Markup, XML, TEI

Character, structure, guidelines

Content

- Markup language ↓ (3-8)
- Text Encoding Initiative ↓ (9-25)
- TEI Guidelines ↓ (26-31)
- References ↓ (30)

Procedural Markup

- Visual, typographic, representation;
- WYSWYG: What You See is *ALL* you Get.

Descriptive Markup

- Semantic, funcional, structured;
- Meaning before format.

"**Miguel de Cervantes** wrote *Don Quixote*"

Procedural

```
<bold>Miguel de Cervantes</bold> wrote <italic>Don Quixote</italic>
```

Descriptive

Markup
```
<name>Miguel de Cervantes</name> wrote <title>Don Quixote</title>
```

Visualization
```
name {font-weight: bold;} title {font-style: italic;}
```

Process
```
<xsl:for-each select="//name">
```

XML Golden Rules

An element consists of a **start-tag,** optional **content,** and an **end-tag.**
`<persname>Miguel de Cervantes</persname>`

An element may have **attributes,** each consisting of a name and a value.
`<name type="person">Miguel de Cervantes</name>`

An XML document may contain **processing** instructions or **comments.**
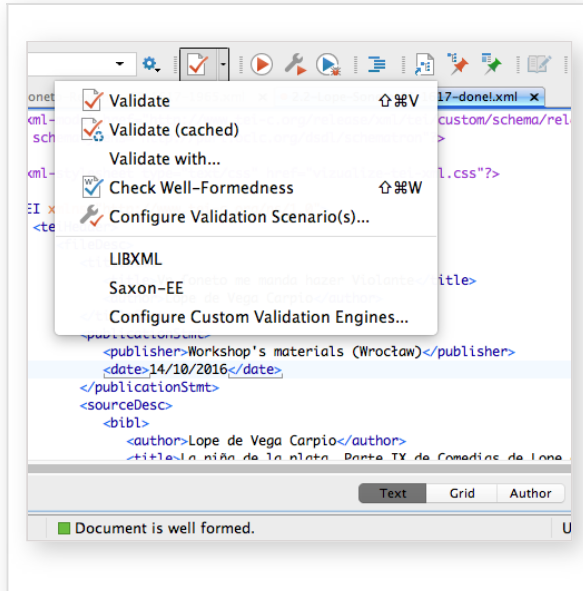`<? xml-stylesheet type="text/css" href="visualize-tei-xml.css"?>`

`<!-- This is a comment: here starts the first sonnet -->`

Elements and attributes are **case sensitive.**
`<name>` ≠ `<Name>`

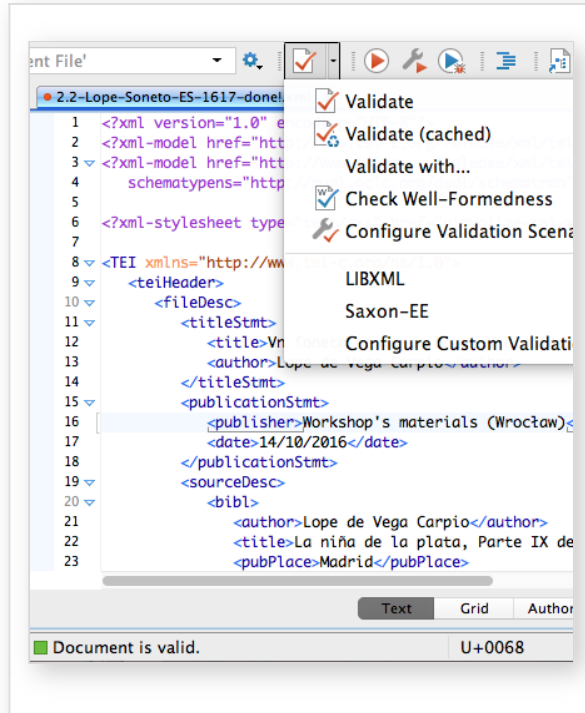**Empty elements** can use a minimized tag syntax.
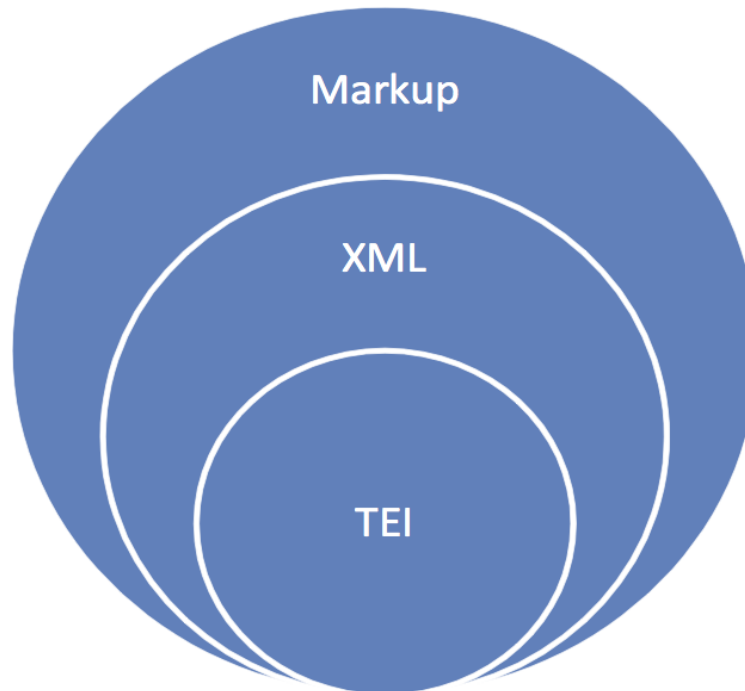`<pb/>` = `<pb></pb>`

Well-formed (good syntax)

- XML-Version, character set;
- A single root element (first node) contains the XML document;
- Each element is nested within the root element in a hierarchical structure: no overlapping;
- Elements, attributes, etc., are case sensitive;
- Start-tags and end-tags are mandatory;
- Attribute values are quoted.

Valid (against the schema)

- Well-formed;
- Follows a structural definition (schema), e.g. TEI;
- Which elements and attributes are used;
- Which elements and attributes can appear where;
- Names, datatypes, and default values for all attributes.

Markup > XML > TEI

Markup

XML

TEI

2

C. Schöch – Digitale Textedition mit TEI (Tutorial)

– TEI: Text Encoding Initiative

- Development of markup for texts;
- International standard;
- Interchange of data;
- Platform independence;
- Preservation;
- Research community.
- TEI P5: Version 3.5.0. Last updated on 29th January 2019

Modular system

Each module defines a group of elements and attributes

TEI basic structure

```
<TEI>
    <teiHeader> ⬤⬤⬤ </teiHeader>
    <text> ⬤⬤⬤ </text>
</TEI>
```

TEI basic structure

<teiHeader>

```
<TEI>
    <teiHeader>
        <fileDesc> ••• </fileDesc>
    </teiHeader>
    <text> ••• </text>
</TEI>
```

TEI basic structure

<teiHeader>

```
<TEI>
    <teiHeader>
        <fileDesc>
            <titleStmt> ▭ </titleStmt>
            <publicationStmt> ▭ </publicationStmt>
            <sourceDesc> ▭ </sourceDesc>
        </fileDesc>
    </teiHeader>
    <text> ▭ </text>
</TEI>
```

Content in the <teiHeader>

```
<teiHeader>
  <fileDesc>
      <titleStmt>
            <title>EXERCISE 01 · Encodage de base</title>
            <author>José Luis Losada</author>
       </titleStmt>
      <publicationStmt>
            <publisher></publisher>
            <date when="2019-04-25">25/04/2019</date>
       </publicationStmt>
      <sourceDesc>
            <bibl><title>La perle de Séville</title></bibl>
      </sourceDesc>
  </fileDesc>
</teiHeader>
```

TEI basic structure

<text>

```
<TEI>
    <teiHeader> ⬬ </teiHeader>
    <text>
        <front> ⬬ </front>
        <body> ⬬ </body>
        <back> ⬬ </back>
    </text>
</TEI>
```

TEI basic structure

TEI basic structure

<text>

```
<TEI>
 <teiHeader> ••• </teiHeader>
 <text>
        <body>
            <div>
                <head></head>
                <p></p>
            </div>
        </body>
    </text>
</TEI>
```

Content in the <text>

```
<text>
    <body>
        <div>
            <head>AVERTISSEMENT</head>
            <p>Je demande pardon aux lecteurs français
d'avoir choisi, parmi le nombreuses pièces de Lope
de Vega, celle-ci, où le précepte des trois unités
semble avoir été plus scandaleusement violé</p>
        </div>
    </body>
</text>
```

Classes of elements

- grouped semantically: names, places, titles, …
- grouped structurally: divisions, paragraphs, verses,…

Verse and name annotation

```
<body>
<lg>
  <l><persName>Doris</persName>, qui sait qu'aux vers quelquefois je me pl
ais,</l>
  <l>Me demande un sonnet, et je m'en désespère:</l>
  <l>Quatorze vers, grand Dieu! Le moyen de les faire?</l>
  <l>En voilà cependant déjà quatre de faits.</l>
</lg>
</body>
```

Atributtes inside of elements

Provides a unique identifier for any element:
  @xml:id

Provides a name or number for any element:
  @n

Specifies the language of any element:
  @xml:lang

Characterizes the element in some sense:
  @**type**

Specifies the visual appearance (rendition):
  @rend, @style, @rendition

Atributtes and elements

```
<body>
   <lg type="quatrain">
      <l n="1">Doris, qui fait qu'aux vers quelquefois je me plais,</l>
      <l n="3">Me demande un Sonnet; & je m'en deſeſpere:</l>
      <l n="4">Quatorze vers, grand Dieu! le moyen de les faire!</l>
      <l n="5">En voilà cependant desja quatre de faits.</l>
   </lg>
</body>
```

Metatextual markup for editions

Enriched with corrections, regularizations, expansions

Groups alternative encodings:
`<choice> </choice>`

Apparent error:
`<sic> </sic>`

Corrected error:
`<corr> </corr>`
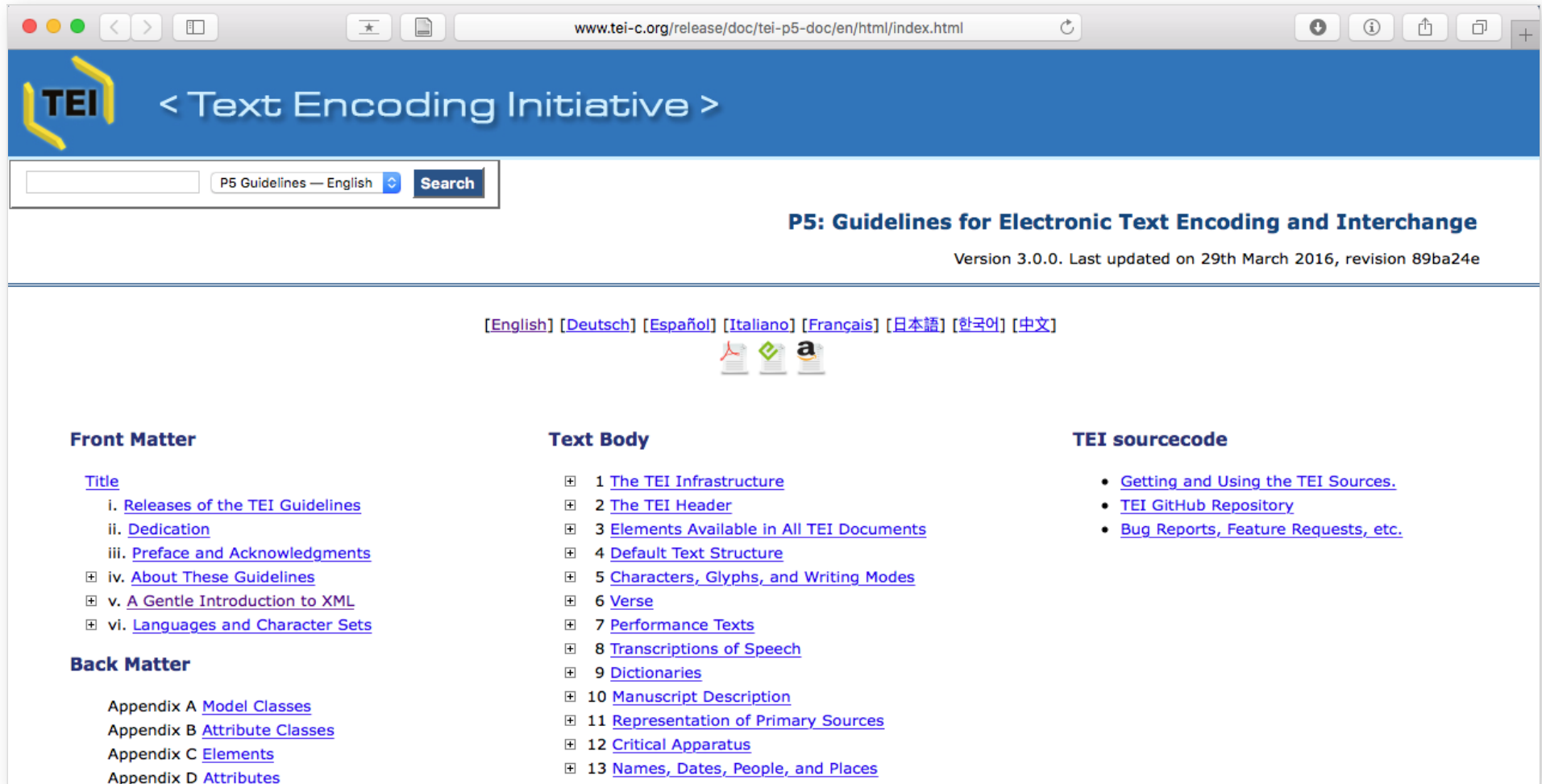
Original form:
`<orig> </orig>`

Regularized form:
`<reg> </reg>`

Markup for corrections

```
1   <body>
2   <lg type="quatrain">
3       <l n="1">Doris, qui fait qu'aux vers quelquefois je me plais,</l
    >
4       <l n="3">Me demande un Sonnet; & je m'en defefpere:</l>
5       <l n="4">Quatorze vers, grand Dieu! le moyen de les faire!</l>
6       <l n="5">En voilà cependant
7               <choice>
8                       <orig>desja</orig>
9                       <reg>déjà</reg>
10
11              </choice>

            quatre de faits.</l>
12
    </lg>

13  </body>
```

Critical Apparatus, Dictionaries, Manuscript Description, Bibliographies, …

nearly 500 elements

Guidelines

Information, eg., on an element

\<sourceDesc\>

\<sourceDesc\> (source description) describes the source from which an electronic text was derived or generated, typically a bibliographic description in the case of a di phrase such as "born digital" for a text which has no previous existence. [2.2.7 The Source Description]

Module	header — The TEI Header
Attributes	att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corr @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change (att.global.responsibility (@cert, @resp)) att.declarable (@default)
Contained by	**header:** biblFull fileDesc
May contain	**core:** bibl biblStruct list listBibl p **figures:** table **header:** biblFull **linking:** ab **msdescription:** msDesc **namesdates:** listEvent listNym listOrg listPerson listPlace **spoken:** recordingStmt scriptStmt **textcrit:** listApp listWit
Example	``` <sourceDesc> <bibl> <title level="a">The Interesting story of the Children in the Wood</title>. In <author>Victor E Neuberg</author>, <title>The Penny Histories</title>. <publisher>OUP</publisher> <date>1968</date>. </bibl> </sourceDesc> ```

Quick look up in the TEI *Guidelines*.

TEI-Look up

TEI Roma (tool)

Generating customizations for the TEI

- possibility to choose your modules;
- possibility to create new elements;
- possibility to generate your schema;
- possibility to generate your documentation.

References

- *A Gentle Introduction to XML*, Guidelines for Electronic Text Encoding and Interchange, TEI P5: Version 3.5.0.
- *digital.humanities @ Oxford*, TEI @ Oxford Teaching Pages.
- Lou Burnard (2014), *What is the Text Encoding Initiative? How to add intelligent markup to digital resources* , Marseille: OpenEdition Press.
- Christof Schöch (2014), *Digitale Textedition mit TEI*, DARIAH TEI-Tutorial.

José Luis Losada Palenzuela, *Markup, XML, TEI. Character, structure, guidelines,* <http://editio.github.io/presentations/xml-tei>, 2019.