TEI and other Schemas

Elli Mylonas
elli.mylonas@gmail.com

BærUt!
Sustainable Digital Scholarly Editions

This Presentation: http://tiny.cc/BaerUT-TEI

The Text Encoding Initiative began in 1987 as an effort to develop guidelines for the encoding of literary and historical texts in order to facilitate interchange and re-use. It is designed to be customizable, so projects may use different subsets, or even variations of the TEI schema.

The TEI is:

- An XML schema (588 Elements)
- Guidelines that explain how to use the elements defined in the schema
- A community of practice

Exploring the TEI—Website and Community

- TEI home: https://tei-c.org/
- TEI Guidelines: https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html
- TEI Listserv: https://tei-c.org/support/#tei-l
- Annual Meeting
- Journal of the TEI: https://journals.openedition.org/jtei/
- Also, Community calls, SIGS, membership.
- The TEI schema and guidelines are maintained by a volunteer Technical Council, and the TEI is governed by a volunteer Board.

Exploring the TEI—Guidelines

- Modules: groupings of elements for different types of document genres and different disciplinary approaches
- Narrative description (semantics): what each element represents
- Authority: how to use the elements.
- Examples: The Guidelines provide many small examples throughout

Exploring the TEI—Guidelines TOC

1. The TEI Infrastructure
2. The TEI Header
3. Elements Available in All TEI Documents
4. Default Text Structure
5. Characters, Glyphs, and Writing Modes
6. Verse
7. Performance Texts
8. Transcriptions of Speech
9. Computer-mediated Communication
10. Dictionaries
11. Manuscript Description
12. Representation of Primary Sources
13. Critical Apparatus
14. Names, Dates, People, and Places
15. Tables, Formulæ, Graphics, and Notated Music
16. Language Corpora
17. Linking, Segmentation, and Alignment
18. Simple Analytic Mechanisms
19. Feature Structures
20. Graphs, Networks, and Trees
21. Non-hierarchical Structures
22. Certainty, Precision, and Responsibility
23. Documentation Elements
24. Using the TEI

Exploring the TEI—Schema (sort of)

The Guidelines are a narrative, contextualized definition of the elements in the TEI

Appendix C of the Guidelines also has reference pages for each element which provide a formal definition of the element, its attributes, its parents and children.

https://tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html

Don't rely on the Element Specification pages.
Read the Guidelines as well.

The Text Encoding Initiative is an XML schema whose semantics are provided as a set of prose guidelines, used by a community of practice.

The Structure of a TEI Document

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <teiHeader>
        [metadata about the edition and encoding]
    </teiHeader>
    <facsimile>
        [links to graphics]
    </facsimile>
    <standOff>
        [annotations, authority lists, linked data]
    </standOff>
    <text>
        [the encoded document]
    </text>
</TEI>
```

The <teiHeader>

- **<fileDesc>** Contains information about the digital document and about the source document including the bibliographic description of the source.
- **<encodingDesc>** and **<profileDesc>** provide information about the encoding and about usages such as calendars and languages used in the document
- **<revisionDesc>** contains the revision history of the document. Who made which change, and when. Very important!!
- And more...

The <text>

<text>
  <front>[front matter]<front>
  <body>[body of the text]</body>
  <back>[back matter]</back>
</text>

Customization

- TEI is a very large and permissive schema
- It is designed to be customizable
- It is possible to eliminate modules, elements and attributes
  - encoding is more constrained
- It is possible to add new elements, attributes and attribute values
  - To meet project needs
- TEI is described in a TEI language called ODD (One Document Does it all)
  - Roma is software that allows you to make TEI customizations

If there is an appropriate TEI element, you should use that element rather than creating a new one.

If you use a TEI element, you should use it with the semantics described in the Guidelines.

EpiDoc

A widely used customization designed for texts that are written on physical objects , but adopted by philologists as well - for example, Digital Latin Library.

- Documentation, a TEI schema customization and a community of practice
- Stylesheets that will format many EpiDoc compliant documents
- Tutorials and workshops
- EFES, a platform for publishing EpiDoc corpora

Guidelines: https://epidoc.stoa.org/gl/latest/
Markup List: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=EPIDOC-MARKUP
Tutorials: https://github.com/EpiDoc/Tutorials/wiki/All-tutorials
EFES: https://github.com/EpiDoc/EFES
Epidoc Home: https://sourceforge.net/p/epidoc/wiki/Home/

Some EpiDoc Projects

- Inscriptions of Roman Tripolitania: https://irt2021.inslib.kcl.ac.uk/en
- US Epigraphy: https: usepigraphy.brown.edu
- SigiDoc (Byzantine Sigillography): https://ifeb.sigidoc.huma-num.fr/fr/
- Corpus of the Inscriptions of Campā (Viet Nam): https://isaw.nyu.edu/publications/inscriptions/campa/inscriptions
- Damos is being transformed after the fact to EpiDoc
- Egyptian, Coptic, Runic and other inscribed texts are also in the works.

Project Best Practices

- Document encoding practices and decisions
- Document workflows and software
- Make documentation public
- Make code and data public as much as possible
- Be clear about permissions and acknowledgements
- Archive data and code in trusted repositories

The process is research as well as the product.

Explore a Well Documented Project

Livingstone Online
https://livingstoneonline.org/in-his-own-words/the-livingstone-online-digital-collection

Menota
https://www.menota.org/HB3_index.xml

CoReMa
https://gams.uni-graz.at/context:corema (although for different purposes)