



TEI
Core

TEI als Modulset

- Quellenbeschreibung
 - Metadaten
- Textbeschreibung
 - strukturell, topographisch, typologisch
- Textannotation
 - Editionsmethoden

TEI als Modulset

- Module werden zu einem TEI-Schema kombiniert
- TEI all
= alle Module zusammen
- daneben TEI Lite
- eigene Schemata: ROMA

TEI Core

- **Kapitel 3** der Guidelines
- TEI Core = Elemente, die in allen TEI-Dokumenten verfügbar sind
- 12 Unterkapitel

TEI Core

- Absätze und Zeichensetzung
- Hervorhebungen
- Zitationen
- editorische Eingriffe
- Semantik und Links
- Listen, Annotationen und Abbildungen
- Referenzen, Bibliographie
- Vers- und Damentexte (!?)

3.1 Absätze

- `<p>` *paragraph*
 - geht fast immer, fast überall
 - weitere Unterteilungen:
 - `<s>` *sentence-like division* (Modul "analysis")
 - `<seg>` *segment* (Modul "linking")
- Konzept "Absatz" zu speziell?
 - `<ab>` *anonymous block* (Modul "linking")

3.2 Zeichensetzung

- Anführungszeichen:
 - `<q>` *quoted* (Text in Gänsefüßchen)
 - aber: `<quote>` *quotation* (Zitat)
- Trennungszeichen
 - als Attribut: `<lb break="no">`
- generische Interpunktion:
 - `<pc>` *punctuation character* (Modul "analysis")

3.3.1 Hervorhebung

- **<hi>** *highlighted* (generisch)
 - **<emph>** *emphasized*
 - beides wird gern kombiniert mit:
 - **@rend** = *rendition* (beschreibt die Hervorhebung)
 - **@rendition** (Verweis auf allgemeine Beschreibung)
- **<head>** Überschrift
 - muss das allererste Kind-Element sein
 - Pendant ist **<trailer>** (Modul "textstructure")

3.3.2 Sprache

- `<foreign>` etwas in anderer Sprache
- `<distinct>` "linguistically distinct"
- wird gern kombiniert mit:
 - `@xml:lang` = *language*
 - hier meist normierte Bezeichnungen, z.B. Two- oder Three-Letter-Code

3.3.3 Zitiertes

- `<q>` (Gänsefüßchen)
 - spezieller: `<said>` (Rede)
- `<quote>` (Erwähnung)
 - spezieller: `<cit>` (Zitation)
- `<mentioned>` (Erwähnung eines Begriffs)
 - `<soCalled>` ('einfache Anführungszeichen')

3.3.4 Fachtermini

- `<term>` Fachbegriff
- `<gloss>` Beschreibung eines Begriffs
- Elemente für die Beschreibung von XML:
 - `<altIdent>`
 - `<desc>`
 - `<equiv>`

3.4 [a] "Textweichen"

- `<sic>` und `<corr>` (seltsame Schreibung)
- `<orig>` und `<reg>` (Normalisierung)
- `<abbr>` und `<expand>` (Abkürzung)
 - darin noch `<ex>` und `<am>`
- `<choice>` (Weiche)

3.4 [b] Eingriffe am Text

- durch den Autorin oder Autor:
 - `<add>` (Hinzufügung), gern mit `@place`, z.B. "above"
 - `` (Tilgung), gern mit `@rend`, z.B. "strikethrough"
- durch den Bearbeiterin oder Bearbeiter:
 - `<supplied>` und `<surplus>`
 - `<gap>` und `<unclear>`
- von irgendjemandem: `<note>` (Anmerkung)
 - gern dazu verwendet: `@resp`

3.5.1 Namen

- `<rs>` *referencing string* (generisch)
- `<name>` Eigename
 - speziellere Elemente finden sich in Kap. 13: "Names, Dates, People, and Places"
- beides gern in Verbindung mit
 - `@key` (z.B. eine projektinterne ID)
 - `@ref` (Verweis auf `@xml:id` bzw. URI)

3.5.2 Adressen

- `<address>` (Anschrift), darunter:
 - `<email>`
 - `<addrLine>` einzelne Zeile einer Adresse
 - `<street>`
 - `<postCode>`
 - `<postBox>`
- dazu kommen meist Elemente aus Kap. 13: "Names, Dates, People, and Places"

3.5.3 Zahlen und Maße

- `<num>` Zahl
- `<measure>` Maßangabe
- `<measureGrp>` mehrere `<measure>`s
- `<unit>` Einheit
- gern in Verbindung mit:
 - `@value` = Zahlwert
 - `@type` = z.B. Prozent, Ordinalzahl, Kardinalzahl
 - `@unit` = Symbol (!)

3.5.4 Datum und Zeit

- `<date>` Datum
- `<time>` Uhrzeit
- gern in Verbindung mit:
 - `@when` = Standarddatum
 - `@notBefore/@notAfter` = nicht vor / nicht nach
 - `@from/@to` = von / bis
- auch mit dem Zusatz `-iso`

3.6 Hyperlinks

- `<ptr>` (Verweis ohne Text)
- `<ref>` (Verweis mit Text)
- `@target` gibt das "Ziel" an:
 - `@target="#abc"`:
verweist auf Element mit `@xml:id="abc"`
 - `@target="https://example.com"`:
verweist auf ein externes Dokument
 - ... und `@target="http://example.com#abc" ?`

3.7 Listen

- `<list>` "Container" für die Liste
- `<item>` ein Listeneintrag
- `<label>` Titel eines Eintrags
- `<head>` Titel einer Liste
- speziell: `<headLabel>` und `<headItem>`
= Spaltenüberschriften für Glossare

3.8 Annotierung

- **<note>** Anmerkung
 - gerne mit **@resp** (responsible)
 - gerne mit **@place**
z.B. für Fußnoten, Marginalien, ...
- **<index>** eine zu indexierende Stelle
 - z.B. für beliebige Registereinträge

3.9 Abbildungen

- `<figure>` irgendeine Abbildung
- `<media>` Audio und Video
- `<graphic>` Abbildung mit URL
- `<binaryObject>` Binärdaten (?!)

3.10 "Milestones"

- `<gb>` Buch/Bundeanfang
- `<pb>` Seitenanfang
- `<cb>` Spaltenanfang
- `<lb>` Zeilenanfang
- und generisch: `<milestone>`
- immer gerne verwendet
 - `@n` (number) = Zählung
 - `@xml:id` = ID, eindeutig im Dokument

3.11 Bibliographie

- `<bibl>` bibliographischer Eintrag
 - darunter: `<title>` `<author>` `<series>` ...
- `<biblFull>` dasselbe plus TEI-Elemente
- `<biblStruct>` streng strukturierter Eintrag
- `<listBibl>` Liste von Einträgen

3.12 Verstexte

- `<l>` *line* (ein Vers)
- `<lg>` *line group* (eine Versgruppe)
- und speziell für Dramentexte:
 - `<sp>` *speech* (direkte Rede)
 - `<speaker>` Sprecher
 - `<stage>` Regieanweisung



TEI
Dokument-
aufbau

TEI Dokumentaufbau

- `<text>`
- `<body>`
 - optional: `<front>` und `<back>`
- optional: `<div>` oder `<div1..9>`
 - diese gerne mit `@xml:id` versehen
- dann `<head>` `<p>` etc. etc. p. p.

1 Datei mit n Texten

- `<group>`
 - steht innerhalb von `<text>`
 - darunter beliebig viele `<text>`
- `<TEICorpus>`
 - als Root-Element
 - darunter beliebig viele `<TEI>`
- wo ist der Unterschied?