Encoding for ELTeC : a summary guide

1 The checklist

Here's a list of the information you must supply in your header:

- Identification of title and language (*@xml:id* and *@xml:lang* on `<TEI>`)

- Title of the work, followed by the phrase ': ELTeC edition' (`<title>`)

- Author, in format 'Surname, Forename/s (birthYear-deathYear)' (`<author>`)

- Statements of responsibility in format 'encoded by Name' (`<respStmt>`)

- Date of publication in ELTeC (`<publicationStmt>`)

- Source description containing one or more `<bibl>` elements

- Encoding level : use the *@n* attribute of `<encodingDesc>`

- Profile description specifying languages used and sampling criteria values (`<langUsage>`, `<textDesc>`)

- Revision description containing at least one `<change>`

Most but not all requirements are enforced by the schema

2 Identification

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="de" xml:id="deu025">
```

- *@xml:id* values must be unique across the corpus, preferably short, and valid XML names

- recommended format: 'xxxyyyyd' where 'xxx' is ISO 639-3 language code, 'yyyy' the year of publication, and 'd' an extra digit for disambiguation

- the *@xmlns* attribute is needed for TEI conformance

- the value of *@xml:lang* is the ISO 639-2 language code

3 The title

```
<title> Ghiță Cătănuță, căpitan de haiduci : ELTeC edition</title>
<title ref="viaf:180462896">L'Éducation sentimentale : édition ELTeC </title>
<title>Господа сељаци: приповетка : ELTeC издање</title>
```

- Use standard form of title, simplified if necessary, followed by the phrase 'ELTeC edition' in appropriate language

- Optionally supply VIAF or other identifying code for title

4 The author

```
<author ref="viaf:26018946">Skene,
Felicia (1821-1899)</author>
<author ref="viaf:52610012">Gozsdu Elek (1849-1919) </author>
<author ref="viaf:37053113">Dombre, Roger (Marguerite Sisson) (1859-1914)</author>
```

- VIAF codes available from **??**ttps://viaf.org

- Standard forms for names should (or for E Europe may not) include comma

- Alternative name may be provided between names and dates

- Dates must be separated by hyphen

Check details on Wikipedia, Worldcat, National library etc.

5 The encoder(s)

```
<respStmt>
 <resp>ELTeC conversion</resp>
 <name>Your Name Here</name>
</respStmt>
<respStmt>
 <resp>Original data capture</resp>
 <name>Their Name
   Here</name>
</respStmt>
```

- provide a `<resp>` containing 'ELTeC conversion' to indicate responsibility for converting from non-ELTeC digital source

- optionally also indicate responsibility for original data capture (e.g. in HTML) where this is not indicated in source description

- `<resp>` containing 'principal' may be used to indicate a principal researcher if required

- supply the `<name>` in a consistent form

6 Extent statement

```
<extent>
 <measure unit="words">99479</measure>
 <measure unit="pages">631</measure>
 <measure unit="vols">2</measure>
</extent>
```

- oXygen will calculate a word count for you

- page and volume counts are optional; wordcount is mandatory

- figures must be supplied without punctuation or spaces

7 Publication statement

```
<publicationStmt>
 <p>First ElTeC edition <date>2019-02-3</date>
 </p>
</publicationStmt>
```

- The date of publication in ELTeC must be supplied in form "YYYY-MM-DD"

- When the text is published on Zenodo, this is automatically converted to a fuller standard form:

```
<publicationStmt>
 <distributor>COST Action ELTeC</distributor>
 <date when="2019-10-03"/>
 <availability>
  <licence target="https://creativecommons.org/licenses/by/4.0/"> Licenced under
     CC-BY 4.0 </licence>
 </availability>
 <ref type="doi"
  target="10.5281/zenodo.8468"/>
</publicationStmt>
```

8 Source description (1)
The ELTeC edition is derived from a known print edition

```
<sourceDesc>
 <bibl type="printEdition">
  <author>Panait Macri</author>
  <title>Ghiță Cătănuță, căpitan de haiduci</title>
  <pubPlace>Bucharest</pubPlace>
  <publisher>Editura Librăriei și Tipografiei H. STEINBERG</publisher>
  <date>1923</date>
 </bibl>
</sourceDesc>
```

- Add an extra `<bibl type="firstEdition">` if this is not the first edition

- Add an extra `<bibl type="digitalSource">` if another digital version of the text is also available e.g. as page images

9 Source description (2)
The ELTeC edition is derived from a pre-existing digital version which is in turn derived from the first printed edition.

```
<sourceDesc>
 <bibl type="digitalSource">
  <title>Wuthering Heights (1st edition) : wikisource edition</title>
  <ref target="https://en.wikisource.org/wiki/Wuthering_Heights_(1st_edition)"/>
 </bibl>
 <bibl type="firstEdition">
  <title>Wuthering Heights. A novel by</title>
  <author>Ellis Bell</author>
  <publisher>London: T. C. Newby</publisher>
```

```
  <date>1847</date>
 </bibl>
</sourceDesc>
```

- Add an extra `<bibl type="printEdition">` if the digital version does not derive from the first edition

10   Encoding description

```
<encodingDesc n="eltec-1">
 <p/>
</encodingDesc>
```

- The value for *@n* must be one of 'eltec-0', 'eltec-1', or 'eltec-2'

- An empty `<p>` is required as content :-(

- We may add more here later

11   Profile description

```
<profileDesc
    xmlns:eltec="http://distantreading.net/eltec/ns">
 <langUsage>
  <language ident="eng"/>
 </langUsage>
 <textDesc>
  <eltec:authorGender key="F"/>
  <eltec:size key="long"/>
  <eltec:canonicity key="high"/>
  <eltec:timeSlot key="T1"/>
 </textDesc>
</profileDesc>
```

- Supply a `<language>` for each language present in significant amounts in the text

- Use ISO 646 three-letter codes to identify the language, optionally extended with script or dialect

- The namespace prefix `eltec` is arbitrary and can be defined wherever you please

- Values for the four balance criteria must be supplied

12   Balance criteria

authorGender	**F** (female)		**M** (male)	**U** (unknown or multiple)	
	author's biological sex, not claimed gender				
size	**short** (10-50k)		**medium** (50-100k)	**long** (>100k)	
	word count excluding header and all markup				
timeSlot	**T1** (1840-1859)		**T2** (1860-1879)	**T3** (1880-1899)	**T4** (1900-1920)

	date first published as a book		
canonicity	**low** (reprinted infrequently or never)	**high** (reprinted very frequently)	**unmarked** (reprint information not available)
	number of times reprinted between 1970 and 2009		

13 Revision description

```
<revisionDesc>
 <change when="2018-11-01">Initial conversion to ELTeC1 using CLIGStoELTeC
   stylesheet</change>
</revisionDesc>
```

- *@when* attribute is required and must have value in form YYYY-MM-DD

- Give <change>s in reverse chronological order (most recent first)

- Include only major changes