

Oasis Legal XML eContracts TC

Document Title: Evaluation of TEILite DTD
(<http://www.tei-c.org/Lite/DTD/teilitex.dtd>)
Text Encoding Initiative Homepage (<http://www.tei-c.org>)

Author: Rolly Chambers

Version: Draft 0.01

Date submitted: 2005-06-10

1. Summary Comparison of TEILite DTD against Schema Evaluation Criteria Draft 0.01 (2005-05-22)

The Text Encoding Initiative (TEI) was founded in 1987 to develop guidelines for encoding machine-readable texts in the humanities and social sciences. It was re-established in 2001 as a membership consortium, with hosts at two North American and two European sites, and nearly a hundred member institutions worldwide.

The TEI DTD allows for different element sets for different types of text documents (e.g. prose, verse, drama, speech). An XML version of the TEI Guidelines (and DTD), P4, was produced in 2002. The TEILite DTD offers an existing, customized subset of the full P4 TEI DTD. It includes elements and attributes common to a wide range of text documents. The numbered rows in the table below correlate to the criteria in the Schema Evaluation Criteria Draft 0.01 (2005-05-22) by Peter Meyer of this TC.

<i>“Clause” Pattern</i>	<i>Candidate Elements</i>	<i>Comments</i>
2. Generic numbered object container/representation	<div>	<ul style="list-style-type: none">- descendant of <TEI.2><text><body>- #PCDATA not allowed- recursive (<div> can contain a <div>)- TEI DTD also allows numbered <div>s (<div0>, <div1>, . . . <div7>) as an alternative- allows sequence for optional <head> element, a mixed content element which can contain #PCDATA or <title>; <head> must precede <body> in a <div>- hidden recursion (<text> can be child of <p>, <head>, or descendant of <div>)

<i>“Clause” Pattern</i>	<i>Candidate Elements</i>	<i>Comments</i>
Heading?	<head>*	<ul style="list-style-type: none"> - <head> is optional for each <div> level - #PCDATA allowed (mixed content) in <head> - <title> is allowed - <head>, if used, must precede <p>
Number?	<num>*	<ul style="list-style-type: none"> - no sequence required (<num> can precede or follow <title>) - <num> also can occur within <p>, <hi> - <emph>-<seg>, and <item>)
Title?	<title>*	<ul style="list-style-type: none"> - <title> can contain #PCDATA
Paragraph+	<p>*	<ul style="list-style-type: none"> - <p> is a parent, sibling, and grandchild of <list> (<p> can contain lists or be contained within list items) - <p> can be represented as a grammatical paragraph that contains lists - <p> also can be represented as a structural paragraph that does not contain lists)
List*	<list>* <item>*	<ul style="list-style-type: none"> - <list> is child and sibling of <p> - <item> can contain <p> or <list> (hidden recursion of <p> and <list>) - <head> is optional within <list> (<list>s can have headings)

<i>“Clause” Pattern</i>	<i>Candidate Elements</i>	<i>Comments</i>
Hierarchy	<TEI.2> <text> <body> <div>* or <p>* <p>* <s>* or <hi>* or <seg>* or <emph>* or <kw>* <kw>*	- no phrase element; <s>, <hi>, <seg>, or <emph> are candidates - tighter DTD would provide a stricter hierarchy (allow <p> only within <div>; allow <kw> only within <s>, <hi>, <seg>, or <emph>, but not within <p>)
Additional infrastructure within <div> and <p>	- <div> and <p> can contain lists, tables, figures - content by reference (id and target (IDREFS) attributes) - version tracking <revision> element	- no image element - is <figure> a candidate?

<i>Consideration</i>	<i>For</i>	<i>Against</i>
3. Metadata	<TEIHeader> and <head>	- <TEIHeader> must contain file, publication, and revision metadata - TEI required metadata not particularly useful for contract documents - <head> contains metadata for a <div> - needs to be extended for contracts
4. Processing technologies/license terms	- open source (GNU General Public License) - TEI XML documents can be processed with open source or proprietary XML tools and applications	
5. Numbering	<num>	- allows authors to use <num> within <head>, <p>, <hi> - <emph> - <seg>

<i>Consideration</i>	<i>For</i>	<i>Against</i>
6. Complete contract documentation	Authors can explicitly represent narrative contract terms	
7. Variables	<code> can contain a fragment for a programming language	- no named element for defining variables - is <ident> a possibility?
8. Ease of use	- TEILite DTD offers an existing customized subset of the full TEI DTD and includes elements common to a wide range of text documents - TEI documentation is clear, well-developed, and up-to-date	- hidden recursion (e.g. <text> can re-occur at “phrase” level) - too much element and attribute choice - loose and deep hierarchy (author can choose <body>-<div> or <body>-<p> hierarchy; <div> and <p> are great-grandchild elements of the root <TEI.2>) - DTD used in academia but not in businesses
9. Schema syntax	- TEI P4, current version, is a DTD - experimental Relax NG schema exists for TEI P4 - TEI P5 is under development and will be expressed in Relax NG schema, XML schema, and DTD	- no developed RNG or W3C XML schema
10. Adaptability to contracts	- <front>, <back>, and <index> within <text> - DTD allows extension elements - TEI tools exist for creating a restricted subset	- no named elements for recitals, attachments, exhibits, schedules

<i>Consideration</i>	<i>For</i>	<i>Against</i>
11. Developer/vendor support and skill set	<ul style="list-style-type: none"> - Well documented XML DTD - ongoing development through Text Encoding Initiative Consortium (http://www.tei-c.org) - XSL stylesheets for outputting TEI documents as HTML and PDF - CSS stylesheets - XSLT stylesheets for importing to and exporting from OpenOffice - search and retrieval, publishing, and versioning applications under development (teiPublisher, PhiloLogic, Xaira, Versioning Machine) 	<ul style="list-style-type: none"> - TEI processing applications are alpha and beta quality

2. Conclusions

- ease of use of the full TEI DTD and TEILite DTD for authors and developers is a major obstacle; required TEI metadata not particularly useful for contract documents
- creation of a radically simplified and reduced subset of TEI or TEILite DTD's customized for contract documents is absolutely necessary, but possible
- little support in existing XML document authoring tools for directly creating TEI encoded XML documents
 - OpenOffice import/export of TEI encoded XML documents via XSLT stylesheets is possible
- Open source processing applications for search and retrieval, document management, version tracking, and publishing of TEI encoded XML documents exist and are under development
- XML namespaces and data typing capabilities are not available via TEI (an issue for any DTD)
- Lack of defined phrase level element for clear hierarchy of grammatical structures
- Lack of defined elements for images and variables
- Additional metadata elements needed?