

Cover Pages

Hosted by OASIS

Online resource for markup language technologies

[SEARCH](#) | [ABOUT](#) | [INDEX](#) | [NEWS](#) | [CORE STANDARDS](#) | [TECHNOLOGY REPORTS](#) | [EVENTS](#) | [LIBRARY](#)

SEARCH

[Advanced Search](#)

ABOUT

[Site Map](#)
[CP RSS Channel](#)
[Contact Us](#)
[Sponsoring CP](#)
[About Our Sponsors](#)

NEWS

[Cover Stories](#)
[Articles & Papers](#)
[Clippings](#)
[Press Releases](#)

CORE STANDARDS

[XML](#)
[SGML](#)
[Schemas](#)
[XSL/XSLT/XPath](#)
[XLink](#)
[XML Query](#)
[CSS](#)
[SVG](#)

TECHNOLOGY REPORTS

[XML Applications](#)
[General Apps](#)
[Government Apps](#)
[Academic Apps](#)

EVENTS

LIBRARY

[Introductions](#)
[FAQs](#)
[Bibliography](#)
[Technology and Society](#)
[Semantics](#)
[Tech Topics](#)
[Software](#)
[Related Standards](#)
[Historic](#)

Last modified: December 21, 2002

Technology Reports

Text Encoding Initiative (TEI)

"Initially launched in 1987, the TEI is an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent."

[June 15, 2002] In June 2002 the Text Encoding Initiative (TEI) Consortium [announced](#) the "publication of a new, updated version of their Guidelines for Electronic Text Encoding and Interchange, known as [P4](#). The Consortium, now in its second year, is an international non-profit corporation set up to maintain and develop the TEI system, which has become the de facto standard for scholarly work with digital text since its first publication in 1994. The launch of a fully XML-compliant version of the TEI Guidelines is a significant advance, placing the TEI firmly in the mainstream of current digital library and World Wide Web developments. The new edition has been available online for a few months, and will continue to be so, but the print edition now available from the University of Virginia Press (URL) marks a new milestone in the history of this long standing exercise in scholarly communication and international co-operation. In simple terms, the TEI Guidelines define a language for describing how texts are constructed and propose names for their components. By defining a standard set of names the Guidelines make it possible for different computer representations of texts to be combined into vast databases, and they also provide a common language for scholars wishing to work collaboratively. There are many such standard vocabularies in the industrial world -- in banking, in aircraft maintenance, or in chemical modelling, for example. The TEI's achievement has been to try to do the same thing for textual and linguistic data -- both for those working with the written culture of the past and for those studying the development of language itself. Membership in the TEI Consortium has climbed steadily during its first year of operation, standing at 56 members worldwide in May 2002, ranging from small university research projects to major academic libraries and institutions. The consortium

HOSTED BY



SPONSORED BY



COVER PAGES NEWSLETTER

Receive [weekly news](#) updates from Managing Editor, Robin Cover.

[Subscribe >>](#)

[View Archives >>](#)

offers a range of membership benefits including participation in TEI elections, special access to training, consultation on grant proposals, and free or discounted copies of the TEI Guidelines."

[January 09, 2002] [Updated XML DTDs for the Text Encoding Initiative Guidelines.](#) A posting from [Lou Burnard](#) invites comment on the publication of updated XML DTDs for the Text Encoding Initiative Guidelines. Based upon extensive public review, the XML DTDs have been improved and corresponding revised documentation has been created in HTML and PDF format for the TEI Guidelines. Approval of the new P4 edition by TEI Technical Council and final publication is expected within the near future. Already widely adopted for use in digital library projects, the TEI Guidelines are "intended for use in interchange between individuals and research groups using different programs and computer systems over a broad range of applications... The Guidelines apply to texts in any natural language, of any date, in any literary genre or text type, without restriction on form or content. They treat both continuous materials ('running text') and discontinuous materials such as dictionaries and linguistic corpora. The primary goal of the P4 revision has been to make available a new and corrected version of the TEI Guidelines which: (1) is expressed in XML and conforms to a TEI-conformant XML DTD; (2) generates a set of DTD fragments that can be combined together to form either SGML or XML document type definitions; (3) corrects blatant errors, typographical mishaps, and other egregious editorial oversights; (4) can be processed and maintained using readily available XML tools instead of the special-purpose ad hoc software originally used for TEI P3. A second major design goal of this revision has been to ensure that the DTD fragments generated would not break existing documents: in other words, that any document conforming to the original TEI P3 SGML DTD would also conform to the new XML version of it. Although full backwards compatibility cannot be guaranteed, we believe our implementation is consistent with that goal."

[August 01, 2001] [Text Encoding Initiative Consortium Releases P4 Draft Guidelines in XML and SGML.](#) TEI editors Lou Burnard and Steve DeRose have announced the official release of version 4 draft *Guidelines for Electronic Text Encoding and Interchange*. The third edition of the *Guidelines* known as 'P3' has been edited by participants in the Text Encoding Initiative Consortium (TEI-C); the third edition "has been heavily used since its released in April of 1994 for developing richly encoded

and highly portable electronic editions of major works in philosophy, linguistics, history, literary studies, and many other disciplines. The fourth edition, 'P4' will be fully compatible with XML, as well as remaining compatible with SGML (XML's predecessor and the syntactic basis for P3). XML-compatible versions of the TEI DTDs have been available for some time by means of an automatic generation process using the TEI ['pizza chef' tool](#) on the project's website. The first stage in the production of P4 has been to remove the need for this process; accordingly, a preliminary set of dual-capability XML or SGML DTDs was made available for testing at the ACH-ALLC Conference in New York in June. The next stage was to apply a series of systematic changes to the associated documentation, which is now complete: the results may be read online." The TEI editors invite participation in public review of the new P4 draft *Guidelines*.

[June 30, 1999] In June 1999, The Text Encoding Initiative (TEI) entered a significant new phase with the official publication of the XML DTD for TEI Lite, available with supporting resources on the [TEI Web site](#). Since 1987, the international [Text Encoding Initiative](#) has sponsored a major effort to "develop guidelines for the preparation and interchange of electronic texts for scholarly research, and to satisfy a broad range of uses by the language industries more generally." The published TEI Guidelines have gone through three major editions under the editorship of C. Michael Sperberg-McQueen and Lou Burnard, and the current TEI-P3 print volumes *TEI Guidelines for Electronic Text Encoding and Interchange* are also publicly available in SGML format. The TEI Guidelines have been used for SGML encoding [in some sixty-nine \(69\) significant projects worldwide](#).

Though the TEI is a large and complex specification, a unique tool known as the [Pizza Chef TEI Tag Set Selector](#) supports the creation of user-defined DTD subsets via an HTML form interface. The [online Pizza Chef](#) will "help you design your own TEI-conformant document type definition (DTD). The TEI Guidelines define several hundred elements and associated attributes, which can be combined to make many different DTDs, suitable for many different purposes, either simple or complex. With the aid of the Pizza Chef, you can build a DTD that contains just the elements you want, suitable for use with any XML processing system. [To use the tool] you need to understand a little about how the TEI DTD is organized. In particular, you need to understand that the TEI scheme is organized into base and additional tagsets (groups of elements), and that

each element in a tagset can be suppressed, or redefined... First, decide whether you need to use one base tagset or several base tagsets (Prose, Verse, Drama, Speech, Dictionaries, Terminology, General). Whichever base you use, you can add as many additional tagsets as you want. There are twelve to choose from. If you wish, your DTD can include declarations for one or more of the ISO public entity sets. If you want to discard or modify elements from the selected tagsets making up your DTD you can do this... you pass the names of your modification files to the pizza chef, along with the tagsets you chose originally... [press the button and] build your personalized DTD..."

The TEI Lite XML DTD's public identifier is "-//TEI//DTD TEI Lite XML ver. 1.0//EN" (or: "-//TEI//DTD TEI Lite XML ver. 1//EN"). The principal resources supporting this XML release of the TEI DTD are described in a recent [announcement](#) from C. M. Sperberg-McQueen (North American TEI Editor) and are referenced in the TEI document ['TEI, SGML and XML Resources.'](#)

Principal References

- [TEI Consortium web site](#)
- [The TEI Guidelines](#). Overview document.
- [The XML Version of the TEI Guidelines](#). HTML browsable. Download the HTML version as a [single archive file](#) for offline browsing. [[cache](#)]
- [Version P4 XML DTDs](#). Browsable. See also the [ZIP archive](#), [cache](#)
- [Projects Using the TEI Guidelines](#)
- [TEI Lite: An Introduction to Text Encoding for Interchange](#). See the [TEI Lite XML DTD](#), [[cache](#)]
- [PizzaChef for creating TEI XML DTDs](#)
- TEI Guidelines in print: *TEI P4: Guidelines for Electronic Text Encoding and Interchange*. Edited by C.M. Sperberg-McQueen and Lou Burnard. Text Encoding Initiative Consortium. XML Version: Oxford, Providence, Charlottesville, Bergen. March 2002. Published for the TEI Consortium by the Humanities Computing Unit, University of Oxford, 2002. Distributed by the University of Virginia Press. XML-compatible edition prepared by Syd Bauman, Lou Burnard, Steve DeRose, and Sebastian Rahtz. ISBN: 0-952-33013-X. Printed in two parts. Volume One: Chapters 1-23, pages i-xviii, 1-572. Volume Two: Chapters 24-36, Index, Appendices, pages 573-1067. Available for [purchase](#).
- Note that the standard TEI DTDs are generated

and maintained using a "literate programming style" system (originally) called ODD ['One Document Does It All']. For details, see the [separate discussion of ODD](#) and [excerpted comments](#) from TEI List postings of Sebastian Rahtz and Lou Burnard.

- [TEI Tutorials](#)
- [TEI News Page](#)
- [The TEI FAQ document](#)
- [About the TEI Consortium](#)
- [1997-1999] [Previous TEI database entry](#) in the SGML/XML Web Page. This document section (though outdated) references software especially applicable to the creation/use of TEI-encoded texts.

Software

2002-12-21 Note: This section under construction/revision

- [TEI Software page](#). Maintained by the TEI Consortium.
- [TEItools](#). By Boris Tobotras.
- PSGML
 - ["First steps with PSGML and TEI."](#) By Christian Wittern (Chung-Hwa Institute of Buddhist Studies). December 2000.
- [Pizza Chef](#) tool for creating custom DTDs. See in this connection the document ["Construction of an XML Version of the TEI DTD."](#) This paper, while potentially difficult for a casual reader to understand, discusses ("sometimes in tedious detail"), what choices the TEI editors have made in creating an XML version of the TEI DTD -- *e.g.*, "drop exclusions, propagate inclusions downward into the content model of every possible descendant, and redefine the attributes as NMTOKEN(S). It combines both a reasonably easy top-level introduction to what has to happen when an SGML DTD is rewritten for XML and a long, 'expose-every-detail' discussion of every single content model in the TEI DTD that needs changing." The online version of the TEI Pizza Chef [2002-12] was developed by Lou Burnard, but all the clever stuff backstage is still done using Michael Sperberg-McQueen's [carthage](#) (DTD pre-

processor). An alternative version, developed by Sebastian Rahtz, is available under the name of [maketeitd](#). [[EDW69 cache](#)]

- [SARA \(SGML Aware Retrieval Application\)](#) from the [British National Corpus Project](#) can be used to index TEI documents. "[Indexing a corpus for SARA](#)" includes examples for indexing TEI. From Lou Burnard 2002-07-05: "we are currently making good progress in implementing a new version of the SARA text indexing software originally developed for the BNC (www.natcorp.ox.ac.uk) which is fully Unicode aware. The intention is to produce efficient indexing of either fully-marked up TEI-XML texts, or of texts which entirely lack any markup, but which are encoded using Unicode... We are using the Xerces parser and the ICU components for Unicode support..."
- [December 20, 2002] [The Versioning Machine \(VM\) 1.0](#). "The Versioning Machine is a software tool designed by a team of programmers, designers, and literary scholars at [Maryland Institute for Technology in the Humanities \(MITH\)](#) for displaying and comparing multiple versions of texts. The display environment seeks not only to provide for features traditionally found in codex-based critical editions, such as annotation and introductory material, but to take advantage of opportunities of electronic publishing, such as providing a frame to compare diplomatic versions of witnesses side by side, allowing for manipulatable images of the witness to be viewed alongside the diplomatic edition, and providing users with an enhanced typology of notes... Because the TEI critical apparatus tagset offers the most efficient and thorough methodology for inscribing variants in a structured, machine-readable format, the Versioning Machine (VM) has adopted it in version 1.0 as its foundation. Using this tagset, then, allows an editor to encode in one document multiple versions of that text; VM 1.0 is able to reconstruct multiple witnesses from the single XML-encoded document and display them, side-by-side, as individual documents. The critical apparatus tagset supports three different types of encoding variation: location-referenced, double-end-point, and parallel-segmentation; however, only parallel-segmentation is currently supported by VM 1.0..." MSIE 6.0+ (only as of 2002-12) See the [announcement](#).
- [teixlite Conversion Scripts](#) (Jean-Daniel Fekete). "A small perl script to translate from TEI Lite to teixlite..."

- [TEI and Other Software](#) [Older list]

Articles, Papers, News

[This section under revision.]

- [September 05, 2003] "[XML Matters: TEI -- the Text Encoding Initiative. An XML Dialect for Archival and Complex Documents.](#)" By [David Mertz](#) (Encoder, Gnosis Software, Inc). From IBM developerWorks, XML zone. [XML is usually thought of as a markup technique utilized by programmers to encode computer-oriented data. Even DocBook and similar document-oriented DTDs focus on preparation of technical documentation. However, the real roots of XML are in the SGML community, which is largely composed of publishers, archivists, librarians, and scholars. The Text Encoding Initiative uses XML in the markup of literary and linguistic texts. TEI allows useful abstractions of typographic features of source documents, but in a manner that enables effective searching, indexing, comparison, and print publication -- something not possible with publications archived as mere photographic images.] "The Text Encoding Initiative (TEI) is a decade older than XML itself, and older than other common documentation encoding XML schemas like DocBook. Specifically, TEI was developed -- in initial SGML form -- in 1987, almost an eternity in Internet time. Despite its age, TEI works at a different level than any other markup format that I am aware of, and remains the best solution to a certain class of problems... TEI aims to [enable encoding of] all the semantically significant aspects of literary texts, both old ones that predate XML technology, or indeed, computers in general, and newly created ones. Certainly the words themselves are the most important semantic feature of prose or poetical texts. But throughout the history of print -- or of writing in general -- other typographic features have been added to texts to encode subsidiary aspects of their meaning. The use of presentation elements -- such as various types of emphasis, indentation and margins, tables, pagination, line breaks (as in verse), graphics, and decorations -- has enhanced, elaborated, or modified the meanings of the words in books, essays, pamphlets, flyers, bills, poems, liturgicals,

and all the other forms literary works take. Moreover, mere typographic features sometimes require an interpretive effort to fully decipher. As a trivial example, many books use italics to mark both foreign words and to mark the titles of other books. The semantic aspect of italicization depends on the verbal context, but clearly authors usually use such marks with distinct intentions. TEI aims to allow the markup of texts in a way that distinguishes all such meaningful aspects. TEI is not really just an 'XML schema'; it is more like a whole family of schemas, related in their general goal but varying in details of the tags and attributes used. In part, these schemas differ in being supported by different DTDs (or RELAX NG schemas). For example, TEI-Lite is a greatly simplified form of TEI that aims to support '90% of the needs of 90% of the TEI user community' (according to the TEI Web site). And other specializations are available as well. But even apart from actual specializations or subsets of the full TEI tag set, most users will utilize only a few of the tags available in the TEI DTD they are using. Different documents demand different markup, and different projects allow differing degrees of granularity... any tool that can work with XML can work with TEI. DTDs are available for several TEI variations, as are XSLT stylesheets of various sorts. Naturally, customizations for working with TEI in Emacs, Framemaker, and MS-Word can be found at the TEI Web site. An XMetal customization is also downloadable. An interesting online tool provided by the initiative lets you customize an XSLT stylesheet to produce just the HTML output you desire. A Web form lets you select a variety of options, then returns a stylesheet reflecting your customizations..."

- [December 24, 2002] ["DALF Guidelines for the Description and Encoding of Modern Correspondence Material."](#) By Edward Vanhoutte & Ron Van den Branden. Version 1.0. Discussion Draft. Centrum voor Teksteditie en Bronnenstudie [Centre for Scholarly Editing and Document Studies], Royal Academy for Dutch Language and Literature, Gent, Belgium. "In view of its assignment to study and valorize the Flemish literary and musical heritage, the Centre for Scholarly Editing and Document Studies (Centrum voor Teksteditie en Bronnenstudie - CTB) has launched the DALF project. DALF is an acronym for "Digital Archive of Letters by Flemish authors and composers from the 19th & 20th

century. It is envisioned as a growing textbase of correspondence material which can generate different products for both academia and a wider audience, and thus provide a tool for diverse research disciplines ranging from literary criticism to historical, diachronic, synchronic, and sociolinguistic research. The input of this textbase will consist of the materials produced in separate electronic edition projects. The DALF project can be expected to stimulate new electronic edition projects, as well as the international debate on electronic editions of manuscripts. In order to ensure maximum flexibility and (re)usability of each of the electronic DALF editions, a formal framework is required that can guarantee uniform integration of new projects in the DALF project. Therefore, the project is from the start aimed at adherence to international standards for electronic text encoding. An important formal standard used in the DALF project is XML, that enables the definition of structural text-grammars as Document Type Definitions (DTD). Also in the construction of such a DTD that is suitable for scientific markup of correspondence material, we tried to align with international efforts to define markup schemes. Without going into detail here, the insights and practices presented in international projects like [TEI](#) (Text Encoding Initiative), [Master](#) (Manuscript Access through Standards for Electronic Records), and [MEP](#) (Model Editions Partnership) were taken into consideration for the implementation of following requirements in a DTD for correspondence material..." See the [announcement](#) "DALF Guidelines and DTD Discussion Draft Version."

- [November 26, 2002] A [posting](#) from [Sebastian Rahtz](#) (Oxford University Computing Services Information Manager) announces updated [Relax NG Schemas for the TEI](#). "There are RelaxNG schemas for MathML and SVG and a demonstration of how to include them in a TEI Relax NG schema and document. I have devised a crude way to 'flatten' a Relax NG schema to remove inclusions and redundant definitions, yielding a single portable file with no dependencies. For each of my example TEI Schemas, I have used James Clark's [trang program](#) to generate a W3C Schema (.xsd schema file). The next stage in this exercise will be to rewrite the TEI "pizzachef" tool to work with the RelaxNG version of the TEI, and generate DTD Relax and W3C constraints according to the

users specifications. Comments on any of the above very welcome... [The relevant directory] contains a set of Relax NG Schema specifications corresponding to TEI P4. They were created automatically from the [ODDs source of the TEI](#), and are kept in sync; you can download all the .rng files in a [zip file](#)."

- [September 28, 2002] "[The Music Encoding Initiative \(MEI\)](#)." By [Perry Roland](#) (Digital Library Research & Development Group, University of Virginia Library). Paper presented at [MAX 2002 - International Conference Musical Application using XML](#), September 19 - 20, 2002. State University of Milan, Italy. [This paper draws parallels between the Text Encoding Initiative (TEI) and the proposed Music Encoding Initiative (MEI), reviews existing design principles for music representations, and describes an Extensible Markup Language (XML) document type definition (DTD) for modeling music notation which attempts to incorporate those principles.] "... TEI is mute regarding the 'proper' way to compose text. Even when texts are initially created using the TEI DTD, they are still essentially transcriptions of an ur-text. Similarly, the MEI does not attempt to encode all musical expression, but instead limits itself to the written form of music, *i.e.*, common music notation (CMN). Like the TEI, the MEI must also remain unconcerned with how music is created. It is not primarily an aid to musical composition just as the TEI does not function as an aid in the creation of text. Some may see the adoption of CMN as the basis for encoding as too limiting. Legitimate arguments could be made for an entirely new form of music notation for the purpose of electronic transcription. However, common music notation is applicable to a wide range of contemporary and, perhaps more importantly, historical music. It has been eloquently described by Selfridge-Field as 'the cornerstone of all efforts to preserve a sense of the musical present for other and later performers and listeners'. Given its expressiveness, extensibility, nearly universal usage, and longevity, there seems to be little reason not to adopt CMN as the starting point for the MEI. The fact that the MEI fundamentally conceives of music as notation does not limit its usefulness for encoding performance and analytical information. While it cannot rival a human rendition, a basic performance suitable for many purposes may be mechanically derived from the notation. Of course, any additional information necessary to

complete this process may also be encoded. Likewise, descriptive and critical information may be included to assist bibliographic and analytical applications. Ultimately, a limited scope makes the design of a representation easier. For example, both the pitch and rhythm models can be greatly simplified when non-CMN requirements are not considered... Because progress toward an encoding standard for music notation is much more feasible when not locked into constant re-invention of past wheels, large parts of the design of the MEI DTD are drawn from existing standards. On the largest scale, the MEI is modeled upon the TEI. At lower levels, the Acoustical Society of America (ASA) system is used to record pitch information, performancespecific data is encoded using elements which have similar names and functions as those in the Musical Instrument Digital Interface (MIDI) standard, most of the mark up for text is designed to be familiar to users of HTML, and TEI header and Dublin Core elements form the basis of the meta-data components. Of course, the Unicode standard underlies the character encoding model for XML, obviating the need to re-invent special character encoding schemes. Finally, while it is not a formal standard, a well-known, authoritative source [Gardner Read, *Music Notation: A Manual of Modern Practice*, 2nd ed., 1979] has been used as the basis for the grammar for music notation parts of the MEI..." An [alpha version XML DTD](#) is available. See: (1) "[Music Encoding Initiative \(MEI\)](#)"; (2) general references in "[XML and Music.](#)" [[cache](#), [conference reference](#)]

- [June 15, 2002] "[New XML Edition Of Text Encoding Guidelines Published.](#)" - "The Text Encoding Initiative (TEI) Consortium (www.tei-c.org) announces publication of a new, updated version of their Guidelines for Electronic Text Encoding and Interchange, known as P4. The Consortium, now in its second year, is an international non-profit corporation set up to maintain and develop the TEI system, which has become the de facto standard for scholarly work with digital text since its first publication in 1994. The launch of a fully XML-compliant version of the TEI Guidelines is a significant advance, placing the TEI firmly in the mainstream of current digital library and World Wide Web developments. The new edition has been available online for a few months, and will continue to be so, but the print edition now available from the University of

Virginia Press (URL) marks a new milestone in the history of this long standing exercise in scholarly communication and international co-operation. In simple terms, the TEI Guidelines define a language for describing how texts are constructed and propose names for their components. By defining a standard set of names the Guidelines make it possible for different computer representations of texts to be combined into vast databases, and they also provide a common language for scholars wishing to work collaboratively. There are many such standard vocabularies in the industrial world -- in banking, in aircraft maintenance, or in chemical modelling, for example. The TEI's achievement has been to try to do the same thing for textual and linguistic data -- both for those working with the written culture of the past and for those studying the development of language itself. Membership in the TEI Consortium has climbed steadily during its first year of operation, standing at 56 members worldwide in May 2002, ranging from small university research projects to major academic libraries and institutions. The consortium offers a range of membership benefits including participation in TEI elections, special access to training, consultation on grant proposals, and free or discounted copies of the TEI Guidelines. The Consortium is actively recruiting and welcomes inquiries at info@tei-c.org. The Consortium is now planning its second annual members' meeting, to be held at the Newberry Library in Chicago on October 11 and 12, 2002. At the annual meeting members have the opportunity to learn about new developments and future plans for the TEI Guidelines, share research with other TEI members, and attend special training sessions... [Print] Copies of P4 may be ordered from the University of Virginia Press (or via the [TEI website](#))..." The Guidelines are also available online.

- [March 22, 2002] ["Relax NG Schemas for TEI P4."](#) Prepared by [Sebastian Rahtz](#) (OUCS Information Manager). "... Relax NG schemata for the TEI which are up to date with the latest version of P4 (now effectively frozen), and are derived automatically from the [ODD] source of [TEI P4...](#)" See also the [ZIP package](#). [[cache 2002-03-23](#)]
- [February 19, 2002] [Updated TEI PizzaChef Tool Supports XML DTD Generation.](#) A posting from [Lou Burnard](#) (Oxford Computing Services)

announces the release of an updated TEI 'PizzaChef' tool, accessible online from the TEI Consortium websites at the University of Virginia and Oxford University. The updated tool uses the *P4 XML Edition* DTD modules from the *TEI Guidelines* and produces only XML DTDs. Using a baking metaphor, the PizzaChef tool enables the designer to create a personalized TEI-conformant document type definition (DTD) simply by clicking radio buttons and check-boxes. "The TEI Guidelines define several hundred elements and associated attributes, which can be combined to make many different DTDs, suitable for many different purposes, either simple or complex. With the aid of the PizzaChef, you can build a DTD that contains just the elements you want, suitable for use with any XML processing system." The *Text Encoding Initiative Guidelines* themselves support SGML and XML, representing an "international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent."

- [The Pizza Chef: a TEI Tag Set Selector](#) - A tool which "allows you to select the TEI tagsets you want from a menu, and also to pick out individual elements for inclusion, exclusion, or modification. You can then download a customized DTD subset, or a completely compiled (*i.e.*, non parameterized) DTD for use by *e.g.* Softquad's Rulesbuilder. This last function is accomplished by means of Michael Sperberg-McQueen's carthage program." [Possible alternate URL](#).
- [April 09, 2001] [Funding for TEI Guidelines into XML Format](#). National Endowment For The Humanities Grants, April 2001. Preservation and access. Charlottesville, TEI Consortium. Grant amount: \$131,963. PROJECT DIRECTOR: Steve J. DeRose, (301) 315-0232. PROJECT TITLE: Converting Text Encoding Initiative Guidelines and Documentation into the XML Format. DESCRIPTION: Conversion of the Text Encoding Initiative guidelines to the Extensible Markup Language format, which will allow easier use and distribution of structured humanities documents via the Web.
- [TEI Lite DTD in XML](#) with public identifier: "-//TEI//DTD TEI Lite XML ver. 1.0//EN" (or "-//TEI//DTD TEI Lite XML ver. 1//EN") [[local archive copy](#)]
- [TEI XML Extensions file for Entities](#) [[local archive copy](#)]

- [TEI XML Extensions file DTD \[local archive copy\]](#)
- [October 24, 2001] "[TEI and XML: A Marriage Made in Heaven. An Introduction to The Future of Digital Information.](#)" By Lou Burnard (Manager of the Humanities Computing Unit at Oxford University Computing Services; TEI Editor, Europe). Presented at "Computing Arts 2001: Digital Resources for Research in the Humanities," University of Sydney 26 - 28 September 2001. "One of the more striking features of XML, by comparison with its progenitor SGML, is the fact that you can use XML without having to know what a document type definition (DTD) is. Documents need only be syntactically valid, and there is no longer any requirement of an application to understand their structure in advance. One consequence of this is that the most effective ways of using XML are in well-defined application areas, or well-defined user communities, where there is a pre-existing consensus as to the meaning of elements and their attributes. Another is that in larger application areas and less well-defined user communities would-be XML users continue to need ways of defining DTDs if they are to benefit from the claimed advantages of XML for document interchange and re-usability. If XML is to be the basis of a new digital demotic, in which a thousand distributed applications can share access to a pool of distributed digital resources, we need to define something more than structure and syntax for that demotic. In this talk I will outline how the Text Encoding Initiative (TEI) 'Recommendations' of 1994 attempted to define an effective framework for the construction of user- and application- specific DTDs. The abbreviation 'DTD' actually has two expansions -- document type Declaration, and document type Definition -- which should not be confused. Originally expressed as a large but modular Document Type Definition, the TEI Guidelines consist essentially of intended semantics for several hundred element types. This definitional work, by setting out formally a broad based consensus as to the topoi of scholarly encoding, is probably one of the more significant contributions to scholarly work made by the TEI. In addition, the Guidelines may be seen as a very loose and generic Document Type Declaration, providing a syntactic framework within from almost any desired document grammar can be constructed. Separating these two aspects helps us see how the TEI recommendations are

peculiarly apt for the XML age. They define a number of distinct vocabularies, simplifying where simplification is appropriate, but allowing for any required depth of semantic complication or enrichment. Because these vocabularies share a common syntactic basis, their exploitation by a wide range of open software tools and systems is greatly facilitated. Because they are robustly and formally defined, moreover, it is possible to add new vocabularies which can build on existing fragments in a controlled and compatible way. With the establishment of a new Consortium to manage and promote the further development of the scheme, and in particular with the publication of the new XML based version (P4), the groundwork has been laid for a new chapter in this attempt to apply the creative energies of the humanistic research community to its traditional task of communicating and preserving cultural heritage and its interpretation. The presentation will briefly review the history and motivation of the design of the TEI system and will also give a flavour of the range of application areas in which it has been successful. The main focus however will be on the future development of the TEI in its new guise as an environment for the construction of compatible XML vocabularies appropriate to many different research areas."

- [October 29, 2001] "Descriptive Meta Data Strategy for TEI Headers: A University of Michigan Library Case Study." By Lynn Marko and Christina Powell (University of Michigan, Ann Arbor). In [OCLC Systems And Services](#) Volume 17, Number 3 (2001), pages 117-120. ISSN: 1065-075X. "The Text Encoding Initiative (TEI) standard was developed for humanities scholars to encode textual documents for data interchange and analytic research. Its header segment contains rich tag sets, which can sufficiently support library cataloging practice with AACR2 rules and authority control. This article presents a strategy that is currently used by the Making of America (MoA) project for transferring complete MARC data created on the library's online system to the header of the TEI encoded documents. It also describes the cooperation for achieving this task between the Digital Library Production Services (DLPS) and Monograph Cataloging Division at the University of Michigan library." See with DLPS the [University of Michigan Digital Library eXtension Service \(DLXS\)](#) which "provides the foundation and the framework for educational and non-profit institutions to fully develop their

digital library collections. The newest DLXS enhancement, XPAT, is a powerful, SGML-aware search engine, and an ultra-versatile tool for the development of digital libraries. XPAT provides excellent support for word and phrase searching, indexing of SGML elements and attributes, fast retrieval, and open systems integration... The XPAT engine is an XML/SGML-aware search engine that the University of Michigan has deployed with an extremely diverse set of digital library resources. XPAT is based on the search engine previously marketed by Open Text as OT5, and sometimes referred to as 'Pat' and 'Pat5.0.' Because of XPAT's origins and the extent to which it has been employed in University of Michigan digital library projects, we are confident about the search engine's reliability, its core functionality, and many aspects of its scalability. XPAT provides excellent support for word and phrase searching, indexing of XML and SGML elements and attributes, extremely fast retrieval, and open systems integration. For example, among the many collections that use XPAT is the 3 million page, 7Gb, 1.5 billion word Making of America collection. As part of the [UM DLXS](#), the University of Michigan Digital Library Production Service has launched a continuous development process in which we have added a number of features to XPAT. We have introduced support for valid and well-formed XML, Linux binaries, better error handling, and improved indexing performance for XML/SGML elements, attributes, and tags." Contact [John Price-Wilkin](#)." See also "[The Making of America II Project](#)."

- [March 23, 2001] [Converting Leiden-style editions to TEI Lite XML](#). By T. J. Finney. Draft, 2001. Unofficial. "These recommendations concern the translation into TEI Lite documents of printed editions that employ the Leiden conventions defined in Chronique D'Egypte 13-14 (1932), pages 285-7. They may also be applied where a transcription is made directly from a manuscript. TEI Lite is an extensible markup language (XML) version of the TEI Lite document type definition. TEI Lite (TEI U5) represents a subset of the full Text Encoding Initiative guidelines (TEI P3). The recommendations should be read in conjunction with the TEI Lite specification. Although TEI Lite is adequate for most features encountered in a printed edition, there are situations where the encoding methods of the full TEI guidelines are better. Following TEI Lite allows the present recommendations to use a widely adopted

framework that is relatively well supported. This in turn should maximize the utility of Leiden-style editions that have been translated into TEI Lite documents according to these recommendations. However, the gain is achieved at a cost of bending less appropriate features of TEI Lite to purposes for which entirely appropriate features exist in the full TEI guidelines. This set of recommendations takes a minimalist approach to rendering features likely to be encountered in Leiden-style transcriptions. A more comprehensive approach that used an XML version of the full TEI guidelines would be less vulnerable to charges of 'tag abuse'..." [[cache](#)]

- [October 19, 2000] "[Text Encoding for Interchange: A New Consortium.](#)" By [Lou Burnard](#) (Manager of the Humanities Computing Unit at Oxford University; European Editor of the Text Encoding Initiative since 1990, University of Oxford). In [Ariadne](#) [ISSN: 1361-3200] Issue 24 (June 2000). ['Lou Burnard on the creation of the TEI Consortium which has been created to take the TEI Guidelines into the XML world.']. . . The goal of the new [TEI Consortium](#) is to establish a permanent home for the TEI as a democratically constituted, academically and economically independent, self-sustaining, non-profit organization. This will involve putting the Consortium on solid legal and organizational footing, developing training and consulting services that will attract paying members, and providing the administrative support that will allow it to continue to exist while income from membership grows. In the immediate future, the Consortium will launch a membership and publicity campaign the goal of which is to bring the new Consortium and the opportunity to participate in it to the attention of libraries, publishers, and text-encoding projects worldwide. Its key message is that the TEI Guidelines have a major role to play in the application of new XML-based standards that are now driving the development of text-processing software, search engines, Web-browsers, and indeed the Web in general. . . The future usefulness of vast collections of electronic textual information now being created and to be created over the coming decades will continue to depend on the thoughtful and well-advised application of non-proprietary markup schemes, of which the TEI is a leading example. We may expect that in the future some of the more trivial forms of markup will be done by increasingly sophisticated software, or even

implied from non-marked-up documents during processing. As XML and related technologies become ever more pervasive in the wired world, we may also expect to see a growing demand for interchangeable markup standards. What is needed to facilitate all of these processes is a sound, viable, and up-to-date conceptual model like that of the TEI. In this way, the TEI can help the digital library, scholar's bookshelf, and humanities textbooks survive into a future in which they can respond intelligently to our queries, can combine effectively with conceptually related materials, and can adequately represent what we know about their structure, content, and provenance."

- [September 29, 2000] "[Guidelines for Markup of Electronic Texts.](#)" Edited by Peter C. Gorman, UW-Madison TEI Markup Guidelines Working Group; Endorsed by the UW-Madison Libraries Digital Steering Committee September 11, 2000. September 29, 2000. "This document is intended for use by staff using the Text Encoding Initiative (TEI) Guidelines TEIP3 to mark up electronic texts for inclusion in the UW-Madison Libraries' digital collections. It is not relevant to other types of projects using SGML encoding, e.g., page-image projects or digital finding aids. Some of the content has been quoted or adapted from other published guidelines, which are referenced in each case. The purpose of this document is not to teach or otherwise document the TEI itself, but rather to create a profile of the TEI for use in the UW-Madison digital library collections. It is assumed that the user is already familiar with TEI markup. The motivation for creating these guidelines is a desire to create a consistent and scalable infrastructure for text encoding projects, whereby new works can be created and added to the collection with minimal development effort on the part of project leaders, text encoders, and technical staff. At the same time, text encoded according to these guidelines should provide a suitable base for further elaboration or expansion by future encoders with minimal restructuring. At any point in this document, you can click on the magnifying glass icon to see examples of the point being discussed. The examples will open in a new window. [...] The primary motivation for creating this document was a desire to define encoding standards for a 'base' level: the minimal level of markup we would accept for locally-produced collections. The result, a 'Reading Level', falls somewhere between the poles of 'use

nothing but <div0>, <p>, and <lb>' and 'TEILite is useless for real documents'. But why define a minimal level at all? For us, the answer is that we want to provide basic ('reading') access to as many materials as possible (as appropriate for the curricular and research needs of our campus), but the production of marked-up texts can be expensive."

- [September 12, 2000] [OCLC Systems & Services, TEI Special Issue, Call for Papers](#) - OCLC *Systems & Services* journal v.17, no.3 issue to be devoted to TEI applications.
- [August 28, 2000] "[The Relationship Between General and Specific DTDs: Criticizing TEI Critical Editions.](#)" By djb@clover.slavic.pitt.edu David J. Birnbaum (Department of Slavic Languages and Literatures, 1417 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260 USA). Paper presented at the [Extreme Markup Languages 2000](#) (August 13 - 18, 2000, Montréal, Canada). Published as pages 9-27 (with 13 references) in *Conference Proceedings: Extreme Markup Languages 2000. 'The Expanding XML/SGML Universe'*, edited by Steven R. Newcomb, B. Tommie Usdin, Deborah A. Lapeyre, and C. M. Sperberg-McQueen. "The present study discusses the advantages and disadvantages of general vs specific DTDs at different stages in the life of an SGML document based on the example of support for [textual critical editions in the TEI](#). These issues are related to the question of when to use elements, attribute, or data content to represent information in SGML and XML documents, and the article identifies several ways in which these decisions control both the degree of structural control and validation during authoring and the generality of the DTDs. It then offers three strategies for reconciling the need for general DTDs for some purposes and specific DTDs for others. All three strategies require no non-SGML structural validation and ultimately produce fully TEI-conformant output. The issues under consideration are relevant not only for the preparation of textual critical editions, but also for other element-vs-attribute decisions and general design issues pertaining to broad and flexible DTDs, such as those employed by the TEI. [...] General Conclusions: Any of the three strategies discussed (processing a modified TEI DTD with respect to TEIform attribute values, transformation of a custom DTDs to a TEI structure, and architectural forms) provides a

solution to the issues posed by a score-like edition. Specifically, these strategies all permit much greater structural control than is available in the standard TEI DTDs, rely entirely on SGML for all validation, and produce a final document that is fully TEI-conformant." See also ["Text Encoding Initiative \(TEI\)."](#)

- [August 28, 2000] ["A TEI-Compatible Edition of the Rus' Primary Chronicle."](#) By djb@clover.slavic.pitt.edu>[David J. Birnbaum](#) (Department of Slavic Languages and Literatures, 1417 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260 USA). To be published in *Medieval Slavic Manuscripts and SGML: Problems and Perspectives*, Anisava Miltenova and David J. Birnbaum, ed), Sofia: Institute of Literature, Bulgarian Academy of Sciences, Marin Drinov Publishing House. 1999. In press [2000-08-26]. "This report describes the development of a TEI-conformant SGML edition of the Rus' Primary Chronicle (*Povest' vremennykh let*) on the basis of an electronic transcription of the text that originally had been prepared for paper publication using troff. The present report also discusses strategies for browsing, indexing and querying the resulting SGML edition. Selected electronic files developed for this project are available at a web site maintained by the author. . . The Rus' Primary Chronicle (PVL) tells the history of Rus' from the creation of the world through the beginning of the twelfth century. It was based on both Byzantine chronicles and local sources and underwent a series of redactions before emerging in the early twelfth century in the form that scholars currently identify as the PVL. This text was then adopted as the foundation of later East Slavic chronicle compilations. [. . .] I decided to use the Text Encoding Initiative (TEI) document type description (DTD) for the SGML edition of the PVL for two reasons. First, the TEI DTD is widely used, which means that a TEI-conformant edition of the PVL can be processed using existing tools and can easily be incorporated into existing TEI-oriented digital libraries. Second, the support for critical editions in the TEI DTD was developed with input from an international committee of experienced philologists from different disciplines, and it was clearly sensible to take advantage of their careful analysis of issues confronting the preparation of critical editions, particularly in an electronic format. In fact, the [TEI DTD supports three different encoding strategies](#) for critical editions (the location-referenced method, the

double-end-point-attached method, and the parallel segmentation method), and my decision to adopt a TEI approach required me to evaluate and choose among those strategies. . .

[Conclusions:] In general, any electronic edition will provide faster searching and retrieval than a paper edition. If one wishes to take the structure of a document into consideration, an SGML document will support more sophisticated structural queries than plain text or text with procedural markup (such as troff). The present report has documented the generation of a TEI-conformant SGML edition of the PVL from troff source using free tools. It has also illustrated the convenience of browsing and searching the text in Panorama, which includes support for queries that refer to the SGML element structure. This report has also described the use of Pat in a web-based environment to retrieve and render only selected portions of the document. Although Pat does not support regular expressions directly, this report has outlined a method for overcoming this limitation." See also the [PVL reference document](#).

On TEI: ["Text Encoding Initiative \(TEI\)." \[cache\]](#)

- [August 02, 1999] TEI Recommendation for 'Best Encoding Practices'. A posting from [C. Perry Willett](#) (Indiana University) [announced](#) the publication of a draft document on best encoding practices in library applications of TEI's *Guidelines*. The draft document is meant for projects using the TEILite DTD, now available for [XML encoding](#) as well as SGML. The document is: [TEI Text Encoding in Libraries. Draft Guidelines for Best Encoding Practices](#). The guidelines provide for encoding at five levels, depending upon project scope and user requirements. "Encoding levels 1-4 require no expert knowledge of content. Level 5, in contrast, requires scholarly analysis. Levels 1-4 allow the conversion and encoding of texts to be performed without the assistance of content experts and can be enriched with more markup at any time. Recommendations for Levels 1-4 are intended for projects wishing to create encoded electronic text with structural markup, but minimal semantic or content markup. Also, the encoding levels are cumulative: encoding requirements at each level incorporate the requirements of lower levels. The recommendations are concerned with the text portion of a TEI-encoded document. The levels are: (1) Fully Automated Conversion and Encoding: create electronic text with the primary purpose of keyword searching and linking to page

images. The primary advantage in using the TEILite DTD at this level is that a TEI Header is attached to the text file. (2) Minimal Encoding: create electronic text for keyword searching, linking to page images, and identifying simple structural hierarchy to improve navigation. (3) Simple Analysis: create text that can stand alone as electronic text and identifies hierarchy and typography without content analysis being of primary importance. (4) Basic Content Analysis: create text that can stand alone as electronic text, identifies hierarchy and typography, specifies function of textual and structural elements, and describes the nature of the content and not merely its appearance. This level is not meant to encode or identify all structural, semantic or bibliographic features of the text. (5) Scholarly Encoding Projects : Level 5 texts are those that require subject knowledge, and encode semantic, linguistic, prosodic or other elements beyond a basic structural level."

- [June 17, 1999] "[Construction of an XML Version of the TEI DTD.](#)" By Lou Burnard and C. M. Sperberg-McQueen. TEI Document TEI ED W69. 'June 17, 1999'. Abstract: "This document describes issues involved in creating an XML version of the SGML document type definition (DTD) created by the Text Encoding Initiative, and proposes solutions. It defines a TEI extensions file which incorporates those solutions, in order to allow experimentation. The discussion of inclusion exceptions defines a method of rewriting SGML content models so as to achieve effects similar to those provided by inclusion exceptions. To make an SGML document type definition compatible with XML, inclusion exceptions must be eliminated. The simplest method of ensuring that this change does not invalidate existing documents is to modify the content model of every element which can occur as a descendant of any element with inclusion exceptions in its content model, in the manner described here. That will ensure that elements named in inclusion exceptions remain legal in all the locations where they are currently legal. The methods of changing content models described in this paper are believed to preserve determinism (what ISO 8879 calls lack of *ambiguity*) and to simulate the effects of inclusion exceptions properly. At this point, however, no proof of either conjecture is offered." See also the [SGML version](#). [[local archive copy](#)]
- [Guidelines for Electronic Text Encoding and Interchange](#). Revised Reprint, Oxford, May 1999.

- See also, as an example of an XML application based upon TEI, "[Manuscript Access through Standards for Electronic Records \(MASTER\).](#)" More than [ninety](#) projects worldwide use the *TEI Guidelines* as a basis for SGML/XML encoding for literary and linguistic texts.

Early History of TEI XML Version

[1998 description] C. Michael Sperberg-McQueen (University of Illinois) is both an Editor of the TEI Project, and XML co-editor. The [TEI Extended Pointer](#) language plays a significant role in the design of XLink and XPointer - the two major components in XML's linking language. The W3C's "[XML Specification DTD](#)" is based in part on the [TEI Lite](#) and Swab DTDs, the latter being an effort largely of Michael Sperberg-McQueen. While the TEI P3 Guidelines now provide DTDs for SGML encoding, effort is underway to make the Guidelines accessible to XML users as well. The TEI has recently chartered a workgroup on architectural issues, chaired by Frank Tompa, where one of its specific charges is the development of an XML version of the full TEI DTD. A conference "[TEI and XML in Digital Libraries](#)" is to be held in the summer of 1998, sponsored by the Digital Library Federation and held Library of Congress, Washington, DC.; one of the goals is to "explore the impact of Extensible Markup Language (XML), and XML-conformant TEI, on digital library efforts."

References:

- [Announcement on TEI-L, "XML 1.0 Is Official."](#)
From TEI Editor, C. M. Sperberg-McQueen. Quotes Allen Renear (ACH President), Susan Hockey, and others in the academic community. Also: [TEI Web site](#)
- Unofficial [work on an XML version](#) of the TEI Lite DTD
- Conference: "[TEI and XML in Digital Libraries.](#)"
- [TEI, SGML and XML Resources](#)
- [May 13, 1999] [Computers and the Humanities](#) [The Official Journal of The Association for Computers and the Humanities.] Volume 33 Nos. 1-2, April 1999. ISSN: 0010-4817. Special Double Issue: Tenth Anniversary of the Text Encoding Initiative. Edited by Nancy Ide [Dept. of Computer Science, Vassar College, USA] and Dan

Greenstein [Arts and Humanities Data Services, King's College, UK]. This issue contains an article by Steve DeRose, "XML and the TEI" (pages 11-30). Also: Jon Bosak, "XML Ubiquity and the Scholarly Community" (pages 199-206). See the [Table of Contents](#)

- [May 13, 1999] Lou Burnard wrote on TEI-L, 11-May-1999, in response to a question by Fotis Jannidis ("...Does anybody know whether the long announced work on a conversion/adaption of the TEI dtDs to XML dtDs has begun, whether a working group has started on this task or whether P. Bonhomme's trial version is still the only thing around?..."): Michael [Sperberg-McQueen] and I have been working on this for the last few months. We have a working draft, almost complete, of a set of TEI extension files which will enable us to generate XML-compatible of any view of the TEI dtd. The first thing we produce with it will be a real XML version of TEI Lite (Patrice B.'s version is only a toy) and we hope to have this available by the ACH-ALLC conference next month [= June 1999]." On the unofficial work, see: (1) [Patrice Bonhomme](#) ('A personal XML release of the TEI Lite DTD') and (2) [Rick Jelliffe](#) ('TEI Lite and Loose DTD, a version of the TEI Lite DTD suitable for use with XML documents'. [\[local archive copy\]](#)

Document URL: <http://xml.coverpages.org/tei.html> — [Legal stuff](#)
