

DIGITALNO ZGODOVINOPISJE

Uvod v ustvarjanje digitalnih virov



REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO,
ZNANOST IN INOVACIJE



UNIVERZA
V LJUBLJANI



NAČRT ZA
OKREVANJE
IN ODPORNOST



Financira
Evropska unija
NextGenerationEU

Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo,
znanost in inovacije ter Evropska unija – NextGenerationEu.

Uvod v ustvarjanje digitalnih virov (1.dan)

- Digitalizacija virov
- Zgodovina
- Kaj je razširljiv označevalni jezik (eXtensible Markup Language) XML
 - Struktura in komponente XML
- Uvod v označevanje besedilnih virov
 - Kaj je Text Encoding Initiative / TEI
 - Nekatera orodja
 - Primeri zaključenih projektov, ki so uporabljali TEI XML (za različne cilje)
- Programska oprema



Uvod v ustvarjanje digitalnih virov (2.dan)

- Zapis metapodatkov v teiHeader (organizacija teiHeader, opis datoteke, opis kodiranja, opis profila, opis revizije)
- Označevanje
 - Proze
 - Dramskih tekstov
 - Primarnih virov
 - Imenskih entitet



- Digitizacija virov
 - Proces pretvarjanja analognega gradiva v digitalno obliko
 - Skeniranje, video pretvorbe, zvočni posnetki
 - Namen ustvariti točno digitalno kopijo
- Digitalizacija virov
 - Ne le pretvorba temveč tudi uporabo drugih orodij za ustvarjanje novih načinov analiz, obdelav ali dostopa do gradiva
 - Uporaba metapodatkov
 - Sprememba delovnega procesa



- Proces:
 - Skeniranje
 - optično prepoznavanje znakov
 - Pretvorba različnih formatov v xml
 - Uporaba programske opreme : Abbyy, tesseract
 - Slikovno gradivo – docx – TEI
 - PDF – docx – TEI
 - Popravljanje napak pri OCR
 - metapodatkovni opis



Zgodovina

- Začetki v 60ih letih
- Prvi poskusi s SGML (Standard Generalized Markup Language) jezikom
- S pojavom XML (1998) nastal nov označevalni jezik, ki velja še danes
- Leta 1987 prvi sestanek in ustanovitev TEI
 - Cilj: uveljaviti način označevanja virov oziroma humanističnih besedil/zapisov
 - Najprej označevanje s SGML, določili majhen nabor oznak
 - Več verzij – P#
 - Pomemben mejnik P3
 - Leta 1999 ustanovljen konzorcij TEI
 - 2001-2003 pretvorba iz SGML v XML sintakso
 - 2007 izdana priporočilna shema P5, ki velja še danes



Kaj je razširljiv označevalni jezik (eXtensible Markup Language) XML

- Označevalni jezik
- Enostaven, vesplošen
- XML je neodvisen od platforme, programske opreme in sistema
- Ni orientiran k prikazu podatkov, izgledu na zaslonu, ampak k njihovi strukturi in notranjim razmerjem.
- Je *metajezik* – jezik za opisovanje drugih jezikov



Kaj je razširljiv označevalni jezik (eXtensible Markup Language) XML

- Namenjen transportu in shranjevanju podatkov
 - Zapisani podatki, ki jih vsebuje xml, so lahko prenosljivi v druge programske jezike
- Dokument .xml je besedilna datoteka
- Ima hierarhično strukturo – pregledna zgradba



Kaj je razširljiv označevalni jezik (eXtensible Markup Language) XML

- XML je berljiv za ljudi in stroje
- Mogoče ga je ustvariti in urejati z enostavnimi urejevalniki besedil
- Omogoča izmenjavo podatkov med različnimi sistemi in platformami



Kaj je razširljiv označevalni jezik (eXtensible Markup Language) XML

- Je široko sprejet in standardiziran s strani World Wide Web Consortium (W3C)
- XML omogoča uporabniku tudi ustvarjanje lastnih značk
 - Fleksibilnost pri definiranju podatkov



Nekatera pravila

- Deklaracija XML

- `<?xml version="1.0" encoding="UTF-8"?>`
- V prvi vrstici
- Zapisana tudi vrsta kodiranja dokumenta (UTF-8)

- Struktura:

- Osnovni gradniki jezika so elementi, zaprti v oglatih oklepajih
- Atributi nudijo dodatne informacije o elementih. Zapisani so znotraj začetne značke

```
<element>Vsebina</element>
```

- Značke (angl. Tag)

- Začetna značka (angl. starting tag)
 - `<div>`
- Zaključna značka (angl. ending tag)
 - `<div>vsebina</div>`
- Prazna značka (angl. empty tag)
 - `<graphic url="#TEI"/>`

```
<element atribut="vrednost atributa">vsebina</element>
```



Nekatera pravila (2)

- XML element predstavlja vse od začetne do zaključne značke. Element lahko vsebuje :
 - Besedilo
 - Atribute
 - Druge elemente
 - Kombinacijo vsega naštetega
- Dokument mora biti dobro oblikovan
- slediti mora specifični sintaksi
- pravilno gnezdene in zaprte značke

```
<root>  
  <child>  
    <subchild>.....</subchild>  
  </child>  
</root>
```



Zakaj XML

- XML prednosti:
 - Poudarja opisno označevanje
 - Je neodvisen od strojne ali programske opreme
 - Je razširljiv – ne vsebuje vnaprej določenega nabora značk/oznak
- Dokument XML mora biti skladen z določeno sintakso
- XML bolj kot na prikaz podatkov daje poudarek njihovemu pomenu (!)
- V zapisu XML in z oznakami TEI je besedilo varno pred komercialnimi tehnološkimi spremembami, kar je velikega pomena za prosto mednarodno sodelovanje, trajno uporabnost besedil
- Na tak način označeno besedilo je uporabno na več načinov:
 - (kompleksnejših) prezentacij znanstveno-kritičnih izdaj
 - Avtomatsko generirane konkordance in leksikoni
 - Statistične analize



Najosnovnejši primer XML

- ```
<?xml version="1.0" encoding="UTF-8"?>
<note>
 <to>Janez</to>
 <from>Ana</from>
 <heading>Pozdrav</heading>
 <body>Lep pozdrav!</body>
</note>
```



# Kaj je TEI - Text Encoding Initiative

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

TEI



**DIGITALNO  
ZGODOVINOPISJE**



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE



UNIVERZA  
V LJUBLJANI



NAČRT ZA  
OKREVANJE  
IN ODPRANOST



Financira  
Evropska unija  
NextGenerationEU

Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo,  
znanost in inovacije ter Evropska unija – NextGenerationEU.

## Kaj je Text Encoding Initiative - TEI (1)

- TEI ni standard, temveč smernice, pravila in priporočila TEI skupnosti
- te smernice so produkt skupnosti raziskovalcev, predvsem s področja humanistike, družboslovja in jezikoslovja, ki so organizirani v konzorciju TEI (*TEI Consortium*)
- TEI konzorcij je neprofitna članska organizacija, ki skrbi za razvoj, objavo in vzdrževanje smernic za kodiranje besedila, dokumentiranega v smernicah TEI (*TEI Guidelines*)
- TEI smernice predstavljajo označevalni jezik in nabor značk, ki ga je predlagal konzorcij TEI, kot tudi njegovo spletno ali tiskano dokumentacijo.  
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>





## Kaj je Text Encoding Initiative - TEI (2)

- vsak raziskovalec mora imeti svobodo pri interpretaciji besedila
- množica elementov za kodiranje je prikazana v smernicah TEI
- Smernice TEI kot referenčni priročnik
- Skupaj 587 elementov in 274 atributov
- 21 modulov, ki združujejo med seboj povezane elemente



# Različni moduli

## 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 [Deprecations](#)
- Appendix H [Prefatory Notes](#)
- Appendix I [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 [Computer-mediated Communication](#)
- 10 [Dictionaries](#)
- 11 [Manuscript Description](#)
- 12 [Representation of Primary Sources](#)
- 13 [Critical Apparatus](#)
- 14 [Names, Dates, People, and Places](#)
- 15 [Tables, Formulæ, Graphics, and Notated Music](#)
- 16 [Language Corpora](#)
- 17 [Linking, Segmentation, and Alignment](#)
- 18 [Simple Analytic Mechanisms](#)
- 19 [Feature Structures](#)
- 20 [Graphs, Networks, and Trees](#)
- 21 [Non-hierarchical Structures](#)
- 22 [Certainty, Precision, and Responsibility](#)
- 23 [Documentation Elements](#)
- 24 [Using the TEI](#)

## TEI sourcecode

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

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



## TEI moduli (1):

- **Infrastruktura TEI (tei):** Definicija skupnih podatkovnih tipov in modularnih struktur razredov, ki se uporabljajo za definiranje elementov in atributov v drugih moduli.
- **Glava TEI (header):** Oprelitev elementov, ki sestavljajo glavo dokumentov TEI. Njegovi glavni deli zagotavljajo elemente za kodiranje podrobnih metapodatkov o bibliografskih vidikih elektronskih besedil, njihovem odnosu z izvirnim gradivom, iz katerega so morda izpeljana, nebibliografskih podrobnostih in popolni zgodovini revidiranja dokumentov.
- **Elementi, ki so na voljo v vseh dokumentih TEI (core):** Oprelitev elementov in atributov, ki se lahko pojavijo v katerem koli besedilu TEI, ne glede na zvrst besedila. Ti elementi zajemajo besedilne pojave, kot so odstavki, poudarjanje in citati, uredniške spremembe (označevanje napak, ureditve, dodatki), podatkovne strukture (imena, naslovi, datumi, številke, okrajšave), mehanizmi navzkrižnega sklicevanja, sezname, opombe, grafični elementi, bibliografske reference in odlomki verzov ali dram.
- **Privzeta struktura besedila (textstructure):** Oprelitev elementov in atributov, ki opisujejo strukturo besedil TEI, kot so sprednja stran in naslovne strani, telo besedila in hrbtna stran. Ti lahko vsebujejo nadaljnje delitve, po možnosti uvedene z naslovi, pozdravi, začetnimi formulami in/ali zaključene z zaključnimi formulami, zaključnimi pozdravi, končnim gradivom in prispevki.



## TEI moduli (2):

- **Znaki, glifi in načini pisanja (gaiji):** Opredelitev posebnih določb za predstavitev znakov, za katere ne obstaja standardizirana predstavitev (kot je definirana s strani Unicode Consortium).
- **Verz (verse):** Opredelitev posebnih elementov in atributov za namensko analizo verznihih materialov, kot so cezure, metrični sistemi, sheme rim in enjambementi.
- **Uprizoritvena besedila (drama):** Opredelitev specifičnih elementov in atributov za namensko analizo dramskih materialov. Ti vključujejo določbe za kodiranje določenih pojavov v sprednji in zadnji strani, kot so podrobnosti o predstavah, prologi, epilogi, dramatično okolje in sezname igralcev. Druge strukture, specifične za dramo, vključujejo govore in odrska navodila. Za multimedijske predstave so predvideni elementi za opis vsebine zaslona, koti kamere, napisi in zvok.
- **Prepisi govora (spoken):** Opredelitev elementov in atributov za (splošno) transkripcijo različnih vrst govorjenega gradiva. Ti zajemajo pojave, kot so izgovori, premori, neleksikalni zvoki, kretnje in premiki v glasovni kakovosti. Poleg tega so na voljo posebni elementi glave za opis vokalnega vira transkripcije.
- **Slovarji (dictionaries):** Opredelitev elementov in atributov za predstavitev slovarjev, z določbami za nestrukturirane in strukturirane slovarske sestavke (po možnosti združene). Slovarski vnosi so lahko strukturirani s številnimi posebnimi elementi, ki označujejo homonime, pomen, besedno obliko, slovnične informacije, definicije, citate, rabo in etimologijo.
- **Opis rokopisa (msdescription):** Definicija posebne glave in strukturnih elementov ter atributov za kodiranje rokopisnih virov. Elementi glave vključujejo določbe za podrobno dokumentacijo identifikacije rokopisa ali dela rokopisa, informacije o naslovu, vsebino, fizični opis, zgodovino in dodatne informacije. Namenski elementi besedila pokrivajo pojave, kot so gesla, dimenzije, heraldika, vodni žigi itd.
- **Predstavitev primarnih virov (transcr):** Opredelitev elementov in atributov za podrobno transkripcijo primarnih virov. Obravnavani pojavi so faksimile, zahtevnejši dodatki, črtanja, zamenjave in obnove, rokopis dokumentov, poškodbe izvirnega gradiva in nečitljivost besedila.
- **Kritični aparat (textcrit):** Opredelitev elementov in atributov za predstavitev (različnih različic besedil kot) znanstvenih izdaj, ki navajajo vse razlike med različicami v variantnem aparatu.



## TEI moduli (3):

- **Imena, datumi, osebe in kraji (namesdates):** Opredelitev elementov in atributov za podrobnejšo analizo imen oseb, organizacij in krajev, njihovih referentov (oseb, organizacij in krajev) ter vidikov časovnih analiz.
- **Tabele, formule, grafike in notni zapis (figures):** Definicija specifičnih elementov in atributov za podrobno predstavitev grafičnih elementov v besedilih, kot so tabele, formule, slike in glasbena notacija.
- **Jezikovni korpus (corpus):** Opredelitev elementov in atributov za kodiranje korpusov besedil, zbranih po posebnih kriterijih. Večina teh elementov velja za dokumentacijo teh meril vzorčenja in kontekstualnih informacij o besedilih, udeležencih in njihovem komunikacijskem okolju.
- **Povezovanje, segmentacija in poravnava (linking):** Opredelitev elementov in atributov za predstavitev kompleksnih sistemov navzkrižnih referenc med identificiranimi sidrišči v besedilih TEI. Priporočila so podana bodisi za linijske ali samostoječe referenčne mehanizme.
- **Enostavni analitični mehanizmi (analysis):** Definicija elementov in atributov, ki omogočajo povezovanje preprostih analiz in interpretacij z elementi besedila. Obravnavani so mehanizmi za reprezentacijo tako generičnih kot posebno jezikovnih analiz.
- **Funkcijske strukture (iso-fs):** Opredelitev elementov in atributov za konstruiranje kompleksnih analitičnih okvirov, ki jih je mogoče uporabiti za predstavitev specifičnih analiz v besedilih TEI.
- **Grafi, omrežja in drevesa (nets):** Definicija elementov in atributov za analitično predstavitev shematskih odnosov med vozlišči v grafih in grafikonih.
- **Gotovost, natančnost in odgovornost (certainty):** Opredelitev elementov za podrobno dodelitev gotovosti za kodiranje v besedilu TEI, kot tudi identifikacijo odgovornosti za ta kodiranja.
- **Elementi za dokumentacijo (tagdocs):** Definicija elementov in atributov za dokumentacijo sheme kodiranja, ki se uporablja v besedilih TEI. Ta modul nudi sredstva za definiranje elementov, atributov, elementov in razredov atributov, bodisi s spreminjanjem obstoječih definicij ali z ustvarjanjem novih.



## TEI najosnovnejši primer:

```
1 <?xml version="1.0" encoding="UTF-8"?> <!-- xml deklaracija in določitev kodiranja znakov -->
2 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="TEI"> <!-- korenski element vseh TEI dokumentov. atribut xmlns določa -->
3 <teiHeader><!-- obvezen TEI element, ki vsebuje metapodatke o dokumentu in ne vsebine -->
4 <fileDesc><!-- obvezen element, opis datoteke -->
5 <titleStmt>
6 <title>Naslov</title>
7 </titleStmt>
8 <publicationStmt>
9 <p>Informacije o publikaciji</p> <!-- <p> označuje navaden odstavek -->
10 </publicationStmt>
11 <sourceDesc>
12 <p>Informacije o viru</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text><!-- vsebuje vsebino našega obravnavanega besedila -->
17 <body><!-- glavni del dokumenta -->
18 <p>Transkribcija/delo na tekstu</p>
19 </body>
20 </text>
21 </TEI>
22
23
```





## TEI pomembno:

- Smernice in ne navodila
  - Možno različno kodiranje iste vsebine – raziskovalčeva svoboda
  - TEI zagotavlja obsežen nabor smernic za označevanje strukture in pomena
  - Prilagodljivost TEI sheme specifičnim projektnim potrebam
  - Olajša ustvarjanje, analizo in dolgoročno shranjevanje digitalnih virov



# Orodje za pretvorbo dokumentov TEI Garage

TEI < Text Encoding Initiative >

TEIGarage Conversion

Choose the file, upload images and press convert

Select file to convert: 2

Izberite datoteko Uvod v ustv...ih virov.docx

- Hide advanced options 2

☐ Convert text only, don't include any

☒ Download images located in the document

☒ Copy and rename all pictures

Conversion: Microsoft Word (.docx)

Language: en

Select conversion profile: jsi

Upload images: 2

Upload image files and .zip files containing images

Need to upload any images, as images included in the document file.

URL for conversion with current properties:

`https://teigarage.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/conversion?properties=<conversions><conversion index="0"><property id="oxgarage.getImages">true</property><property id="oxgarage.getOnlineImages">true</property><property id="oxgarage.lang">en</property><property id="oxgarage.textOnly">false</property><property id="pl.psnc.dl.ege.tei.profileNames">jsi</property></conversions>`

URL for conversion with default properties:

`https://teigarage.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/`

<https://teigarage.tei-c.org/>



DIGITALNO  
ZGODOVINOPISJE



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE



UNIVERZA  
V LJUBLJANI



NAČRT ZA  
OKREVANJE  
IN ODPOORNOST



Financira  
Evropska unija  
NextGenerationEU

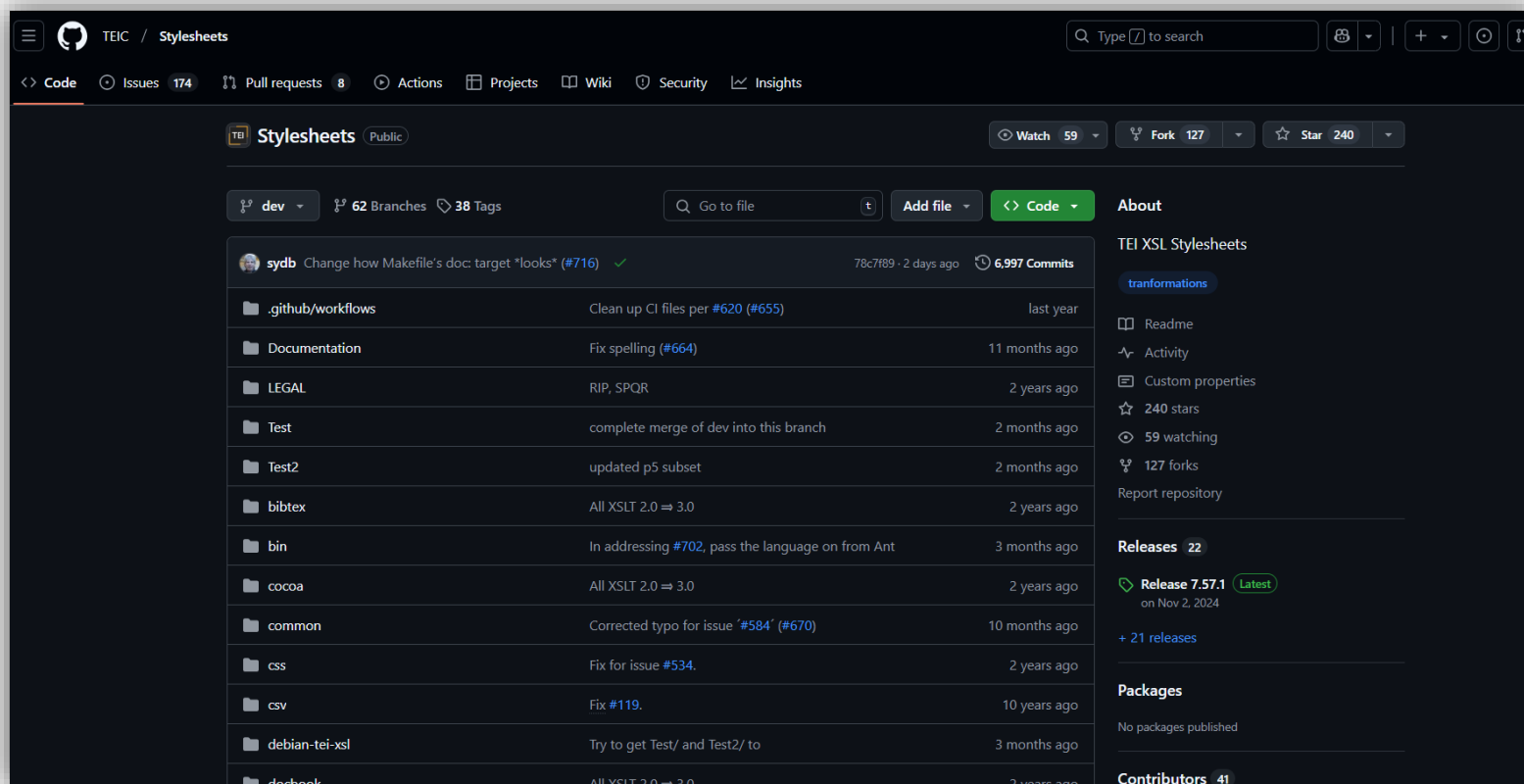
Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo,  
znanost in inovacije ter Evropska unija – NextGenerationEU.



# XSL Stylesheets za pretvorbo dokumentov TEI v različne formate

TEI vzdržuje knjižnico XSL Stylesheets (pretvorbe XSLT), ki lahko pretvori datoteke TEI XML v dokumente HTML, LaTeX XSL:FO ipd. Te pretvorbene datoteke so oblikovane za posebne namene in niso mišljene kot orodja za splošno pretvorbo.

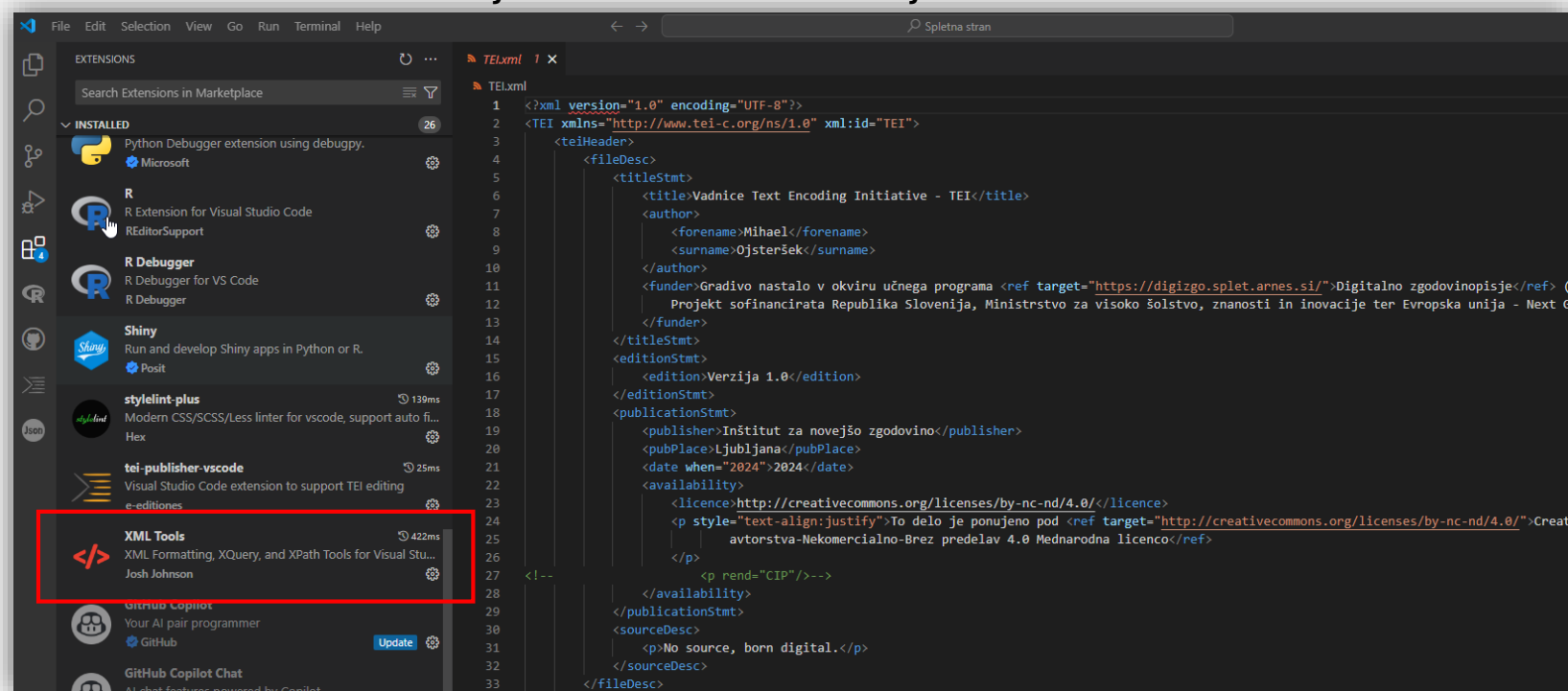
<https://github.com/TEIC/Stylesheets>



# Orodja za ustvarjanje in urejanje TEI XML

Brezplačna orodja Notepad++ in Visual Studio

Orodje VS Code z razširitvijo XML Tools:



<https://code.visualstudio.com/>



**DIGITALNO  
ZGODOVINOPISJE**



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE



UNIVERZA  
V LJUBLJANI



NAČRT ZA  
OKREVANJE  
IN ODPRANOST



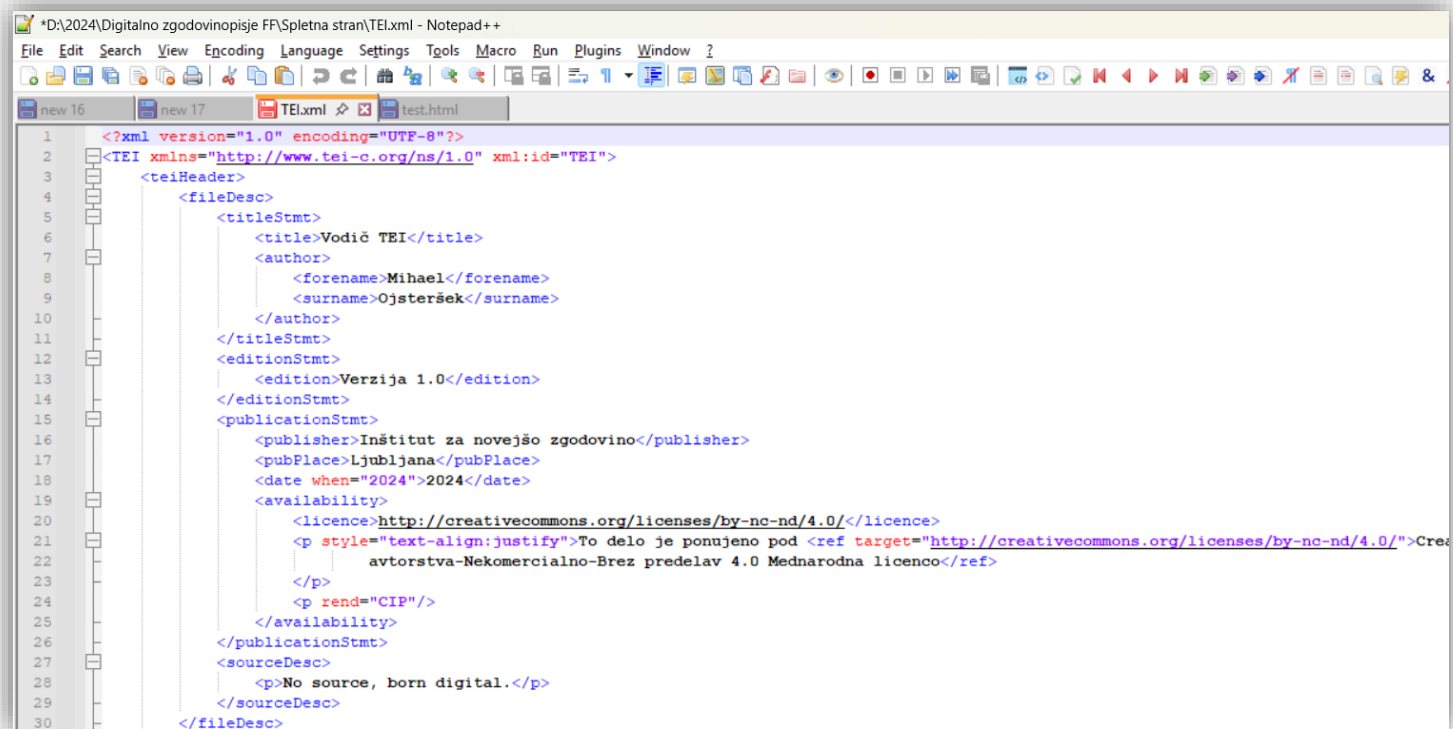
Financira  
Evropska unija  
NextGenerationEU

Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo, znanost in inovacije ter Evropska unija – NextGenerationEU.

# Orodja za ustvarjanje in urejanje TEI XML

Brezplačna orodja Notepad++ in Visual Studio

Orodje notepad++ z razširitvijo XML Tools:



The screenshot shows the Notepad++ text editor with the XML Tools extension installed. The editor displays a TEI XML document. The XML structure includes a header section with file description, title, author, edition, and publication information. The main content area contains a paragraph of text with a Creative Commons license reference. The XML Tools extension provides a sidebar on the left with a tree view of the document structure, allowing for easy navigation and editing of the XML elements.

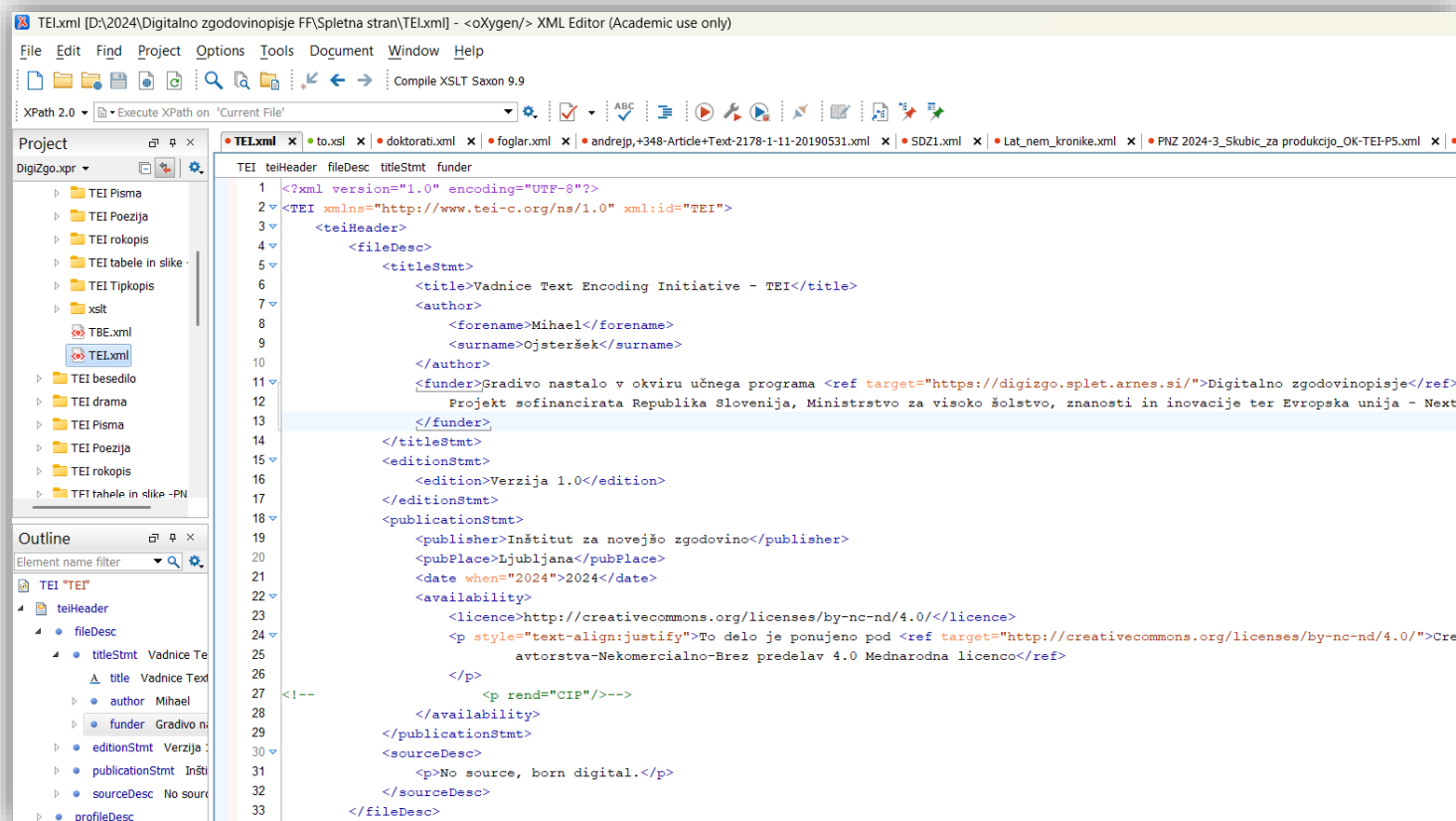
```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="TEI">
3 <teiHeader>
4 <fileDesc>
5 <titleStmt>
6 <title>Vodič TEI</title>
7 <author>
8 <forename>Mihael</forename>
9 <surname>Ojsteršek</surname>
10 </author>
11 </titleStmt>
12 <editionStmt>
13 <edition>Verzija 1.0</edition>
14 </editionStmt>
15 <publicationStmt>
16 <publisher>Inštitut za novejšo zgodovino</publisher>
17 <pubPlace>Ljubljana</pubPlace>
18 <date when="2024">2024</date>
19 <availability>
20 <licence>http://creativecommons.org/licenses/by-nc-nd/4.0/</licence>
21 <p style="text-align:justify">To delo je ponujeno pod <ref target="http://creativecommons.org/licenses/by-nc-nd/4.0/">Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License</ref> avtorstva-Nekomercialno-Brez predelav 4.0 Mednarodna licenco</p>
22 <p rend="CIP"/>
23 </availability>
24 </publicationStmt>
25 <sourceDesc>
26 <p>No source, born digital.</p>
27 </sourceDesc>
28 </fileDesc>
```

<https://notepad-plus-plus.org/>



# Orodja za ustvarjanje in urejanje TEI XML

## Lastniško orodje OxygenXML



<https://www.oxygenxml.com/>



**DIGITALNO  
ZGODOVINOPISJE**



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE



UNIVERZA  
V LJUBLJANI



NAČRT ZA  
OKREVANJE  
IN ODPOORNOST

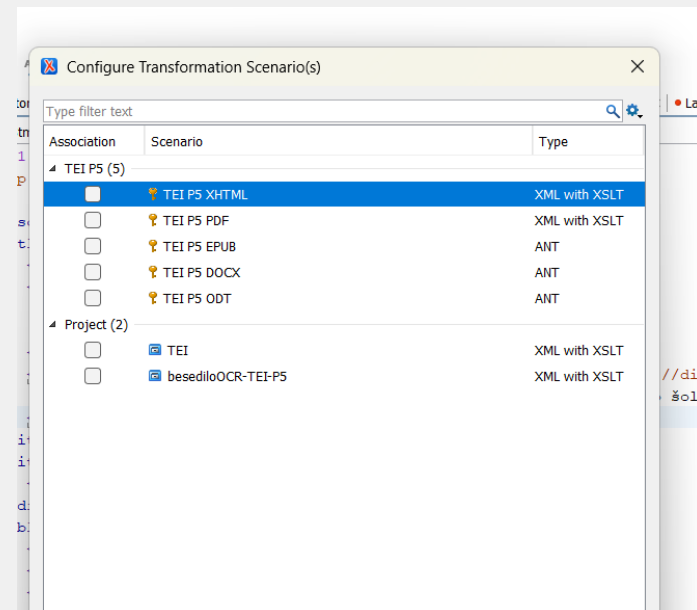
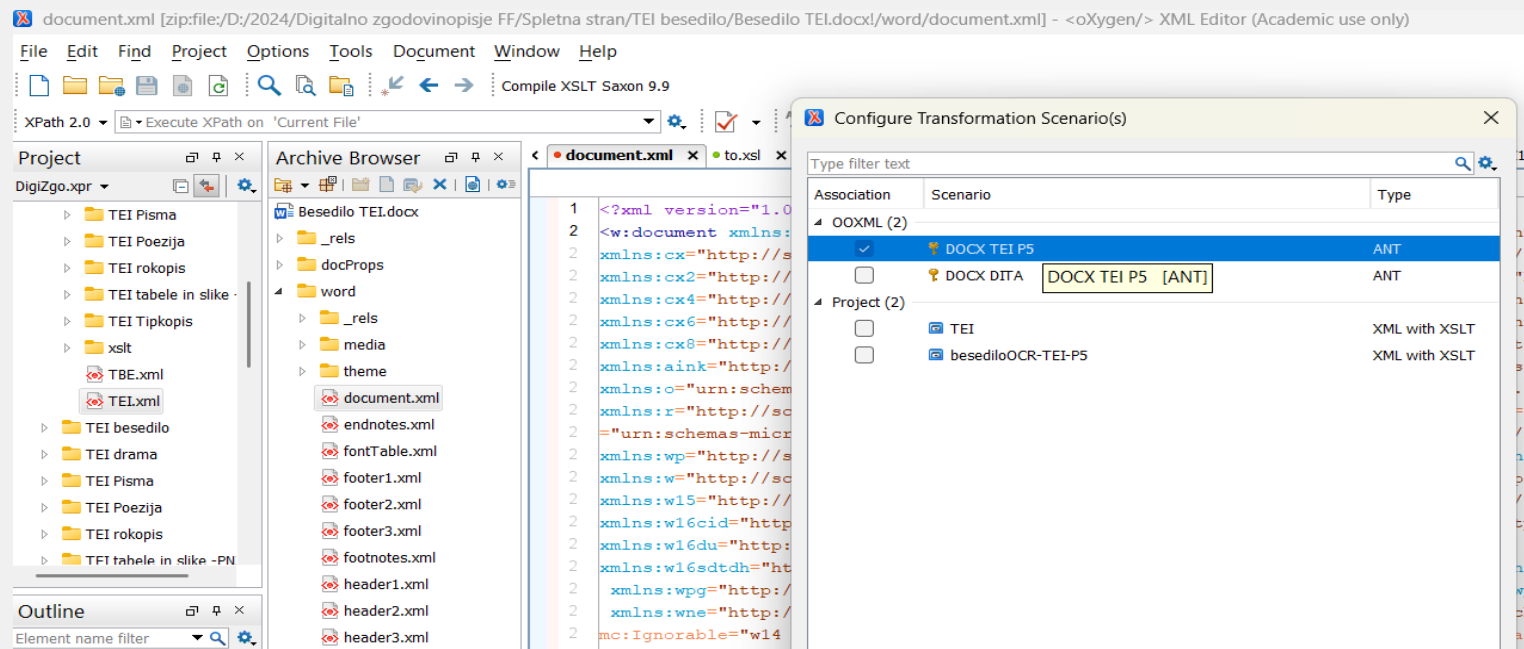


Financira  
Evropska unija  
NextGenerationEU

Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo,  
znanost in inovacije ter Evropska unija – NextGenerationEU.

# Orodja za ustvarjanje in urejanje TEI XML

## Lastniško orodje OxygenXML



Pri odpiranju dokumentov .docx ponudi pretvorbo v TEI P5

Avtomatsko pretvori  
označene dokumente v  
druge željene formate



# Orodja za ustvarjanje in urejanje TEI XML

Lastniško orodje OxygenXML

- Autofill
- Kratka razlaga TEI elementa
- Povezava do dokumentacije elementa

```
<div type="chapter" xml:id="franz.1837">
 <head>Leto 1837</head>
 <div xml:id="franz.1837.001" type="letter">
 <pb n="1837.01.01_b._100_9363" break="hr"/>
 <head>
 <hi rend="bold">1837/1</hi>
 </head>
 <head style="text-align:right;">Ljubljana, 1. januar 1837</head>
 <opener>
 <date>
 <pa>
 <dateline>
 </da>
 </pa>
 </date>
 </opener>
 <p style="text-align:right;">
 <hi>
 </hi>
 </p>
 <p style="text-align:right;">
 <hi rend="italic">Bar. Richard</hi> von Lazarini <hi rend="italic">wurde heute
 wegen seiner vorhabenden Vermählung - zum 2</hi>
 <hi rend="italic superscript">ten</hi>
 <hi rend="italic"> mal kirchlich verkündigt.</hi>
 </p>
 </div>
 </div>
```

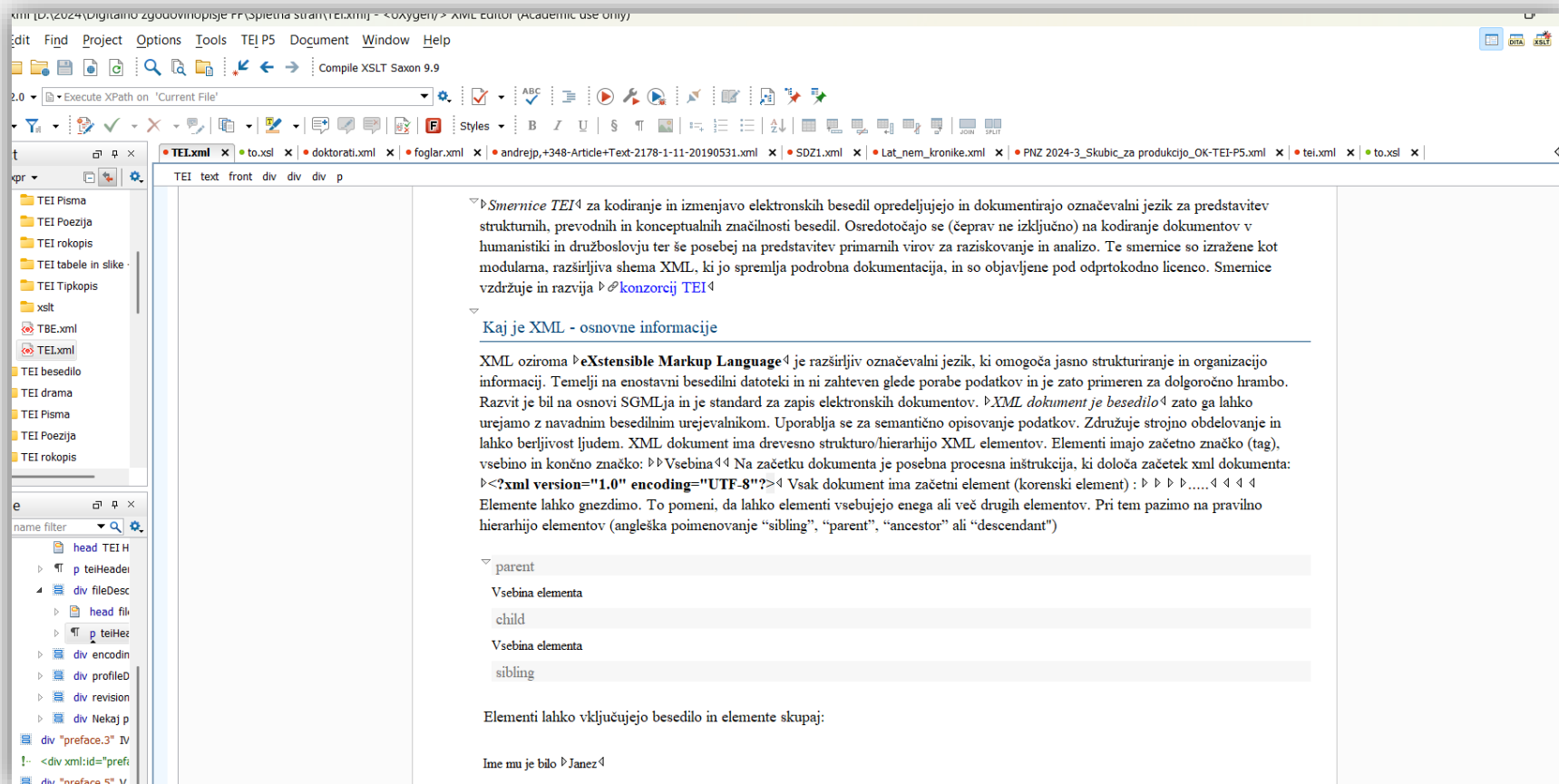
contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer. [4.2.2. Openers and Closers]  
TEI Guidelines





# Orodja za ustvarjanje in urejanje TEI XML

## Lastniško orodje OxygenXML – Author pogled



# Ustvarjanje digitalnih virov

- Zapis metapodatkov v teiHeader (organizacija teiHeader, opis datoteke, opis kodiranja, opis profila, opis revizije)
- Označevanje
  - Proze
  - Dramskih tekstov
  - Imenskih entitet





# teiHeader

- teiHeader je strukturirana komponenta xml dokumenta, ki vsebuje metapodatke o besedilu. Je ključen del dokumenta, saj omogoča razumljivost in uporabnost besedil.
  - Je glavni element za vse metapodatke, povezane z dokumentom
  - Je prvi element v TEI dokumentu
  - Vsebuje več podrejenih delov. Najpomembnejši so:
    - <fileDesc>
    - <encodingDesc>
    - <profileDesc>
    - <revisionDesc>



- <fileDesc>
  - Opis datoteke, vključno z informacijami o naslovu, avtorjih, izdajah itd.
  - Sestavljen iz več podrejenih elementov
    - <titleStmt> (Title Statement):
      - Vsebuje naslov dela in povezane informacije, kot so avtorji.
      - Elementi:
        - <title>: Naslov besedila.
        - <author>: Avtor ali avtorji besedila.
        - <editor>: Urednik besedila (če je relevantno).
    - <publicationStmt> (Publication Statement):
      - Informacije o publikaciji, kot so založnik, kraj in datum izdaje.
      - Elementi:
        - <publisher>: Izdajatelj besedila.
        - <pubPlace>: Kraj izdaje.
        - <date>: Datum izdaje.
    - <sourceDesc> (Source Description):
      - Opis vira, iz katerega je besedilo izpeljano.



- **<encodingDesc>** (Encoding Description):
  - opis kodiranja, označen z encodingDesc, opisuje razmerje med elektronskim besedilom in njegovim virom ali viri.
  - Omogoča podroben opis, ali (oziroma kako) je bilo besedilo normalizirano med transkripcijo, kako je označevalec razrešil dvoumnosti v izvoru, katere ravni kodiranja ali analize so bile uporabljene itd.
  - Pojasnilo o kodiranju besedila, uporabljeni metodologiji in standardih.
  - Opis kodiranja lahko vsebuje poljubno kombinacijo odstavkov besedila, označenih z elementom **<p>**, skupaj z bolj specializiranimi elementi kot npr.:
    - **<projectDesc>**: Opis projekta, ki je povezano z kodiranjem besedila.
    - **<editorialDecl>**: Uredniške smernice, ki so bile uporabljene



- <profileDesc> (Profile Description)
  - Vsebuje informacije o vsebini besedila, kot so jezik, slog in predmet.
  - Elementi:
    - <langUsage>: Opis uporabe jezika v besedilu.
    - <textClass>: Razvrstitev besedila po zvrsti, tematiki itd.



- <revisionDesc> (Revision Description):
  - Evidenca sprememb, ki so bile narejene v dokumentu
  - Pomembno za sledljivost sprememb
  - Element <change>: Podrobnosti o posamezni spremembi, vključno z datumom, avtorjem spremembe in opisom spremembe

# Primer teiHeader

Foglarjeva pesmarica,  
Elektronska znanstvenokritična izdaja

Meni

Foglarjeva pesmarica

Elektronska  
znanstvenokritična izdaja

Inštitut za slovensko literaturo in literarne  
vede ZRC SAZU  
Ljubljana

## Foglarjeva pesmarica

### Elektronska znanstvenokritična izdaja

Soustvarjalec: Nina Ditmajer (Diplomatični prepis, kritični aparat, kritični prepis, opombe in komentar)  
Nina Ditmajer (Uredila)  
Natalija Ulčnik (Strokovni pregled)  
Luka Vidmar (Strokovni pregled)  
Tomaž Erjavec (Redakcija zapisa TEI)  
Andrej Pančur (Pretvorba v XHTML, postavitve na splet)  
Marko Zaplatil (Fotografije rokopisa)  
Matija Ogrin (Pri urejanju sodeloval)

Distributer: DARIAH-SI

fe.si4\_field\_releaseAnnot: Izdaja za slovensko literaturo in literarne vede, ZRC SAZU

Datum: 2021-01-28

fe.si4\_field\_si4Behavior: izdaja

Vir: Elektronska izdaja je pripravljena po izvorniku, ki je predstavljen v digitalnem faksimilu in opisan v Registru slovenskih rokopisov 17. in 18. stoletja (NRSS, Ms 123).

Izdaja: Objava 1.0

Založnik: Založba ZRC, ZRC SAZU

Kraj izdaje: Ljubljana

Jezik: slv  
lat

Izveček: Foglarjev rokopis je baročna romarska pesmarica, ki je nastala na Spodnjem Štajerskem med letoma 1757 in 1762. To je doslej najstarejša znana rokopisna pesmarica, ki so jo napisali Štajerci, vsebuje tudi prve doslej znane slovenske marijinocejske pesmi. Elektronska znanstvenokritična izdaja rokopisa predstavlja z naslednjimi glavnimi elementi: digitalni faksimile predstavlja dokumentarni vir celotne izdaje; jamči za verodostojnost tako prepisov kot kritičnega aparata ter interpretacij, podanih v opombah in študiji; diplomatični prepis predstavi besedilo Foglarjevega rokopisa z vsemi historičnimi besedilnimi posebnostmi, vključno z napakami; diplomatičnemu prepisu je dodan kritični aparat, ki zajema variantna mesta nekaterih verzij teh pesmi, ohranjenih v osmih rokopisnih in tiskanih slovenskih pesmaricah 18. in 19. stoletja; kritični prepis predstavi besedilo rokopisa v prilagojeni, branju namenjeni podobi, kar najbolj zvesti domnevni historični izreki; je jezikoslovna interpretacija, pripravljena po pretehtanih jezikovnih načelih; dodatek vsebuje na prvem mestu seznam rokopisov ter tiskov, po katerih je narejen kritični aparat variantnih mest; sledi mu znanstveni komentar z razlago okoliščin nastanka rokopisa, jezikovnih in literarnih posebnosti ter uredniških načel prepisa.

Tema: pesmarice, rokopisi, barok, slovenska književnost, Lovrenc Foglar

Avtorske pravice: To delo je ponujeno pod Creative Commons Priznanje avtorstva-Nekomercialno-Brez predelav 4.0 Mednarodna licenco

Licenca: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Identifikator: <http://hdl.handle.net/20.500.12325/2>

fe.si4\_field\_si4Type: Intellectual entity

Vsebovane entitete

- Foglar, Lovrenc: [Foglarjeva pesmarica](#)



```

<fileDesc>
 <titleStmt>
 <title xml:lang="slv">Foglarjeva pesmarica</title>
 <title xml:lang="slv">Elektronska znanstvenokritična izdaja</title>
 <respStmt>
 <resp>Diplomatični prepis, kritični aparat, kritični prepis, opombe in komentar</resp>
 <name>Nina Ditmajer</name>
 </respStmt>
 <respStmt> [3 lines]
 <respStmt> [4 lines]
 <respStmt> [3 lines]
 <respStmt> [3 lines]
 <respStmt> [3 lines]
 <resp>Pretvorba v XHTML, postavitve na splet</resp>
 <name>Andrej Pančur</name>
 </titleStmt>
 <editionStmt>
 <edition>Objava 1.0</edition>
 </editionStmt>
 <publicationStmt>
 <publisher>Založba ZRC, ZRC SAZU</publisher>
 <authority>Inštitut za slovensko literaturo in literarne vede, ZRC SAZU</authority>
 <distributor>DARIAH-SI</distributor>
 <pubPlace>Ljubljana</pubPlace>
 <date when="2021-01-28">2021-01-28</date>
 <availability>
 <licence>http://creativecommons.org/licenses/by-nc-nd/4.0/</licence>
 <p>To delo je ponujeno pod
 <ref target="http://creativecommons.org/licenses/by-nc-nd/4.0/">Creative Commons Priznanje avtorstva-Nekomercialno-Brez predelav 4.0 Mednarodna licenco</ref>
 </p>
 </availability>
 </publicationStmt>
 <sourceDesc>
 <p>Elektronska izdaja je pripravljena po izvirniku, ki je predstavljen v digitalnem faksimilu in opisan v Registru slovenskih rokopisov 17. in 18. stoletja (NRSS, Ms 123).</p>
 </sourceDesc>
</fileDesc>

```



```

<profileDesc>
 <abstract xml:lang="slv">
 <p>Foglarjev rokopis je baročna romarska pesmarica, ki je nastala na Spodnjem Štajerskem med letoma 1757 in 1762.
 </abstract>
 <abstract xml:lang="eng">
 <p>Foglar's manuscript is a Baroque pilgrimage hymnal composed in Lower Styria between 1757 and 1762. It is the ol
 </abstract>
 <textClass>
 <keywords xml:lang="slv">
 <term>pesmarice</term>
 <term>rokopisi</term>
 <term>barok</term>
 <term>slovenska književnost</term>
 <term>Lovrenc Foglar</term>
 </keywords>
 <keywords xml:lang="eng">
 <term>hymnals</term>
 <term>manuscripts</term>
 <term>Baroque</term>
 <term>Slovenian literature</term>
 <term>Lovrenc Foglar</term>
 </keywords>
 </textClass>
 <langUsage>
 <language ident="slv">Slovenščina</language>
 <language ident="lat">Latinščina</language>
 </langUsage>
</profileDesc>

```





# Označevanje imenskih entitet (TEI modul za označevanje imen, datumov, krajev in ljudi)

- TEI omogoča več načinov označevanja imen in imenskih izrazov
  - <rs> vsebuje niz, ki se nanaša na osebo ali kraj, npr. "njegova gospa", "ona" itd.
  - <name> vsebuje lastni samostalnik ali samostalniško besedno zvezo, npr. „France Prešeren“ in „Vrba na Gorenjskem“
  - Z atributom @type razlikujemo osebe in kraje



# Označevanje imen

## Kodiranje imen z uporabo elementa <name>

špekulante.<note place="end" n="18">Uredba o ukrepih, ki se nanašajo na Žide glede izvrševanja obratov s predmeti ljudske prehrane. Službeni list banske uprave 1940, str. 862.</note> V zvezi z omenjeno uredbo je ljubljansko mestno poglavarstvo poročalo banski upravi, da so v <name>Ljubljani</name> tri židovske firme, ki so trgovale s predmeti ljudske hrane na debelo in sicer <name>Karel Bolaffio</name>, <name>Oskar Ebenspanger</name> ter "<name>Alko</name>", katere lastnika sta bila <name>Adolf</name> in <name>Oton Lorant</name>. Vsem trem firmam je bila nato obrtna pravica razveljavljena in obratovanje prepovedano.<note place="end" n="19">MALJ, Reg.



# Razlikovanje med različnimi vrstami imen

- Dodan atribut `@type`, za natančnejše označevanje

```
špekulante.<note place="end" n="18">Uredba o ukrepih, ki se nanašajo na žide
glede izvrševanja obratov s predmeti ljudske prehrane. Službeni list banske
uprave 1940, str. 862.</note> V zvezi z omenjeno uredbo je ljubljansko
mestno poglavarstvo poročalo banski upravi, da so v <name type="place"
>Ljubljani</name> tri židovske firme, ki so trgovale s predmeti ljudske
hrane na debelo in sicer <name type="person">Karel Bolaffio</name>, <name
type="person">Oskar Ebenspanger</name> ter "<name type="org">Alko</name>",
katere lastnika sta bila <name type="person">Adolf</name> in <name type="person"
>Oton Lorant</name>. Vsem trem firmam je bila nato obrtna pravica
razveljavljena in obratovanje prepovedano.<note place="end" n="19">MALJ, Reg.
```



# TEI modul 14 Names, Dates, People, and Places

- Ta TEI modul omogoča kodiranje imen in drugih besednih zvez, ki opisujejo osebe, kraje ali organizacije
- <persName> vsebuje lastno ime ali lastnoimensko besedno zvezo, ki se nanaša na osebo, po možnosti vključuje enega ali več osebnih imen, priimkov, častnih izrazov, dodanih imen itd. Element <persName> je „zmogljivejši“ od elementov <rs> in <name>, ker je mogoče komponente značilnega imena, ki se pojavljajo v njem, označiti kot take. Vsebuje lahko
  - <surname> vsebuje priimek oz. družinsko ime
  - <forename> vsebuje krstno oz. osebno ime<placeName>
- Z atributom @type lahko natančneje označimo vsebino



# Označevanje osebnih in krajevnih imen

špekulante.<note place="end" n="18">Uredba o ukrepih, ki se nanašajo na Žide glede izvrševanja obratov s predmeti ljudske prehrane. Službeni list banske uprave 1940, str. 862.</note> V zvezi z omenjeno uredbo je ljubljansko mestno poglavarstvo poročalo banski upravi, da so v

<placeName>Ljubljani</placeName> tri židovske firme, ki so trgovale s predmeti ljudske hrane na debelo in sicer <persName><forename>Karel</forename><surname>Bolaffio</surname></persName>, <persName><forename>Oskar</forename><surname>Ebenspanger</surname></persName> ter "Alko", katere lastnika sta bila <persName><forename>Adolf</forename></persName> in <persName><forename>Oton</forename><surname>Lorant</surname></persName>. Vsem trem firmam je bila nato obrtna pravica razveljavljena in obratovanje prepovedano.<note place="end" n="19">MALJ,



# Več komponent za osebna imena

- <roleName> sestavina imena, ki označuje vlogo ali položaj v družbi (npr. cesar, predsednik, general)
- <genName> generacijski ali dinastični podatki (npr. starejši, mlajši, II)
- <addName> dodatno ime, na primer vzdevek ali vzdevek (npr. Rdeči, Veliki)
- <nameLink> povezovalni izraz delov imena (npr. van, of, de)



# Imena in entitete, na katere se sklicujejo

- TEI razlikuje med oznakami v besedilu (pojavitev imena) in predmetom, na katerega se nanašajo (predmet v resničnem svetu).
- Tako se element:
  - `<person>` se ujema z `<persName>`
  - `<place>` se ujema z `<placeName>`
  - `<org>` se ujema z `<orgName>`
- Podrobne informacije o osebi so običajno shranjene v TEI header (`<teiHeader>`)



# Imena organizacij

- Element `<orgName>` vsebuje imena organizacij
- Organizacije so poimenovane skupine ljudi, ki se obravnavajo kot ena enota: univerzitetne ustanove, radijske in televizijske družbe, podjetja, rasne ali etnične skupine, politične skupine itd.

```
špekulante.<note place="end" n="18">Uredba o ukrepih, ki se nanašajo na Žide
glede izvrševanja obratov s predmeti ljudske prehrane. Službeni list banske
uprave 1940, str. 862.</note> V zvezi z omenjeno uredbo je ljubljansko
mestno poglavarstvo poročalo banski upravi, da so v
 <placeName>Ljubljani</placeName> tri židovske firme, ki so trgovale s
predmeti ljudske hrane na debelo in sicer <persName><forename>Karel</forename>
 <surname>Bolaffio</surname></persName>, <persName><forename>Oskar</forename>
 <surname>Ebenspanger</surname></persName> ter "<orgName>Alko</orgName>",
katere lastnika sta bila <persName><forename>Adolf</forename></persName> in
 <persName><forename>Oton</forename>
 <surname>Lorant</surname></persName>. Vsem trem firmam je bila nato obrtna
pravica razveljavljena in obratovanje prepovedano.<note place="end" n="19">MALJ,
```





# Podrobne informacije o osebi so običajno shranjene v TEI header

- `<person>` združi informacije o osebi.
- `<personGrp>` skupina posameznikov, ki se obravnavajo kot ena enota, npr. "občinstvo")
- `<listPerson>` vsebuje vrsto elementov `<person>` in `<personGrp>` ter je naveden znotraj elementa `<particDesc>` v elementu `<profileDesc>` v glavi TEI (`<teiHeader>`).



# Hvala za pozornost

Mail: [Mihael.Ojstersek@inz.si](mailto:Mihael.Ojstersek@inz.si)



DIGITALNO  
ZGODOVINOPISJE



DIGITALNO  
ZGODOVINOPISJE



REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE



UNIVERZA  
V LJUBLJANI



NAČRT ZA  
OKREVANJE  
IN ODORNOST



Financira  
Evropska unija  
NextGenerationEU

Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo,  
znanost in inovacije ter Evropska unija – NextGenerationEU.