

# TEI Basic Structure and Core Elements

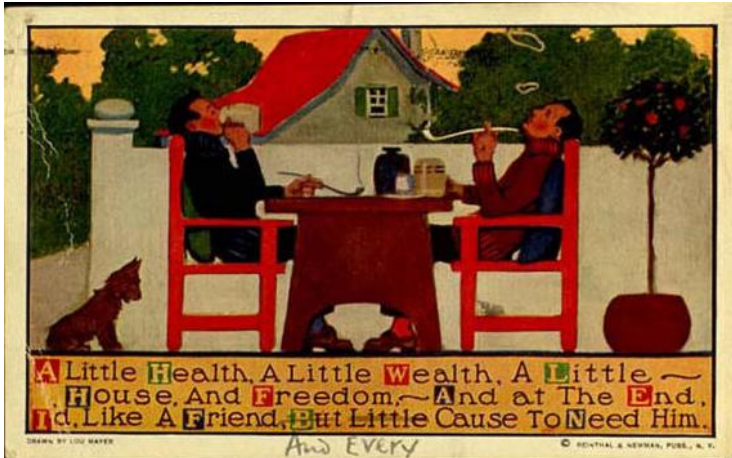
James Cummings  
@jamescummings

12 November 2014

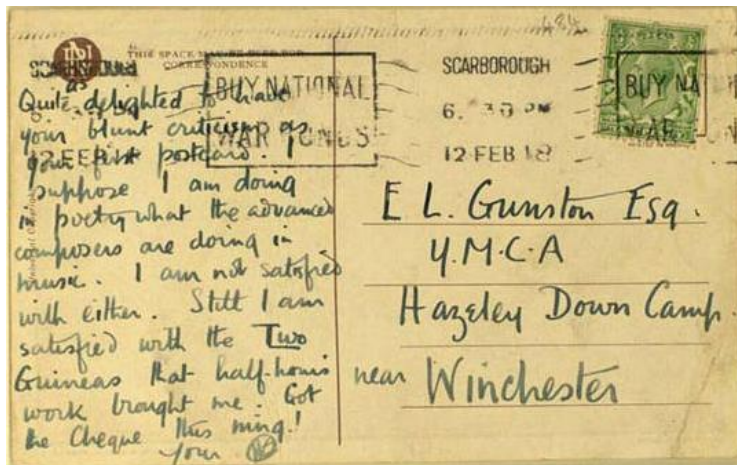
# Default Text Structure

All TEI documents are structured in a particular manner. This section attempts to describe the different variations on this as briefly as possible.

## Postcard Front



# Postcard Back



## Letter 1917-01-10 (page 1)

Jan. 10. 1917.  
My own Mother,             
I was  
censoring letters all afternoon.  
After tea <sup>among</sup> commenced a big  
commotion ~~of~~ my friendly  
neighbours & the Houtzgers,  
in the midst of which I  
wrote a ~~special~~ ~~particular~~ distracted  
note to Leslie, but the  
concussion blew out my  
candle so many times that  
I lost heart.  
I am kept pretty  
busy, tho' there is only a  
short 'harass'. The men do  
practically nothing all day  
but write letters, ~~per~~ but officers  
have frequent meetings over  
schemes, maps, instructions,  
and a thousand cares.  
Yesterday I took  
a turn out the line which

## Letter 1917-01-10 (page 2)

2  
we shall occupy. Our <sup>little</sup> party  
was shelled going up  
across the open country.  
It was not at all frightful  
and only one 4.7 got  
anywhere near, falling  
plump in the road but  
quite a minute after we  
had passed the spot.  
I tell you these things  
because afterwards they will  
sound less exciting. If I leave  
all my exploits for recitation  
after the war without  
mentioning them now, they  
will be appearing bomb-shell-bastic.  
Now I am not  
so uncomfortable as last  
week for my new servant  
who had been a chemist's  
assistant, has turned out not  
only clean & smart, but

## Letter 1917-01-10 (page 5)

5/1  
I am not allowed to send a  
sketch, but you must know  
I am transformed now, wearing  
a steel helmet, buff jerkin of  
leather, & rubber-waders up  
to the hips, & gauntlets.  
But for the rifle we are  
exactly like Cromwellian  
Troopers. The waders are  
of course indispensable.  
In 2½ miles of trench which I  
waded yesterday there was  
not one inch of dry  
ground. There is a mean  
depth of 2 feet of water.  
It seems an era since  
Christmas, Day, and Goose,  
Carols, Dickens & Mistletoe.  
Assuming the war lasts another  
year I should get leave twice,  
or three times, for we get it,  
or should get it every 3 months.  
Be sure to have no chloride

Not so bad!

Preface.

I have no hesitation in making public  
the phrase "Hypocrites are not <sup>to</sup> generally such  
~~consultations~~" is in no sense a consultation  
with <sup>this</sup> generation. They may be to the  
next.

nest. ~~book limits~~ All last ~~I~~ <sup>was</sup> ~~ought to have~~  
 used proper names. <sup>is</sup> ~~was~~ <sup>ought to have</sup>  
 That is why the ~~man~~ <sup>man</sup> ~~was~~ <sup>was</sup> ~~Post~~ <sup>Post</sup> ~~must be~~ <sup>must be</sup> ~~truthful~~  
 If I thought the ~~cells of~~ <sup>cells of</sup> ~~the book would~~ <sup>the book would</sup> ~~last~~ <sup>last</sup> ~~I~~  
 we might have ~~used~~ <sup>used</sup> ~~proper names~~ <sup>proper names</sup>; but if the spirit of  
 it ~~survives~~ <sup>survives</sup> ~~Proper~~ <sup>Proper</sup> ~~names will~~ <sup>names will</sup>  
 be content, ~~for the~~ <sup>for the</sup> ~~future~~ <sup>future</sup> ~~fields~~ <sup>fields</sup> ~~the~~ <sup>the</sup> ~~flowers~~ <sup>flowers</sup>  
 for be, all of way, will be



# Preface Edited

## PREFACE

This book is not about heroes. English poetry is not yet fit to speak of them.

- 1 Nor is it about deeds, or lands, nor anything about glory, honour, might, majesty, dominion, or power, except War.

Above all I am not concerned with Poetry.

My subject is War, and the pity of War.

The Poetry is in the pity.

- 2 Yet these elegies are to this generation in no sense consolatory. They may be to the next. All a poet can do today is warn. That is why the true Poets must be truthful.

(If I thought the letter of this book would last, I might have used proper names; but if the spirit of it survives—survives Prussia—my ambition and those names will have achieved fresher fields than Flanders. . . .)

At Ripon, probably in May 1918, WO began this draft Preface for a collection of war poems that he hoped to publish in 1919 (WO, 265–6).

- 1 Cp. the General Epistle of Jude 25, 'To the only wise God our Saviour, be glory and majesty, dominion and power, both now and ever.'  
2 WO had considered – and decided against – calling his book *English Elegies*.

CP&F, 535

## Red Cross or Iron Cross?

**RED CROSS** OR **IRON CROSS?**

**WOUNDED AND A PRISONER**  
**OUR SOLDIER CRIES FOR WATER.**

**THE GERMAN "SISTER"**  
**POURS IT ON THE GROUND BEFORE HIS EYES.**

**THERE IS NO WOMAN IN BRITAIN**  
**WHO WOULD DO IT.**

**THERE IS NO WOMAN IN BRITAIN**  
**WHO WILL FORGET IT.**

## The Kitchen is the Key to Victory



# Looking at this WW1 material, what do we need to mark up?

- Identification information, page numbers, sources
- "chunks" or divisions of text, which may contain a picture, a poem, some prose, or a combination
- within the chunks, we can identify formal structural units such as
  - a picture, a caption
  - stanzas, lines
  - paragraphs
- and more at phrase level...

# Structure of a TEI Document

There are two basic structures of a TEI Document:

- **<TEI>** (TEI document) contains a single TEI-conformant document, comprising a TEI header and a text, either in isolation or as part of a `teiCorpus` element.
- **<teiCorpus>** contains the whole of a TEI encoded corpus, comprising a single corpus header and one or more TEI elements, each containing a single text header and a text.

## TEI basic structures (1)

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

## TEI basic structures (2)

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- required -->
    </teiHeader>
    <facsimile>
    <!-- optional-->
    </facsimile>
    <sourceDoc>
    <!-- optional -->
    </sourceDoc>
    <text>
    <!-- required if no facsimile or sourceDoc-->
    </text>
  </TEI>
```

## <text>

### What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter



<text>

## What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter

## <text>

### What is a text?

- A text may be unitary or composite
  - unitary: forming an organic whole
  - composite: consisting of several components which are in some important sense independent of each other
- a unitary text contains
  - optional front matter
  - <body> (required)
  - optional back matter

## Composite texts

A composite text contains

- optional front matter
- `<group>` with `<text>` inside (required)
- optional back matter

A corpus is a collection of text and header pairs that also has its own header.

`<group>` tags may self-nest.

# TEI text structure (1)

```
<text>
  <front>
<!-- optional -->
  </front>
  <body>
<!-- required -->
  </body>
  <back>
<!-- optional -->
  </back>
</text>
```

## TEI text structure (2)

```
<text>
  <front>
<!-- front matter to an anthology -->
  </front>
  <group>
    <text>
      <body>
        <p>
<!-- content of first text -->
          </p>
        </body>
      </text>
<!-- more texts here -->
    </group>
  <back>
<!-- back matter to an anthology -->
  </back>
</text>
```

## Another Grouped Text Example

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- header information for the whole collection -->
  </teiHeader>
  <text>
    <!-- optional front matter -->
    <group>
      <text>
        <!-- optional front matter -->
        <body>
          <!-- First Body -->
        </body>
      </text>
      <text>
        <!-- optional front matter -->
        <body>
          <!-- Second Body-->
        </body>
      </text>
    </group>
    <!-- optional back matter -->
  </text>
</TEI>
```

## 'Core' elements

The so-called 'Core' module groups together elements which may appear in any kind of text and the tags used to mark them in all TEI documents.

This includes:

- paragraphs
- highlighting, emphasis and quotation
- simple editorial changes
- basic names numbers, dates, addresses
- simple links and cross-references
- lists, notes, annotation, indexing
- graphics
- reference systems, bibliographic citations
- simple verse and drama

# Paragraphs

`<p>` (paragraph) marks paragraphs in prose

- Fundamental unit for prose texts
- `<p>` can contain all the phrase-level elements in the core
- `<p>` can appear directly inside `<body>` or inside `<div>` (divisions)

```
<p>Thanks for yours of this morning. I hope <lb/>you have had my card posted  
last Monday. <lb/>On Mond. next I lecture the <orgName ref="#Fieldclub">Field  
Club</orgName> - <lb/>a Nat. Hist. Association, in the lines of our <lb/>old  
Society - Geological, (you + me) + Botanical <lb/>(New) Do you remember:  
you<supplied>r</supplied> old <lb/>Black Molt?</p>
```



## Highlighting

By *highlighting* we mean the use of any combination of typographic features (font, size, hue, etc.) in a printed or written text in order to distinguish some passage of a text from its surroundings. For words and phrases which are:

- distinct in some way (e.g. foreign, archaic, technical)
- emphatic or stressed when spoken
- not really part of the text (e.g. cross references, titles, headings)
- a distinct narrative stream (e.g. an internal monologue, commentary)
- attributed to some other agency inside or outside the text (e.g. direct speech, quotation)
- set apart in another way (e.g. proverbial phrases, words mentioned but not used)

## Highlighting Examples

- `<hi>` (general purpose highlighting); `<distinct>` (linguistically distinct)

```
<p>Last week I wrote (to order) a strong <lb/>bit of Blank: on  
<hi rend="ul">Antaeus v.  
    Heracles</hi>. <lb/>These are the best lines, methinks: <lb/>(N.B. Antaeus  
deriving  
strength from his Mother Earth <lb/>nearly licked old  
<distinct>Herk</distinct>.) </p>
```

- Other similar elements include: `<emph>`, `<mentioned>`,  
`<soCalled>`, `<term>` and `<gloss>`

## Quotation

Quotation marks can be used to set off text for many reasons, so the TEI has the following elements:

- `<q>` (separated from the surrounding text with quotation marks)
- `<said>` (speech or thought)
- `<quote>` (passage attributed to an external source)
- `<cit>` (groups a quotation and citation)

```
<cit>
  <quote>
    <l>... How Earth herself empowered him with her trick,</l>
    <l>Gave him the grip and stringency of Winter,</l>
    <l>And all the ardour of th' invincible Spring;</l>
  </quote>
  <bibl>
    <author>Wilfred Owen</author>
    <title ref="works.xml#W0123">Letter to Leslie Gunston / The Wrestler</title>
    <date when="1917-07">July 1917</date>
  </bibl>
</cit>
```

## Simple Editorial Changes: <choice> and Friends

- <choice> (groups alternative editorial encodings)
- Errors:
  - <sic> (apparent error)
  - <corr> (corrected error)
- Regularization:
  - <orig> (original form)
  - <reg> (regularized form)
- Abbreviation:
  - <abbr> (abbreviated form)
  - <expn> (expanded form)

## Choice Example

```
<p>...any might, <unclear reason="scribbled">majesty</unclear>,  
<choice>  
  <abbr>domin</abbr>  
  <expn>domin<ex>ion</ex>  
  </expn>  
</choice> or power...</p>
```

## Additions, Deletions, and Omissions

- `<add>` (addition to the text, e.g. marginal gloss)
- `<del>` (phrase marked as deleted in the text)
- `<gap>` (indicates point where material is omitted)
- `<unclear>` (contains text unable to be transcribed clearly)

## Example of <add>, <del>, <gap>, and <unclear>

```
<p>  
  <add place="left">My </add>  
  <del rend="stoked">It's </del>  
  <add place="above">  
    <del rend="stoked">The </del>  
  </add> subject <del rend="stoked">of</del> is War, and the <unclear>pity  
</unclear>of <del rend="stoked">it</del> War. <lb/> The Poetry is in the pity.  
</p>
```

## Basic Names

- `<name>` (a name in the text, contains a proper noun or noun phrase)
- `<rs>` (a general-purpose name or referencing string )

The `@type` attribute is useful for categorizing these, and they both also have `@key`, `@ref`, and `@nymRef` attributes.



## Addresses

- `<email>` (an electronic mail address)
- `<address>` (a postal address)
- `<addrLine>` (a non-specific address line)
- `<street>` (a full street address)
- `<postCode>` (a postal (or zip) code)
- `<postBox>` (a postal box number)
- `<name>` can also be used
- and the 'namesdates' module extends this with more geographic names

## Basic Address Example

WW1 Poetry Archive Project:

```
<email>ww1collections@it.ox.ac.uk</email>
```

Shell-shock hospital 'Craiglockhart' that Wilfred Owen stayed in:

```
<address>  
  <street>14 Frederick Street</street>  
  <postCode>EH2 2HB</postCode>  
  <settlement>Edinburgh</settlement>  
  <country>United Kingdom</country>  
</address>
```

(or should that 'country' be 'Scotland'? Encoding is a form of interpretation and thus sometimes political.)

## Basic Numbers and Measures

- `<num>` (marks a number of any sort)
- `<measure>` (marks a quantity or commodity)
- `<measureGrp>` (groups specifications relating to a single object)
- While `<num>` has simple *@type* and *@value* attributes, `<measure>` has *@type*, *@quantity*, *@unit* and *@commodity* attributes

## Number and Measure examples

```
<l>With a <num value="1000">thousand</num> pains that vision's face was  
grained;</l>
```

```
... only <measure type="distance" unit="m"  
quantity="3218.69">two miles</measure> from the front....
```

## Dates

- **<date>** (contains a date in any format and includes a *@when* attribute for a regularised form and a *@calendar* attribute to specify what calendar system)
- **<time>** (contains a time in any format and includes a *@when* attribute for a regularised form)

```
<date when="1917-07">July 1917.<lb/> Wednesday</date>
```

## Simple Linking

- `<ptr>` (defines a pointer to another location)
- `<ref>` (defines a reference to another location, with optional linking text)
- Both elements have:
  - `@target` attribute taking a URI reference
  - `@cRef` attribute for canonical referencing schemes
- If the linking text is able to be generated, `<ptr>` and `<ref>` might be used in the same place.

## Simple Linking Example

See `<ref target="#Section12">`section 12 on page 34`</ref>`.

See `<ptr target="#Section12"/>`.

# Lists

- `<list>` (a sequence of items forming a list)
- `<item>` (one component of a list)
- `<label>` (label associated with an item)
- `<headLabel>` (heading for column of labels)
- `<headItem>` (heading for column of items)



## Simple List Example

The previous slide contained only:

```
<div>
  <head>Lists</head>
  <list>
    <item>
      <gi>list</gi> (a sequence of items forming a list)</item>
    <item>
      <gi>item</gi> (one component of a list)</item>
    <item>
      <gi>label</gi> (label associated with an item)</item>
    <item>
      <gi>headLabel</gi> (heading for column of labels)</item>
    <item>
      <gi>headItem</gi> (heading for column of items)</item>
  </list>
</div>
```

## Notes

- `<note>` (contains a note or annotation)
- Notes can be those existing in the text, or provided by the editor of the electronic text
- A *@place* attribute can be used to indicate the physical location of the note
- Notes should usually be encoded where its identifier/mark first appears; notes can also be kept separately and point back to their location with a *@target* attribute

`<note>Painted by <persName>John Singer Sargent</persName>, 1918</note>`

# Indexing

- If converting an existing index, use nested lists. For auto-generated indexes:
- `<index>` (marks an index entry) with optional `@indexName` attribute
- The `<term>` element is used to mark a term inside an `<index>` element
- The `<index>` element can self-nest for hierarchical index entries

# Indexing Example

```
<p>Last week I wrote (to order) a strong <lb/>bit of Blank<index>  
  <term>Verse</term>  
  <index>  
    <term>Blank Verse</term>  
  </index>  
</index>:</p>
```

# Graphics

- `<graphic>` (indicates the location of an inline graphic, illustration, or figure)
- `<binaryObject>` (encoded binary data embedding a graphic or other object)
- The *figure* module provides `<figure>` and `<figDesc>` for more complex graphics with `<figDesc>`

```
<figure>
  <graphic url="materials/postcard-front.jpg"/>
  <figDesc>A postcard image of two men relaxing at a table, smoking pipes and
drinking. A dog and potted fruit tree are nearby with a house over the wall in
the distance.</figDesc>
</figure>
```

## Simple Verse

```
<lg type="stanza">  
  <l>It seemed that out of battle I escaped</l>  
  <l>Down some profound dull tunnel, long since scooped</l>  
  <l>Through granites which titanic wars had groined.</l>  
</lg>
```

# Simple Drama

```
<sp>
  <speaker>The reverend Doctor Opimiam</speaker>
  <p>I do not think I have named a single unrepresentable fish.</p>
</sp>
<sp>
  <speaker>Mr Gryll</speaker>
  <p>Bream, Doctor: there is not much to be said for bream.</p>
</sp>
```