current Projects and Initiatives**
  - Work continues on the [Scaife Viewer](#), our first new reading environment in nearly 15 years. For more, please read [About the Scaife Viewer](#) and send us your comments.
  - The Perseus Digital Library is a partner and supporter of [Open Greek and Latin](#), an international collaboration committed to creating an open educational resource featuring a corpus of digital texts, deep-reading tools, and open-source software. Look for new OGL materials in the Scaife Viewer.
  - News, help and support-related content for this site ("Perseus 4.0") will be updated periodically, but the site collections and infrastructure are no longer under active development as we begin the transition to the next phase of Perseus.

### Release Announcements

- October 2013

New texts: the [English Bohn](#) and [Greek Kaibel](#) editions of Athenaeus' *Deipnosophists* and *Harpocration*.

Corrections to Greek and Latin lexicons, Oppian, Smith's Geography, Pausanias, Cassius Dio.

CIDOC RDF download links added for Art and Architecture data.

- June 17, 2013

The navigation bar and text sidebars now include links to the [Perseus Catalog](#).

Search  
("Agamemnon", "Hom. Od. 9.1", "denarius")  
[All Search Options](#) [view abbreviations]

### Popular Texts

- Caesar, *Gallic War* (English, Latin)
- Catullus, *Carmina* (English, Latin)
- Cicero, *In Catilinam I* (English, Latin)
- Vergil, *Aeneid* (English, Latin)
- Herodotus, *Histories* (English, Greek)
- Homer, *Odyssey* (English, Greek)
- Plato, *Republic* (English, Greek)
- Tom Martin, *Overview of Classical Greek History from Mycenae to Alexander* (English)

### Art and Archaeology



Aegina, Temple of Aphaia



Silver obol from Athens



Satyr on Attic red figure vase



The Bartlett Head

### Exhibits



# MEDIEVAL NORDIC TEXT ARCHIVE

Medieval Nordic Text Archive  
www.menota.org ■■■■■

Arkiv for nordiske middelaldertekster  
Arkiv för nordiska medeltidstexter  
Safn norrænna miðaldatexta

Norsk

## Background

Menota is a network of leading Nordic archives, libraries and research departments working with medieval texts and manuscript facsimiles. The aim of Menota is to preserve and publish medieval texts in digital form and to adapt and develop encoding standards necessary for this work. The archive will contain texts in the Nordic languages as well as in Latin. There are now 15 members of Menota (see bottom of the page) and new members are welcome to join the network. Since its founding in 2001, the archive has been led by Odd Einar Haugen at the University of Bergen

## News (in Norwegian)

- ▶ Konungs skuggsjá ferdig annotert (19.09.2018).
- ▶ Menota på YouTube (16.08.2018).
- ▶ Egils saga Skalla-Grímssonar i Wolfenbüttelbók i arkivet (22.06.2018).
- ▶ Kalkars ordbok er på nettet med fulle sokemuligheter (31.03.2017).
- ▶ Fjorten nye tekster i arkivet (27.07.2016).
- ▶ Avtale mellom Menota og Universitetsbiblioteket i Bergen (02.06.2016).

## Text archive



Menota can now offer over 40 Medieval Nordic texts (around 1.7 million words), several of which are fully lemmatised. The majority of the texts are Old Icelandic or Old Norwegian, but there are also some Old Swedish texts. Old Danish texts as well as Latin ones (from the Nordic countries) are most welcome. A catalogue with full search facilities was opened on 29 June 2007.

The texts have been encoded on one or more levels. The most widely used level is the diplomatic level (as seen in many Arnamagnæan editions), while some texts have also been encoded on a facsimile level (i.e. in a very close transcription) and some on a normalised level (as in the Íslensk fornrit series). These levels are specified under Facs, Dipl and Norm in the catalogue. By clicking the file name in the second column of the catalogue, you will be able to read the text on one or more levels, depending on how many levels have been encoded.

## Foundation

Menota was established on 10 September 2001 at a meeting in Oslo between major Nordic institutions.

- ▶ Initial meeting, Oslo 10 September 2001

## Statutes

The Menota statutes were passed at the council meeting in Reykjavík on 6 September 2003.

- ▶ The Menota statutes (Norwegian text)

## Council

The council is an advisory body in which each participating institution is represented by one member. Different departments or sections at the same university are recognised as individual institutions. In addition to the permanent members, other representatives may join the meeting at the discretion of the board.

- ▶ Menota's council 2019-2021
- ▶ Menota's council 2016-2018
- ▶ Menota's council 2013-2015
- ▶ Menota's council 2010-2012
- ▶ Menota's council 2007-2009
- ▶ Menota's council 2004-2006
- ▶ Menota's council 2002-2003

## Notice of meeting

- ▶ Council meeting, Bergen 26.08.2019

# BECKETT ARCHIVE

SAMUEL BECKETT  
DIGITAL MANUSCRIPT PROJECT

KRAPP'S LAST TAPE / LA DERNIÈRE BANDE

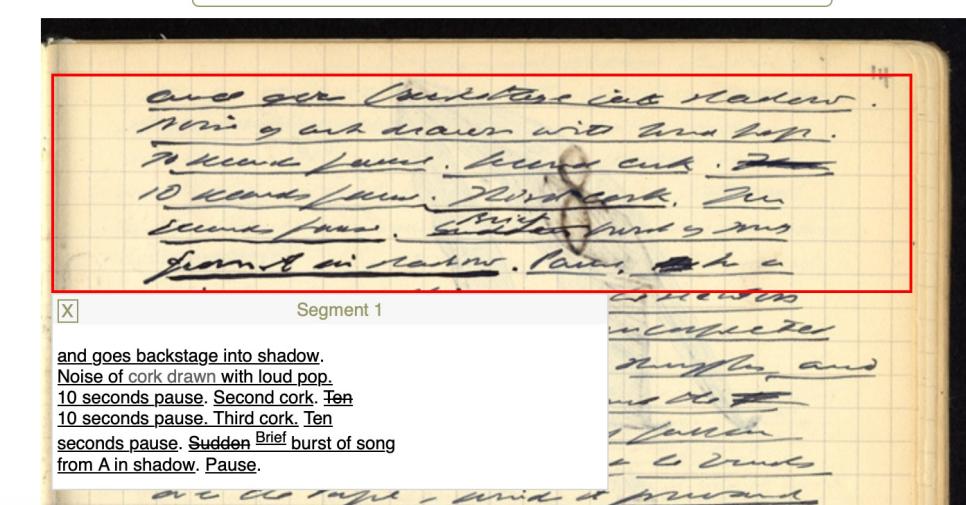


ABOUT ▾ DOCUMENTS ▾ TOOLS ▾ COMPARE SENTENCES



MS-UoR-1227-7-7-1

14r ▾



[www.beckettarchive.org](http://www.beckettarchive.org)

## And many more

- ▶ Base de Français Médiéval (<https://txm-bfm.huma-num.fr>)
- ▶ Le mariage sous L'Ancien Régime (<https://mariage.uvic.ca>)
- ▶ Newton Manuscript Project (<https://www.newtonproject.ox.ac.uk/>)
- ▶ etc.

# But not everyone

# THE WILLIAM BLAKE ARCHIVE

The WILLIAM BLAKE ARCHIVE GALLERY MODE

The Marriage of Heaven and Hell Copy G (Printed c. 1818)

Search IT G R ?

100 DPI 300 DPI

Diplomatic Transcription

Illustration Description

**Sector E  
Plate Illustration:**

At top right a female figure in a long swelling gown holds with her left hand to the upper branch of a leafy tree that extends the length of the right margin. With her right hand she seems to hand something, perhaps fruit, down to a second female figure, similarly gowned, who stands below her, holding a small object. Her arms crossed, she reaches upward with her right arm while supporting herself with her raised left arm around the trunk. At lower left a male (left) and possibly a female (right), both nude, lie on leafy vegetation. The figure to the left rests his head on his right hand and has his legs crossed. The female to the right lies with arm outstretched and bent at the elbow. Behind the right figure is the head and upper body of a third figure. Between the two groups of figures is a loopy vine with a few leaves. The colors of the sky

**The Argument.**

Rintrah roars & shakes his fires in the burdene air;  
Hungry clouds swag on the deep  
Once meek, and in a perilous path,  
The just man kept his course along  
The vale of death.  
Roses as planted where thorns grow,  
And on the barren heath  
Sing the honey bees.  
Then the prohbit was planted:  
A red rose, and a spring  
On every cliff and tomb;  
And on the bleaching bones  
Red clay brought forth.  
Till the villain left the paths of ease,

Object 3 (Bentley 3, Erdman 3, Keynes 3)

THE MARRIAGE OF HEAVEN AND HELL  
Copied from Copy G  
Erdman 1, Keynes 1

Rotate Magnify True Size Enlargement Diplomatic Transcription

Objects in Copy Objects from the Same Matrix Copy Information Electronic Edition Information

<http://www.blakearchive.org>

ELECTRONIC BEOWULF

Left/Top

Goto Edition 129r, 1-21 <> Fit frame

129r | HWÆT: WE GAR-DENA IN GEARDAGUM  
þeodcynninga þrym gefrunon.  
Hu ða æfelinges ellen fremedon!  
Oft Scyld Seeling sceafena þreatum  
5 monegum mægþum medosetla ofteah,  
egsode eorl, syððan ærest wearð  
feasceaft funden. He þes frofr gebad,  
weox under wolcnem, weorðmyndum þah,  
oð þat him aghwylc þara ymbsittendra  
10 ofer hronrade hyran scolde,  
gomban gyldan. Þat wæs god cyning.  
Dæm eafera was æfter cenned,  
geong in geardum, þone God sende  
folce to frofne. Efenðearfe ongeat.

Right/Bottom

Goto MS 129r, BL 132r <> Fit frame

129r | HƿEAT PE GARDE

þa mægþum lægum. heod cymmis  
þym se þrunon hūda æfelinges ellen  
fremedon. oft scyld sceafna sceafna  
monegum mægþum medo stel  
oð teah egsode eorl syððan ærest wearð  
feasceaft funden. He þes frofr gebad,  
weox under wolcnem, weorðmyndum þah,  
oð þat him aghwylc þara ymbsittendra  
10 ofer hronrade hyran scolde,  
gomban gyldan. Þat wæs god cyning.  
Dæm eafera was æfter cenned,  
geong in geardum, þone God sende  
folce to frofne. Efenðearfe ongeat.

<https://ebeowulf.uky.edu>

# TEI for Manuscript Cataloguing

Kapitan, Modelling humanities data with TEI-XML



## e-codices - Virtual Manuscript Library of Switzerland

The goal of e-codices is to provide free access to all medieval and a selection of modern manuscripts of Switzerland by means of a virtual library.

At the moment, the virtual library contains **2918 manuscripts** from **99** different collections.  
The virtual library will be continuously updated and extended.

# The Virtual Library of Switzerland: e-codices

► <https://www.e-codices.ch/en/about/metadata>



[Introduction](#) - [Technical Help](#) - [Repositories](#) - [Curated Collections](#) - [Search](#)

**Help us plan the future of OPenn!** We want to learn more about our users, your motivations research projects and goals, and we want to hear from you. Please fill out [this brief survey](#). It should take only 3-5 minutes of your time and will be invaluable to us. — The OPenn Team

[Home](#)

**OPenn: Read Me**

**University of Pennsylvania Libraries: OPenn**

[HTTPS://GITHUB.COM/SIMS-MSS/OPENN-XML](https://github.com/sims-mss/openn-xml)

# MEDIEVAL MANUSCRIPTS

IN OXFORD LIBRARIES

A catalogue of Western manuscripts at the Bodleian Libraries and selected Oxford colleges

Search Medieval Manuscripts



[ADVANCED SEARCH](#)

OR BROWSE BY

MANUSCRIPTS

WORKS

PEOPLE

PLACES

# Bodleian Library, University of Oxford

► <https://github.com/bodleian/medieval-mss>

## And many more

- ▶ FIHRIST, a union Catalogue of Manuscripts from the Islamicate World (<https://www.fihrist.org.uk>)
- ▶ Handrit, a union Catalogue of Icelandic manuscripts (<https://handrit.is>)
- ▶ Cambridge Digital Library (<https://cudl.lib.cam.ac.uk>)
- ▶ etc.

# But not everyone

- ▶ Many archives and libraries use **MARC (machine-readable cataloguing standard)**, maintained by the Library of Congress to create machine-readable description of items catalogued by libraries.
  - ▶ See: <https://www.loc.gov/marc/bibliographic/>
- ▶ Many archives and libraries (notably Calames and BnF) use **Encoded Archival Description (EAD)**, an XML standard for encoding archival finding aids, maintained by the Society of American Archivists.
  - ▶ EAD Documentation: <https://github.com/SAA-SDT/EAD3>

# TEI Structure

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
```



```
  <teiHeader>
```

metadata

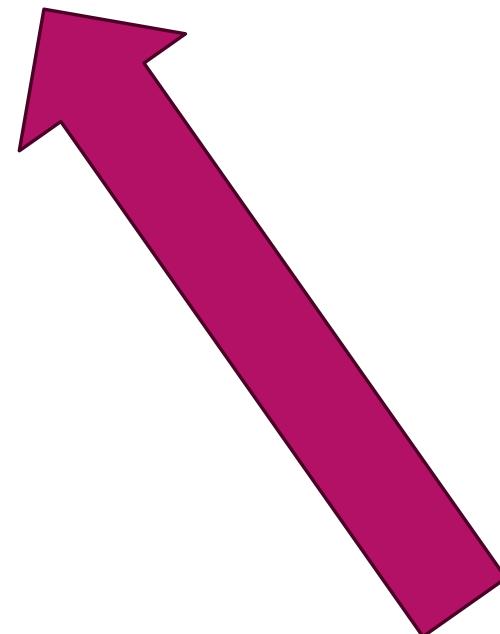
```
  </teiHeader>
```

```
  <text>
```

textual content

```
  </text>
```

```
</TEI>
```



*The root element of a TEI file is TEI (not XML) & it contains an attribute which specifies the TEI namespace*

# teiHeader

- ▶ **fileDesc** (file description) contains a full bibliographic description of an electronic file
- ▶ **encodingDesc** (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.
- ▶ **profileDesc** (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.
- ▶ **revisionDesc** (revision description) summarizes the revision history for a file.

# fileDesc

```
<fileDesc>
  <titleStmt>
    <title> [...] </title>
  </titleStmt>
</fileDesc>
```

- **titleStmt** (title statement) groups information about the title of a work and those responsible for its content

# fileDesc

<fileDesc>

```
<publicationStmt>  
  <p> [...] </p>  
</publicationStmt>
```

</fileDesc>

- **publicationStmt** (publication statement) groups information concerning the publication or distribution of an electronic or other text.

# fileDesc

```
<fileDesc>
  <sourceDesc>
    <p> [...] </p>
  </sourceDesc>
</fileDesc>
```

- **sourceDesc** (source description) describes the source(s) from which an electronic text was derived or generated (a bibliographic description in the case of a digitized text, or a phrase ‘born digital’).

# Example: teiHeader

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Thomas Paine: Common sense, a
          machine-readable transcript</title>
      <respStmt>
        <resp>compiled by</resp>
        <name>Jon K Adams</name>
      </respStmt>
    </titleStmt>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD7>

## Example: teiHeader

```
<!-- -->
<publicationStmt>
  <distributor>Oxford Text Archive.</distributor>
  <address>
    <addrLine>Oxford University Computing Services,</addrLine>
    <addrLine>13 Banbury Road,</addrLine>
    <addrLine>Oxford OX2 6RB,</addrLine>
    <addrLine>UK</addrLine>
  </address>
</publicationStmt>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD7>



```
<sourceDesc>
  <biblStruct>
    <monogr>
      <editor>Foner, Philip S.</editor>
      <title>The collected writings of Thomas Paine</title>
      <imprint>
        <pubPlace>New York</pubPlace>
        <publisher>Citadel Press</publisher>
        <date>1945</date>
      </imprint>
    </monogr>
  </biblStruct>
</sourceDesc>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD7>

```
<encodingDesc>
  <samplingDecl>
    <p>Editorial notes in the Foner edition have not
      been reproduced. </p>
    <p>Blank lines and multiple blank spaces, including paragraph
      indents, have not been preserved. </p>
  </samplingDecl>
  <editorialDecl>
    <correction status="high"
      method="silent">
      <p>The following errors
        in the Foner edition have been corrected:
      <list>
        <item>p. 13 l. 7 cotemporaries contemporaries</item>
        <item>p. 28 l. 26 [comma] [period]</item>
        <item>p. 84 l. 4 kin kind</item>
        <item>p. 95 l. 1 stuggle struggle</item>
        <item>p. 101 l. 4 certainy certainty</item>
        <item>p. 167 l. 6 than that</item>
        <item>p. 209 l. 24 publshed published</item>
      </list>
    </p>
  </correction>
```

```
<profileDesc>
  <creation>
    <date>1774</date>
  </creation>
  <langUsage>
    <language ident="en" usage="100">English.</language>
  </langUsage>
  <textClass>
    <keywords scheme="#lcsh">
      <term>Political science</term>
      <term>United States -- Politics and government –
          Revolution, 1775–1783</term>
    </keywords>
    <classCode scheme="#lc">JC 177</classCode>
  </textClass>
</profileDesc>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD7>



```
<revisionDesc>
  <change when="1996-01-22" who="#MSM"> finished proofreading </change>
  <change when="1995-10-30" who="#LB"> finished proofreading </change>
  <change notBefore="1995-07-04" who="#RG"> finished data entry at end of term </change>
  <change notAfter="1995-01-01" who="#RG"> began data entry before New Year 1995 </change>
</revisionDesc>
```

Source: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD7>

# Exercise 1

- Using the online version of the TEI Guidelines find out how to:
  - specify the language in which the XML document is written. Add the relevant element or attribute in a correct place.
  - add information about the license.
    - Let's assume that this file can be distributed under CC0 license; include a link to the deed  
(<https://creativecommons.org/publicdomain/zero/1.0/deed.e>)
- Find all the dates in the document
  - is it clear what they refer to?
  - how could the encoding of the dates be improved?

# Text

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
```

```
  <teiHeader>
    metadata
  </teiHeader>
  <text>
    textual content
  </text>
</TEI>
```



<**text**>

<**front**> </**front**>

<**body**> </**body**>

<**back**> </**back**>

</**text**>

**front (front matter)** contains any prefatory matter (headers, abstracts, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.

**body (text body)** contains the whole body of a single unitary text, excluding any front or back matter.

**back (back matter)** contains any appendixes, etc. following the main part of a text.

# Div and its attributes

**<div>** (text division) contains a subdivision of the front, body, or back of a text.

**@type** characterizes the element in some sense, using any convenient classification scheme or typology.

**@subtype** provides a sub-categorization of the element, if needed.

**@n** gives a number (or other label) for an element, which is not necessarily unique within the document.

# Div may contain other elements

**<p>** (paragraph) marks paragraphs in prose..

**<lg>** (line group) contains one or more verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

**<l>** (verse line) contains a single, possibly incomplete, line of verse.

**@met** (metrical structure, conventional) contains a user-specified encoding for the conventional metrical structure of the element.

**@rhyme** (rhyme scheme) specifies the rhyme scheme applicable to a group of verse lines.

# Div may contain milestones

**<gb>** (gathering beginning) marks the beginning of a new gathering or quire in a transcribed codex.

**<pb>** (page beginning) marks the beginning of a new page in a paginated document.

**<lb>** (line beginning) marks the beginning of a new (typographic) line in some edition or version of a text.

**<cb>** (column beginning) marks the beginning of a new column of a text on a multi-column page.

# Div: Example

```
<div type="letter">
  <opener>
    <dateline>
      <placeName>Rimaone</placeName>
      <date when="2006-11-21">21 Nov 06</date>
    </dateline>
    <salute>Dear Susan,</salute>
  </opener>
  <p>Thank you very much for the assistance splitting those
    logs. I'm sorry about the misunderstanding as to the size of
    the task. I really was not asking for help, only to borrow the
    axe. Hope you had fun in any case.</p>
  <closer>
    <salute>Sincerely yours,</salute>
    <signed>Seymour</signed>
  </closer>
  <postscript>
    <label>P.S.</label>
    <p>The collision occurred on <date when="2001-07-06">06 Jul 01</date>.</p>
  </postscript>
</div>
```

## Exercise 2

- ▶ Open Ex2\_Text.txt and Ex2\_Image.jpg (image & text of page 101 of the First Folio of Shakespeare's plays, Bodleian Arch. G c.7)
- ▶ In Oxygen Editor create a new TEI document using the TEI All template.
- ▶ Encode the text on Encode all relevant metadata in the teiHeader
- ▶ Compare your file with the official encoding of the First Folio of Shakespeare's plays  
<https://firstfolio.bodleian.ox.ac.uk/download/xml/F-ado.xml>