



# Text Encoding Initiative

AN INTRODUCTORY  
WORKSHOP

# What is TEI?

- ❖ TEI is a subset of eXtensible Markup Language (XML) that was developed specifically for humanistic research.
- ❖ Although TEI files must include certain elements, TEI's structure is intentionally malleable to make it adaptable to the unique needs of a particular research project.
- ❖ Like all XML, TEI is first and foremost descriptive.
- ❖ TEI files may be transformed for countless digital representations and platforms.
- ❖ TEI is both mature and enduring: a great standard to use when creating a research project that can stand up to peer review.

# How is TEI structured?

```
<text>
  <body>
    <div type="section" n="1">
      <head>Section 1</head>
      <p>Some text here.</p>
    </div>
    <div type="section" n="2">
      <head>Section 2</head>
      <p>Some text here.</p>
    </div>
  </body>
</text>
```

- ❖ Information (plain text) is contained in XML elements, delimited by start tags (e.g.: <TEI>) and end tags (e.g.: </TEI>). **With TEI files, <TEI/> is the root element.**
- ❖ Reminders: XML/TEI follows a “tree” structure, and tags are “nested” one within another. Tags must be closed and cannot overlap.

# How is TEI structured?

A full TEI document consists of a

- ❖ `<teiHeader>`: documents all the metadata describing a document; and a
- ❖ `<text>`: containing the document proper.
- ❖ This common structure is mandatory for all TEI documents. This basic structural pair is contained by the root `<TEI/>` element.

# The TEI Element

- ❖ The TEI Element (<TEI>) is the outermost tag (or frame) of your TEI document.
- ❖ It indicates that you are using the language TEI and the following structuring (the tags or element and attribute lexicon) is drawn from a particular namespace. Here is the opening tag for a TEI document:

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

- ❖ Reminder: an XML namespace defines a set of rules for encoding documents.

# The TEI Header

The TEI header (<teiHeader>) is mandatory. It minimally contains a description of the electronic file inside (<fileDesc>). The <fileDesc> consists of three mandatory components:

- ❖ **the title statement** (<titleStmt>), which includes the title (<title>) and those responsible for the electronic text (e.g. <author> or <editor>)
- ❖ **the publication statement** (<publicationStmt>), providing publication details about the electronic text (e.g., <publisher>, <pubPlace>, and <date>)
- ❖ **a description of the source** (<sourceDesc>), documenting bibliographic details about the electronic text's material source (sometimes just in a <p>)

# The TEI Header, with some basic elements added

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title></title>
      <author>
    </titleStmt>
    <publicationStmt>
      <publisher></publisher>
      <pubPlace></pubPlace>
      <date></date>
    </publicationStmt>
    <sourceDesc></sourceDesc>
  </fileDesc>
</teiHeader>
```

# The Text

- ❖ The actual text of a TEI document contains any single text. There is only one mandatory tag for the text (<text>).
- ❖ For many texts, elements to differentiate between front matter (<front>); the body (<body>); and back matter (<back>) are used.
- ❖ Within these sections, one often finds paragraphs (<p>), divisions (<div>) with attributes to distinguish individual components such as @type="contents" or @type="epilogue".
- ❖ The form and content of your text will determine what elements and attributes you use to further delineate the nature of your text.



# The Text

```
<text>
```

```
  <body>
```

```
    <p>This is an example of a (prose) text.</p>
```

```
  </body>
```

```
</text>
```

# Reminder: Well-Formed vs. Valid

- ❖ If a document is well-formed, it conforms to the basic rules of XML.
- ❖ If a document is valid, it conforms to the rules of a DTD (Document Type Definition) or schema. **A document can be well-formed but not valid.**
- ❖ Since we're just learning today, we won't be validating against anything at first.

# Resources

- ❖ TEI by Example - <http://teibyexample.org>
- ❖ TEI Consortium - <http://www.tei-c.org/>
- ❖ Getting Started Using TEI - <http://tei.oucs.ox.ac.uk/GettingStarted/html/in.html>

# Encoding: An Exercise

Use your handout to guide you through encoding the following Emily Dickinson poem from poets.org.

My life closed twice before its close—  
It yet remains to see  
If Immortality unveil  
A third event to me

So huge, so hopeless to conceive  
As these that twice befell.  
Parting is all we know of heaven,  
And all we need of hell.

# Official TEI Guidelines for Poetry

Use the TEI Guidelines for poetry to help you move through the exercise. You can be as detailed as you wish.

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/VE.html>

\*\*In particular, I'd like to see if you can use the Guidelines to teach yourself how to encode **rhyme**. The purpose of this is to give you some experience using the Guidelines to find out how to do something.