

Modelling humanities data with TEI-XML

SCHOLARLY EDITING AND MANUSCRIPT CATALOGUING IN THE DIGITAL AGE

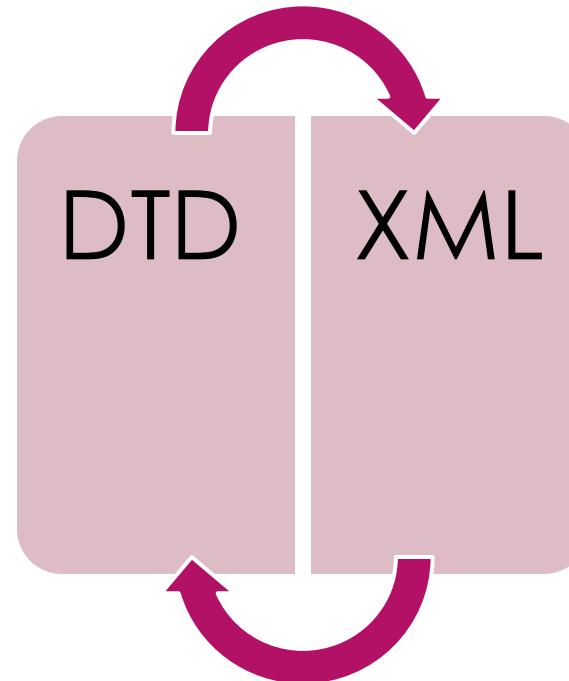
Dr Katarzyna Anna Kapitan
3 December 2024

Customisation and Documentation

Recap DTD

DTD

- Define a fixed set of elements, attributes and attribute values needed to encode information in **XML**
- Define relationship between these elements



XML

- Encode information in a consistent and structured way
- Validate against **DTD**

Other Schema Languages

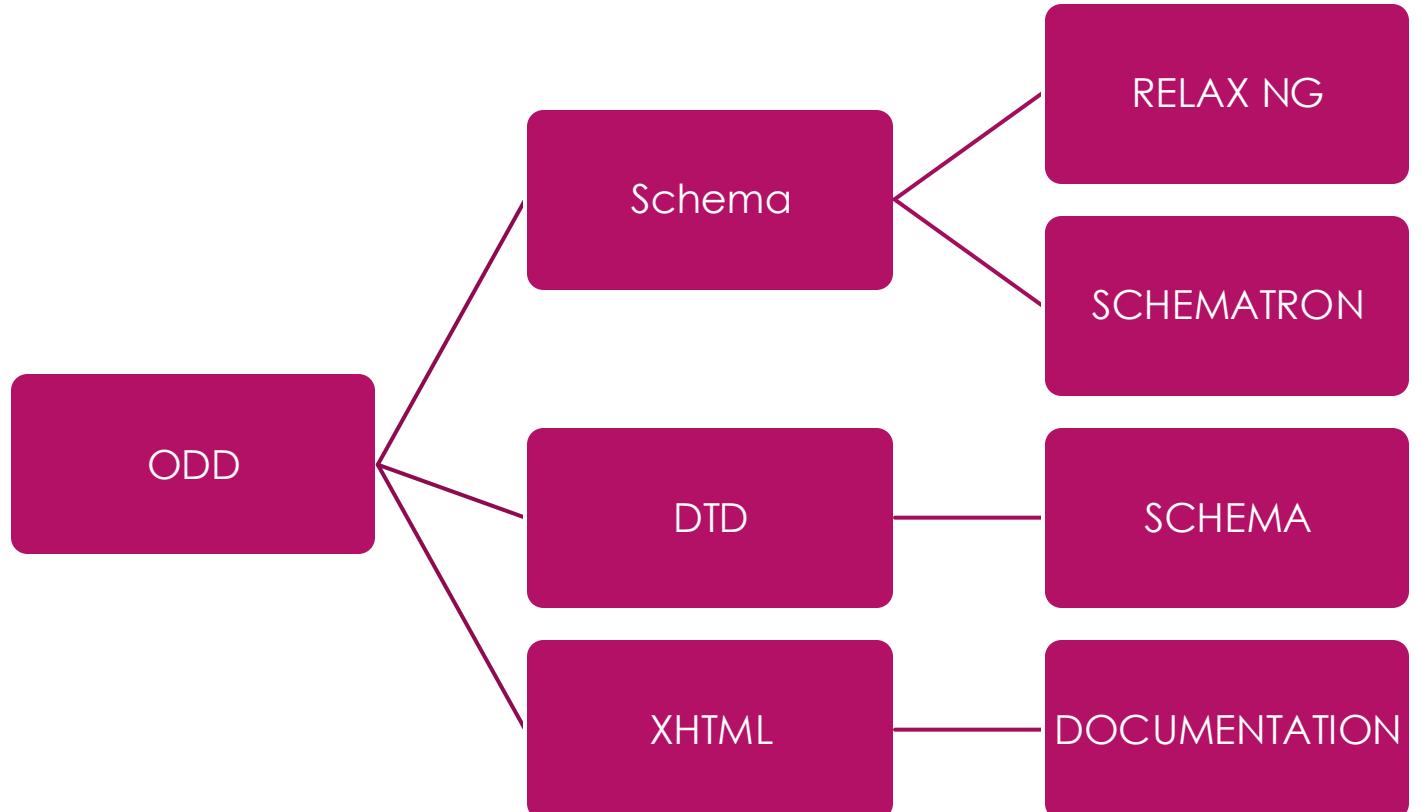
► Source: Syd Bauman, « A TEI customization for writing TEI customizations », Journal of the Text Encoding Initiative [Online], Issue 12 | July 2019 -, Online since 15 November 2019. URL: <http://journals.openedition.org/jtei/2573>

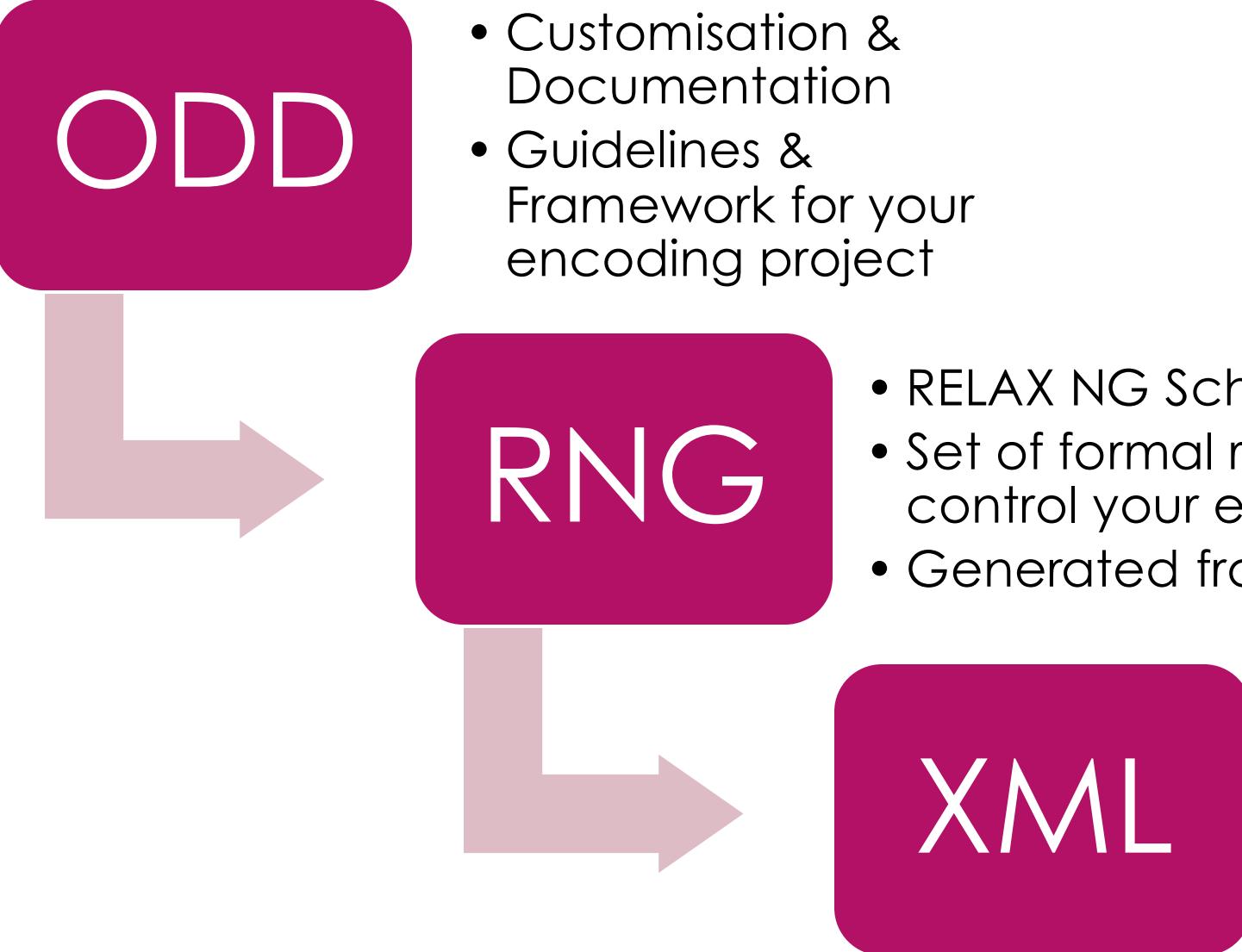
Table 1. Some of the schema languages for XML documents, arranged roughly by family of language.

| SGML Document Type Declaration Family | W3C Schema Language Family | Regular Expression Family | Others |
|---|--------------------------------------|-------------------------------------|-------------------|
| DTD | XML-Data | RELAX | DSD |
| XDTD | XDR | XDUCE | <i>Schematron</i> |
| DTD++ | DCD | TREX | Examplatron |
| DTD++ 2.0 | SOX | <i>RELAX NG</i> | X-definition |
| | DDML | | <i>TEI ODD</i> |
| | XSD | | |

TEI ODD (One Document Does it All)

- ▶ Read:
 - ▶ Bauman, Syd. 'Freedom to Constrain: where does attribute constraint come from, mommy?',
<https://www.balisage.net/Proceedings/vol1/html/Bauman01/BalisageVol1-Bauman01.html>
 - ▶ Syd Bauman, 'A TEI customization for writing TEI customizations',
<http://journals.openedition.org/jtei/2573>





ODD

- Customisation & Documentation
- Guidelines & Framework for your encoding project

RNG

- RELAX NG Schema
- Set of formal rules to control your encoding
- Generated from ODD

XML

- Your valid TEI-XML document with your consistently encoded data

Writing ODD files by hand

Let's bear in mind: An ODD file is just a specialized TEI file, with elements that say things like:

- “Include this element!”
- “Delete this element!”
- “Change this attribute!”
- “Replace the default definition with my custom definition!”
- ... and of course “Here’s what it means and why I did it!”



Introduction to Writing ODDs, slide 1 of 24

© 2010 Syd Bauman, Julia Flanders, and the Women Writers Project This TEI-encoded XML file is available under the terms of the Creative Commons Attribution-ShareAlike 3.0 (Unported) license.

How?

Chapter 24 of the Guidelines

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#IM-unified>

TEI ODD (One Document Does it All)

- ▶ ODD is a TEI-XML file.
- ▶ ODD **specifies** & **describes** the rules which the other XML documents in your project must obey.
- ▶ ODD consists of two parts: **teiHeader** for metadata, and **text** (with **body**) for content.
- ▶ The **body** element consists of two parts:
 - ▶ the first part is a description in prose of your customisation (in elements such as **head** and **p**)
 - ▶ the second part is schema specification that is included in **schemaSpec** element

```
<TEI xml:lang="en">
  <teiHeader> ... Metadata... </teiHeader>
  <text>
    <body>
      <head> Title of my project / customisation </head>
      <p> Description of my project and customisation </p>

      <schemaSpec ident="test" targetLang="en">
        Actual specification of your customization
      </schemaSpec>

    </body>
  </text>
</TEI>
```

```
<TEI xml:lang="en">
  <teiHeader> Metadata </teiHeader>
  <text>
    <body>
      <head> Title of my project / customisation </head>
      <p> Description of my project and customisation </p>
<schemaSpec ident="test" targetLang="en">
  Actual specification of your customization
</schemaSpec>

  </body>
  </text>
</TEI>
```

```
<TEI xml:lang="en">
  <teiHeader> Metadata </teiHeader>
  <text>
    <body>
      <head> Title of my project /
        customisation </head>
      <p> Description of my project
        and customisation </p>
      [...]
    </body>
  </text>
</TEI>
```

```
<text>
  <body>
    <!-- HERE STARTS THE DESCRIPTION OF OUR CUSTOMISATION -->
    <head>A TNAH TEI Customization</head>
    <p>This TEI ODD defines a TEI customization for TNAH. It includes the seven elements from the
      header module: <list rend="numbered">
        <item><gi>teiHeader</gi> from the "header" module to store required metadata</item>
        <item><gi>fileDesc</gi> from the "header" module to record information about this file</item>
        <item><gi>titleStmt</gi> from the "header" module to record information about the title</item>
        <item><gi>publicationStmt</gi> from the "header" module to detail how it is published</item>
        <item><gi>authority</gi> from the "header" module to specify who published it</item>
        <item><gi>availability</gi> from the "header" module to specify how the file can be
          reused</item>
        <item><gi>sourceDesc</gi> from the "header" module to record where it is derived from</item>
      </list> It also contains the entire "core" module and the entire "tei" module. Finally, it
      also includes almost everything from the "textstructure" module, but with a modification to
      exclude the front matter <gi>front</gi>, back matter <gi>back</gi>, and the numbered divs
      1-7 (<gi>div1</gi>, <gi>div2</gi>, <gi>div3</gi>, <gi>div4</gi>, <gi>div5</gi>,
      <gi>div6</gi>, <gi>div7</gi>). We rather want to use simple <gi>div</gi>s with
      <att>type</att> and encode everything in the <gi>body</gi>. </p>
```

A TNAH TEI Customization

This TEI ODD defines a TEI customization for TNAH. It includes the seven elements from the header module:

1. [`<teiHeader>`](#) from the "header" module to store required metadata
2. [`<fileDesc>`](#) from the "header" module to record information about this file
3. [`<titleStmt>`](#) from the "header" module to record information about the title
4. [`<publicationStmt>`](#) from the "header" module to detail how it is published
5. [`<authority>`](#) from the "header" module to specify who published it
6. [`<availability>`](#) from the "header" module to specify how the file can be reused
7. [`<sourceDesc>`](#) from the "header" module to record where it is derived from

It also contains the entire "core" module and the entire "tei" module. Finally, it also includes almost everything from the "textstructure" module, but with a modification to exclude the front matter `<front>`, back matter `<back>`, and the numbered divs 1-7 (`<div1>`, `<div2>`, `<div3>`, `<div4>`, `<div5>`, `<div6>`, `<div7>`). We rather want to use simple `<div>`s with `@type` and encode everything in the `<body>`.

```

51
52 <!-- HERE STARTS THE ACTUAL CUSTOMISATION -->
53 <schemaSpec ident="TNAH_customisation" targetLang="en">
54   <desc>A minimal customisation for the students of TNAH at ENC.</desc>
55   <moduleRef key="core"/>
56   <!-- This loads the entire "core" module, described in Chapter 3 of the Guidelines https://tei-c.org/release/doc/tei-p5-doc/en/html/C0.html -->
57
58   <moduleRef key="tei"/>
59   <!-- This load the entire "tei" module, described in Chapter 1 https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html -->
60
61 <moduleRef key="header"
62   include="teiHeader fileDesc titleStmt publicationStmt authority availability sourceDesc revisionDesc listChange change"/>
63   <!-- This loads only selected elements from the "header" module, described in Chapter 2 https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html -->
64
65   <!-- !!! Note that for the customisation from Exercise 1 to work you ALSO need to include "revisionDesc" "listChange" and "change" HERE !!! -->
66
67   <moduleRef key="textstructure" except="back front div1 div2 div3 div4 div5 div6 div7"/>
68   <!-- This loads the "entire" textstructure module, described in Chapter 4 https://tei-c.org/release/doc/tei-p5-doc/en/html/DS.html except back, front and numbered divs-->
69
70
71 <!-- START ELEMENT SPECIFICATION 1 -->
72   <!-- Here we replace the specification of <publicationStmt> from "header". -->
73 <elementSpec ident="publicationStmt" mode="replace" module="header">
74   <!-- Here we describe our element -->
75   <desc>[Note Change!] contains info about publication of this file. In our project <gi>publicationStmt</gi> must contain a sequence of
76     two elements: <gi>authority</gi> and <gi>availability</gi> as children. No other elements are allowed. </desc>
77   <content>
78     <!-- Here we define content of our element -->
79     <sequence><!-- We want a sequence of the elements listed below -->
80       <elementRef key="authority" minOccurs="1" maxOccurs="1"/><!-- We want exactly one authority -->
81       <elementRef key="availability" minOccurs="1" maxOccurs="1"/><!-- We want exactly one availability -->
82     </sequence>
83   </content>
84   <exemplum>
85     <!-- Here you create a minimal encoding example that uses this element -->

```

```

<TEI xml:lang="en">
  <teiHeader> Metadata </teiHeader>
  <text>
    <body>
      [...]
<schemaSpec ident="test" targetLang="en"> Actual specification of your customization </schemaSpec>
    </body>
  </text>
</TEI>

```

Description and Documentation

- **<desc>** - (description) contains a short description of the purpose, function, or use of its parent element
- **<gloss>** - (gloss) identifies a phrase or word used to provide a gloss or definition for some other word or phrase.
- **<gi>** - (generic identifier) contains the name of an element within running text
- **<att>** - (attribute) contains the name of an attribute within running text
- **<val>** - (value) contains a single attribute value.

```
<attDef ident="columns">
  <gloss xml:lang="en"
    versionDate="2007-06-12">columns</gloss>
  <desc versionDate="2005-01-14"
    xml:lang="en">specifies the number of columns per page</desc>
  <datatype minOccurs="1" maxOccurs="2">
    <dataRef key="teidata.count"/>
  </datatype>
  <remarks xml:lang="en"
    versionDate="2017-07-09">
    <p>If a single number is given, all pages referenced
      have this number of columns. If two numbers are given,
      the number of columns per page varies between the
      values supplied. Where <att>columns</att> is omitted
      the number is assumed to be <val>1</val>. </p>
  </remarks>
</attDef>
```

moduleRef

```
<moduleRef key="core"/>
<moduleRef key="tei"/>
<moduleRef key="header"/>
<moduleRef key="textstructure"/>
```

TEI Guidelines ▾ Activities ▾ Tools ▾ Membership ▾ Support ▾ About ▾ News Search Search ▾

TEI: Guidelines for Electronic Text Encoding and Interchange
P5 Version 4.8.1. Last updated on 1st November 2024, revision 0a2bff95a

<moduleRef>

Home C Elements

<moduleRef> (module reference) references a module which is to be incorporated into a schema. [23.2 Modules and Schemas]

| Module | tagdocs — Documentation Elements |
|------------|--|
| Attributes | <ul style="list-style-type: none">att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:spaceatt.global.rendition: @rend, @style, @renditionatt.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @selectatt.global.analytic: @anaatt.global.facs: @facsatt.global.change: @changeatt.global.responsibility: @cert, @respatt.global.source: @source |

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

TEI Modules

► Source: <https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ST.html>

| Module name | Formal public identifier | Where defined |
|---------------|--------------------------------------|--|
| analysis | Analysis and Interpretation | 18 Simple Analytic Mechanisms |
| certainty | Certainty and Uncertainty | 22 Certainty, Precision, and Responsibility |
| core | Common Core | 3 Elements Available in All TEI Documents |
| corpus | Metadata for Language Corpora | 16 Language Corpora |
| dictionaries | Print Dictionaries | 10 Dictionaries |
| drama | Performance Texts | 7 Performance Texts |
| figures | Tables, Formulae, Figures | 15 Tables, Formulae, Graphics, and Notated Music |
| gaiji | Character and Glyph Documentation | 5 Characters, Glyphs, and Writing Modes |
| header | Common Metadata | 2 The TEI Header |
| iso-fs | Feature Structures | 19 Feature Structures |
| linking | Linking, Segmentation, and Alignment | 17 Linking, Segmentation, and Alignment |
| msdescription | Manuscript Description | 11 Manuscript Description |
| namesdates | Names, Dates, People, and Places | 14 Names, Dates, People, and Places |
| nets | Graphs, Networks, and Trees | 20 Graphs, Networks, and Trees |

<div>

Module: textstructure

[Home](#)
[C Elements](#)

<div> (text division) contains a subdivision of the front, body, or back of a text. [\[Divisions of the Body\]](#)

| Module | textstructure — Default Text Structure |
|------------|---|
| Attributes | <ul style="list-style-type: none">• att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:space<ul style="list-style-type: none">◦ att.global.rendition: @rend, @style, @rendition◦ att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select◦ att.global.analytic: @ana◦ att.global.facs: @facs◦ att.global.change: @change◦ att.global.responsibility: @cert, @resp◦ att.global.source: @source• att.divLike: @org, @sample<ul style="list-style-type: none">◦ att.metrical: @met, @real, @rhyme◦ att.fragmentable: @part• att.typed: @type, @subtype• att.declaring: @decls• att.written: @hand |
| Member of | model.divLike |

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

| Module name | Formal public identifier | Where defined |
|--------------------|--------------------------------------|--|
| analysis | Analysis and Interpretation | 18 Simple Analytic Mechanisms |
| certainty | Certainty and Uncertainty | 22 Certainty, Precision, and Responsibility |
| core | Common Core | 3 Elements Available in All TEI Documents |
| corpus | Metadata for Language Corpora | 16 Language Corpora |
| dictionaries | Print Dictionaries | 10 Dictionaries |
| drama | Performance Texts | 7 Performance Texts |
| figures | Tables, Formulae, Figures | 15 Tables, Formulae, Graphics, and Notated Music |
| gaiji | Character and Glyph Documentation | 5 Characters, Glyphs, and Writing Modes |
| header | Common Metadata | 2 The TEI Header |
| iso-fs | Feature Structures | 19 Feature Structures |
| linking | Linking, Segmentation, and Alignment | 17 Linking, Segmentation, and Alignment |
| msdescription | Manuscript Description | 11 Manuscript Description |
| namesdates | Names, Dates, People, and Places | 14 Names, Dates, People, and Places |
| nets | Graphs, Networks, and Trees | 20 Graphs, Networks, and Trees |
| spoken | Transcribed Speech | 8 Transcriptions of Speech |
| tagdocs | Documentation Elements | 23 Documentation Elements |
| tei | TEI Infrastructure | 1 The TEI Infrastructure |
| textcrit | Text Criticism | 13 Critical Apparatus |
| textstructure | Default Text Structure | 4 Default Text Structure |
| transcr | Transcription of Primary Sources | 12 Representation of Primary Sources |
| verse | Verse | 6 Verse |

Chapter 1: The TEI Infrastructure (<https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html>)

What else is in the **textstructure** module?

TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.8.1. Last updated on 1st November 2024, revision 0a2bff95a

Table of contents

- 4.1 Divisions of the Body
- 4.2 Elements Common to All Divisions
- 4.3 Grouped and Floating Texts
- 4.4 Virtual Divisions
- 4.5 Front Matter
- 4.6 Title Pages
- 4.7 Back Matter
- 4.8 Module for Default Text Structure

« 3 Elements Available in All TEI Documents

» 5 Characters, Glyphs, and Writing Modes

Home

4 Default Text Structure

This chapter describes the default high-level structure for TEI documents. A full TEI document combines metadata describing it, represented by a [teiHeader](#) element, with the document itself, represented by one or more [text](#) elements or other elements taken from the [model.resource](#) class. That is, the [TEI](#) element is used to group together metadata about an encoded resource (in [teiHeader](#), specified by the [header](#) module, which is fully described in chapter [2 The TEI Header](#)) with an encoded resource. Possible encoded resources are

- a logical transcription of a source document in a [text](#) element; the [text](#) element is specified along with its high-level constituents in the [textstructure](#) module and described in the remainder of the current chapter
- a diplomatic transcription of a source document in a [sourceDoc](#) element, which is specified in the [transcr](#) module and described in chapter [12 Representation of Primary Sources](#)
- an encoded representation of a text-bearing object as images in a [facsimile](#) element, which is also specified in the [transcr](#) module and described in chapter [12 Representation of Primary Sources](#)
- a collection of contextual information or annotations that provides more detail about another encoded resource (whether in the same or a different TEI document) in a [standOff](#) element, which is specified in the [linking](#) module and described in section [17.10 The standOff Container](#)
- a feature system declaration which can be used to declare the use of [fs](#) elements in the rest of the document, which is specified in the [iso-fs](#) module and described in section [19.11 Feature System Declaration](#)

Chapter 4: Default Text Structure (<https://tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>)

Because the [TEI](#) can be a child of itself, a set of collection of documents may be represented by an outermost [TEI](#) element that contains a [teiHeader](#) with metadata that is applicable to the entire set of collection of transcriptions, and then a complete [TEI](#) element for each document in the collection or set; each of these [TEI](#) elements contains a [teiHeader](#) with metadata that is applicable to the individual document, and one or more [text](#) or other elements taken from the [model.resource](#) class.

Customisation: Elements

Trimming your modules

To delete specific elements from the modules you selected:

```
<moduleRef key="namesdates" except="addName affiliation bloc climate"/>
```

To include specific elements from the modules you selected:

```
<moduleRef key="textstructure" include="TEI text div back body"/>
```

Two important tips:

- you can't use both at once for the same <moduleRef>!
- be careful that the elements listed are actually in the module specified!



Customizing your elements: elementSpec

► <elementSpec

ident="Element_name"

mode="mode_of_change"

module="module_containing_this_element">

My specification

</elementSpec>

Mode = "delete" – delete this element

Mode = "add" – add this new element

Mode = "change" – change only these parts I am specifying

Mode = "replace" – replace everything with my specification

Customizing your elements: Description

```
<elementSpec ident="publicationStmt" mode="replace" module="header">  
  <desc>  
    contains info about publication of this file.  
    In our project <gi>publicationStmt</gi> must contain a sequence of two  
    elements: <gi>authority</gi> and <gi>availability</gi> as children.  
    No other elements are allowed.  
  </desc> [...]  
</elementSpec>
```

Customizing your elements: Content

```
<content>
  <sequence>
    <elementRef key="authority" minOccurs="1" maxOccurs="1"/>
    <elementRef key="availability" minOccurs="1" maxOccurs="1"/>
  </sequence>
</content>
```

Customizing your elements: Content

```
<elementSpec ident="availability" mode="change" module="header">  
    <content>  
        <textNode/>  
    </content>  
</elementSpec>
```

elementSpec in ODD

```
<!-- START ELEMENT SPECIFICATION 1 -->
<!-- Here we replace the specification of <publicationStmt> from "header".  -->
<elementSpec ident="publicationStmt" mode="replace" module="header">
<!-- Here we describe our element -->
<desc>[Note Change!] contains info about publication of this file. In our project <gi>publica</gi> two elements: <gi>authority</gi> and <gi>availability</gi> as children. No other elements are allowed.
<content>
<!-- Here we define content of our element -->
<sequence><!-- We want a sequence of the elements listed below -->
<elementRef key="authority" minOccurs="1" maxOccurs="1"/><!-- We want exactly one authority -->
<elementRef key="availability" minOccurs="1" maxOccurs="1"/><!-- We want exactly one availability -->
</sequence>
</content>
<exemplum>
<!-- Here you create a minimal encoding example that uses this element -->
<egXML xmlns="http://www.tei-c.org/ns/Examples">
<publicationStmt>
<authority>Katarzyna Anna Kapitan</authority>
<availability>Free to reuse under CC-BY 4.0 licence.</availability>
</publicationStmt>
</egXML>
</exemplum>
</elementSpec>
<!-- END ELEMENT SPECIFICATION 1-->
```

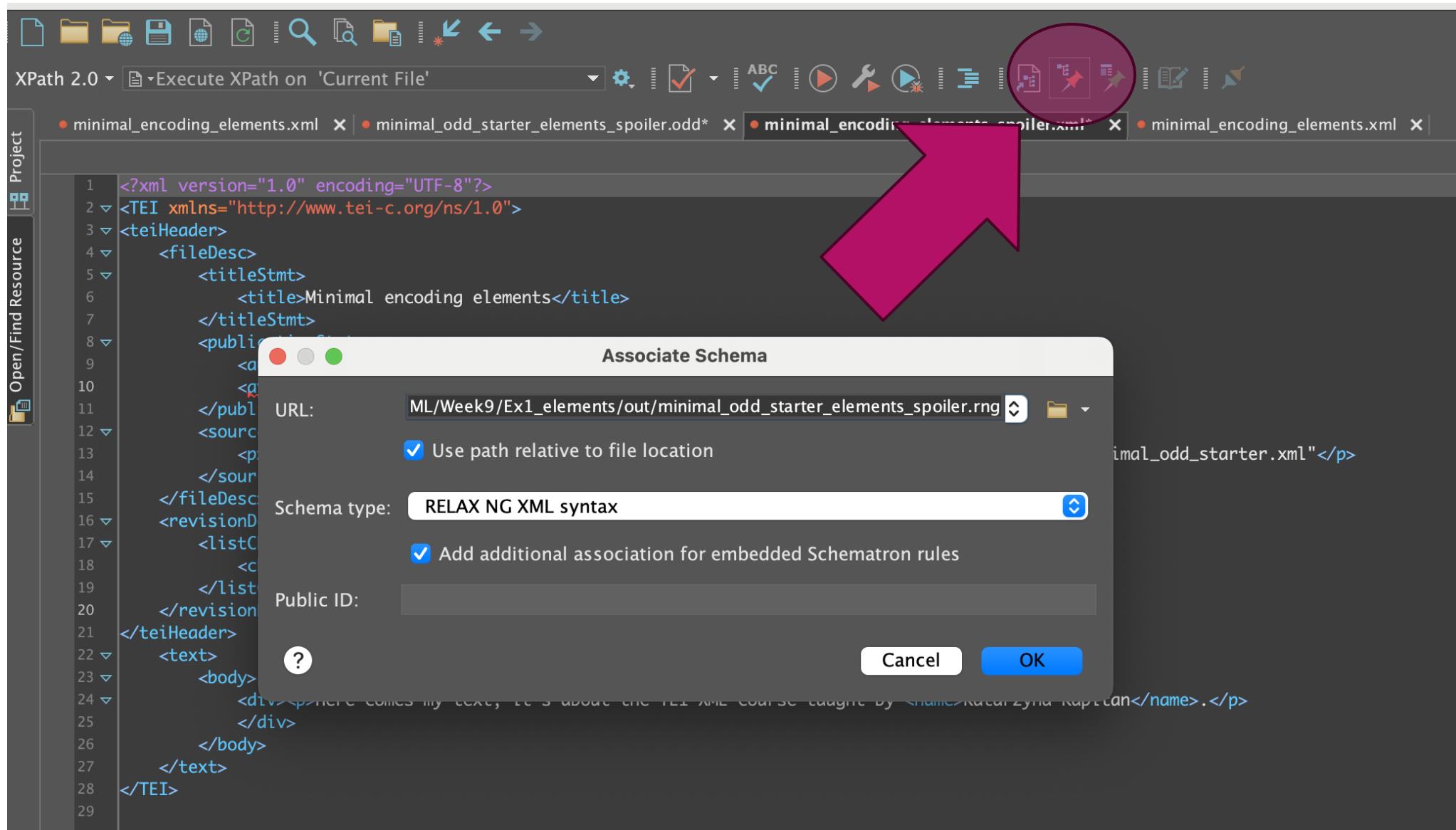
Representation of elementSpec in HTML

1.85. <publicationStmt>

| | |
|---|---|
| <publicationStmt> [Note Change!] contains info about publication of this file. In our project <publicationStmt> must contain a sequence of two elements: <authority> and <availability> as children. No other elements are allowed. | |
| Module | header |
| Contained by | header: fileDesc |
| May contain | header: authority availability |
| Example | <pre><publicationStmt> <authority>Katarzyna Anna Kapitan</authority> <availability>Free to reuse under CC-BY 4.0 licence.</availability> </publicationStmt></pre> |
| Content model | <pre><content> <sequence minOccurs="1" maxOccurs="1"> <elementRef key="authority" minOccurs="1" maxOccurs="1"/> <elementRef key="availability" minOccurs="1" maxOccurs="1"/> </sequence> </content> +-----+</pre> |
| Schema Declaration | <pre>element publicationStmt { authority, availability };</pre> |

Representation of elementSpec in RELAX NG

```
5165 <define name="publicationStmt">
5166   <element name="publicationStmt">
5167     <@documentation xmlns:ot="http://relaxng.org/ns/compatibility/annotations/1.0">[Note Change!] contains info about publication of this file. In our project <code> xmlns="http://www.tei-c.org/ns/Examples"</code> must contain a sequence of two elements: <code><authority></code> and <code><availability></code> as children. No other elements are allowed.
5168   <group>
5169     <ref name="authority"/>
5170     <ref name="availability"/>
5171   </group>
5172 </element>
5173 </define>
5174 <define name="authority">
5175   <element name="authority">
5176     <@documentation xmlns:ot="http://relaxng.org/ns/compatibility/annotations/1.0">(release authority) supplies the name of a person or other agency responsible for making a work available for reuse.
5177     <ref name="macro.phraseSeq.limited"/>
5178     <ref name="att.global.attributes"/>
5179     <ref name="att.canonical.attributes"/>
5180   </element>
5181 </define>
5182 <define name="availability">
5183   <element name="availability">
5184     <@documentation xmlns:ot="http://relaxng.org/ns/compatibility/annotations/1.0">[Note Change!] contains info about availability of the file. In our project <code> xmlns="http://www.tei-c.org/ns/Examples"</code> must contain a sequence of two elements: <code><authority></code> and <code><availability></code> as children. No other elements are allowed.
5185     <text/>
5186     <ref name="att.global.attributes"/>
5187     <ref name="att.declarable.attributes"/>
5188     <optional>
5189       <attribute name="status"/>
5190     </optional>
5191   </element>
5192 </define>
```



Exercise 1, Part 1

- ▶ Using the ODD file called **minimal_odd_starter_elements.odd** and pre-defined TEI transformation scenarios in Oxygen
 - ▶ generate XHTML guidelines (with TEI ODD XHTML)
 - ▶ generate RELAX NG schema (with TEI ODD to RELAX NG XML)
- ▶ Associate the RELAX NG schema with your XML file **minimal_encoding_elements.xml** and validate the XML file
 - ▶ Follow the steps described in: **Step_by_step_validation_in_Oxygen_XML_Editor.pdf**)
- ▶ Edit the XML file (**minimal_encoding_elements.xml**) so it validates with the RNG schema created from our ODD (**minimal_odd_starter_elements.odd**)
- ▶ **Consult the XHTML to decide which elements you must remove?**

Before associating a custom schema: 1 error (violating TEI All)

```
1 XML version='1.0' encoding='UTF-8'?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <title>Minimal encoding elements</title>
6     </titleStmt>
7     <publicationStmt>
8       <authority>Katarzyna Anna Kapitan</authority>
9       <availability>Free to reuse under CC-By 4.0.</availability>
10    </publicationStmt>
11    <sourceDesc>
12      <p><!-- Born digital --></p>
13    </sourceDesc>
14  </fileDesc>
15  <revisionDesc>
16    <listChange>
17      <change>
18        <!-- Info about my revision history -->
19      </change>
20    </listChange>
21  </revisionDesc>
22 </teiHeader>
23 <text>
24   <front>
25     <p><!-- S --></p>
26   </front>
27   <body>
28     <div><p><!-- A --></p>
29   </div>
30   </body>
31   <back>
32     <p><!-- A --></p>
33   </back>
</text>
```

element "availability" incompatible with element "fileDesc"

element "revisionDesc" not allowed anywhere; expected the element end-tag

After associating a custom schema: 9 errors (violating custom rules)

```
3 <TEI xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <title>Minimal encoding elements</title>
7     </titleStmt>
8     <publicationStmt>
9       <authority>Katarzyna Anna Kapitan</authority>
10      <availability>Free to reuse under CC-By 4.0.</availability>
11    </publicationStmt>
12    <sourceDesc>
13      <p><!-- Born digital, created on <date>28/11/2024</date> --></p>
14    </sourceDesc>
15  </fileDesc>
16  <revisionDesc>
17    <listChange>
18      <change><!-- Info about my revision history --></change>
19    </listChange>
20  </revisionDesc>
21 </teiHeader>
22 <text>
23   <front>
24     <p><!-- Some text --></p>
25   </front>
26   <body>
27     <div><p>Here comes my text, it's about the TEI XML course taught by <name>Katarzyna Kapitan</name>. --></p>
28   </div>
29   </body>
30   <back>
31     <p><!-- And even more text --></p>
32   </back>
</text>
```

After editing XML according to a custom schema: 0 errors

```
3 <TEI xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <title>Minimal encoding elements</title>
7     </titleStmt>
8     <publicationStmt>
9       <authority>Katarzyna Anna Kapitan</authority>
10      <availability>Free to reuse under CC-By 4.0.</availability>
11    </publicationStmt>
12    <sourceDesc>
13      <p><!-- Born digital, created on <date>28/11/2024</date> to test the ODD customisation in "minimal_odd_starter.xml" --></p>
14    </sourceDesc>
15  </fileDesc>
16
17 </teiHeader>
18 <text>
19   <body>
20     <div><p>Here comes my text, it's about the TEI XML course taught by <name>Katarzyna Kapitan</name>. --></p>
21   </div>
22   </body>
23 </text>
24
25 </TEI>
```

Elements not allowed:

- revisionDesc in teiHeader
- front, back, numbered divs in body
- persName in p

Exercise 1, Part 2

- ▶ Revise the ODD file called **minimal_odd_starter_elements.odd** so that it will **ONLY** allow for the following encoding of **revisionDesc** in the teiHeader of **minimal_encoding_elements.xml**:
- ▶

```
<revisionDesc><! Can have only one child called listChange>
  <listChange><! Can have multiple childred called change>
    <change><!– Takes only plain text, no elements allowed--></change>
  </listChange>
</revisionDesc>
```
- ▶ Generate RELAX NG schema & associate it with your XML file **minimal_encoding_elements.xml**
 - ▶ Follow the steps described in: **Step_by_step_validation_in_Oxygen_XML_Editor.pdf**)

- ▶ First include the elements when loading your modules.
- ▶ Elements **revisionDesc** **listChange** **change** belong to **header**, so:
 - ▶ <moduleRef key="header" include="teiHeader fileDesc titleStmt publicationStmt authority availability sourceDesc **revisionDesc** **listChange** **change**" />

- ▶ First include the elements when loading your modules.
- ▶ Elements **revisionDesc** **listChange** **change** belong to **header**, so:
 - ▶ <moduleRef key="header" include="teiHeader fileDesc titleStmt publicationStmt authority availability sourceDesc **revisionDesc** **listChange** **change**" />
- ▶ Then define the relationship between your elements:
 - ▶ <elementSpec ident="**revisionDesc**" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <elementRef key="**listChange**" minOccurs="1" maxOccurs="1"/></content> [...] </elementSpec>

- ▶ First include the elements when loading your modules.
- ▶ Elements **revisionDesc** **listChange** **change** belong to **header**, so:
 - ▶ `<moduleRef key="header" include="teiHeader fileDesc titleStmt publicationStmt authority availability sourceDesc revisionDesc listChange change" />`
- ▶ Then define the relationship between your elements:
 - ▶ `<elementSpec ident="revisionDesc" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <elementRef key="listChange" minOccurs="1" maxOccurs="1"/></content> [...] </elementSpec>`
 - ▶ `<elementSpec ident="listChange" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <elementRef key="change" minOccurs="1" maxOccurs="unbounded"/></content>
 [...] </elementSpec>`

- ▶ First include the elements when loading your modules.
- ▶ Elements **revisionDesc** **listChange** **change** belong to **header**, so:
 - ▶ <moduleRef key="header" include="teiHeader fileDesc titleStmt publicationStmt authority availability sourceDesc **revisionDesc** **listChange** **change**" />
- ▶ Then define the relationship between your elements:
 - ▶ <elementSpec ident="**revisionDesc**" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <elementRef key="**listChange**" minOccurs="1" maxOccurs="1"/></content> [...] </elementSpec>
 - ▶ <elementSpec ident="**listChange**" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <elementRef key="**change**" minOccurs="1" maxOccurs="unbounded"/></content>
 [...] </elementSpec>
 - ▶ <elementSpec ident="**change**" mode="replace" module="header">
 <desc>[...]</desc>
 <content>
 <textNode/>
 </content> [...] </elementSpec>

Customisation: Attributes

Changing attributes: Overview

Several ways for elements to possess attributes:

- By being a member of an attribute class
- By having attributes privately

Hence, several ways to remove attributes from elements:

- By deleting the attribute class from the schema altogether
- By deleting the attribute from the attribute class
- By unsubscribing the element from that attribute class
- By removing a privately held attribute from the element that owns it



<div>

[Home](#)
[C Elements](#)

<div> (text division) contains a subdivision of the front, body, or back of a text. [4.1 Divisions of the Body]

| | |
|-------------------|---|
| Module | textstructure — Default Text Structure |
| Attributes | <ul style="list-style-type: none">• att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:space<ul style="list-style-type: none">◦ att.global.rendition: @rend, @style, @rendition◦ att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select◦ att.global.analytic: @ana◦ att.global.facs: @facs◦ att.global.change: @change◦ att.global.responsibility: @cert, @resp◦ att.global.source: @source• att.divLike: @org, @sample<ul style="list-style-type: none">◦ att.metrical: @met, @real, @rhyme◦ att.fragmentable: @part• att.typed: @type, @subtype• att.declaring: @decls• att.written: @hand |
| Member of | model.divLike |



Attributes

grouped by attribute classes

Attribute classes

- ▶ A group of attributes that can occur in the same place in a TEI document.
- ▶ The names of attribute classes all start with **att.**, followed by a name that gives an indication of the group of attributes it contains.
- ▶ **att.global** defines **global attributes**
- ▶ **@xml:id** and **@n** are defined in the attribute class **att.global**
- ▶ Source:
<https://teibyexample.org/exist/tutorials/TBED08v00.htm>

att.global

[Home](#)
[B Attribute Classes](#)

att.global provides attributes common to all elements in the TEI encoding scheme. [1.3.1.1 Global Attributes]

| | |
|----------------|--|
| Module | tei — The TEI Infrastructure |
| Members | TEI ab abbr abstract accMat acquisition activity actor add addName addSpan additional additions addrLine address adminInfo affiliation age alt altGrp altIdent altIdentifi annotation annotationBlock anyElement app appInfo application arc argument att attDef attList attRe back bibl biblFull biblScope biblStruct bicond binary binaryObject binding bindingDesc birth bloc body broadcast byline c cRefPattern caesura calendar calendarDesc camera caption case castGroup castItem castList catDesc catRef catchwords category cb cell certainty change channel charDecl choice cit citeData citeStructure citedRange cl classCode classDecl classRef classSpec classes climate closer code collation collection colloc colophon cond condition constitution constraint constraintSpec content conversion corr correction correspAction correspContext correspDesc country creation custEvent custodialHist damage damageSpan dataFacet dataRef dataSpec datatype date dateline death decoDesc decoNote def default defaultVal delSpan depth derivation desc dictScrap dim dimensions distinct distributor district div div1 div2 div3 div4 div5 div6 div7 divGen docAuthor docDate docEdition docImprint docTitle domain eLeaf eTree edition editionStmt editor editorialDecl education eg egXML |

| | | | | | | | |
|-------------------------|--|---------------|----------|-----------------|------------------------------|-------------|---|
| Attributes | <ul style="list-style-type: none"> att.global.analytic: @ana att.global.change: @change att.global.fac: @fac att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select att.global.rendition: @rend, @style, @rendition att.global.responsibility: @cert, @resp att.global.source: @source | | | | | | |
| @xml:id | (identifier) provides a unique identifier for the element bearing the attribute. <table border="1"> <tr> <td>Status</td><td>Optional</td></tr> <tr> <td>Datatype</td><td>ID</td></tr> <tr> <td>Note</td><td>The @xml:id attribute may be used to specify a canonical reference for an element; see section 3.11 Reference Systems.</td></tr> </table> | Status | Optional | Datatype | ID | Note | The @xml:id attribute may be used to specify a canonical reference for an element; see section 3.11 Reference Systems . |
| Status | Optional | | | | | | |
| Datatype | ID | | | | | | |
| Note | The @xml:id attribute may be used to specify a canonical reference for an element; see section 3.11 Reference Systems . | | | | | | |
| @n | (number) gives a number (or other label) for an element, which is not necessarily unique within the document. <table border="1"> <tr> <td>Status</td><td>Optional</td></tr> <tr> <td>Datatype</td><td>teidata.text</td></tr> <tr> <td>Note</td><td>The value of this attribute is always understood to be a single token, even if it</td></tr> </table> | Status | Optional | Datatype | teidata.text | Note | The value of this attribute is always understood to be a single token, even if it |
| Status | Optional | | | | | | |
| Datatype | teidata.text | | | | | | |
| Note | The value of this attribute is always understood to be a single token, even if it | | | | | | |

Source:
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.html>

TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.8.1. Last updated on 1st November 2024, revision 0a2bff95a

Table of contents

[Appendix B.1 About the Attribute Classes Appendix](#)

[« Appendix A Model Classes](#)

[» Appendix C Elements](#)

[Home](#)

Appendix B Attribute Classes

⚓ **Appendix B.1 About the Attribute Classes Appendix**

This appendix gives you a list of attribute classes and links to the reference pages for them. There are 84 distinctly-named attribute classes in revision [0a2bff95a](#) of TEI P5 [Version 4.8.1](#) of the TEI Guidelines.

Sorted alphabetically

[a](#) [b](#) [c](#) [d](#) [e](#) [f](#) [g](#) [h](#) [i](#) [l](#) [m](#) [n](#) [p](#) [r](#) [s](#) [t](#) [w](#) [Show all](#)

[Show by module](#)

a

[att.anchoring](#) [att.ascribed](#) [att.ascribed.directed](#)

b

[att.breaking](#)

c

[att.calendarSystem](#) [att.canonical](#) [att.citeStructurePart](#) [att.citing](#) [att.cmc](#) [att.combinable](#) [att.coordinated](#)

Source: <https://tei-c.org/release/doc/tei-p5-doc/en/html/REF-CLASSES-ATTS.html>

Changing attributes: Examples

To delete an entire attribute class from the schema (scorched earth approach):

```
<classSpec type="atts" ident="att.personal" mode="delete"/>
```

To delete a single attribute from an attribute class (selective thinning):

```
<classRef key="att.global" except="xml:base xml:space"/>
```

To unsubscribe an element from an attribute class (gentle bureaucratic approach):

```
<elementSpec module="core" ident="pb" mode="change">
  <classes mode="change">
    <memberOf key="att.typed" mode="delete"/>
  </classes>
</elementSpec>
```

To remove a specific attribute from an element that holds it privately (repossession):

```
<elementSpec module="core" ident="note" mode="change">
  <attList>
    <attDef ident="anchored" mode="delete"/>
  </attList>
</elementSpec>
```



Changing attribute values <attDef>

```
<elementSpec ident="title" mode="change" module="core">
  <desc>This specification changes the element "title", which is part [...] </desc>
  <attList>
    <attDef ident="type" mode="replace">
      <desc>In this part we replace the specification of type attribute [...] </desc>
      <datatype>
        <dataRef key="teidata.enumerated"/>
      </datatype>
      <valList type="closed">
        <valItem ident="uniform"/>
        <valItem ident="supplied"/>
      </valList>
    </attDef></attList></elementSpec>
```

Datatype

- ▶ **<datatype>** (datatype) specifies the declared value for an attribute, by referring to any datatype defined by the chosen schema language
- ▶ In the TEI scheme, most datatypes are expressed using pre-defined TEI macros, which map a name in the form **teidata.xxxx**
- ▶ For example, **teidata.enumerated**, which defines the range of attribute values expressed as a single XML name taken from a list of documented possibilities.

Table of contents

[Appendix E.1 About the Datatypes and Macros Appendix](#)

[« Appendix D Attributes](#)

[» Appendix F Bibliography](#)

[Home](#)

Appendix E Datatypes and Other Macros

[**Appendix E.1 About the Datatypes and Macros Appendix**](#)

This appendix gives you a list of datatypes and links to the reference pages for them. There are 35 distinctly-named data specifications in revision [0a2bff95a](#) of TEI P5 [Version 4.8.1](#) of the TEI Guidelines.

Alphabetical list

[cmc] Computer-mediated communication

[macro.specialPara.cmc](#)

[tei] Declarations for classes, datatypes, and macros available to all TEI modules

[macro.abContent](#) [macro.limitedContent](#) [macro paraContent](#) [macro phraseSeq](#) [macro phraseSeq.limited](#) [macro specialPara](#) [macro.xtext](#) [teidata.authority](#) [teidata.certainty](#) [teidata.count](#) [teidata.duration.iso](#) [teidata.duration.w3c](#) [teidata.enumerated](#) [teidata.gender](#) [teidata.interval](#) [teidata.language](#) [teidata.name](#) [teidata.namespace](#) [teidata.namespaceOrName](#) [teidata.nullOrName](#) [teidata.numeric](#) [teidata.outputMeasurement](#) [teidata.pattern](#) [teidata.point](#) [teidata.pointer](#) [teidata.prefix](#) [teidata.probability](#) [teidata.probCert](#) [teidata.replacement](#) [teidata.sex](#) [teidata.temporal.iso](#) [teidata.temporal.w3c](#) [teidata.temporal.working](#) [teidata.text](#) [teidata.truthValue](#) [teidata.unboundedCount](#) [teidata.version](#) [teidata.versionNumber](#) [teidata.word](#) [teidata.xmlName](#) [teidata.xpath](#) [teidata.xTruthValue](#)

[↑ TEI P5 Guidelines](#) [« Appendix D Attributes](#) [» Appendix F Bibliography](#)

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

Source: <https://tei-c.org/release/doc/tei-p5-doc/en/html/REF-MACROS.html>

Exercise 2: Attributes

- ▶ Using **minimal_odd_starter_attributes.odd**
 - ▶ Create RELAX NG schema from ODD
 - ▶ Create XHTML guidelines from ODD.
- ▶ Associate the RNG schema with **minimal_encoding_attributes.xml** and see whether the file validates.
- ▶ **Revise the ODD file** so that:
 - ▶ divisions **<div>** must have the **type** attribute.
 - ▶ the only allowed attribute values of **@type** are **book** and **chapter**.
- ▶ Save your ODD with a new name and generate a RNG and XHTML from it.
- ▶ Validate your XML with the new RNG schema

▶ <elementSpec ident="div" mode="change" module="core">
 <desc>[Note Change!] ... </desc>

 <attList>
 <attDef ident="type" mode="replace" usage="req">
 <desc>[Note Change!] In our project we use divs only to distinguish between books and chapters of a work, and we express it through the values of attribute
 <att>type</att>. </desc>
 <datatype>
 <dataRef key="teidata.enumerated"/>
 </datatype>
 <valList type="closed">
 <valItem ident="book"/>
 <valItem ident="chapter"/>
 </valList>
 </attDef>
</attList>

</elementSpec>
</schemaSpec>

**@type on <title> and @type on <div> can now have different pre-defined values!
Test your file. Where else can you put @type? Which values can it have there?**

Exercise 3: Generate ODD

- ▶ Create an ODD file from **minimal_encoding.xml** by using the **oddbyexample.xsl**.
 - ▶ Tutorial: Burnard_2013_How_to_Make_an_ODD_Automagically.pdf
- ▶ Generate **RELAX NG** schema and **XHTML** guidelines from your ODD.
- ▶ Associate the RNG schema with your XML file (**minimal_encoding.xml**),
 - ▶ Does it validate correctly?
 - ▶ What if you tag “Paris” with **placeName**?
- ▶ Edit the ODD file so that it additionally allows the **placeName** element in the main body of the text.
- ▶ Document this change by providing a brief **description** in **<desc>** and one usage **example** in **<exemplum> <egXML>** (See ODD from Exercise 1 for inspiration).

More about elements

HOMEWORK (OPTIONAL)

Model classes

TEI: Guidelines for Electronic Text Encoding and Interchange

P5 Version 4.8.1. Last updated on 1st November 2024, revision 0a2bff95a

<div>

Module: textstructure

Home
C Elements

<div> (text division) contains a subdivision of the front, body, or back of a text. [See [Divisions of the Body](#).]

| Module | textstructure — Default Text Structure |
|------------|---|
| Attributes | <ul style="list-style-type: none">• att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:space<ul style="list-style-type: none">◦ att.global.rendition: @rend, @style, @rendition◦ att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select◦ att.global.analytic: @ana◦ att.global.facs: @facs◦ att.global.change: @change◦ att.global.responsibility: @cert, @resp◦ att.global.source: @source• att.divLike: @org, @sample<ul style="list-style-type: none">◦ att.metrical: @met, @real, @rhyme◦ att.fragmentable: @part• att.typed: @type, @subtype• att.declaring: @decls• att.written: @hand |
| Member of | model.divLike |

Model class: Model.divLike

Div

The only member of
model.divLike

model.divLike

model.divLike groups elements used to represent un-numbered generic structural divisions.

| | |
|----------------|---|
| Module | tei — The TEI Infrastructure |
| Used by | back body div front lem rdg |
| Members | div |

[[English](#)] [[Deutsch](#)] |

Source: <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-model.divLike.html>

Model classes

- ▶ A group of elements that can occur in the same place in a TEI document.
- ▶ The names of model classes all start with **model** . , followed by a name that gives an indication of the group of elements it contains.
- ▶ For example, **<p>** and **<ab>** are being grouped in the model class **model.pLike**, which holds all paragraph-like elements.
- ▶ Source:
<https://teibyexample.org/exist/tutorials/TBED08v00.htm>

model.pLike

model.pLike regroupe des éléments de type paragraphe.

| | |
|--------------------|--|
| Module | tei — The TEI Infrastructure |
| Utilisé par | abstract additional application availability back binding I correspDesc custodialHist decoDesc editionStmt editoria hyphenation interpretation langKnowledge langUsage la msContents msDesc msFrag msItem msItemStruct msP persona physDesc place population post prefixDef proje refsDecl remarks samplingDecl scriptDesc scriptStmt se stdVals styleDefDecl supportDesc terrain trait transcript |
| Membres | ab p |

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

Reminder: Namespace in XML

- ▶ Naming conflict in XML can be avoided by using a **name prefix**. When using prefixes in XML, a **namespace** for the prefix must be defined. The namespace declaration has the following syntax. **xmlns:prefix="URI"**
 - ▶ Example: <root **xmlns:w=http://www.w3.org/TR/html4/**
xmlns:k="https://www.kakapitan.com/myEncoding">
- ▶ Defining a default namespace for an element saves us from using prefixes in all the child elements. It has the following syntax:
xmlns="namespaceURI"
 - ▶ Example: <TEI **xmlns="http://www.tei-c.org/ns/1.0"**>
Read more: https://www.w3schools.com/xml/xml_namespaces.asp

New elements in your own name space

- ▶ Analyse files in the folder **Ex4_Homework_optional**

```
<!-- BEGINNING OF YOUR ELEMENT SPEC -->
<elementSpec ident="cat" ns="http://www.kakapitan.com//ns/1.0">
    <desc>This element allows me to tag all cats mentioned in my document.
        It can only contain text and has no other children elements.</desc>
    <classes>
        <memberOf key="model.nameLike.agent"/><!-- By subscribing your new element to "model.nameLike.agent" it will behave just
        like all other nameLike elements in the model -->
        <memberOf key="att.global"/><!-- By subscribing your new element to "att.global" you allow it to have all the global attributes
        of the att.global element -->
    </classes>
    <content>
        <!-- Here below you specify the content of your element, in our case, a text node. -->
        <textNode/>
    </content>
    <attList>
        <!-- Here below you specify your custom attributes -->
        <attDef ident="breed" ns="http://www.kakapitan.com//ns/1.0" usage="req">
            <desc>The attribute <att>breed</att> used on <gi>cat</gi> elements allows me to group cats by breed.
                This attribute is required (usage = "req") on all <gi>cat</gi> elements.
                The value of the <att>breed</att> must fit the datatype "teidata.word", but the values aren't pre-defined. </desc>
            <datatype minOccurs="1" maxOccurs="1">
                <dataRef key="teidata.word"/>
            </datatype>
        </attDef>
    </attList>
    <example>
```