

1. Host Schema Evaluation

This evaluates: *TEI Lite*

The Text Encoding Initiative has a very large standard which can be subsetted. Unfortunately, one must pay to purchase the full standard as PDF for about ninety dollars. But, they make available the documentation for a very popular subset called TEI Lite. And, they have an HTML version of their specification on line.

(Interestingly enough, the TEI consortium uses the TEI Lite for their own documentation.)

They also have a "bare bones TEI" which is an even smaller subset.

1.1. Object Representation

Q: Does the host schema use a generic structural markup model?

A: Yes. It supports a `p` tag, as well as `<div0>`, `<div1>`, `<div2>`... `<div n >` tags. These have a `type` attribute. This specifies that the **div** element would represent "Book," "Chapter," section, etc.

Or one can use `div` elements and nest them arbitrarily.

Q: Does the host schema define a "clause" object?

A: One would have to use a `div` element with the appropriate `type` attribute.

Q: Does the host schema define a paragraph level object that represents a structural or grammatical paragraph?

A: Yes, as indicated above, it supports a `p` for a paragraph and `list` which consists of `list`, `head` and `item`. One could number the `item`'s using a `label` or the `n` attribute.

Q: Using the host schema, can the clause equivalent object be inserted at arbitrary levels in the document hierarchy without transformation?

A: There is no mention in the TEI lite documentation of being able to put `div` elements inside a `p` tag or `list` tag.

Review of the HTML documentation for the full TEI indicates that one cannot do this.

Q: In the host schema, are element names and the structure sufficiently flexible that the clause and paragraph level objects can be used for other legal and business documents?

A: Absolutely! The user picks the naming convention they wish to use and uses this with the `type` attribute for the `div`.

1.2. Metadata

Q: Does the host schema provide a mechanism to add semantic information about:

- whole documents
- distinct objects, such as clauses, within documents?

A: The TEI Lite documentation includes the following in the titlepage element: docTitle, docAuthor, docDate, projectDesc, editorialDecl and a profile-Desc which includes a creation and a Revision containing date and changes elements.

Q: If so, is the metadata model for the host schema sufficient for contracts or will it be necessary to extend it?

A: These metatags appear insufficient for contracts.

Q: Does the host schema allow embedded values to be represented and semantic information to be added to these values?

A: The full TEI supports symbolic values, numbers, ranges, rate of change, string values, and binary values for "features" There is also a structuring mechanism for things such as employee's, architectural descriptions, and the like.

1.3. Processing Technologies

Q: Does the host schema require use of a particular processing technology?

A: The TEI is dependent upon DTD's or Schema's. However, the resultant XML should be easy to work with using style sheets or Java programs.

Q: Does the design of the host schema preclude use of particular currently available processing technologies?

A: There appears to be no interest in converting TEI to use Schema or Relax-NG.

1.4. Number of Content Objects

Q: Does the host schema permit the numbering of clauses, paragraphs, lists and other objects to be represented in the markup?

A: Yes, this is provided for with the **n** attribute. It is particularly needed since TEI is used to encode antiquarian and scholarly texts where it is important to refer to things by the original numbering convention.

1. Host Schema Evaluation

Q: Does the host schema provide a mechanism to define the numbering schema applied to the document so that two applications could apply the same numbering, if desired?

A: See above.

1.5. Complete Document Representation

Q: Using the host schema, will it be possible for the contract author to explicitly represent all parts of the narrative contract terms or will it be necessary to imply some parts?

A: There are tags for the textual divisions but none that would be useful for things particular to contracts such as recitals.

Q: Does the host schema represent the relationship between all significant components in a way that allow high quality print and web rendition of of contact documents?

A: There are extensive tags for italics, emphasis, etc. Also, there are tags for marking off lines and blocks, intended for typesetting poetry or other verses.

The TEI web page includes XSL style sheets, customizations for Passive Tex, and an open source tool called Anastasia.

1.6. Variables Definition

Q: Does the host schema include a mechanism for defining variables for embedded data values?

A: The full TEI supports symbolic values, numbers, ranges, rate of change, string values, and binary values for "features" There is also a structuring mechanism for things such as employee's, architectural descriptions, and the like.

The TEI lite does support xptr, xref, ref, ptr. However, these appear intended for cross references and not variable values.

Q: If the host schema does not include such a mechanism, is there any obstacle to adding it?

A: See 12.2

1.7. Ease of use for authors

Q: Based on the following factors is the host schema easy for contract authors to use:

- Does it require authors to know only a small number of elements (positive factor)?
- Does it require authors make unnecessary or subtle distinctions that will be applied inconsistently (negative factor?)

- Does it have a clear logical structure that can be quickly explained to new users (positive factor)?
- Does it allow authors to re-locate content objects within a document hierarchy with minimal or no need for transformation of markup (positive factor)?

A: The concepts of `list`, `p` and `div` tags are straight forward. However, it appears that if one sets up a structure with `div`, it could not be moved into list item or header.

1.8. Schema Syntax

Q: Is the host schema a DTD only or can it also be expressed as an XML Schema or other schema type?

A: The documentation for TEI all reference DTD or SGML. However the new version, P5, is expressed in both Relax NG and W3C schema as ell. There is an application to generate "P5-compatible schemas and docuemntation."

1.9. Adaptability to contracts

Q: Does the host schema provide for the complete representation for the distinct structures commonly found in contracts?

A: TEI Lite does not have these structures.

Q: If not, does the host schema explicitly allow additional distinct structures to be added?

A: The full TEI has an extension mechanism. It apparently is based upon the parameter entry of DTD's.

Q: Does the host schema allow elements not considered necessary for contracts markup to be removed without contract documents being incompatible in a disadvantageous way with other documents using the host schema?

A: Yes, there is a well-used and documented subsetting mechanism.

Q: If distinct contract structures are added to the host schema, will this result in contracts documents being incompatible in a disadvantageous way with other documents using the host schema?

A: There are definite guidelines on how to extend the TEI. This includes the `TEIFORM` attribute to specify that a new element is used in the role of an existing TEI XML tag. Presumably, if they are followed, one could still use the TEI software with our contracts.

1.10. Vendor and Developer Support

- Q: Is the host schema already in widespread or general use for markup of narrative documents?
- A: By viewing the Cover Pages, there appear to be many software projects to do TEI. Fifty-six organizations are members of the TEI Consortium.
- Q: Are the already developed applications that Will make it easy of for organizations to implement the TC's specification based around the host schema?
- A: Their web page lists preloaded versions of oXygen, Gnu Emacs, Open Office Adobe FrameMaker and Softquad's XML. loaded with TEI templates ready to work with. Also, there is a utility available for converting "well-structured Microsoft Word Documents" into TEI.
- Q: Is there any reason to expect that the host schema will prove any particular advantages in gaining market support.
- A: The TEI standard does have some adherants with 123 projects on their web page. However, they are mostly projects concerned with traditional scholarly research and academic work. (The Cover page for the TEI is categorized under "academic.") It has been supported by the United States National Endowment for the Humanities, the European Community,k the Mellon foundation, and the Social Science Humanities Research Council.

1.11. Other Factors

- Q: Does the host schema provide any other advantages for use in the TC's specification?
- A: Norman Walsh and others reported at XML Europe 2004 on an effort to integrate TEI and Docbook.
- Q: Does the host schema have any other disadvantages that make it undesirable for use in the TC's specification?
- A: There is a special charge to obtain the specification for the full TEI standard as PDF. The HTML version is less convenient.