

TEI BY EXAMPLE



MODULE 1: COMMON STRUCTURE, ELEMENTS, AND ATTRIBUTES

Edward Vanhoutte

Ron Van den Branden

Melissa Terras

Centre for Scholarly Editing and Document Studies (CTB) , Royal
Academy of Dutch Language and Literature, Belgium, Gent, 9 July 2010

Last updated September 2020

Licensed under a Creative Commons Attribution ShareAlike 3.0 License

TABLE OF CONTENTS

2. General TEI Document Structure.....	1
2.1 TEI Header.....	3
2.2 Text.....	4
2.2.1 Body.....	4
2.2.2 Front.....	4
2.2.3 Back.....	5
2.2.4 Full Example <text>.....	6
2.2.5 Unitary or Composite Texts.....	7
2.2.6 Summary.....	7

2. General TEI Document Structure

The TEI makes use of XML as its governing metalanguage. This means that all TEI metadata are expressed as XML elements and thus comply with the World Wide Web Consortium XML Recommendation. Information (plain text) is contained in XML elements, delimited by start tags (e.g., `<TEI>`) and end tags (e.g., `</TEI>`). Additional information to these XML elements can be given in attributes, consisting of a name (e.g., `xml:id`) and a value (e.g., `"text1"`). XML comments are delimited by start markers (`<!--`) and end markers (`-->`).¹

A full TEI document consists of one single `<TEI>` element, which consists of two major components:

- `<teiHeader>`: an element containing all the metadata describing the document.
- `<text>`: an element containing the actual document

This common structure is mandatory for all “standard” TEI documents.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
  </teiHeader>
  <text>
    <!-- ... -->
  </text>
</TEI>
```

Example 1. The minimal structure of a TEI document.

.....

1 In these TEI by Example tutorials, names of TEI components are formatted in a specific way:

- *Element* names are printed in monospace between pointy brackets, e.g., `<TEI>`
- *Attribute* are displayed in monospace, and prefixed with the “at” sign, e.g., `@n`
- *Class*, *datatype*, and *macro* names are displayed in monospace, e.g., `att:global`

All of these components are being presented as hyperlinks to their declaration in the TEI Guidelines. This should make it easier to look up the reference documentation.

This is an example of a TEI XML text, containing both information and meta-information. This example, as any TEI text, is recognizable as a TEI text by the outermost `<TEI>` element, which is declared in the dedicated TEI namespace (<http://www.tei-c.org/ns/1.0>). Before proceeding, let's first have a look at the namespace declaration. In the previous example, the TEI namespace is declared as the “default” namespace, i.e., without any *prefix*. It could have been expressed equally as follows:

```
<tei:TEI xmlns:tei="http://www.tei-c.org/ns/1.0">
  <tei:teiHeader>
    <!-- ... -->
  </tei:teiHeader>
  <text xmlns="http://www.tei-c.org/ns/1.0">
    <!-- ... -->
  </text>
</tei:TEI>
```

Example 2. A TEI document with mixed namespace prefixes.

Here, the namespace declaration `xmlns:tei="http://www.tei-c.org/ns/1.0"` on the `<TEI>` element binds the TEI namespace URI (<http://www.tei-c.org/ns/1.0>) to the namespace prefix `tei`. All descendant elements using that prefix before the actual element name belong to this namespace (e.g., `<tei:teiHeader>`). Yet, the `<text>` element contains its own namespace declaration: `xmlns="http://www.tei-c.org/ns/1.0"`, only this time it is binding it to an *empty* namespace prefix. All descendant elements without a namespace prefix (in the “default” namespace), will belong to this namespace. Since both namespace declarations in the previous example are referencing the same namespace URI, the previous example is equivalent to the first.

Because the TEI namespace is vital to any TEI element, the examples in these TEI by Example tutorials will explicitly render their top-level element(s) with a “default” (i.e., without namespace prefix) namespace declaration for the TEI namespace URI. In order not to hamper legibility, no namespace prefix will be used, and the namespace declaration won't be repeated on any descendant elements.

2.1 TEI Header

The TEI header (<[teiHeader](#)>) is mandatory and contains descriptive meta-information about the document. The <[teiHeader](#)> minimally contains a description of the electronic file inside a (<[fileDesc](#)>). The latter element consists of three mandatory components:

- the title statement (<[titleStmt](#)>), providing information about the title (<[title](#)>), author (<[author](#)>), and others responsible for the electronic text
- the publication statement (<[publicationStmt](#)>), providing publication details about the electronic text in a structured way or as prose inside a paragraph (<[p](#)>)
- a description of the source (<[sourceDesc](#)>), documenting bibliographic details about the electronic text's material source (if any) in a structured way or in a prose paragraph (<[p](#)>)

```
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title>The Strange Adventures of Dr. Burt Diddledygook: a machine-readable
        transcription</title>
      <respStmt>
        <resp>editor</resp>
        <name xml:id="EV">Edward Vanhoutte</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <p>Not for distribution.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Transcribed from the diaries of the late Dr. Roy Offire.</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Example 3. A minimal TEI header.

REFERENCE

See [Module 2: The TEI Header](#) for detailed information on `<teiHeader>`.

2.2 Text**2.2.1 Body**

The actual text (`<text>`) contains a single text of any kind. This commonly contains the actual text and other encodings. A text `<text>` minimally contains a text body (`<body>`). The body contains lower-level text structures like paragraphs (`<p>`), or different structures for text genres other than prose: lines (`<l>`) for poetry, speeches (`<sp>`) for drama.

```
<text xmlns="http://www.tei-c.org/ns/1.0">
  <body>
    <p>For the first time in twenty-five years, Dr Burt Diddledygook decided not to
      turn up to the annual meeting of the Royal Academy of Whoopledywhaa (RAW). It
      was a sunny day in late September 1960 bang on noontime and Dr Burt was looking
      forward to a stroll in the park instead. He hoped his fellow members of theRAW
      weren't even going to notice his absence.</p>
  </body>
</text>
```

Example 4. A `<body>` element with paragraphs.**2.2.2 Front**

Next to the `<body>`, a text can optionally contain front matter which may be encoded with `<front>`. Clear examples are title pages, headers, prefaces, or dedications. Prologues in drama or forewords and introductions in prose may also be considered prefatory material. May, because the encoder may choose simply not to encode the front matter of a text as such. With exception of the title page, for which the TEI defines specific elements, front matter should be encoded using the same elements as the rest of a text. This means that there are no specific

elements to encode prefaces, dedications, abstracts, frontispieces etc. Instead, either numbered or un-numbered divisions `<div>` with an attribute `@type` are used to distinguish between the different components of a `<front>` section. The following suggested values for the `@type` attribute may be used for this purpose:

- "preface": a foreword or preface addressed to the reader
- "ack": a formal declaration of acknowledgement by the author
- "dedication": a formal offering or dedication of a text by the author
- "abstract": a summary of the content of a text as continuous prose
- "contents": a table of contents. A `<list>` element should be used to mark its structure
- "frontispiece": a pictorial frontispiece, possibly including some text

```
<front xmlns="http://www.tei-c.org/ns/1.0">
  <div type="dedication">
    <p>In memory of Lisa Wheeman.</p>
  </div>
  <div type="contents">
    <head>Table of Contents</head>
    <list>
      <item>I. The Decision</item>
      <item>II. The Fuss</item>
      <item>III. The Celebration</item>
    </list>
  </div>
</front>
```

Example 5. A `<front>` section with a dedication and table of contents.

2.2.3 Back

All back matter to a text may be grouped within `<back>`. As is the case with `<front>`, either numbered or un-numbered divisions `<div>` with a `@type` attribute are used to distinguish the different components. The following attribute values may be supplied for the `@type` in order to distinguish various kinds of division characteristic of back matter:

- "appendix": an appended self-contained section of a work, often providing additional information or text
- "glossary": contains a list `<list>` of terms and their explanations

- "notes": a section in which textual or other kinds of notes are gathered together
- "bibliogr": contains a list of bibliographical citations <[listBibl](#)>
- "index": any form of index to the work
- "colophon": a statement appearing at the end of a book describing the conditions of its physical production

```
<back xmlns="http://www.tei-c.org/ns/1.0">
  <div type="colophon">
    <p>Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South
      Africa.</p>
  </div>
</back>
```

Example 6. A <back> section with a colophon.

2.2.4 Full Example <text>

```
<text xmlns="http://www.tei-c.org/ns/1.0">
  <front>
    <div type="dedication">
      <p>In memory of Lisa Wheeman.</p>
    </div>
    <div type="contents">
      <head>Table of Contents</head>
      <list>
        <item>I. The Decision</item>
        <item>II. The Fuss</item>
        <item>III. The Celebration</item>
      </list>
    </div>
  </front>
  <body>
```

```
<p>For the first time in twenty-five years, Dr Burt Diddledygook decided not to
turn up to the annual meeting of the Royal Academy of Whoopledywhaa (RAW). It
was a sunny day in late September 1960 bang on noontime and Dr Burt was looking
forward to a stroll in the park instead. He hoped his fellow members of the RAW
weren't even going to notice his absence.</p>
</body>
<back>
  <div type="colophon">
    <p>Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South
    Africa.</p>
  </div>
</back>
</text>
```

Example 7. A full <text> structure.

2.2.5 Unitary or Composite Texts

Apart from simple texts, TEI provides means to encode composite texts, either by grouping structurally related texts in a <group> element inside <text>, or treating them as a corpus of diverse texts, using <teiCorpus> as the outermost element.

2.2.6 Summary

The following example shows the empty framework of a basic TEI document structure:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          <!--Title-->
        </title>
      </titleStmt>
      <publicationStmt>
        <p>
          <!--Publication Information-->
        </p>
```

```
</publicationStmt>
<sourceDesc>
  <p>
    <!--Information about the source-->
  </p>
</sourceDesc>
</fileDesc>
</teiHeader>
<text>
  <body>
    <!--Some structural division, paragraph, line group, speech, ...-->
  </body>
</text>
</TEI>
```

Example 8. A minimal structure for the <TEI> element.

The following example fills this empty framework with the text of the examples:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>The Strange Adventures of Dr. Burt Diddledygook: a machine-readable
          transcription</title>
        <respStmt>
          <resp>editor</resp>
          <name xml:id="EV">Edward Vanhoutte</name>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <p>Not for distribution.</p>
      </publicationStmt>
      <sourceDesc>
        <p>Transcribed from the diaries of the late Dr. Roy Offire.</p>
      </sourceDesc>
    </fileDesc>
```

```
</teiHeader>
<text>
  <front>
    <div type="dedication">
      <p>In memory of Lisa Wheeman.</p>
    </div>
    <div type="contents">
      <head>Table of Contents</head>
      <list>
        <item>I. The Decision</item>
        <item>II. The Fuss</item>
        <item>III. The Celebration</item>
      </list>
    </div>
  </front>
  <body>
    <p>For the first time in twenty-five years, Dr Burt Diddledygook decided not to turn up to the annual meeting of the Royal Academy of Whoopledywhaan (RAW). It was a sunny day in late September 1960 bang on noontime and Dr Burt was looking forward to a stroll in the park instead. He hoped his fellow members of the RAW weren't even going to notice his absence.</p>
  </body>
  <back>
    <div type="colophon">
      <p>Typeset in Haselfoot 37 and Henry 8. Printed and bound by Whistleshout, South Africa.</p>
    </div>
  </back>
</text>
</TEI>
```

Example 9. The example text encoded as a TEI text with <TEI>.