

Appendix C Prefatory Notes

This Appendix contains (in reverse chronological order) the 'Introductory Notes' prefixed to each revision of the TEI Guidelines since its first publication in 1994.

Appendix C1. Introductory Note (November 2001)

To complete the work started in June of this year, the TEI Editors asked for volunteers from the TEI community to proof-read the preliminary XML version. 24 volunteers responded to this call during August, and gave invaluable help both by identifying a number of previously un-noticed errors, and by suggesting areas in which more substantial revision should be undertaken in the future. The Editors gratefully acknowledge the assistance of the following individuals during this exercise:

Jimmy Adair, Syd Bauman, Michael Beddow, Steven Bird, Lisa Charlong, Matthew Driscoll, Patrick Durusau, Tomaz Erjavec, Nick Finke, Tim Finney, Julia Flanders, Mike Fraser, Pankaj Kamthan, François Lachance, Terry Langendoen, Anne Mahoney, Gregory Murphy, Daniel Pitti, Rafal Prinke, Laurent Romary, Stewart Russell, Gary Simons, Elisabeth Solopova, Christian Wittern, Martin Wynne.

In addition to error correction, and clear delineation of those sections in which substantial revision is yet to be undertaken for TEI P5, the present draft differs from earlier ones in the following respects:

- Formal Public Identifiers have been introduced as a means of constructing TEI DTDs and an SGML Open Catalog is now included with the standard release;
- Some systematic errors and omissions in the reference section have been removed; the format of this section has been substantially changed, we hope for the better;
- The chapters on obtaining the TEI DTDs and WSDs have been brought up to date; the chapter on modification has been expanded to include a discussion of the TEI Lite customization;
- All examples and cited markup has been checked for XML validity against the published DTDs, and corrected where faulty; examples have been formatted in a (more or less) consistent style.

Lou Burnard and Syd Bauman (Editors)

Oxford and Providence, November 2001.

Appendix C2. Introductory Note (June 2001)

This is a preliminary version of a revised and fully XML-compliant edition of the TEI Guidelines. Although work on revising and correcting the text of the document is incomplete, by making available this preliminary version we hope to facilitate testing of the XML document type declarations which it describes by as wide a range of TEI users as possible.

The primary goal of this revision is to make available the corrected (May 1999) edition of the Guidelines in a new version which:

- is expressed in XML and itself conforms to a TEI-conformant XML DTD;
- generates a set of XML DTD fragments that can be combined together in the same way as the existing TEI (P3) SGML DTD fragments to form true TEI XML DTD fragments without loss of functionality;
- can be processed and maintained using readily available XML tools instead of the special-purpose ad hoc software originally used for TEI P3.

As noted elsewhere, a number of errors were corrected in the May 1999 edition. A (much) smaller number of errors have also been corrected in this edition, but no new material has been added. We expect the expansion and modification of the Guidelines to become a real possibility in the context of the newly formed TEI Consortium, which has funded the preparation of this present edition.

A major design goal of both this and the previous 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.

In making this new version, we relied extensively on preliminary work carried out by the outgoing North American editor of the TEI Guidelines, Michael Sperberg-McQueen. In a TEI working paper written in 1999, TEI ED W69, Michael sketched out a precise blueprint for the conversion of the TEI from SGML to XML, which we have implemented, with only slight modification. The current TEI editors wish to express here our admiration for the detailed care put into that paper, without which our task would have been forbiddingly difficult, if not impossible. We would also like to express our thanks to Sebastian Rahtz of Oxford University Computing Services, for his invaluable assistance in preparing this new edition.

We list here in summary form all the changes made in the present edition. Full technical details are provided in documents TEI EDW69 and TEI EDW70, available from the TEI website.

- A new keyword TEI.XML has been added. By setting its value to INCLUDE, rather than the default IGNORE, the user can request generation of an XML rather than an SGML DTD;
- The content models of all elements have been checked, and, where necessary, changed so that they are equally valid as SGML or as XML;
- The declared value for all attributes has been changed to a form which is equally valid as SGML or as XML;
- All the examples have been checked for conformance and converted to use XML syntax, where possible. (This process is currently incomplete.)
- Some errors and duplications in the class membership of elements from the names and dates tagsets have been corrected.

To implement the first of these, we have parameterized the *tag omissibility* indicators “- o” and “- -” used within element declarations in the DTD. When XML is to be generated, the parameter entities concerned are redeclared with the null string as their value.

The second change was achieved by removing SGML-specific features (ampersand connectors, inclusion and exclusion exceptions, various types of attribute content) from the DTD and revising the syntax of the DTD to conform to XML requirements (specifically in the representation of mixed-content models, and by removing redundant parentheses). In making these changes, we took care to ensure that the resulting content model would continue to accept existing valid documents, though in the nature of things it could not be guaranteed to reject the same set of documents. As further discussed in EDW69 and EDW70, some constraints (exclusion exceptions, for example) which could be carried out by a generic SGML parser using TEI P3 will have to be implemented by a special purpose TEI validator using TEI P4.

Much work remains to be done, firstly in testing the new DTD fragments against as wide a range of TEI materials as possible, secondly in revising the discussion of markup theory and practice within the text to reflect current thinking. A few sections of the current text (the Gentle Introduction to SGML and the discussion of Extended Pointer syntax are two examples) will need substantial rewriting. For the most part, however, we think the Guidelines have stood the test of time well and can be recommended to a new generation of text encoders scarcely born at the time they were first formulated.

Lou Burnard and Steve De Rose (Editors)

Oxford and Providence, May 2001.

Appendix C3. Introductory Note (May 1999)

No work of the size and complexity of the TEI *Guidelines* could reasonably be expected to be error-free on publication, nor to remain long uncorrected. It has however taken rather longer than might have been anticipated to complete production of the present corrected reprint of the first edition, for which we present our apologies, both to the many individuals and institutions whose enthusiastic adoption and promotion of the TEI encoding scheme have ensured its continued survival in the rapidly changing world of digital scholarship, and also to the many helpfully critical users whose assiduous uncovering and reporting of our errors have made possible the present revision.

At its first meeting in Bergen, in June 1996, the TEI Technical Review Committee (TRC) approved the setting up of a small working committee to oversee the production of a revised edition of the TEI *Guidelines*, to include corrections of as many as possible of the ‘corrigible errors’ notified to the editors since publication of the first edition in May 1994, the bulk of which are summarized in a TEI working paper (TEI EDW67, available from the TEI website).

During the spring of 1997, this TRC Core Subcommittee reviewed nearly 200 comments and proposals which the editors had collected from public debate and discussion over the preceding two years, and provided invaluable technical guidance in disposition of them. We are glad to take this opportunity of expressing our thanks to this subcommittee, whose members were Elli Mylonas, Dominic Dunlop, and David T. Barnard.

The work of making the corrections and regenerating the text proceeded rather fitfully during 1998 and 1999, largely because of increasing demands on the editors’ time from their other responsibilities. With the establishment of the new TEI Consortium, it is hoped that maintenance of the *Guidelines* will be placed on a more secure footing. Some specific areas in which we anticipate future revisions being carried out are listed below.

Appendix C3.1. *Typographic corrections made*

- examples of TEI markup throughout the text were all checked against the relevant DTD fragment and an embarrassingly large number of tagging errors corrected;
- various minor typographic and spelling errors were corrected;
- the ‘corrigible errors’ listed in working paper TEI EDW67 were all corrected: some of these required specific changes to the DTD which are listed in the next section.

Appendix C3.2. *Specific changes in the DTD*

A major goal of this revision was to avoid changes which might invalidate existing data, even where existing constructs seemed erroneous in retrospect. To that end, wherever changes have been made in content models for existing elements, they have as far as possible been made so that the DTD will now accept a superset of what was previously legal. Only one new element (<ab>) has been added.

Where possible, a few content models have been changed in such a way as to facilitate conversion to XML, but XML compatibility is *not* a goal of this revision.

Brief details of all changes made in the DTD follow:

- Several changes were made in class membership, in order to correct unreachability problems. Specifically:
 - elements <geogName>, <persName>, <placeName> were added to the m.data class;
 - <geogName> and <placeName> were removed from the m.placepart class;
 - the elements <addSpan>, <delSpan>, <gap>, were added to the m.Edit class;
 - a new class m.editIncl was defined, with members <addSpan>, <delSpan>, and <gap>; this class was then added to the global inclusion class m.globlIncl along with <anchor> (erroneously a member of the m.Seg class, from which it is now removed), m.metadata and m.refsys;

- added <name> element to m.addrPart class;
- added <dateLine> to m.divtop and m.divbot classes;
- added <epilogue> and <castList> to m.dramafront class;
- added <divGen> to m.front class;
- added <dateLine> to m.divtop and m.divtop classes;
- added <u> element to a.declaring class;
- defined new class m.fmchunk (front matter chunk), comprising <argument>, <byline>, <docAuthor>, <docDate>, <docEdition>, <docImprint>, <docTitle>, <epigraph>, <head>, and <titlePart> for use in simplification of the content model for <front> element;
- defined new element <ab> (anonymous block), and added it to the m.chunk class;
- corrected an error whereby global attributes were not properly defined for elements specifying a non-default value for any of the a.global attributes: elements affected include: <foreign>, <hi>, , <pb>, <lb>, <cb>, <language>, <anchor>, and <when>;
- changed content models to permit empty <list> and empty <availability> elements;
- changed content model for <series> element to permit #PCDATA;
- changed content model for <setting> element to permit <date> element as a direct child;
- added a key attribute to the <distance> element, for consistency with other elements in its class;
- changed content model for <orgName> element to make it more consistent with e.g. <persname>;
- changed content model for <opener> element to include <argument>, <byLine>, and <epigraph>;
- changed content models for <app>, <rdgGrp>, and <wit> elements;
- revised attributes on <hand> element.

A number of content models were changed with a view to easing the creation of an XML compatible version of the Guidelines. Specifically:

- removed ampersand connectors from <cit>, <respStmt>, <publicationStmt>, and <graph>;
- changed the mixed content models for <sense>, <re>, <persName>, <placeName>, <geogName>, <dateStruct>, <timeStruct>, and <dateLine> to make them XML-conformant.

Appendix C3.3. Outstanding errors

A small number of other known problems remain uncorrected in this version and are briefly listed below. Please watch the TEI mailing list for announcements of their correction.

- elements of class inter don't always behave as they should (e.g. one cannot insert a <table> before anything else in a <div>);
- some mixed-content problems consequent on the definition of specialPara need to be addressed systematically; in particular, the treatment of list items or notes which contain several paragraphs continues to surprise many users: no white space is allowed between the paragraphs;
- the resp attributes on editorial elements are not consistently defined;
- the discussions of DTD invocation, and the DTD itself, all use system identifiers instead of formal public identifiers.

Our next priority however will be the production of a fully XML-compliant version of the TEI DTD, work on which is already well advanced.

C.M. Sperberg McQueen and Lou Burnard, May 1999

Appendix C4. Preface (April 1994)

These Guidelines are the result of over five years' effort by members of the research and academic community within the framework of an international cooperative project called the Text Encoding Initiative (TEI), established in 1987 under the joint sponsorship of the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing.

The impetus for the project came from the humanities computing community, which sought a common encoding scheme for complex textual structures in order to reduce the diversity of existing encoding practices, simplify processing by machine, and encourage the sharing of electronic texts. It soon became apparent that a sufficiently flexible scheme could provide solutions for text encoding problems generally. The scope of the TEI was therefore broadened to meet the varied encoding requirements of any discipline or application. Thus, the TEI became the only systematized attempt to develop a fully general text encoding model and set of encoding conventions based upon it, suitable for processing and analysis of any type of text, in any language, and intended to serve the increasing range of existing (and potential) applications and use.

What is published here is a major milestone in this effort. It provides a single, coherent framework for all kinds of text encoding which is hardware-, software- and application-independent. Within this framework, it specifies encoding conventions for a number of key text types and features. The ongoing work of the TEI is to extend the scheme presented here to cover additional text types and features, as well as to continue to refine its encoding recommendations on the basis of extensive experience with their actual application and use.

We therefore offer these Guidelines to the user community for use in the same spirit of active collaboration and cooperation with which they have so far been developed. The TEI is committed to actively supporting the wide-spread and large-scale use of the Guidelines which, with the publication of this volume, is now for the first time possible. In addition, we anticipate that users of the TEI Guidelines will in some instances adapt and extend them as necessary to suit particular needs; we invite such users to engage in the further development of the Guidelines by working with us as they do so.

Like any standard which is actually used, these Guidelines do not represent a static finished work, but rather one which will evolve over time with the active involvement of its community of users. We invite and encourage the participation of the the user community in this process, in order to ensure that the TEI Guidelines become and remain useful in all sorts of work with machine-readable texts.

This document was made possible in part by financial support from the U.S. National Endowment for the Humanities, an independent federal agency; Directorate General XIII of the Commission of the European Communities; the Andrew W. Mellon Foundation; and the Social Science and Humanities Research Council of Canada. Direct and indirect support has also been received from the University of Illinois at Chicago, the Oxford University Computing Services, the University of Arizona, the University of Oslo and Queen's University (Kingston, Ont.), and Ohio State University.

The production of this document has been greatly facilitated by the willingness of many software vendors to provide us with evaluation versions of their products. Most parts of this text have been processed at some time by almost every currently available SGML-aware software system. In particular, we gratefully acknowledge the assistance of the following vendors:

- Berger-Levrault AIS s.a. (for Balise);
- E2S n.v. (for E2S Advanced SGML Editor);
- Electronic Book Technology (for DynaText);
- SEMA Group and Yard Software (for Mark-It and Write-It);
- Software Exoterica (for CheckMark and Xtran);
- SoftQuad, Inc., (for Author/Editor and RulesBuilder);
- WordPerfect Corporation (for Intellitag);
- Xerox Corporation (for Ventura Publisher).

Details of the software actually used to produce the current document are given in the colophon at the end of the work.

Appendix C5. Acknowledgments

Many people have given of their time, energy, expertise, and support in the creation of this document; it is unfortunately not possible to thank them all adequately. Below are listed those who have served as formal members of the TEI's Work Groups and Working Committees during its six-year history; others not so officially enfranchised also contributed much to the quality of the result.

The editors take this opportunity to acknowledge our debt to those who have patiently endured and corrected our misunderstandings of their work; we hope that they will feel the wait has not been in vain. For any errors and inconsistencies remaining, we must accept responsibility; any virtue in what is here presented, we gladly ascribe to the energies of the keen intellects listed below.

C. M. Sperberg McQueen and Lou Burnard

Appendix C5.1. TEI Working Committees (1990-1993)

(Not all members listed were able to serve throughout the development of the Guidelines.)

Committee on Text Documentation: Chair: Dominik Wujastyk (Wellcome Institute for the History of Medicine)

Members 1990–1992: J. D. Byrum (Library of Congress); Marianne Gaunt (Rutgers University); Richard Giordano (Manchester University); Barbara Ann Kipfer (Independent Consultant); Hans Jørgen Marker (Danish Data Archive, Odense); Marcia Taylor (University of Essex);

Committee on Text Representation Chair: Stig Johansson (University of Oslo)

Members 1990–1992: Roberto Cencioni (Commission of the European Communities); David R. Chesnutt (University of South Carolina); Robin C. Cover (Dallas Theological Seminary); Steven J. DeRose (Electronic Book Technology Inc); David G. Durand (Boston University); Susan M. Hockey (Oxford University Computing Service); Claus Huitfeldt (University of Bergen); Francisco Marcos-Marin (University Madrid); Elli Mylonas (Harvard University); Wilhelm Ott (University of Tübingen); Allen H. Renear (Brown University); Manfred Thaller (Max-Planck-Institut für Geschichte, Göttingen)

Committee on Text Analysis and Interpretation Chair: D. Terence Langendoen (University of Arizona)

Members 1990–1992: Robert Amsler (Bell Communications Research); Stephen Anderson (Johns Hopkins University); Branimir Boguraev (IBM T. J. Watson Research Center); Nicoletta Calzolari (University of Pisa); Robert Ingria (Bolt Beranek Newman Inc); Winfried Lenders (University of Bonn); Mitch Marcus (University of Pennsylvania); Nelleke Oostdijk (University of Nijmegen); William Poser (Stanford University); Beatrice Santorini (University of Pennsylvania); Gary Simons (Summer Institute of Linguistics); Antonio Zampolli, University of Pisa.

Committee on Metalanguage and Syntax Chair: David T. Barnard (Queen's University)

Members 1990–1994: David G. Durand (Boston University); Jean-Pierre Gaspard (Associated Consultants and Software Engineers sa/nv); Nancy M. Ide (Vassar College); Lynne A. Price (Software Exoterica / Xerox PARC); Frank Tompa (University of Waterloo); Giovanni Battista Varile (Commission of the European Communities).

In addition, the two TEI editors served *ex officio* on each committee.

Following publication of the first draft of the TEI Guidelines (P1) in November 1990, a number of specialist work groups were charged with responsibility for drafting revisions and extensions, which, together with material already presented in P1, constitute the basis of the present work.

In addition, many members of the work groups listed below met on three occasions to review the emerging proposals in detail at technical review meetings convened by the TEI Steering Committee. These meetings, held in Myrdal, Norway (November 1991), Chicago (May 1992) and Oxford (May 1993), were largely responsible for the technical content and organization of the present work.

Attendants at these meetings are starred in the list below.

- TR1 Character sets** Chair: Harry Gaylord* (University of Groningen); Syun Tutiya* (Chiba University).
- TR2 Text criticism** Chair: Peter Robinson* (Oxford University); David Chesnutt* (University of South Carolina); Robin Cover* (Dallas Theological Seminary); Robert Kraft (University of Pennsylvania); Peter Shillingsburg (Mississippi State University).
- TR3 Hypertext and hypermedia** Chair: Steven J. DeRose* (Electronic Book Technologies Inc); David Durand (Boston University); Edward A. Fox (Virginia State University); Eve Wilson (University of Kent).
- TR4 Formulæ, Tables, figures, and graphics** Chair: Paul Ellison* (University of Exeter); Anders Berglund (Independent Consultant); Dale Waldt (Thompson Professional Publishing).
- TR6 Language corpora** Chair: Douglas Biber* (University of Northern Arizona); Jeremy Clear (Birmingham University); Gunnel Engwall (University of Stockholm).
- TR9 Manuscripts and codicology** Chair: Claus Huitfeldt* (University of Bergen); Dino Buzzetti (University of Bologna); Jacqueline Hamesse (University of Louvain); Mary Keeler (Georgetown University); Christian Kloesel (Indiana University); Allen Renear* (Brown University); Donald Spaeth (Glasgow University).
- TR10 Verse** Chair: David Robey* (University of Manchester); Elaine Brennan* (Brown University); David Chisholm (University of Arizona); Willard McCarty (University of Toronto).
- TR11 Drama and performance texts** Chair: Elli Mylonas* (Harvard University); John Lavagnino* (Brandeis University); Rosanne Potter (University of Iowa).
- TR12 Literary prose** Chair Thomas N. Corns* (University of Wales); Christian Delcourt (University of Liège).
- AI1 Linguistic description** Chair: D. Terence Langendoen* (University of Arizona); Stephen R. Anderson (Johns Hopkins University); Nicoletta Calzolari (University of Pisa); Geoffrey Sampson* (University of Sussex); Gary Simons* (Summer Institute of Linguistics).
- AI2 Spoken text** Chair: Stig Johansson* (University of Oslo); Jane Edwards (University of California at Berkeley); Andrew Rosta (University College London).
- AI3 Literary studies** Chair: Paul Fortier* (University of Manitoba); Christian Delcourt (University of Liège); Ian Lancashire (University of Toronto); Rosanne Potter (University of Iowa); David Robey* (University of Manchester).
- AI4 Historical studies** Chair: Daniel Greenstein* (University of Glasgow); Peter Denley (Queen Mary Westfield College, London); Ingo Kropac (University of Graz); Hans Jørgen Marker (Danish Data Archive, Odense); Jan Oldervoll (University of Tromsø); Kevin Schurer (University of Cambridge); Donald Spaeth (Glasgow University); Manfred Thaller (Max-Planck-Institut für Geschichte, Göttingen).¹⁷⁸
- AI5 Print dictionaries** Chairs: Robert Amsler* (Bell Communications Research) and Nicoletta Calzolari (University of Pisa); Susan Armstrong-Warwick (University of Geneva); John Fought (University of Pennsylvania); Louise Guthrie (University of New Mexico); Nancy M. Ide* (Vassar College); Frank Tompa (University of Waterloo); Carol Van Ess-Dykema (US Department of Defense); Jean Veronis (University of Aix-en-Provence).
- AI6 Machine lexica** Chair: Robert Ingria* (Bolt Beranek Newman Inc); Susan Armstrong-Warwick (University of Geneva); Nicoletta Calzolari (University of Pisa).
- AI7 Terminological data** Chair: Alan Melby* (Brigham Young University) Gerhard Budin (University of Vienna); Gregory Shreve (Kent State University); Richard Strehlow (Oak Ridge National Laboratory); Sue Ellen Wright (Kent State University).

¹⁷⁸ This Workgroup was jointly sponsored by the Association for History and Computing.

Appendix C5.2. Advisory Board

Members of the TEI Advisory Board during the lifetime of the project are listed below, grouped under the name of the organization represented.

American Anthropological Association: Chad McDaniel (University of Maryland).

American Historical Association: Elizabeth A. R. Brown (Brooklyn College, CUNY).

American Philological Association: Jocelyn Penny Small (Rutgers University).

American Philosophical Association: Allen Renear (Brown University).

American Society for Information Science: Clifford A. Lynch (University of California).

Association for Computing Machinery, Special Interest Group for Information Retrieval:
1989–93: Scott Deerwester (University of Chicago); 1993– : Martha Evens (Illinois Institute of Technology).

Association for Documentary Editing: David Chesnutt (University of South Carolina).

Association for History and Computing: 1989–91: Manfred Thaller, Max-Planck-Institut für Geschichte, Göttingen; 1991– : Daniel Greenstein (Glasgow University).

Association Internationale Bible et Informatique 1989–93: Wilhelm Ott (University of Tübingen); 1993– : Winfried Bader (University of Tübingen).

Canadian Linguistic Association: Anne-Maria di Sciullo (Université du Québec à Montréal)

Dictionary Society of North America: Barbara Ann Kipfer (Independent Consultant).

AAP Electronic Publishing Special Interest Group: 1989–92: Betsy Kiser (OCLC); 1992– : Deborah Bendig and Andrea Keyhani (OCLC).

International Federation of Library Associations and Institutions: J. D. Byrum Jr. (The Library of Congress).

Linguistic Society of America: Stephen Anderson (The Johns Hopkins University)

Modern Language Association: Randall Jones (Brigham Young University) and Ian Lancashire (University of Toronto).

Appendix C5.3. Steering Committee Membership

Members of the Steering Committee of the TEI during the preparation of this work were:

Association for Computational Linguistics:

- 1987–1993: Robert A. Amsler (Bell Communications Research);
- 1987–1993: Donald E. Walker (Bell Communications Research);
- 1993–1994: Susan Armstrong-Warwick (University of Geneva);
- 1994–1999: Judith Klavans (Columbia University).

Association for Computers and the Humanities:

- 1987–1999: Nancy M. Ide (Vassar College);
- 1987–1994: C. M. Sperberg-McQueen (University of Illinois at Chicago);
- 1994–1999: David Barnard (Queen’s University).

Association for Literary and Linguistic Computing:

- 1987–1999: Susan M. Hockey (Center for Electronic Texts in the Humanities);
- 1987–1999: Antonio Zampolli (University of Pisa).