TEI for Transcription

...

This lecture

- How is transcription different from other encoding?
- Encoding from a physical document PoV
- What are the main elements necessary for transcription

How is transcription different?

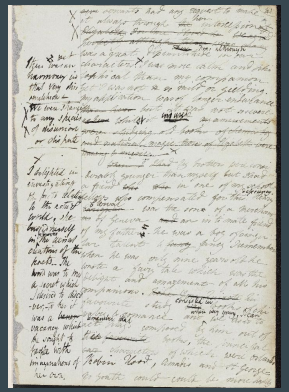
- Transcription aims to:
  - make a primary source accessible
  - make a primary source comprehensible
- Which may entail:
  - adding or making use of additional information

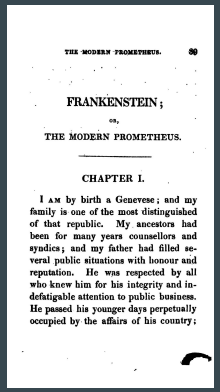
Choices and decisions made by editors

- Transcription is:
  - selective
  - interpretative
  - subjective
  - dependent on editors' choices

How is transcription different?

- Transcription happens in all types of editing, so how is this different?
  - Encoding for transcription foregrounds a different point of view





THE MODERN PROMETHEUS. 89

FRANKENSTEIN;

OR,

THE MODERN PROMETHEUS.

CHAPTER I.

I AM by birth a Genevese; and my family is one of the most distinguished of that republic. My ancestors had been for many years counsellors and syndics; and my father had filled several public situations with honour and reputation. He was respected by all who knew him for his integrity and indefatigable attention to public business. He passed his younger days perpetually occupied by the affairs of his country;

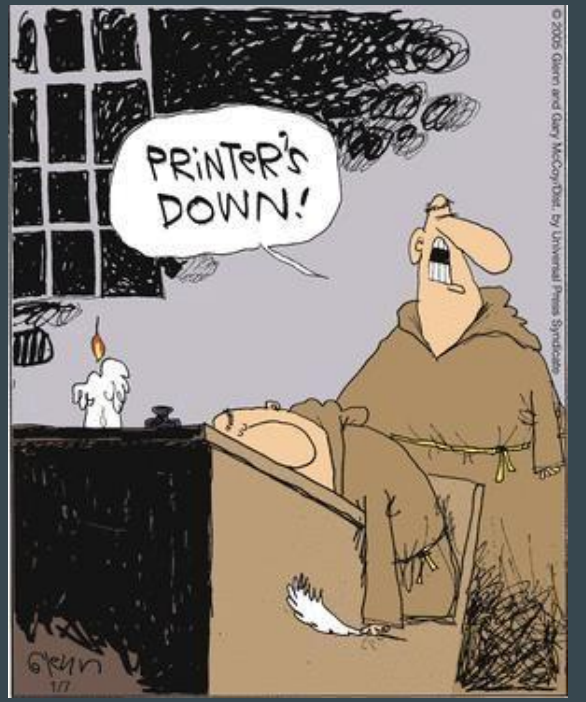
Physical Manifestation

I am by birth a Genevese, and my family is one of the most distinguished of that republic. My ancestors had been for many years counsellors and syndics, and my father had filled several public situations with honour and reputation. He was respected by all who knew him for his integrity and indefatigable attention to public business. He passed his younger days perpetually occupied by the affairs of his country;

Semantic Content

What is Transcription?

- A range of phenomena based on technologies and historical context:
  - Original layout information
  - Abbreviations or other ‘arcana’
  - Errors which invite correction or conjecture
  - Scribal additions, deletions, substitutions, restorations, transpositions
  - Damaged or illegible passages
  - Non-standard orthography which invites normalisation
  - ...



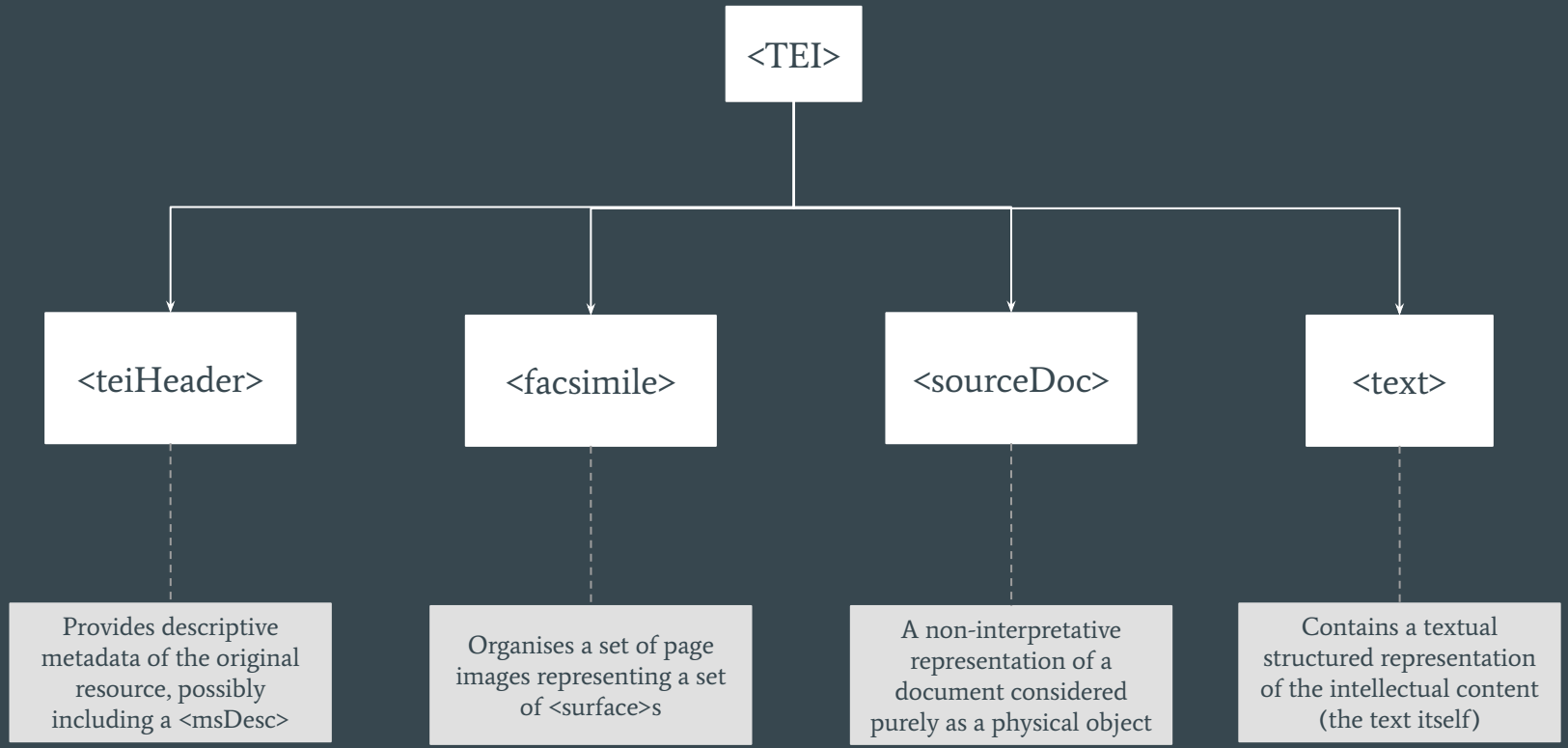
Transcription: how to do it

- Inspect the original document or a surrogate (facsimile)
  - Identify areas in the document (text, graphics, etc)
  - Find the first line
  - Type the first line, identify special characters
  - Decide what kind of diplomatic method you are using; this will determine issues of lineation, and how much can be transcribed.
  - Record textual modifications/interventions
    - Highlighting
    - Additions
    - Deletions
    - Transpositions
    - Notes
  - Identify text-structures
  - Recognise writing activities
  - Identify named entities

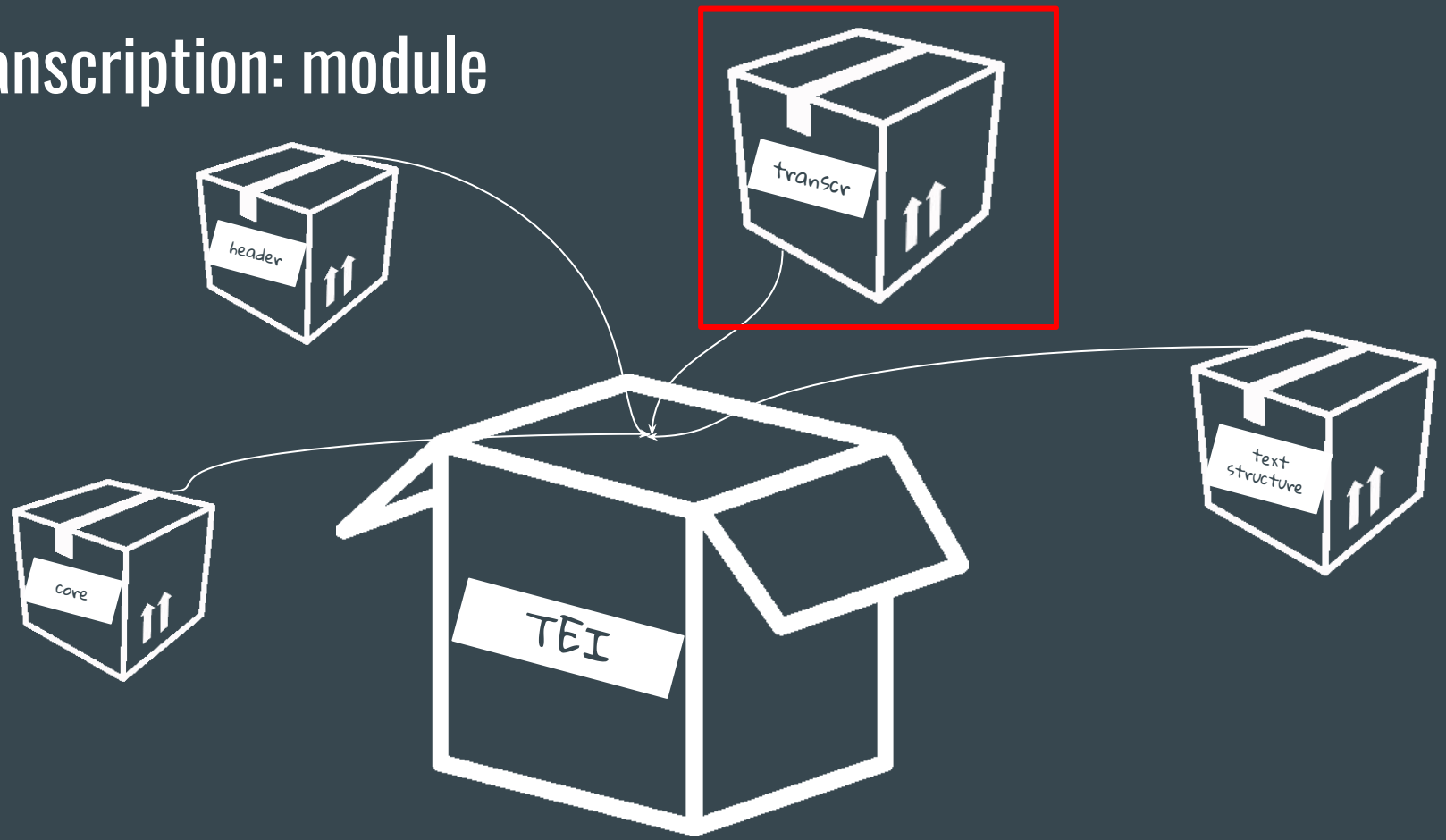
Transcription: getting physical

- Can we record all this? Yes we can!
  - Inspect the original document or a surrogate (digital or otherwise, facsimile)
    - <sourceDoc> or <facsimile>
  - Identify areas in the document (text, graphics, etc)
    - <surface> or <zone>
  - Find the first line
    - <line>
  - Type the first line, identify special characters
    - <g> or <hi>
  - Record textual modifications/interventions
    - <hi>, <mod>, <transpose>, ...
  - Identify text-structures
    - , <head>, <p>, <list>, ...
  - Recognise writing activities
    - <add>, <del>, <subst>, <abbr>, ...
  - Identify named entities
    - <name>, <persName>, <placeName>, ...

Transcription: structure



Transcription: module



Abbreviations

- An abbreviation is used in manuscript materials to shorten labour
- It uses significant marks to replace:
    - Single letters
    - Groups of letters
    - Whole words
    - Whole phrases

Recap: as we've seen, the core model already supports some simple elements to record abbreviations:

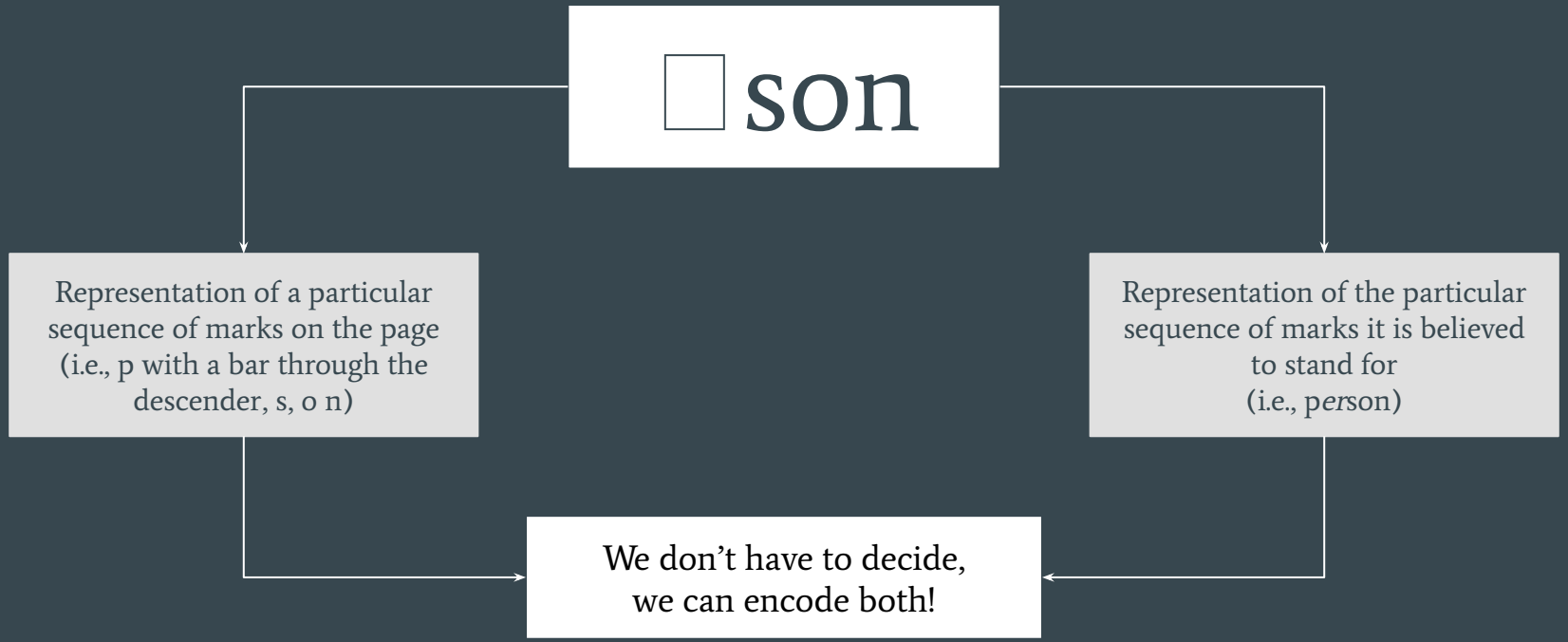
- <choice> —> groups abbreviated and expanded readings
- <abbr> —> abbreviated form
- <expan> —> expanded form

Abbreviations

- Suspension —> only the first letter of a word or phrase, followed by a point
    - <abbr>p.</abbr> —> <expan>page</expan>
- Contraction —> the letters in the middle of a word are omitted
    - <abbr>Dr</abbr> —> <expan>Doctor</expan>
- Brevigraph —> a single character representing two or more letters
    - <abbr>□son</abbr> —> <expan>person</expan>
- Superscript —> characters representing various kinds of contractions
    - <abbr>Ma.^{tie}</abbr> —> <expan>Majestie</expan>

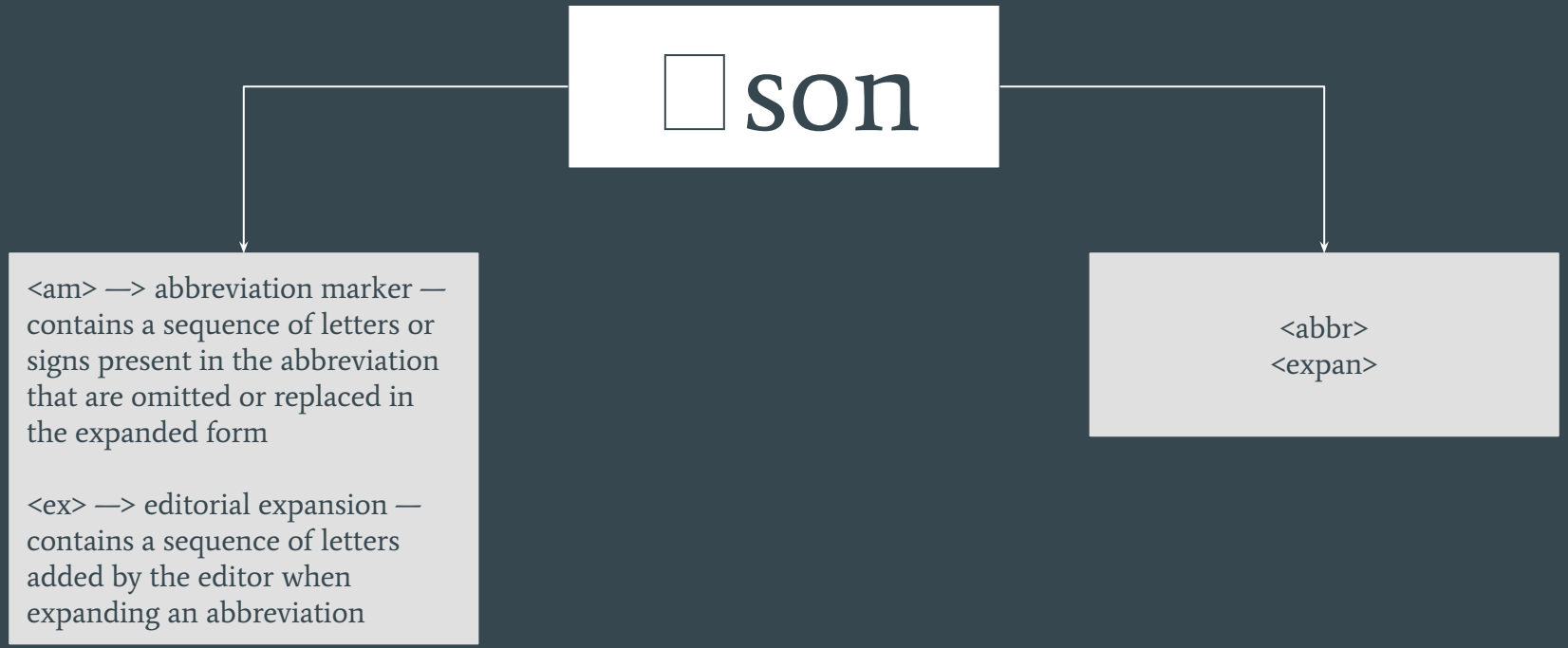
Abbreviations

- How do you view abbreviations?



Abbreviations

- How do you view abbreviations?



Abbreviations

Ma.tie

```
<choice>
   <abbr>
      Ma<am rend="superscript">.tie</am>
   </abbr>
   <expan>
      Ma<ex>jestie</ex>
   </expan>
</choice>
```

Corrections and emendations

- Apparent errors can be recorded:
  - in their original state —> <sic>
  - as corrected text —> <corr>
  - combined —> <choice> (+ <sic> and <corr>)

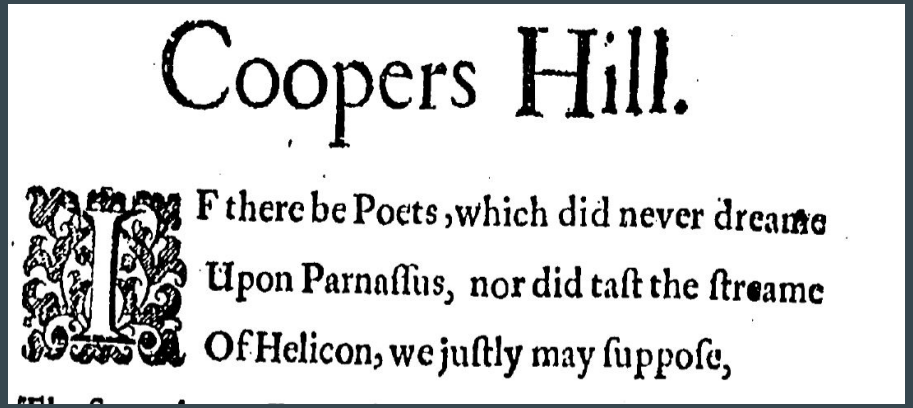
Recap: as we've seen, the core model already supports these elements!

My favourite food is pasghetti

```
<p>
   My favourite food is <corr>spaghetti</corr>
</p>
```

```
<p> My favourite food is
   <choice>
      <sic>pasghetti</sic>
      <corr>spaghetti</corr>
   </choice>
</p>
```

Regularisation



Coopers Hill.

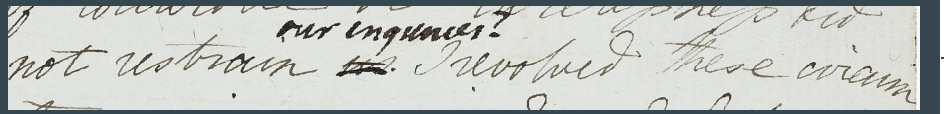
IF there be Poets, which did never dreame
Upon Parnassus, nor did taſt the ſtreame
Of Helicon, we juſtly may ſuppoſe,

```
<div type="poem">
   <head>Coopers Hill.</head>
   <lg>
      <l><hi rend="decorated-capital">I</hi>F
         there be Poets, which did never
         <choice>
            <orig>dreame</orig>
            <reg>dream</reg>
         </choice></l>
      <l>Upon Parnassus, nor did <choice>
            <orig>taſt</orig>
            <reg>taste</reg>
         </choice> the <choice>
            <orig>streame</orig>
            <reg>stream</reg>
         </choice></l>
      <l>Of Helicon, we justly may suppose,</l>
   </lg>
</div>
```

Modifications

- There are several ways of recording modifications in a text:
  - <mod> —> general modifications, without any specific interpretation
    - @type attribute can be used for further specification
  - <add> —> addition to the text
  - <del> —> characters or sequence of characters marked as deleted from the text
  - <subst> —> groups additions and deletions as a single intervention
  - <supplied> —> editorially supplied text

Modifications




```
not restrain
<mod>
   <del>us.</del>
   <add>our enquiries></add>
</mod>
I revolved these circum
```

```
not restrain
<subst>
   <del>us.</del>
   <add>our enquiries></add>
</subst>
I revolved these circum
```

Partially legible text

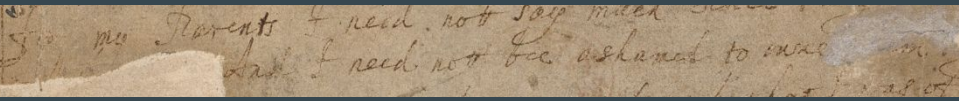
- The text may be only partially legible, due to:
  - cancellation;
  - damage;
  - other environmental factors.
- <unclear> marks the portion of the text that can be read
  - @reason records the cause of uncertainty
  - @resp attributes responsibility for the offered reading
  - @cert records the degree of certainty for the offered reading
  - @agent categorises the cause of any damage

Omitted or damaged material

- The text may be completely illegible due to:
  - thorough cancellation;
  - damage;
  - other environmental factors.
- Or the editor might decide to omit a certain portion of the text
  - <gap> —> records the existence of text that cannot be read or has been purposefully left out
  - <damage> —> records the existence of an area of physical damage to the source

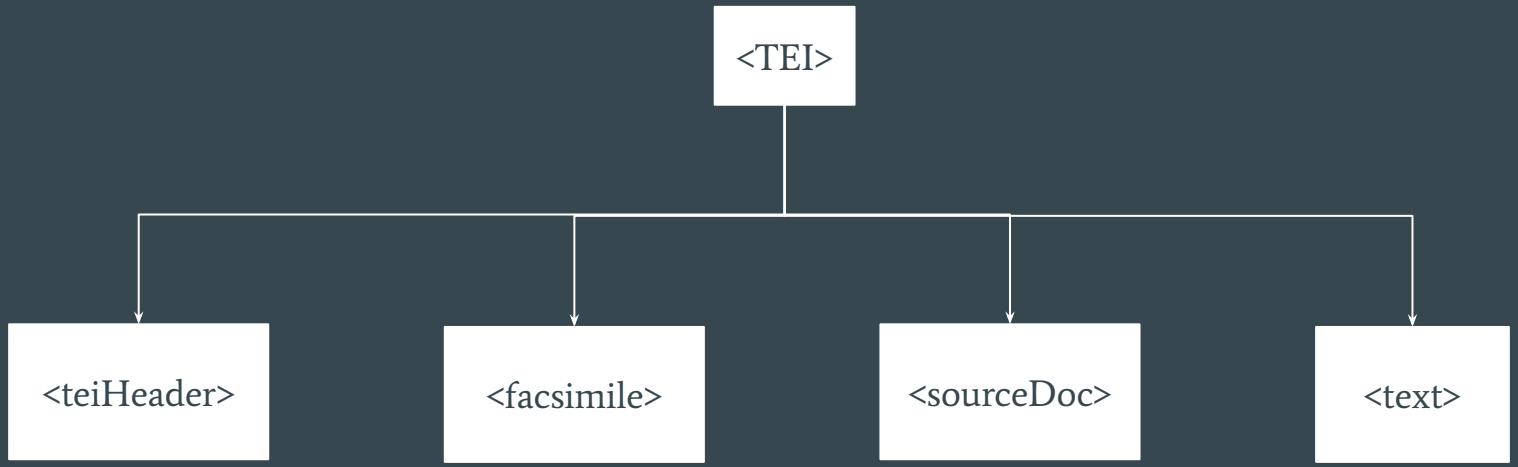
<damage> element often contains <gap>s or <unclear> sections of text

Partly legible and illegible text



```
<damage agent="rip">
   <gap/>
</damage>
And I need not bee ashamed to owne
<damage agent="rip">
   <supplied cert="high">the</supplied>
   <unclear>m.</unclear>
</damage>
```

Transcription: structure



Facsimile

- A digital facsimile contains images of the pages of a source document
- A valid (but not very useful) TEI document may contain only:
  - Metadata
  - Images of the source document
- A valid (and much more useful) TEI document may also contain:
  - Metadata
  - Images of the source document
  - A transcription of those pages
- <facsimile> —> contains the representation of the source as a set of images
- <graphic> —> indicates the location of each image with @url attribute

Facsimile

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
   <teiHeader>
      <!-- METADATA! -->
   </teiHeader>
   <facsimile>
      <graphic url="A1r.png"/>
      <graphic url="A1v.png"/>
      <graphic url="A2r.png"/>
      <graphic url="A2v.png"/>
      <!-- and so on -->
   </facsimile>
</TEI>
```

Facsimile and transcription

- We can combine a facsimile with a transcription:
  - If the transcription is regarded as text in its own right, and organised independently of its physical manifestation, we use the <text> element
  - If the transcription is meant to prioritise its physical manifestation, we use the <sourceDoc> element

Facsimile and transcription

- For simple cases in which one image corresponds to one page transcription:
  - Facsimiles may be referenced directly in <text> by using:
    - <pb/> —> the milestone page break element in conjunction with
    - @facs the facsimile attribute.

```
<text>
   <pb facs="A1r.png"/>
   <!-- transcription of A1r goes here -->
   <pb facs="A1v.png"/>
   <!-- transcription of A1v goes here -->
</text>
```

- However, this approach assumes that
  - only one image exists per page of text
  - the transcribed page corresponds exactly with the image
- Not a very scalable approach

Facsimile and transcription

- A more sustainable approach is to define <surface>s and <zone>s with <facsimile>s
- <surface> —> defines a written surface as a two-dimensional coordinate space
  - i.e., pages, openings, etc.
- <zone> —> defines a single area on a <surface> using coordinates
  - @points —> define the <zone> using point-pairs to build the text area (x,y)
  - @ulx, @uly, @lrx, @lry —> define the upper left and lower right corners of a rectangle
  - You can automate the definition of these coordinates by using additional tools, for example:
    - Oxygen facsimile addon
    - Image Markup Tool

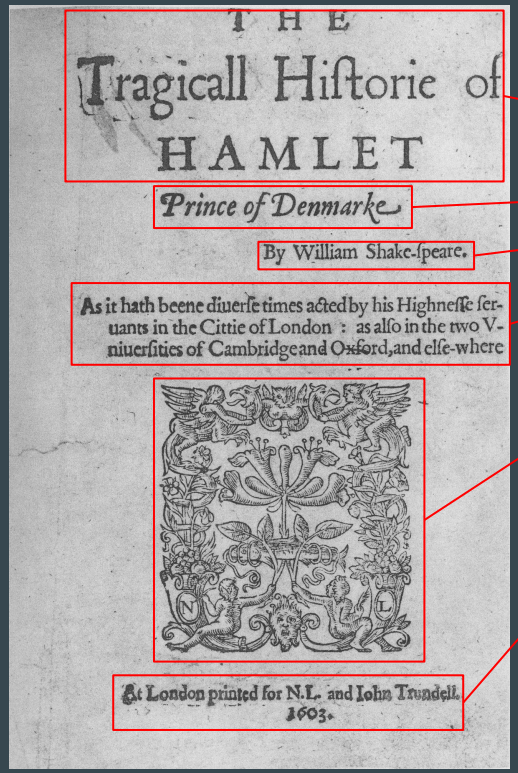
Facsimile and transcription

```
<facsimile>
   <surface xml:id="A1r"><graphic url="A1r.png"/></surface>
   <surface xml:id="A1v"><graphic url="A1v.png"/></surface>
   <surface xml:id="A2r"><graphic url="A2r.png"/></surface>
   <surface xml:id="A2v"><graphic url="A2v.png"/></surface>
</facsimile>
<text>
    <pb facs="#A1r" />
    <!-- transcrption of A1r goes here -->
    <pb facs="#A1v"/>
    <!-- transcription of A1v goes here -->
</text>
```

Facsimile and transcription

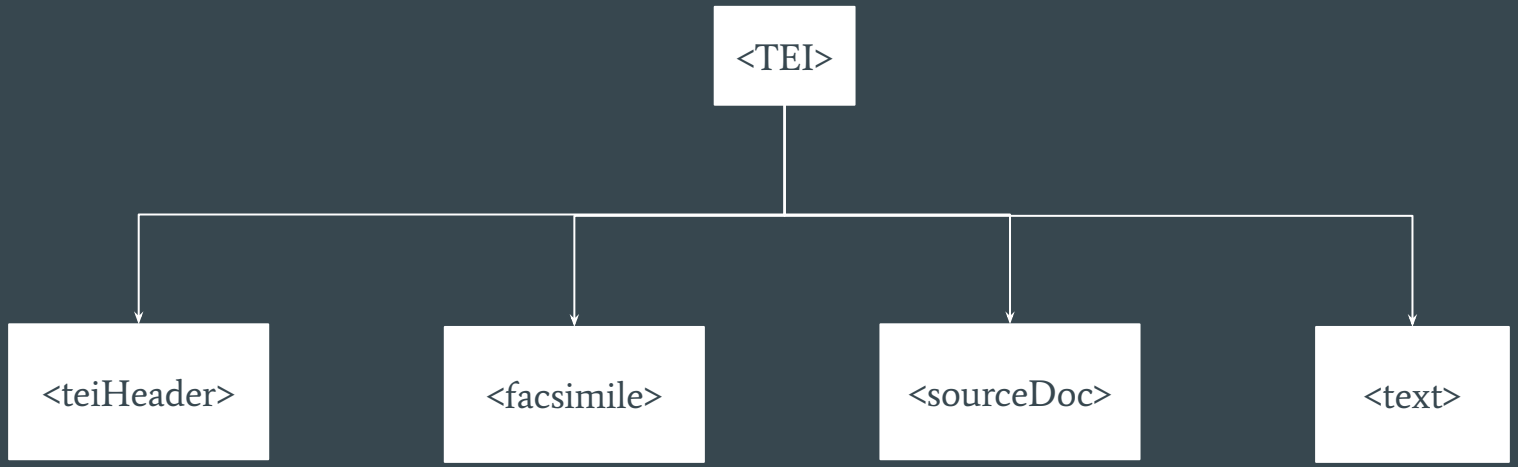
- Defining <zone>s is particularly useful for:
  - Encoding text that is further signified by its position on the page
  - Encoding MSS that have different blocks of composition
  - Encoding text in non-standard positions

Facsimile and transcription




```
<facsimile>
   <surface xml:id="frontispiece">
      <graphic url="1200px-Hamlet_Q1_Frontispiece_1603"/>
      <zone xml:id="title" ulx="152" uly="-16" lrx="1185" lry="405"/>
      <zone xml:id="subtitle" ulx="345" uly="425" lrx="944" lry="528"/>
      <zone xml:id="by-line" ulx="593" uly="553" lrx="1108" lry="624"/>
      <zone xml:id="blurb" ulx="161" uly="661" lrx="1188" lry="856"/>
      <zone xml:id="device" ulx="325" uly="869" lrx="972" lry="1564"/>
      <zone xml:id="publisher" ulx="269" uly="1593" lrx="1080" lry="1724"/>
   </surface>
</facsimile>
<text>
   <body>
      <pb facs="#frontispiece"/>
      <head type="title" facs="#title"><lb/>THE <lb/>Tragicall Historie of<lb/> HAMLET</head>
      <head type="subtitle" facs="#subtitle" rend="italic">Prince of Denmark</head>
      <head type="by-line" facs="#by-line">By William Shakespeare.</head>
      <p facs="#blurb">
         <lb/>As it hath beene diuerse times acted by his Highnesse ser-
         <lb/>uants in the Cittie of London : as also in the two V-
         <lb/>niuersities of Cambridge and Oxford, and else-where
      </p>
      <p facs="#device"><graphic url="device.png"/></p>
      <p facs="#publisher"><lb/>At London printed for N.L. and Iohn Trundell. <lb/>1603</p>
   </body>
</text>
```

Transcription: structure



Documentary transcription

- <sourceDoc> contains a documentary, embedded or faithful transcription of a single document, prioritising its physical disposition and/or its genetic history
  - Like <facsimile>, a <sourceDoc> usually contains one or more <surface> elements, each with potentially multiple <zone>s or <line>s
    - <line> —> contains the transcription of a topographical line
- Some editorial markup is allowed (<add>, <del>, <unclear>, etc.) but any interpretative markup should be avoided

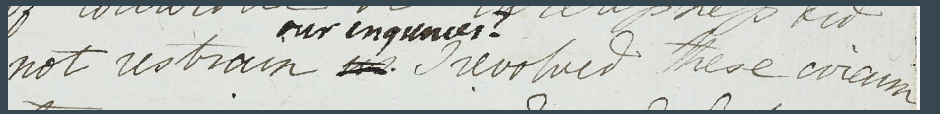
Documentary transcription

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
      <!-- METADATA! -->
  </teiHeader>
  <sourceDoc>
     <surface>
        <zone>
           <line>
              <!-- transcription of first line goes here -->
           </line>
           <line>
              <!-- transcription of second line goes here -->
           </line>
           <!-- and so on -->
        </zone>
     </surface>
  </sourceDoc>
</TEI>
```

Genetic Editing

- Genetic editing is concerned primarily with the order of composition
- There are several elements in the TEI that can be used for genetic editing, such as:
  - <mod> —> a familiar one, general modification element
  - <metamark> —> any kind of mark designed to guide the reading of a document
    - (i.e., arrows, lines, index symbols, etc.)
  - <retrace> —> any writing that has been rewritten or otherwise fixed or reinforced
    - (i.e., for example tracing in pen what had been written in pencil)
  - <undo> and <redo> —> textual modifications that have been reversed and/or reinstated
  - <transpose> and <transposeGrp> —> groups of text that have their order shifted

Genetic editing

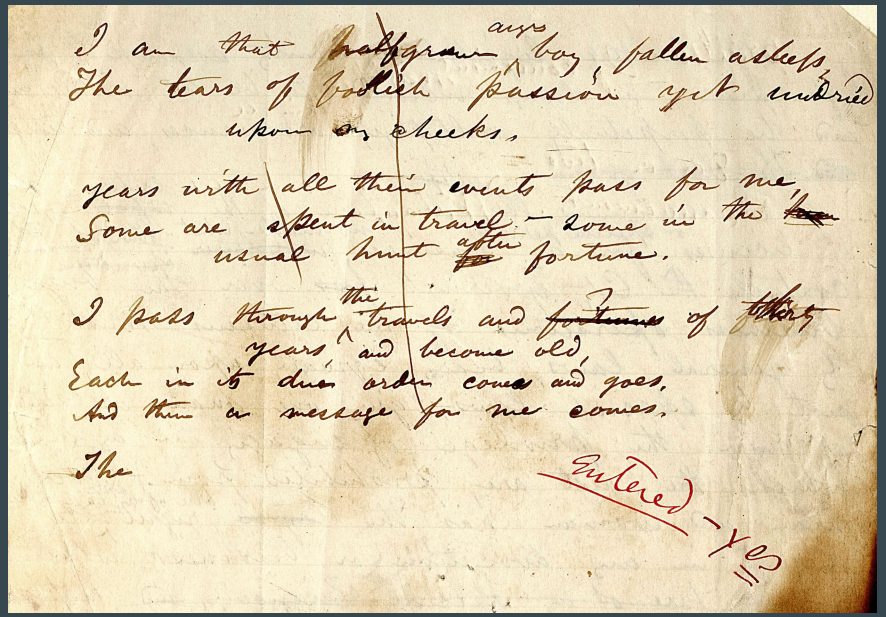



```
<line>
    not restrain
    <mod change="#stage1">
        <del>us.</del>
        <add>our enquiries></add>
    </mod>
    I revolved these circum
</line>
```

Metamark

- Any kind of mark that guides the reading rather than being part of the text
  - Arrows, crosses, asterisks, other symbols
- More specificity can be added by using the attributes:
  - @function —> to specify the function of the metamark
    - transposition, deletion, insertion, status, etc.
  - @target —> identifies one or more elements to which the function indicated by the metamark applies

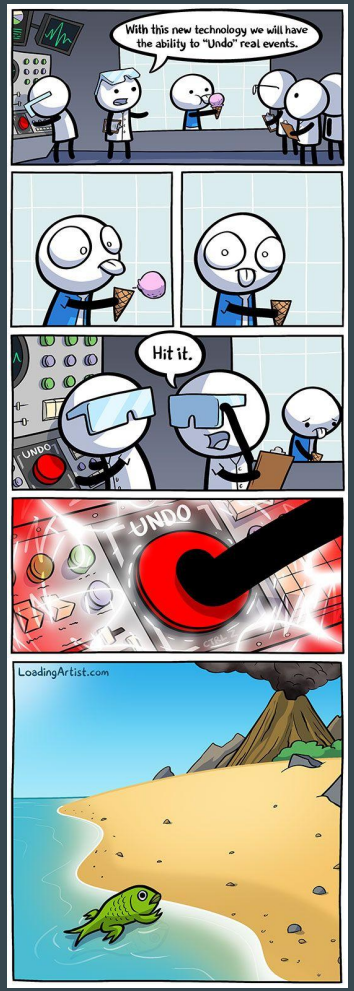
Metamark



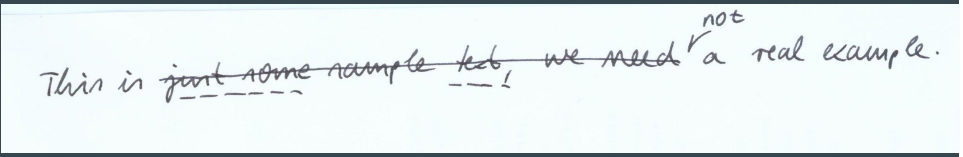

```
<surface>
   <metamark function="used" rend="line"
      target="#X2"/>
   <zone xml:id="X2">
      <line>I am that halfgrown <add>angry</add> boy, fallen asleep</line>
      <line>The tears of foolish passion yet undried</line>
      <line>upon my cheeks.</line>
      <!-- ... -->
      <line>I pass through <add>the</add> travels and <del>fortunes</del> of
         <retrace>thirty</retrace>
      </line>
      <line>years and become old,</line>
      <line>Each in its due order comes and goes,</line>
      <line>And thus a message for me comes.</line>
      <line>The</line>
   </zone>
   <metamark function="used" target="#X2">Entered - Yes</metamark>
</surface>
```

<undo> and <redo>

- An alteration that either has been
  - cancelled —> <undo>
    - @target attribute points to the elements which are to be reverted
  - reinstated —> <redo>
    - @target attribute points to the elements which are to be reasserted



<undo> and <redo>

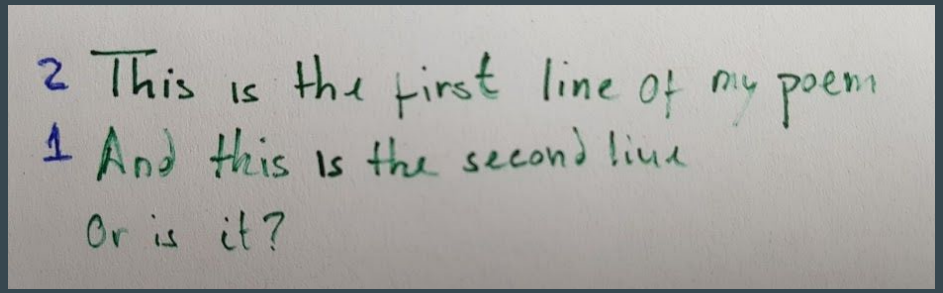



```
<line>
   This is
   <del change="#s2" rend="overstrike">
      <undo spanTo="#Xa" rend="dotted" change="#s3"/>
      just some
      <anchor xml:id="Xa"/>
      sample
      <undo spanTo="#Xb" rend="dotted" change="#s3"/>
      text,
      <anchor xml:id="Xb"/>
      we need
   </del>
   <add change="#s2">not</add>
   a real example.
</line>
```

Transpositions

- Passages that should be moved to a different position
- Often associated with metamarks indicating the change
- <transpose> element indicates a single textual transposition of at least two pointers (<ptr>), recording the order in which the elements should be read
- <transpose> must be included within a <listTranspose> element:
  - <listTranspose> groups a list of <transpose> elements occurring in the document. This list can be:
    - Embedded in the transcription
    - Part of the <teiHeader>, under <profileDesc>

Transpositions

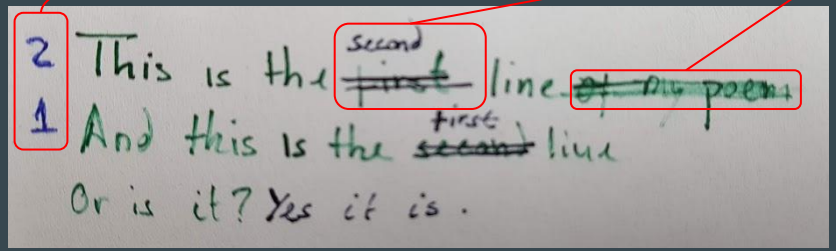



```
<line xml:id="line1">
   <metamark function="transposition" place="margin" change="#stage2">2</metamark>
   This is the first line of my poem
</line>
<line xml:id="line2">
   <metamark function="transposition" place="margin" change="#stage2">1</metamark>
   And this is the second line
</line>
<line xml:id="line3"> Or is it? </line>
<listTranspose>
   <transpose>
      <ptr target="#line2"/>
      <ptr target="#line1"/>
   </transpose>
</listTranspose>
```

Genetic editing — recording the process

- Genetic editing is concerned with order of composition
- Ascertaining that order is the editor's interpretative act
- The editor can record their 'best guess' by indicating different stages of composition:
  - <listChanges> —> groups a list of revision phases
    - @order attribute can record whether the order of changes is significant (if known) or not (if unknown)
  - <change> —> describes a single revision phase
    - @xml:id —> uniquely identifies that phase
  - As the list of changes is metadata, it is part of <profileDesc><creation> in the <teiHeader>

Transpositions

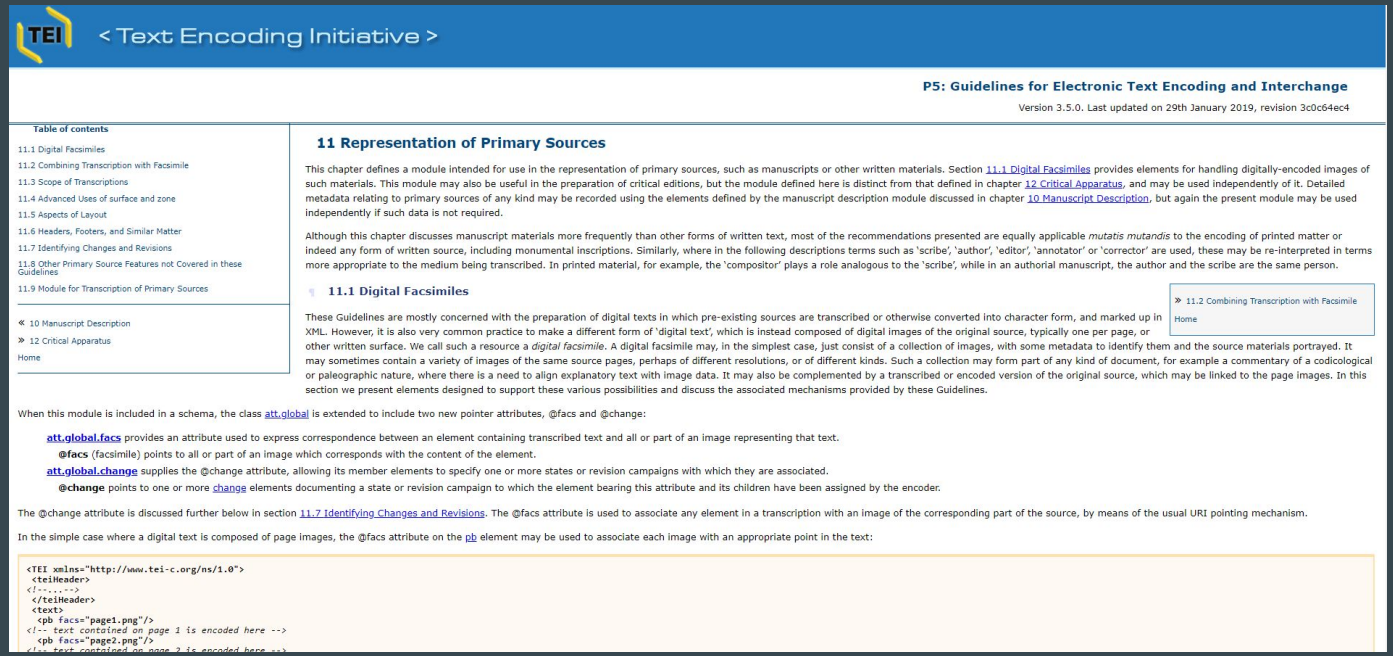



```
<profileDesc>
   <creation>
      <listChange>
         <change xml:id="stage1">First revision, green ink</change>
         <change xml:id="stage2">Second revision, blue ink</change>
         <change xml:id="stage3">Third revision, black ink</change>
      </listChange>
   </creation>
</profileDesc>
```

```
<line xml:id="line1">
   <metamark function="transposition" place="margin" change="#stage2">2</metamark> This is
   the <mod change="#stage3"><del>first</del><add>first</add></mod> line <del
      change="#stage1">of my poem</del>
</line>
<line xml:id="line2">
   <metamark function="transposition" place="margin" change="#stage2">1</metamark> And this
   is the <mod change="#stage3"><del>second</del><add>first</add></mod> line </line>
<line xml:id="line3"> Or is it? <add change="#stage3">Yes it is.</add></line>
<listTranspose>
   <transpose change="#stage2">
      <ptr target="#line2"/>
      <ptr target="#line1"/>
   </transpose>
</listTranspose>
```

I want to know more!

- Chapter 11 of the TEI guidelines for more about representation of primary sources



< Text Encoding Initiative >

P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.5.0. Last updated on 29th January 2019, revision 3c0c64ec4

Table of contents

11.1 Digital Facsimiles
11.2 Combining Transcription with Facsimile
11.3 Scope of Transcriptions
11.4 Advanced Uses of surface and zone
11.5 Aspects of Layout
11.6 Headers, Footers, and Similar Matter
11.7 Identifying Changes and Revisions
11.8 Other Primary Source Features not Covered in these Guidelines
11.9 Module for Transcription of Primary Sources

« 10 Manuscript Description
» 12 Critical Apparatus
Home

11 Representation of Primary Sources

This chapter defines a module intended for use in the representation of primary sources, such as manuscripts or other written materials. Section 11.1 Digital Facsimiles provides elements for handling digitally-encoded images of such materials. This module may also be useful in the preparation of critical editions, but the module defined here is distinct from that defined in chapter 12 Critical Apparatus, and may be used independently of it. Detailed metadata relating to primary sources of any kind may be recorded using the elements defined by the manuscript description module discussed in chapter 10 Manuscript Description, but again the present module may be used independently if such data is not required.

Although this chapter discusses manuscript materials more frequently than other forms of written text, most of the recommendations presented are equally applicable *mutatis mutandis* to the encoding of printed matter or indeed any form of written source, including monumental inscriptions. Similarly, where in the following descriptions terms such as 'scribe', 'author', 'editor', 'annotator' or 'corrector' are used, these may be re-interpreted in terms more appropriate to the medium being transcribed. In printed material, for example, the 'compositor' plays a role analogous to the 'scribe', while in an authorial manuscript, the author and the scribe are the same person.

11.1 Digital Facsimiles

» 11.2 Combining Transcription with Facsimile
Home

These Guidelines are mostly concerned with the preparation of digital texts in which pre-existing sources are transcribed or otherwise converted into character form, and marked up in XML. However, it is also very common practice to make a different form of 'digital text', which is instead composed of digital images of the original source, typically one per page, or other written surface. We call such a resource a *digital facsimile*. A digital facsimile may, in the simplest case, just consist of a collection of images, with some metadata to identify them and the source materials portrayed. It may sometimes contain a variety of images of the same source pages, perhaps of different resolutions, or of different kinds. Such a collection may form part of any kind of document, for example a commentary of a codicological or paleographic nature, where there is a need to align explanatory text with image data. It may also be complemented by a transcribed or encoded version of the original source, which may be linked to the page images. In this section we present elements designed to support these various possibilities and discuss the associated mechanisms provided by these Guidelines.

When this module is included in a schema, the class att.global is extended to include two new pointer attributes, @facs and @change:

att.global.facs provides an attribute used to express correspondence between an element containing transcribed text and all or part of an image representing that text.
@facs (facsimile) points to all or part of an image which corresponds with the content of the element.
att.global.change supplies the @change attribute, allowing its member elements to specify one or more states or revision campaigns with which they are associated.
@change points to one or more change elements documenting a state or revision campaign to which the element bearing this attribute and its children have been assigned by the encoder.

The @change attribute is discussed further below in section 11.7 Identifying Changes and Revisions. The @facs attribute is used to associate any element in a transcription with an image of the corresponding part of the source, by means of the usual URI pointing mechanism.

In the simple case where a digital text is composed of page images, the @facs attribute on the pb element may be used to associate each image with an appropriate point in the text:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!--...-->
 </teiHeader>
 <text>
  <pb facs="page1.png"/>
<!-- text contained on page 1 is encoded here -->
  <pb facs="page2.png"/>
<!-- text contained on page 2 is encoded here -->
```