

# HTML als schlankere Alternative zu TEI-XML

Eckhart Arnold, [arnold@badw.de](mailto:arnold@badw.de)  
12. September 2022

---

- 
1. Ausgangssituation: Es gibt keine Technologie zum Selbstlernen und -machen für digitale Editionen!
  2. Verschiedene verfügbare und denkbare „zweitbeste“ Technologien.
  3. Exkurs: Warum TEI-XML kein Standard ist und nie einer werden wird.
  4. Warum HTML 5 genauso gut, aber einfacher als XML ist.
  5. Wie man HTML für digitale Transkriptionen und Editionen einsetzen und was man dabei besser machen als bei TEI-XML.

---

## Ziele (beinahe) jeder digitalen Edition:

1. Strukturierte Daten
2. Online-Ausgabe, unabhängig vom Ausgabegerät (also kein PDF)
3. Druckausgabe (bei Transkriptionen ggf. verzichtbar)

*Alle verfügbaren Technologien erfordern in jedem einzelnen Projekt dezidierte Techniker-Arbeitskraft, wenn alle drei Ziele (Daten, Online, Druck) erreicht werden sollen!*

Technologie	Daten	Druck	Online	Ökosystem	Modularisiert	Erlernbarkeit
Textverarbeitung	***	**	**	-	-	**
LaTeX	**	**	**	**	**	**
TEI-XML Ediarum?	**	**	**	**	**	**
CTE	**	**	**	-	-	**
RST u. andere	** / **	** / **	** / **	** / **	**	**
HTML 5	**	**	**	**	**	**

---

## Kriterien für einen Standard:

- **Interoperabilität:** Bei TEI-XML **nicht einmal theoretisch gegeben**.  
Grund: Üblicherweise arbeitet jedes Projekt nur mit einer Untermenge von TEI-XML. Unterschiedliche Untermengen sind aber nicht operabel und können in der Regel auch nicht auf eine triviale Weise (etwa durch Bildung der Vereinigungsmenge der zugelassenen Tags oder dergleichen) interoperabel gemacht werden.
- **Breite Unterstützung durch Softwarewerkzeuge:**  
Bei TEI **definitiv nicht gegeben**. Schon bei der Dateneingabe bleibt man praktisch auf ein einziges, kostenpflichtiges Werkzeug verwiesen.

Auch bei weiteren Aspekten (z.B. Verfügbarkeit von Dokumentation und Lehrbuchliteratur, Modularisierung) fällt die Bilanz mindestens durchwachsen aus. Mit XML stützt sich TEI zudem auf eine Technologie mit eher schwachem Software-Ökosystem.

Insbesondere: TEI hat mehr als 30 Jahre auf dem Buckel und in Richtung Standardentwicklung im Sinne der oben genannten Kriterien hat sich nichts getan!

Fazit: **TEI-XML bietet hinsichtlich der Nachnutzung und Interoperabilität keinerlei Vorteile gegenüber anderen Kodierungen, sei es in XML oder anderen strukturierten Formaten!**

Welche Arbeit muss von Technikerinnen erledigt werden, bevor eine digitale Edition im Sinne der Anfangs genannten drei Ziele realisiert werden kann?

Technologie	Daten	Online	Druck	Sofort loslegen
LaTeX	(geschenkt) (Abfallprodukt der HTML-Konvertierung)	HTML-Konvertierung (aufwändiger Teil: LaTeX-Parser)	geschenkt	ja
TEI-XML	geschenkt	HTML-Konvertierung	LaTeX-Konvertierung (Print-CSS?)	nein
HTML	geschenkt	(geschenkt) CSS noch erforderlich	LaTeX-Konvertierung (Print-CSS?)	beinahe

### Was TEI-XML und HTML 5 gleich gut können:

- Strukturierte Daten kodieren und Texte semantisch auszeichnen.
- Nachnutzbarkeit. (Keinerlei Vorteile dadurch, dass TEI-XML vermeintlich Standard ist!)

### Was HTML 5 und Gutes-Altes-XML (ohne TEI) besser können:

- Tipparbeit-reduzierende Tag-Sets!
- Verschachtelungsarme und „orthogonale“ Auszeichnung (heißt: derselbe Name für dieselbe Funktion in unterschiedlichen Kontexten, und ein Name für eine Funktion!)
- Verzicht auf alte Zöpfe (TEI-Header)

### Was HTML 5 besser kann als TEI-XML und Gutes-Altes-XML:

- Breite Unterstützung durch Software-Infrastruktur (z.B. Templating-Systeme als bessere Alternative zu XSLT)
- Jede Menge erstklassiger freier Editoren, mit Vorschau etc.
- Bessere Praktiken

„Speakeasy“-HTML	TEI-XML
<b>Ursprungskennzeichnung</b>	
data-by, data-from, data-place, data-time-Attribute <i>(intuitiv verständlich, differenziert, eindeutig!)</i>	resp-Attribut, source-Attribut <evidence>-Tag mit supplied Attr.
<b>Hervorhebungen</b>	
<i>, <b>, <em> für vielgenutzte Auszeichnungen	<hi rend=...> schnell unübersichtlich, durch Überfrachtung von rend tendenziell mehrdeutig.
<b>Verweise/semantische Auszeichnungen</b>	
<a href=...> Semantik wird durch href- Pfad ausgedrückt, etwa /lit/... /person/... /place/... <i>(Nachnutzung der Semantik bestehender Datenstrukturen!)</i>	<placeName>, <persName>, <index>, <reg>, <ref>



„Speakeasy“-HTML	TEI-XML
<b>Textstruktur (Kapitel und Unterkapitel)</b>	
<pre>&lt;section class="session"&gt;   &lt;section&gt;     ... 1. Unterkapitel ...   &lt;/section&gt;    &lt;section&gt;     ... 2. Unterkapitel ...   &lt;/section&gt; &lt;/section&gt;</pre>	<pre>&lt;div type="session" n="1"&gt;   &lt;div&gt;     ... 1. Unterkapitel ...   &lt;/div&gt;    &lt;div&gt;     ... 2. Unterkapitel ...   &lt;/div&gt; &lt;/div&gt;      (kein Scherz!)</pre>
<b>Textstruktur (besondere Abschnitte)</b>	
<pre>&lt;section class="opener"&gt;   &lt;p class="dateline"&gt;...&lt;/p&gt;   &lt;p&gt;...&lt;/p&gt; &lt;/section&gt;</pre>	<pre>&lt;opener&gt;   &lt;dateline&gt;...&lt;/dateline&gt;   &lt;salute&gt;...&lt;/salute&gt; &lt;/opener&gt;</pre>

„Speakeasy“-HTML	TEI-XML
<b>Sprachangaben</b>	
<p><code>&lt;span lang="la"&gt;...&lt;/span&gt;</code></p> <p><code>&lt;font lang="la"&gt;</code></p> <p>(Sprache wird nur als Attribut ausgedrückt!)</p>	<p><code>&lt;foreign xml:lang="lat"&gt;...&lt;/foreign&gt;</code></p> <p><code>&lt;hi rend="aq"&gt;</code>  <code>&lt;foreign xml:lang="lat"&gt;...&lt;/foreign&gt;</code>  <code>&lt;/hi&gt;</code></p>
<b>Datumsangaben</b>	
<p><code>&lt;time datetime="1789-07"&gt;</code>  heuer im Juli<code>&lt;/time&gt;</code></p> <p><code>&lt;time</code>  <code>datetime="1824-12-24--1824-12-26"&gt;</code>  Weihnachten<code>&lt;/time&gt;</code></p>	<p><code>&lt;date when="1789-07"&gt;</code>  heuer im Juli<code>&lt;/date&gt;</code></p> <p><code>&lt;date from="1824-12-24"</code>  <code>to="1824-12-26"&gt;</code>  Weihnachten<code>&lt;/date&gt;</code></p>

„Speakeasy“-HTML	TEI-XML
<b>Ersetzungen</b>	
<p><code>&lt;s data-by="hg"&gt;</code>, wenn <code>&lt;/s&gt;</code> <code>&lt;ins&gt;</code>. Wenn <code>&lt;/ins&gt;</code></p> <p>3 Tags decken alle Fälle ab: <code>&lt;s&gt;</code>, <code>&lt;del&gt;</code>, <code>&lt;ins&gt;</code></p>	<p><code>&lt;subst resp="hg"&gt;</code> <code>&lt;del&gt;</code>, wenn <code>&lt;/del&gt;</code> <code>&lt;add&gt;</code>. Wenn <code>&lt;/add&gt;</code> <code>&lt;/subst&gt;</code></p> <p>4 Tags, um alle Fälle abzudecken: <code>&lt;subst&gt;</code>, <code>&lt;del&gt;</code>, <code>&lt;add&gt;</code>, <code>&lt;supplied&gt;</code></p>

„Speakeasy“-HTML von Stefan Müller. Referenz: <http://138.246.225.147/corpus/HTML-XML.html>

TEI-XML: <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

DTA-Bf: <https://www.deutschestextarchiv.de/doku/basisformat/>