

Beyond the TEI Trees. Graphical Visualisations of TEI as a path towards seeing the whole forest

Since the first version of the TEI guidelines were published in 1990, and the major P5 release in 2007, a growing number of visualisations of encoded texts has taken place. With programming languages such as XSLT to make complex access to the text rather straightforward, and graphical output languages like SVG and Javascript, it is now relatively simple to “picture” your TEI text (cf. experiments by Wendell Piez such as that at <http://wendellpiez.com/eatyourvegetables/?p=327>, or tools like Sinclair and Rockwell’s Voyant). Increasingly, the powerful Javascript libraries available in common browsers (eg. Bostock’s D3.js) make publishing online dynamic visualisations easy.

While development towards visualisations of TEI encoded texts is gaining pace, a new visitor to the TEI website will be surprised to find very little in the direction of graphical presentations of the TEI schema itself. In the TEI Guidelines the user will find a large number of examples, a reasonable amount of which are accompanied by illustrations, such as manuscript facsimiles. But she or he will not find a full graphical presentation of the TEI system of elements and their relationships, either as an interactive tool or as static illustrations in the documentation of the elements, element classes, or more general subject areas. However, the reference pages of the Guidelines are presented in a formalised structure, including, for example, lists of parents and children for elements, and class memberships; to expect a graphical representation is therefore not astray. In addition, the lack of visual presentation within the TEI Guidelines stands out compared to its common use in data standards adopted within neighbouring areas such as CIDOC CRM and FRBRoo, and in programming languages in general.

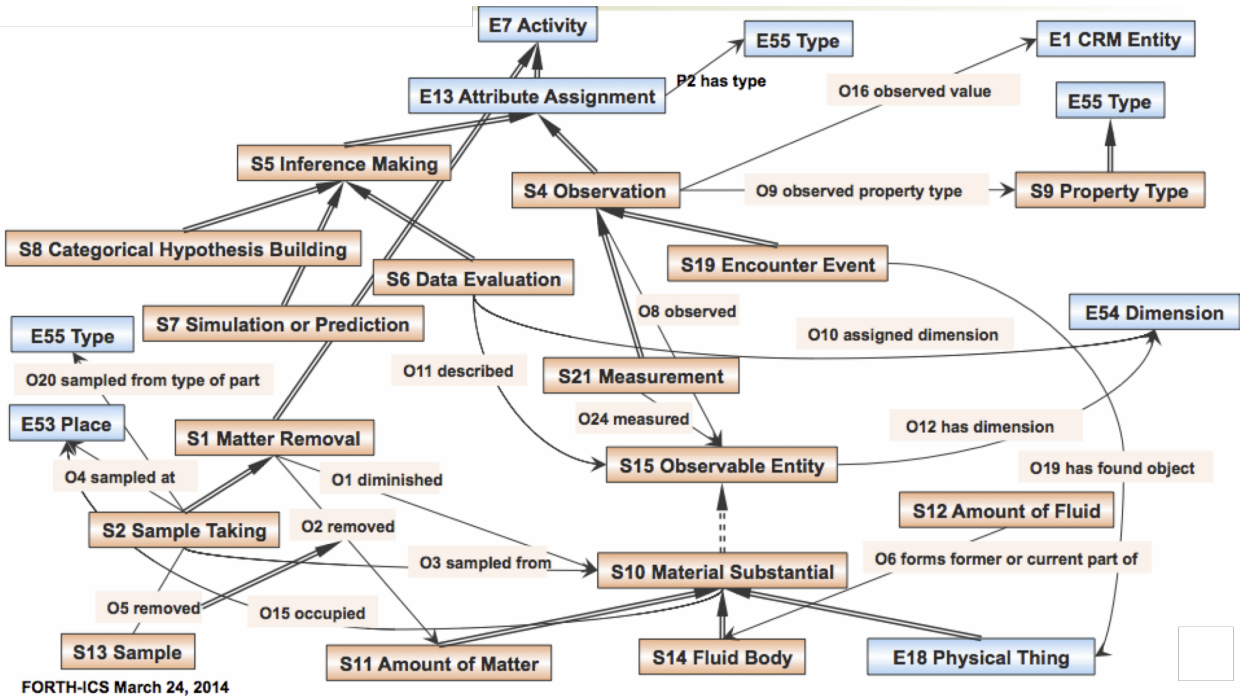


Figure 1 Example of visualisation of a CIDOC CRM data model.

An attempt at a “TEI Visualizer” (Rahtz and Cornelius 2006) was based on a BNF representation of a generated RELAX NG schema, and missed some of the information inherent in an ODD schema. It is clear from trying to use the otherwise excellent graphical representation of schemas in the oXygen editor that something fairly high level is needed to visualise the TEI; as figure 2 shows, the scale and detail of the pictured RELAX NG or XSD schemas makes them unreadable. Other experiments have been carried out but not incorporated in the guidelines (e.g. Middell 2012).

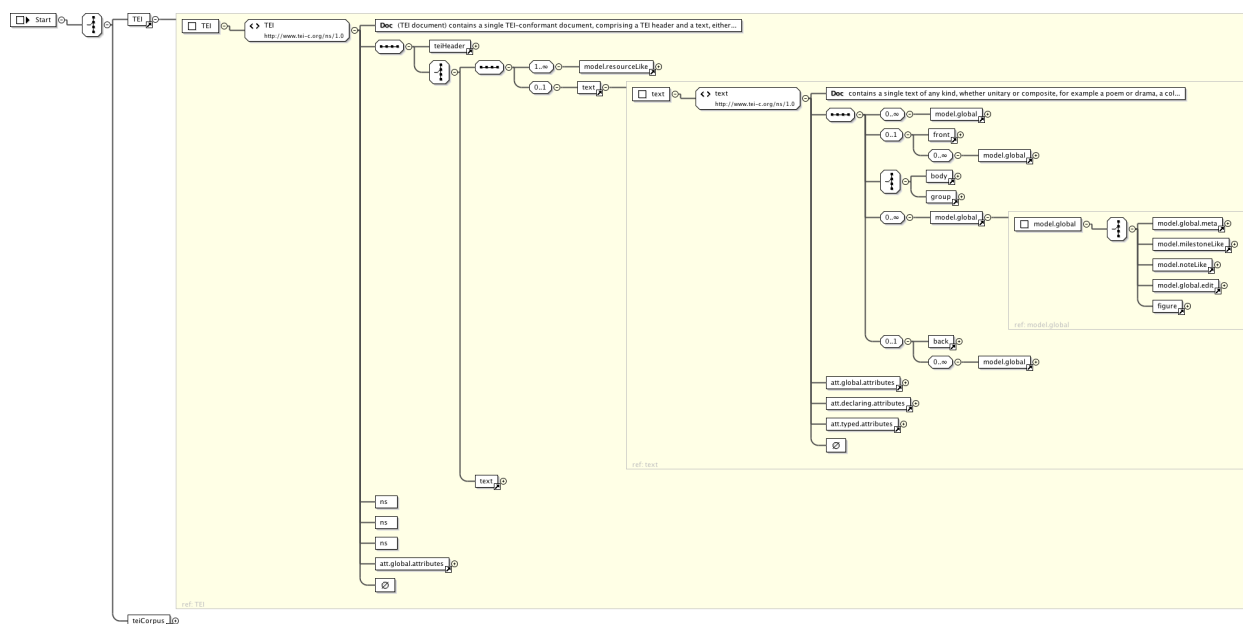


Figure 2 Example of a TEI RELAX schema as visualised in the main oXygen editing panel.

In this paper we will sketch the historical background for the difference in how data standards are presented (see also Ciula and Eide 2014). We will discuss the degree to which the lack of visualisations is related to the possible lack of one clear conceptual model behind TEI. Could it be that the standard cannot be clearly expressed in graphical form because the structure is too complex or not consistent enough? or are we simply lacking the right visual metaphor to show it?

Based on the results of this analysis we will propose a way forward to a visualisation system for the TEI which could be used not only to understand the standard from a new user's perspective, but also to deepen the understanding of how the various modules of the standard are related to each other for the expert user (e.g. engaged in customisation exercises) and developer of the standard itself. Recent experiments with a Javascript-based replacement for Roma (<https://github.com/sebastianrahtz/Byzantium>) started to show some simple graphs from the internal data structure, and representation of the basic information from a TEI ODD in JSON format; this has also been demonstrated by Viglianti (ODDViz). It is to be expected that a graphical editor for TEI ODD, based on a visualization of the source, would be an attractive addition.

We see visualisations as they are described here as tools which can be used both in the creation of *models for* and *models of* (cf. the distinction in McCarty 2005). On the one hand, visualisations can be used as an aiding tool to enhance modelling practices, that is, *models for*. A well developed visual interface will exemplify scenarios for the user and assist in the searching for the right combination of elements. On the other hand, visualisations can also play a role in enhancing the understanding of the TEI abstract model, that is, they play the roles of *models of*. In this context, the visualisations can also be used to better understand the consequences of changing the abstract model. In this way, visualisation as a tool could contribute to further development of TEI into a more coherent system or at least supply a visually-oriented interface

to act like the text-based Roma, and allow a user to manipulate their desired subset of the TEI, seeing the effects visually as changes are applied.

References

- Bostock, Mike. 2013. D3.js <<http://d3js.org/>>
- CIDOC CRM <<http://www.cidoc-crm.org/>>
- Ciula, Arianna and Øyvind Eide. 2014 forthcoming. "Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory." *Proceedings from the DATECH: Digital Access to Textual Cultural Heritage, Madrid, Spain, May 19-20 2014.*
- FRBR-Object Oriented (FRBR-oo) <http://www.cidoc-crm.org/frbr_inro.html>
- McCarty, Willard. *Humanities computing*. Palgrave Macmillan, Basingstoke, 2005.
- Middell, Gregor. 2012. <<http://gregor.middell.net/roma/>>
- Rahtz, Sebastian and Barry Cornelius. 2006. *TV: a TEI Visualizer*. Poster presented at Text Encoding Initiative Conference 2006 (Victoria BC, October 2006).
- Sinclair, Stéfan and Geoffrey Rockwell. 2014. Voyant Tools, v. 1.0 beta (4647) <<http://voyant-tools.org/>>
- Viglianti, Raffaele, 2012 <<https://github.com/raffazizzi/ODDViz>>.