

The TEI Processing Model.

Introduction, limitations and potential extensions

Declarative Amsterdam, 7-8 November 2024

peter.boot@huygens.knaw.nl

<https://gitlab.huc.knaw.nl/edition-publication-model>

Contents

- Text Encoding Initiative (TEI)
- TEI Processing Model (PM)
- Implementation in e-editiones' TEI Publisher
 - PM extensions in TEI Publisher
- PM limitations
 - Nested output
 - Embedding in TEI ODD schema definition
- Conclusion

Text Encoding Initiative



- Tagset 600+ elements for humanities texts
- Consortium with institutional and individual members, Board and Technical Council
- Schema and Guidelines
- Online tool for schema definition
- Schema specification language
- Toolset for schema generation
- oXygen plugin

The question every TEI workshop instructor dreads: “So, how do I publish my TEI data?”

(Flanders and Hamlin 2013)

TEI Processing model

- Yes, every edition has different publication needs
- But most publications share a set of common building blocks
 - Paragraphs, heads, lists, tables, pictures, links, ...
- So what we need is a flexible mapping to HTML?
- No! We need an abstraction level above HTML
 - We are not just talking web output
 - We may want components not available in plain HTML (numbered paragraph, alternate displays, notes, popups)
- Need to describe set of abstract behaviours
- And configurable mapping of TEI elements to these behaviours

TEI Processing Model: behaviours

Behaviour	Parameters
alternate	default, alternate
anchor	id
block	content
body	content
break	type, label
cell	content
cit	content
document	content
figure	title
glyph	uri
graphic	url, width, height, scale, title
heading	content, level
index	type

Behaviour	Parameters
inline	content, label
link	content, uri
list	content
listItem	content
metadata	content
note	content, place, label
omit	
paragraph	content
row	content
section	content
table	content
text	content
title	content

TEI Processing Model: model and elementSpec

In project's ODD document:

```
<schemaSpec ident="mySchema">  
  ...  
  <elementSpec ident="p">  
    ...  
    <model behaviour="paragraph"/>  
    ...  
  </elementSpec>  
  ...  
</schemaSpec>
```

TEI Processing Model: @cssClass

```
<schemaSpec ident="mySchema">  
  ...  
  <elementSpec ident="p">  
    ...  
    <model behaviour="paragraph" cssClass="mypara"/>  
    ...  
  </elementSpec>  
  ...  
</schemaSpec>
```



TEI Processing Model: modelSequence

```
<schemaSpec ident="mySchema">
```

```
...
```

```
<elementSpec ident="p">
```

```
...
```

```
<modelSequence>
```

```
  <model behaviour="anchor"/>
```

```
  <model behaviour="paragraph"/>
```

```
</modelSequence>
```

```
...
```

```
</elementSpec>
```

```
...
```

```
</schemaSpec>
```

TEI Processing Model: param

```
<schemaSpec ident="mySchema">
  ...
  <elementSpec ident="p">
    ...
    <modelSequence>
      <model behaviour="anchor">
        <param name="url" value="concat('para-',@xml:id)"/>
      </model>
      <model behaviour="paragraph"/>
    </modelSequence>
    ...
  </elementSpec>
  ...
</schemaSpec>
```

TEI Processing Model: @predicate

```
<schemaSpec ident="mySchema">
```

```
...
```

```
<elementSpec ident="p">
```

```
...
```

```
<model behaviour="paragraph-with-number" predicate="ancestor::tei:body"/>
```

```
<model behaviour="paragraph"/>
```

```
...
```

```
</elementSpec>
```

```
...
```

```
</schemaSpec>
```

TEI Processing Model: modelGrp, @output

```
<schemaSpec ident="mySchema">  
  ...  
  <elementSpec ident="p">  
    ...  
    <modelGrp output="web">  
      <model behaviour="paragraph-with-number" predicate="ancestor::tei:body"/>  
      <model behaviour="paragraph"/>  
    </modelGrp>  
    <modelGrp output="pdf">  
      <model behaviour="paragraph"/>  
    </modelGrp>  
    ...  
  </elementSpec>  
  ...  
</schemaSpec>
```

TEI Processing Model

- Is just a vocabulary
- Not working software
- List of behaviours: ‘suggested values’
- Possible extensions
 - Changing default processing model for certain elements
 - Introducing new behaviours
 - New attributes or elements to describe processing model

TEI Publisher



- Implementation of TEI PM
- In context of eXist/DB XML database
- Proposes a default ODD (set of processing definitions)
- Facilitates generation of apps for own publishing own XML
- Development coordinated by e-editiones



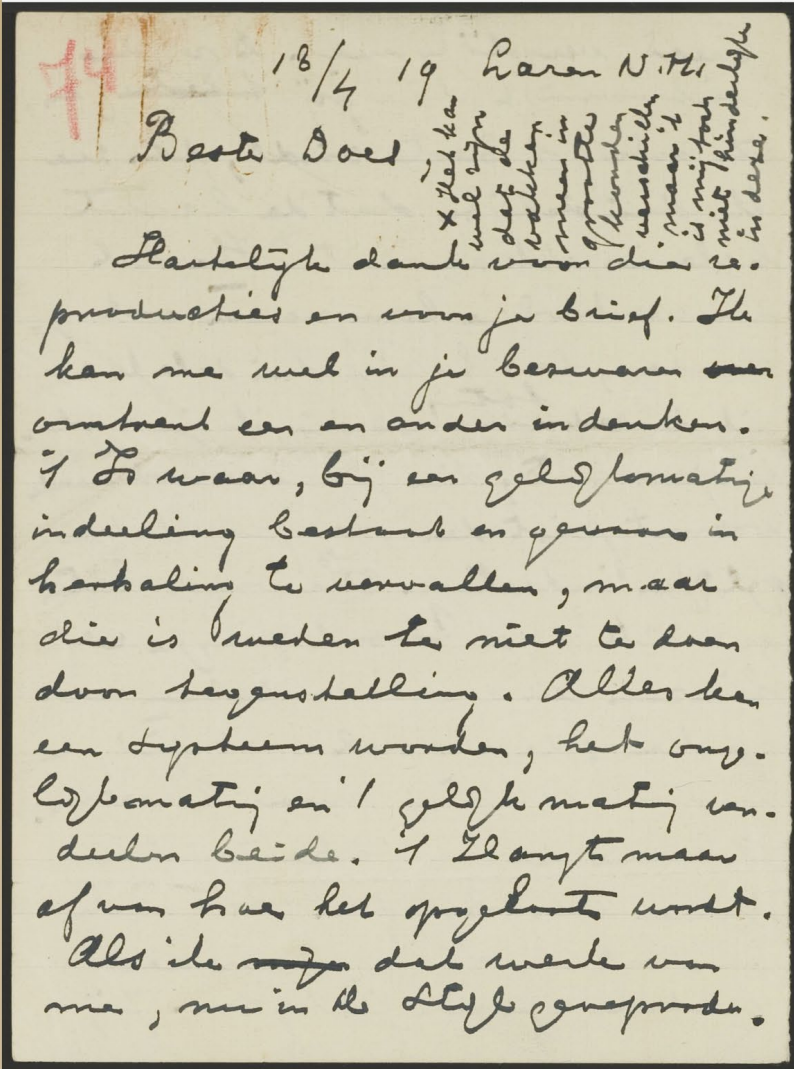
- Also provides facilities for editing XML sources (not discussed)

TEI Publisher extensions to PM

- XQuery rather than XPath in attributes
- Use of external parameters in attributes (e.g. mode)
- User-defined behaviours
- Introduction of Web Components
- Templates in models

```
<elementSpec ident="listAnnotation">
  <model behaviour="section">
    <pb:template>
      <div><h3>[[content_heading]]</h3>[[content]]</div>
    </pb:template>
    <param name="content_heading" value="'Notes'"/>
  </model>
</elementSpec>
```

FACSIMILE TEXT TRANSLATION



FACSIMILE TEXT TRANSLATION POSTAL DATA

18/4 19 Laren N.H.

Beste [Does](#) ,
 Hartelijk dank voor die reproducties en voor je brief. Ik kan me wel in je bezwaren omtrent een en ander indenken.
 't Is waar, bij een gelijkmatige indeeling bestaat er gevaar in herhaling te vervallen, maar die is weder te niet te doen door tegenstelling. Alles kan een systeem worden, het ongelijkmatig en 't gelijkmatig verdeelen beide. 't Hangt maar af van hoe het opgelost wordt.

Als ik dat werk van me, nu in De Stijl gereproduceerd, vergelijk met bv. die ruitvormige die jij (en ik ook) 't beste vindt, dan zie ik duidelijk dat de laatste beter in elkaar zit. Het kan wel zijn dat de vakken meer in grootte konden verschillen maar 't is mij toch niet hinderlijk in deze. Ik heb die werkwijze langzaam ook gevonden, dat kan je duidelijk zien uit de fotos die ik je verleden jaar zond; die je nu reproduceerde vormt juist den overgang tot de gelijke indeeling. Ik vind laten we elk deze soort dingen zelf kiezen en voor ons zelve verwerken; de een zal met dit, de ander met dat meer bereiken.

Als we de hoofdbeginselen maar vasthouden. Zoo ook over het al of niet van een natuurgegeven uitgaan. Jij omschrijft 't nu wel wat eng, dunkt me. De hoofdzaak ben ik wel met je eens, dat 't destructie v.h. natuur-lijke en reconstructie daarvan volgens den geest moet zijn, maar laten we dit heel ruim nemen. Het natuurlijke behoeft toch geen bepaalde voorstelling te zijn. Ik ben nu bezig met een ding dat een reconstructie van een sterrenlucht is, maar toch maak ik 't zonder natuurgegeven.


Een die zegt van natuurgegeven uit te gaan kan dus gelijk hebben en tegelijk hij die zegt van niets uit te gaan! Ik wil maar aantoonen hoe gevaarlijk 't is een systeem aan te nemen. Ik denk dat op den duur je niets tegen *mijn* werkwijze, hoe dan ook, zult hebben. Verder geen nieuws, ik moet 1 Juni hier vertrekken en blijf dan nog, denklijk in A'dam, om de vrienden op te zoeken enz. enz. Ik kom dan ook naar Leiden in de eerste plaats. Ik denk nog 5 stukken gereed te krijgen.


De Valori Plast. interesseert me natuurlijk. Wat ernstig werk vergeleken bij de nat. schilderkunst. De *onze* vind ik natuurlijk nog beter. Wanneer begin je aan 't Holl. nummer? Me dunkt jij moest een korte omschrijving geven, zooals dat ook bij de Cubisten gedaan is. Is er dan nog plaats, dan was 't beter dit aan te vullen met een vertaling van een stukje van mij, en van een architect, b.v., beter dan die Fransche verzen van [Salmon](#) b.v.!


Wie betaalt de kosten van de reproducties? Wilt je me eens melden hoeveel je voor het Fransche nummer van me krijgt. Komen er nog onkosten op en hoeveel?


Dag [Does](#) , ik moet met Paschen naar Arnhem. Veel beste groeten van je vrind Piet.

METADATA NOTES ARTWORKS PERSONS

- 

Composition with color planes and gray lines 3
 Piet Mondriaan, dated 1918
 Den Haag / Harskamp, private collection
 Helene Brückmann-Kröller
- 

Composition with grid 5: lozenge composition with colors
 Piet Mondriaan, dated 1919
 Otterlo (Ede), Kröller-Müller Museum,
 inv./cat.nr. 536-00
- 

Composition with grid 9: checkerboard composition with light colors
 Piet Mondriaan, dated 1919
 The Hague, Kunstmuseum Den Haag,
 inv./cat.nr. 167-1971 / 0334327
- 

Composition with grid 8
 Piet Mondriaan, dated 1919
 The Hague, Kunstmuseum Den Haag,
 inv./cat.nr. 168-1971 / 0334326

TEI PM limitations: Multiple behaviours

```
<elementSpec ident="listAnnotation">  
  <modelSequence>  
    <model behaviour="heading">  
      <param name="content" value=""Notes""/>  
    </model>  
    <model behaviour="section"/>  
  </modelSequence>  
</elementSpec>
```

```
<elementSpec ident="listAnnotation">  
  <model behaviour="section-with-heading">  
    <param name="content-heading" value=""Notes""/>  
    <param name="level" value="3"/>  
  </model>  
</elementSpec>
```

```
<elementSpec ident="listAnnotation">  
  <model behaviour="section">  
    <modelSequence>  
      <model behaviour="heading">  
        <param name="level" value="3"/>  
        <param name="content" value=""Notes""/>  
      </model>  
      <model behaviour="block"/>  
    </modelSequence>  
  </model>  
</elementSpec>
```

TEI PM limitations: ordering

- E.g. list of persons in alphabetic order
- No available behaviour
- New behaviour sorted block with param sort-key?
- But what if multiple sort options?

```
<elementSpec ident="listPerson">  
  <model behaviour="sorted-block">  
    <param name="content" value="tei:person"/>  
    <order name="Last name" value="tei:persName/tei:surname/text()"/>  
    <order name="Date of birth" value="string(tei:birth/@when)"/>  
    <order name="Nationality" value="tei:nationality/text()"/>  
  </model>  
</elementSpec>
```

TEI PM limitations: modularity

Current ODD structure:

```
<schemaSpec ident="abc">
  <elementSpec ident="listPerson">
    ...
    <model behaviour="list"/>
  </elementSpec>

  <elementSpec ident="person">
    ...
    <model behaviour="listItem"/>
  </elementSpec>
</schemaSpec>
```

Proposal for ODD structure:

```
<schemaSpec ident="abc">
  <elementSpec ident="listPerson">
    ...
  </elementSpec>
  <elementSpec ident="person">
    ...
  </elementSpec>
</schemaSpec>
<schemaProc ident="abc">
  <procBlock procmode="default">
    <elementProc ident="listPerson">
      <model behaviour="list"/>
    </elementProc>
    <elementProc ident="person">
      <model behaviour="listItem"/>
    </elementProc>
  </procBlock>
```

```
<procBlock xml:id="popup">
  <elementProc ident="person">
    <model behaviour="popup"/>
  </elementProc>
</procBlock>

<procBlock xml:id="label">
  <elementProc ident="person">
    <model behaviour="text">
      <param name="content"
value="tei:persName[@full='abb']"/>
    </model>
  </elementProc>
</procBlock>
</schemaProc>
```

Conclusions

- TEI Processing Model is useful
- Need for some enhancements
- Practical confrontation with needs of an edition is useful

<https://gitlab.huc.knaw.nl/edition-publication-model>