

Modelling humanities data with TEI-XML

SCHOLARLY EDITING AND MANUSCRIPT CATALOGUING IN THE DIGITAL AGE

Dr Katarzyna Anna Kapitan
5 November 2025

Feedback

Encoding Portfolio Part 1 and looking ahead

- Follow the instructions, if uncertain ask your teacher!
- Respect the word limits for ReadMe & samples!
- Three samples of the same type should be encoded in **separate XML files**, all following the **same schema**.
 - Not three pages of the same book, but **three pages from three different books**.
 - Don't submit entire books, only samples that you'll encode!
 - If uncertain, ask your teacher.

Feedback

Encoding Portfolio Part 1 and looking ahead

- Each XML file (only your template file in Part 1) must be correctly associated with, and validate properly against, the same schema:
 - A **locally stored, custom DTD** (for Part 1), not any TEI schema (!).
 - An **online TEI RNG** (for Part 2).
 - A **locally stored, custom RNG schema** (for Part 3).
- Your DTD must be valid, correctly formatted, and properly associated with your XML template file.
- Remember the **naming conventions**:
 - DTD file → file_name.dtd
 - XML file → file_name.xml
 - etc

Feedback

Encoding Portfolio Part 1 and looking ahead

- Your encoding should demonstrate **critical engagement** with both **metadata** and **content**.
 - Where are your sources coming from? Who encoded them ?
 - How your textual content can be encoded beyond <p> (paragraphs), what can this encoding be used for?
- Your encoding should use **elements**, **attributes**, and **attribute values beyond the examples presented in class**.
- You should **explain your encoding choices** — for example:
 - Why do you propose that the <text> element contain a single <body> element?
 - How is this structure useful for your materials?

Proposed revision for Part 2

Regardless the grade you've received for Part 1, if you submitted a TEI file and a DTD reflecting the default TEI structure, for Part 2 you must:

1. Submit a **DTD** which is not defining the TEI syntax & vocabulary but reflects your own way of thinking about encoding with XML (**Part 1**).
 1. Changing only the name of the root element is not enough.
 2. You can use comments describe what is happening in your files, if you feel like 300 words of ReadMe is not enough.
2. Submit an **XML template** which is associated with your DTD and validates (**Part 1**).
3. Submit **ReadMe file in markdown** describing your project (**Part 1/Part2**).
4. Submit **one sample** encoded in TEI (using TEI all) (**Part 2**).
5. Submit the **reflection note** comparing your initial DTD with TEI (**Part 2**)
6. Three samples : provide plain text transcriptions and photos/pdfs (**Part 2**).

From Transcribing To Digital Scholarly Editing

Kapitan, Modelling humanities data with TEI-XML

Left/Top

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

Right/Bottom

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

OPTIONS

129r | HWÆT: WE GAR-DENA IN GEARDAGUM
þeodcyninga þrym gefrunon.
Hu ða æbelingas ellen fremdon!
Oft Scyld Scefing sceafena þreatum
5 monegum mæghum meodosetla ofteah,
egsode eorl, syððan ærest wearð
feascaft funden. He þas frofre gebad,
weox under wolcnum, weorðmyndum þah,
oð þæt him æghwylc þara ymbsittendra
10 ofer hronrade hyran scolde,
gomban gyldan. Þæt wæs god cyning.
Dæm eafera wæs æfter cenned,
geong in geardum, þone God sende
folce to frofre. Fyrenðearfe ongeat.

Right/Bottom

MS 129r, BL 132r

129r | HWÆT PE GARDEN
na mæghum datum heod cymisal
þrym se Rwanon huda æbelingas ellen
fremdon. oft scyld scepinz sceches
hæatum monegum mæghum meodo scell
ofteah eorl siddan wære regn
e geba
um hal
cen dya
mbar
wæra þa
one god
re on
Lunge
a lær

Glossary

þeodcyninga, gp. 129r2:2; see also: [cyning](#).

þeoden, m., lord, prince, leader, king; ns. 132v16:129, 152v13:1045, 156v13:1211, *167v6:7:1718, 170v20:1874, 192v9:2870, 195v18:3039; **þioden**, 181v19:2337, 191r2:2811; **ðeoden**, 177r4:2134; **þeodnes**, gs. 147r14-15:796, 153v9:1084, 185v11:1629, 170r11:1840; **ðeodnes**, 149v13:909, *178v8:2177 (TN: usu. rest.); 189A(197)r1:2857; **þeodne**, ds. 137v6:345, 187r6:2573, 189r4:2710; **ðeodne**, *163v8:1527, 173v13:1995, *174v12:2035 (TN: MS **ðeoden**, usu. em.); **þeoden**, as. 129v11:34, 134v19:201, 137v13:353, 165v6:1600, *183r1:2385, *189r13-14:2722, 190v4:2787, 192r21:2884, 196r12:3081, 198r15:3143; **þioden**, 190v6:2789;

Electronic Beowulf

<https://ebeowulf.uky.edu>

Volume the First: Diplomatic Display | **Oxford Bodleian Library, MS. Don.e.7.**

Page:

3

[<< Prev](#)[Next >>](#)[View Head Note](#)[View facsimile in a separate window](#)

3

Charlotte, whose character was a willing:

:ness to oblige every one, when she returned

into the Country, brought her Freind the

wished-for Bonnet, & so ended this little

Charlotte, whose character was a willing:
ness to oblige every one, when she returned
into the Country, brought her Freind the
wished-for Bonnet, & so ended this little
adventure, much to the satisfaction of all
parties.¹³

Jane Austen's Fiction Manuscripts

<http://www.janeausten.ac.uk>

Members are welcome to join the network. Since its founding in 2001, the archive has been led by Gunn Eiríði 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 søkemuligheter (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.

Encoded on one or more levels. The most widely used level is the diplomatic level (as in the original manuscript). Some parts of the texts have also been encoded on a facsimile level (i.e. in a

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](#)

Medieval Nordic Text Archive

<https://menota.org>

OPEN ELECTRONIC PUBLISHING, LE BLOG EN ANGLAIS D'OPENEDITION

Meet the OpenEdition team! Autumn 2025
19 SEPTEMBRE 2025

EDCH Training – the Diamond open access online training platform launched by

Schema XML TEI OpenEdition

PAR JEAN-FRANÇOIS RIVIÈRE · PUBLIÉ 19 MAI 2013 · MIS À JOUR 15 SEPTEMBRE 2014



Le 10 avril 2013, OpenEdition a mis en ligne la version 1.1 du schéma XML W3C pour l'encodage TEI OpenEdition.

ARTICLE SUIVANT

OpenEdition recrute son-sa chargé-e de la qualité documentaire

ARTICLE PRÉCÉDENT

OpenEdition recrute un responsable de projets d'édition

Open Edition
<https://www.openedition.org>

Editing Primary Sources

[Facs]

[ms: fol. 2r]

- 1 a þíngínu ᛃ fík ú lígía mundzinn allr ᄂ
- 2 þar legg ek ú aðat þe iammít ᄂ è sa hút
- 3 ueiggía þeit è ag audþ ber eñ eþ þu
- 4 uillt è beriazt ú m̄ þa sktu ag allrí þiāh
- 5 eimtūi* þa þagnaði .m. ᄂ retz ú ú uini si
- 6 na ú holmgunguna hm̄ **fuaði** jórundz godí
- 7 è þarþtu at ea rad ú off ú þetta mal þetþu
- 8 matt þ uíta eþ þu berzt ú .k. at þu munt
- 9 bæði lata líþ Þeit è .k. uel þít. þt h̄ è
- 10 m̄ill ag sialgū fer zm̄ græknaztr þa kuad .m.

[Dipl]

[ms: fol. 2r]

- 1 a þíngínu *ok* *skal* *uid* lígía mundzinn allr *ok*
- 2 þar legg ek *uid* *annat* fe iammikit *ok* *eigi* sa huort.
- 3 ueiggía feit er af *audrum* ber enn ef þu
- 4 uillt *eigi* beriazt *uid* *mik* þa *skaltu* af allrí *fiarh.*
- 5 eimtunni* þa þagnaði *Mordur* *ok* *retz* *um* *uid* *uini* *si.*
- 6 na u holmgunguna honum **fuaradí** jórundz godí
- 7 *eigi* þarftu at eiga rad *uid* *off* *um* *þetta* mal *þvit* þu
- 8 matt þat uíta ef þu berzt *uid* *rut* at þu munt
- 9 bæði lata lifit *ok* feit er *ruti* uel *farit.* *þvit* *hann* *er*
- 10 mikill af sialfum fer *ok* *manna* *fræknaztr* þa kuad *Mordur*

[Norm]

[ms: fol. 2r]

- 1 á þinginu, ok skal við liggja mundrinn allr, ok
- 2 þar legg ek við annat fé jafnmikit, ok eigi sá hvárt.
- 3 veggja féit, er af qðrum berr. En ef þú
- 4 vilt eigi berjask við mik, þá skaltu af allri fjárh.
- 5 eimtunni* .” Þá þagnaði Mørðr ok rézk um við vini sí.
- 6 na um hólmgonguna. Honum **svaraði** Jórundr goði:
- 7 “Eigi þarftu at eiga ráð við oss um þetta mál, því at þú
- 8 mátt þat vita, ef þú bersk við Hrút, at þú munt
- 9 bæði láta lífit ok féit. Er Hrúti vel farit, því at hann er
- 10 mikill af sjálfum sér ok manna frœknastr.” Þá kvað Mørðr

Source: Kapitan, A fragment of Njáls saga, AM 162 B alfa fol.
(<https://clarino.uib.no/menota/text/menota/AM-162-B-alfa-fol>)

<choice>

<**sic**> (Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

<**corr**> (correction) contains the correct form of a passage apparently erroneous in the copy text.

<**reg**> (regularization) contains a reading which has been regularized or normalized in some sense.

<**orig**> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

<**unclear**> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

<**supplied**> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.

Choice

<p>

How are you doing

<choice>

<abbr>Dr<am>.</am></abbr>

<expan>D<ex>octo</ex>r</expan>

</choice>

Jones? </p>

Exercise 1: Transcription with choice

3 mss & 3 groups

- ▶ Use the transcription starter XML (Chapter 14 of *De excidio Troiae historia*)
- ▶ Include menota entities in your DOCTYPE
- ▶ Introduce *numbered* line breaks
- ▶ Introduce `<choice>` with `<am>` and `<ex>`
- ▶ Use at least three different character entities: e.g. `&bar;` and/or `&usbase` for abbreviation markers and `ı` and/or `&slong;` for special characters

x

Texts in Prose: Basic Structure

```
<text>
  <body>
    <div>
      <p>
        <choice>
          <am></am><ex></ex>
        </choice>
      </p>
    </div>
  </body>
</text>
```

Configure Transformation Scenario(s)

Type filter text

Association	Scenario	Type
▼ TEI P5 (5)	TEI P5 XHTML	XML with XSLT
	TEI P5 PDF	XML with XSLT
	TEI P5 EPUB	ANT
	TEI P5 DOCX	ANT
	TEI P5 ODT	ANT
▼ Global (1)	xmlstylesheet processing instruction	XML with XSLT

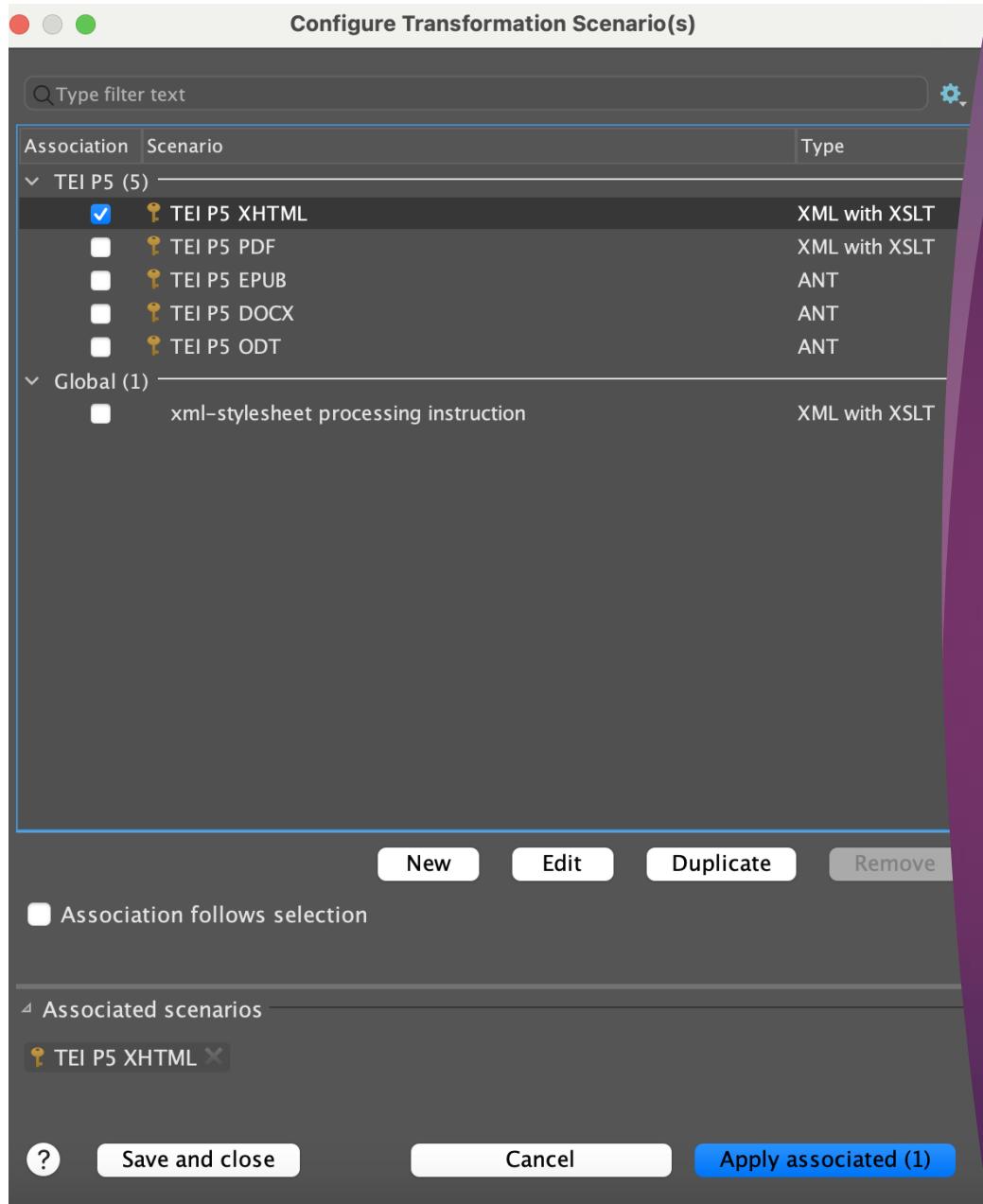
New Edit Duplicate Remove

Association follows selection

Associated scenarios

TEI P5 XHTML X

? Save and close Cancel Apply associated (1)



Default Transformation

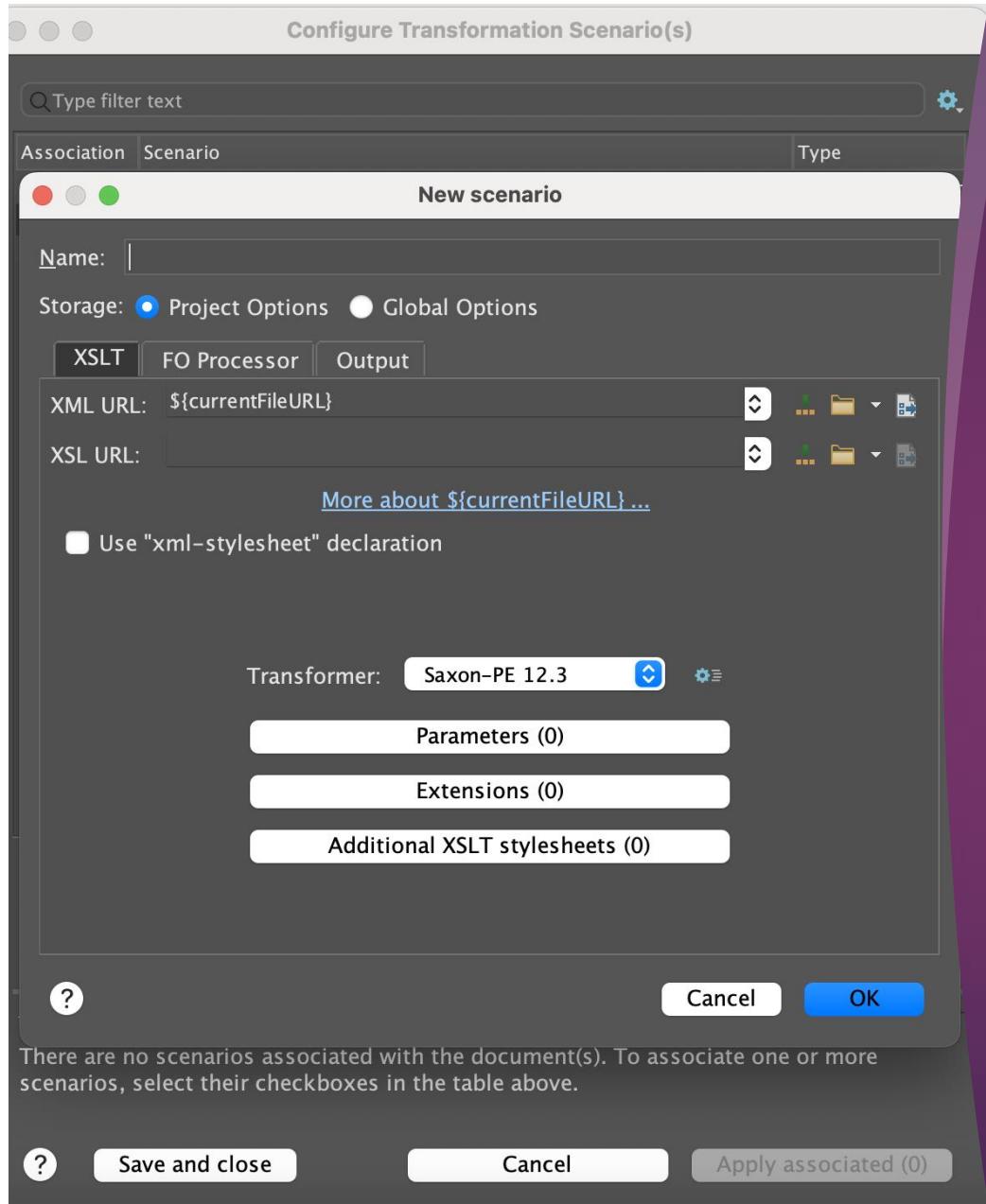
TEI P5 XML to XHTML with XSLT

- ▶ Download and install a MUFI compatible font, e.g. Junicode (<https://sourceforge.net/projects/junicode/files/junicode/junicode-1.002/>)
- ▶ Activate your new font in your browser, e.g. Firefox (see how: <https://support.mozilla.org/en-US/kb/change-fonts-and-colors-websites-use>)

The screenshot shows the MUFI (Medieval Unicode Font Initiative) website. The header is green with the word "MUFI" in white. The top right corner has a search icon. Below the header, there's a navigation menu with links: Home, Browse by character, Browse by code chart, Browse by range, Browse by updates, and Full code chart. The main content area features a large button labeled "Archived old MUFI site". To the right, there's a sidebar titled "subsections" with a list of 9 items, each with a right-pointing arrow icon:

- 1. Background
- 2. MUFI character recommendation
- 3. MUFI fonts
- 4. Medievalist proposals to Unicode
- 5. LINCIA
- 6. Board 2001–2015
- 7. Other Fonts & Projects
- 8. Links
- 9. Notes

The central content area also contains text about the character database and download options, as well as sections for "Disclaimer" and "Board 2016".



Custom Transformation

transf_dipl_basic.xsl & transf_facs_basic.xls
Guidelines_Transformation_Scenario.pdf