TEI – General introduction

Hans Clausen

CLARIAH-AT
TEI Introductory School
10.-13. September 2019

Venue:
Centre for Information Modelling (ZIM)
Elisabethstraße 59/III, SR 81.31
Programm

Tuesday 10.9.
9:00-10:30  Digital Scholarly Editing
            Roman Bleier, Martina Scholger
11:00-12:30 Textencoding with XML
            Hans Clausen, Gerlinde Schneider
14:00-15:30 TEI – General Introduction
            Hans Clausen, Gerlinde Schneider
16:00-17:30 Practice: XML/TEI
            Hans Clausen, Gerlinde Schneider
The Need for a Standard

• You have created different encoding schemes
  — Problems?
    • Different languages
    • Ambiguity (for human and machine)

• November 1987: Meeting at Vassar College, Poughkeepsie
  — Possibility of creating a standard encoding scheme
  — Founding of the Text Encoding Initiative (TEI)
  — Agreed on a set of principles („the Poughkeepsie Principles“)
    • Standard format for data interchange
    • Hardware and Software independent
Text Encoding Initiative (TEI) – a short history

• 1987: The Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange started to work
• Standard Generalized Markup Language (SGML)
  — Encoding method, which could handle many different types of text
• Iterative cycles of work
  — Work groups for specific applications
• 1994: first full version of the TEI Guidelines
  — Published in print and electronic form
• eXtensible Markup Language (XML) instead of SGML
  • Since ca. 1998
  • XML is the successor of SGML
• 2000: TEI Consortium
  • Incorporated as a not-for-profit legal entity
TEI Today

• „Over nearly three decades the TEI has been extraordinarily successful at achieving its objective and it is now widely used by scholarly projects and libraries around the world.“ (Source: https://tei-c.org/about/history/)

• TEI is the *de facto* standard for text encoding in the humanities.

• Place of discussion on text

• Community
  – TEI-L discussion list: TEI-L-subscribe-request@LISTSERV.BROWN.EDU
The TEI Guidelines

- [https://tei-c.org](https://tei-c.org)
- Current Version: P5 (Version 3.6.0)

- Modules
  - Group elements that belong together
  - Represent certain aspects of texts
The TEI Guidelines – General Structure

• Each document has at least the following two aspects:
  – Metadata associated with the digital document:
    <teiHeader>
  – the document itself:
    <text>

Example XML:

```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```
The TEI Schema – Customization

• Often only a subset of the elements and attributes provided by the TEI is necessary
• You can customize the TEI Schema
  – Remove unnecessary elements (and attributes)
  – Add your own elements (and attributes)
  – High flexibility
  – Roma: https://roma2.tei-c.org
Exercise

1. Open a new file in the Oxygen Editor and choose the template "TEI P5" --> "All".
2. Take the structures of the poem from the previous unit.
3. Encode the poem with these structures using TEI elements and attributes.
4. Additional task: Encode some metadata to the poem in the <fileDesc> (<teiHeader>).
Exercise (hints)

• How to find the right elements and attributes?
  — text divisions:
    • Module „Default Text Structure“ → 4.1 Divisions of the body
  — headings:
    • Module „Default Text Structure“ → 4.2.1 Headings and Trailers
  — verse, line group:
    • Module „Elements Available in All TEI Documents“ → 3.12 Passages of Verse or Drama
  — Authorline: <byline>
  — Dedication: <div type=„dedication“>
  — Attributes:
    • Classification of Elements: @type
    • How the text was rendered in the source text: @rend