

The TEI Consortium

The Association for Computers and the Humanities (ACH);
The Association for Computational Linguistics (ACL);
The Association for Literary and Linguistic Computing (ALLC)

TEI P4 Guidelines for Electronic Text Encoding and Interchange XML-compatible edition

*edited by C M Sperberg-McQueen
and Lou Burnard*

*XML conversion by Syd Bauman, Lou Burnard,
Steven DeRose, and Sebastian Rahtz*

Oxford — Providence — Charlottesville — Bergen
March 2002

Contents

I	Introduction	1
1	About These Guidelines	3
1.1	Structure and Notational Conventions of this Document	4
1.1.1	Structure	4
1.1.2	Notational Conventions	4
1.2	Underlying Principles and Intended Use	6
1.2.1	Design Principles of the TEI Scheme	6
1.2.2	Intended Use	8
1.2.2.1	Use in Text Capture and Text Creation	8
1.2.2.2	Use for Interchange	8
1.2.2.3	Use for Local Processing	10
1.3	Historical Background	10
1.3.1	Origin and Development of the TEI	11
1.3.2	Future Developments	12
2	A Gentle Introduction to XML	13
2.1	What's special about XML?	13
2.1.1	Descriptive markup	14
2.1.2	Types of document	14
2.1.3	Data independence	14
2.2	Textual structure	15
2.3	XML structures	16
2.3.1	Elements	16
2.3.2	Content models: an example	16
2.4	Validating a document's structure	17
2.4.1	An example DTD	18
2.4.2	Generic identifier	18
2.4.3	Content model	19
2.4.4	Occurrence indicators	19
2.4.5	Connectors	19
2.4.6	Model groups	19
2.5	Complicating the issue	20
2.6	Attributes	22
2.6.1	Declaring attributes	22
2.6.2	Attribute names	22
2.6.3	Attribute values	23
2.6.4	Default value	23
2.6.5	ID and IDREF attributes	23
2.7	Entities	25
2.7.1	Entity declarations	25
2.7.2	Entity references	26
2.7.3	Character references	26
2.7.4	Unparsed entities and Notations	27
2.7.5	Parameter entities	27
2.8	Marked sections	28
2.8.1	CDATA marked section	28
2.8.2	Conditional marked section	28
2.9	Other components of an XML document	30
2.9.1	Processing instructions	30
2.9.2	Namespaces	31
2.10	Putting it all together	32
2.10.1	SGML and XML declarations	32

2.10.2	The DOCTYPE declaration	32
2.10.3	The Document Instance	33
2.10.4	Ancillary Files	34
3	Structure of the TEI Document Type Definition	35
3.1	Main and Auxiliary DTDs	35
3.2	Core, Base, and Additional Tag Sets	36
3.2.1	The Core Tag Sets	37
3.2.2	The Base Tag Sets	37
3.2.3	The Additional Tag Sets	38
3.2.4	User-Defined Tag Sets	39
3.3	Invocation of the TEI DTD	39
3.4	Combining TEI Base Tag Sets	40
3.5	Global Attributes	42
3.6	The TEI2.DTD File	45
3.6.1	Structure of the TEI2.DTD File	45
3.6.2	Embedding Local Modifications	46
3.6.3	Embedding the Core Tag Sets	48
3.6.4	Embedding the Base Tag Set	48
3.6.5	Embedding the Additional Tag Sets	49
3.7	Element Classes	50
3.7.1	Classes Which Share Attributes	50
3.7.2	Classes Used in Content Models	52
3.7.3	The TEICLAS2.ENT File	53
3.7.4	Low-Level Element Classes	54
3.7.5	High-Level Element Classes	56
3.7.6	Elements Marked for Text Type	57
3.7.7	Standard Content Models	58
3.7.8	Components in Mixed and General Bases	59
3.7.9	Miscellaneous Content-Model Classes	60
3.8	Other Parameter Entities in TEI DTDs	61
3.8.1	Inclusion and Exclusion of Elements	62
3.8.2	Parameter Entities for Element Generic Identifiers	62
3.8.3	Parameter Entities for TEI Keywords	62
3.8.4	Generation of an XML DTD	63
3.8.5	Declaration of TEI keywords	63
II	Core Tags and General Rules	65
4	Languages and Character Sets	67
4.1	A simple character encoding model	67
4.1.1	Some definitions	67
4.1.2	Characters and glyphs	68
4.1.3	Characters and their encoding	70
4.1.4	Character semantics	71
4.1.5	Characters from the Private Usage Area	72
4.2	Entry and display of characters	72
4.2.1	Character input and entity references	72
4.2.2	Transliteration schemes	74
4.3	Code shifting	75
4.4	The Writing System Declaration	76
5	The TEI Header	79
5.1	Organization of the TEI Header	79
5.1.1	The TEI Header and Its Components	79
5.1.2	Types of Content in the TEI Header	81

5.2	The File Description	82
5.2.1	The Title Statement	83
5.2.2	The Edition Statement	85
5.2.3	Type and Extent of File	87
5.2.4	Publication, Distribution, etc.	88
5.2.5	The Series Statement	89
5.2.6	The Notes Statement	91
5.2.7	The Source Description	93
5.2.8	Computer Files Derived from Other Computer Files	94
5.2.9	Computer Files Composed of Transcribed Speech	94
5.3	The Encoding Description	96
5.3.1	The Project Description	98
5.3.2	The Sampling Declaration	98
5.3.3	The Editorial Practices Declaration	99
5.3.4	The Tagging Declaration	102
5.3.5	The Reference System Declaration	104
5.3.5.1	Prose Method	105
5.3.5.2	Stepwise Method	105
5.3.5.3	Milestone Method	106
5.3.6	The Classification Declaration	108
5.3.7	The Feature System Declaration	110
5.3.8	The Metrical Declaration Element	111
5.3.9	The Variant-Encoding Method Element	112
5.4	The Profile Description	113
5.4.1	Creation	114
5.4.2	Language Usage	114
5.4.3	The Text Classification	115
5.5	The Revision Description	117
5.6	Minimal and Recommended Headers	119
5.7	Note for Library Cataloguers	122
6	Elements Available in All TEI Documents	125
6.1	Paragraphs	125
6.2	Treatment of Punctuation	127
6.3	Highlighting and Quotation	128
6.3.1	What Is Highlighting?	128
6.3.2	Emphasis, Foreign Words, and Unusual Language	129
6.3.2.1	Foreign Words or Expressions	130
6.3.2.2	Emphatic Words and Phrases	130
6.3.2.3	Other Linguistically Distinct Material	131
6.3.3	Quotation	132
6.3.4	Terms, Glosses, and Cited Words	136
6.3.5	Some Further Examples	137
6.4	Names, Numbers, Dates, Abbreviations, and Addresses	138
6.4.1	Referring Strings	138
6.4.2	Addresses	139
6.4.3	Numbers and Measures	141
6.4.4	Dates and Times	143
6.4.5	Abbreviations and Their Expansions	145
6.5	Simple Editorial Changes	146
6.5.1	Correction of Apparent Errors	147
6.5.2	Regularization and Normalization	149
6.5.3	Additions, Deletions, and Omissions	150

6.6	Simple Links and Cross References	154
6.7	Lists	156
6.8	Notes, Annotation, and Indexing	160
6.8.1	Notes and Simple Annotation	160
6.8.2	Index Entries	161
6.9	Reference Systems	163
6.9.1	Using the ID and N Attributes	163
6.9.2	Creating New Reference Systems	164
6.9.3	Milestone Tags	166
6.9.4	Declaring Reference Systems	168
6.10	Bibliographic Citations and References	170
6.10.1	Elements of Bibliographic References	170
6.10.2	Components of Bibliographic References	173
6.10.2.1	Analytic, Monographic, and Series Levels	174
6.10.2.2	Authors, Titles, and Editors	175
6.10.2.3	Imprint, Pagination, and Other Details	179
6.10.2.4	Series Information	182
6.10.2.5	Notes and Other Additional Information	182
6.10.2.6	Order of Components within References	183
6.10.3	Bibliographic Pointers	183
6.10.4	Relationship to Other Bibliographic Schemes	184
6.11	Passages of Verse or Drama	184
6.11.1	Core Tags for Verse	185
6.11.2	Core Tags for Drama	187
6.12	Overview of the Core Tag Set	190
7	Default Text Structure	191
7.1	Divisions of the Body	192
7.1.1	Un-numbered Divisions	193
7.1.2	Numbered Divisions	193
7.1.3	Numbered or Un-numbered?	195
7.1.4	Partial and Composite Divisions	197
7.2	Elements Common to All Divisions	198
7.2.1	Headings and Trailers	198
7.2.2	Openers and Closers	199
7.2.3	Arguments and Epigraphs	201
7.2.4	Content of Textual Divisions	202
7.3	Groups of Texts	203
7.4	Front Matter	208
7.5	Title Pages	210
7.6	Back Matter	213
7.7	DTD Fragment for Default Text Structure	214
III	Base Tag Sets	217
8	Base Tag Set for Prose	219
9	Base Tag Set for Verse	221
9.1	Structure of the Base Tag Set for Verse	221
9.2	Structural Divisions of Verse Texts	222
9.3	Components of the Verse Line	225
9.4	Rhyme and Metrical Analysis	227
9.4.1	Sample Metrical Analyses	228
9.4.2	Segment-Level versus Line-level Tagging	230
9.4.3	Metrical Analysis of Stanzaic Verse	230
9.5	Rhyme	231

9.6	Encoding Procedures For Other Verse Features	232
10	Base Tag Set for Drama	235
10.1	Front and Back Matter	235
10.1.1	The Set Element	236
10.1.2	Prologues and Epilogues	237
10.1.3	Records of Performances	239
10.1.4	Cast Lists	240
10.2	The Body of a Performance Text	243
10.2.1	Major Structural Divisions	243
10.2.2	Speeches and Speakers	244
10.2.3	Stage Directions	246
10.2.4	Speech Contents	248
10.2.5	Embedded Structures	250
10.2.6	Simultaneous Action	253
10.3	Other Types of Performance Text	253
10.3.1	Technical Information	255
11	Transcriptions of Speech	257
11.1	General Considerations and Overview	258
11.1.1	Divisions	259
11.2	Elements Unique to Spoken Texts	261
11.2.1	Utterances	263
11.2.2	Pause	264
11.2.3	Vocal, Kinesic, Event	265
11.2.4	Writing	266
11.2.5	Temporal Information	267
11.2.6	Shifts	267
11.2.7	Formal Definition	269
11.3	Elements Defined Elsewhere	270
11.3.1	Segmentation	270
11.3.2	Synchronization and Overlap	272
11.3.3	Regularization of Word Forms	275
11.3.4	Prosody	275
11.3.5	Speech Management	277
11.3.6	Analytic Coding	277
12	Print Dictionaries	279
12.1	Dictionary Body and Overall Structure	279
12.2	The Structure of Dictionary Entries	282
12.2.1	Hierarchical Levels	283
12.2.2	Groups and Constituents	285
12.3	Top-level Constituents of Entries	289
12.3.1	Information on Written and Spoken Forms	289
12.3.2	Grammatical Information	294
12.3.3	Sense Information	296
12.3.3.1	Definitions	296
12.3.3.2	Translation Equivalents	297
12.3.4	Etymological Information	299
12.3.5	Other Information	301
12.3.5.1	Examples	301
12.3.5.2	Usage Information and Other Labels	303
12.3.5.3	Cross References to Other Entries	305
12.3.5.4	Notes within Entries	307
12.3.6	Related Entries	309

12.4	Headword and Pronunciation References	310
12.5	Typographic and Lexical Information in Dictionary Data	312
12.5.1	Editorial View	313
12.5.2	Lexical View	315
12.5.3	Retaining Both Views	316
12.5.3.1	Using Attribute Values to Capture Alternate Views	316
12.5.3.2	Recording Original Locations of Transposed Elements	319
12.5.4	Attributes for Dictionary Elements	319
12.6	Unstructured Entries	319
13	Terminological Databases	323
13.1	The Terminological Entry	323
13.2	Tags for Terminological Data	324
13.3	Basic Structure of the Terminological Entry	329
13.3.1	Nested Term Entries	329
13.3.2	Flat Term Entries Using Rules of Adjacency	329
13.3.3	Flat Term Entries Using Group and Depend Attributes	330
13.3.4	References between Term Entries	332
13.4	Overall Structure of Terminological Documents	332
13.4.1	DTD Fragment for Nested Style	334
13.4.2	DTD Fragment for Flat Style	335
13.5	Additional Examples of Term Entries	336
13.5.1	Example Term Entry from ISO 472	336
13.5.2	The Example Treated as a Single Term Entry in Nested Form	336
13.5.3	The Example Treated as Two Separate Term Entries in Nested Form	337
13.5.4	The Example Treated as a Flat Term Entry Using Adjacency Rules	338
13.5.5	The Example Treated as a Flat Term Entry Not Using Adjacency Rules	338
IV	Additional Tag Sets	341
14	Linking, Segmentation, and Alignment	343
14.1	Pointers	344
14.1.1	Pointers and Links	345
14.1.2	Using Pointers and Links	346
14.1.3	Groups of Links	349
14.1.4	Intermediate Pointers	352
14.2	Extended Pointers	352
14.2.1	Extended Pointer Elements	353
14.2.2	Extended Pointer Syntax	353
14.2.2.1	Location Ladders	354
14.2.2.2	Location Terms	354
14.2.2.3	The ROOT Keyword	356
14.2.2.4	The HERE Keyword	356
14.2.2.5	The ID Keyword	356
14.2.2.6	The REF Keyword	356
14.2.2.7	The CHILD Keyword	356
14.2.2.8	The DESCENDANT Keyword	359
14.2.2.9	The ANCESTOR Keyword	359
14.2.2.10	The PREVIOUS Keyword	360
14.2.2.11	The NEXT Keyword	360
14.2.2.12	The PRECEDING Keyword	360
14.2.2.13	The FOLLOWING Keyword	361
14.2.2.14	The PATTERN Keyword	361
14.2.2.15	The TOKEN Keyword	362
14.2.2.16	The STR Keyword	363

14.2.2.17	The SPACE Keyword	364
14.2.2.18	The FOREIGN Keyword	364
14.2.2.19	The HYQ Keyword	364
14.2.2.20	The DITTO Keyword	365
14.2.3	Using Extended Pointers	365
14.2.4	Representation of HTML links in TEI	367
14.3	Blocks, Segments and Anchors	368
14.4	Correspondence and Alignment	373
14.4.1	Correspondence	373
14.4.2	Alignment of Parallel Texts	374
14.4.3	A Three-way Alignment	376
14.5	Synchronization	379
14.5.1	Aligning Synchronous Events	379
14.5.2	Placing Synchronous Events in Time	381
14.6	Identical Elements and Virtual Copies	383
14.7	Aggregation	384
14.8	Alternation	388
14.9	Connecting Analytic and Textual Markup	393
15	Simple Analytic Mechanisms	395
15.1	Linguistic Segment Categories	395
15.2	Global Attributes for Simple Analyses	400
15.3	Spans and Interpretations	400
15.4	Linguistic Annotation	404
16	Feature Structures	409
16.1	Introduction	409
16.2	Elementary Feature Structures: Features with Binary Values	409
16.3	Feature, Feature-Structure and Feature-Value Libraries	411
16.4	Symbolic, Numeric, Measurement, Rate and String Values	414
16.5	Structured Values	419
16.6	Singleton, Set, Bag and List Collections of Values	421
16.7	Alternative Features and Feature Values	424
16.8	Boolean, Default and Uncertain Values	428
16.9	Indirect Specification of Values Using the rel Attribute	431
16.9.1	The Not-Equals Relation	431
16.9.2	Other Inequality Relations	432
16.9.3	Subsumption and Non-subsumption Relations	432
16.9.4	Relations Holding with Sets, Bags, and Lists	435
16.9.5	Varieties of Subsumption and Non-subsumption	436
16.10	Two Illustrations	437
17	Certainty and Responsibility	445
17.1	Levels of Certainty	445
17.1.1	Using Notes to Record Uncertainty	446
17.1.2	Structured Indications of Uncertainty	446
17.2	Attribution of Responsibility	449
18	Transcription of Primary Sources	453
18.1	Altered, Corrected, and Erroneous Texts	454
18.1.1	Use of Core Tags for Transcriptional Work	454
18.1.2	Abbreviation and Expansion	457
18.1.3	Correction and Conjecture	459
18.1.4	Additions and Deletions	460
18.1.5	Substitutions	464
18.1.6	Cancellation of Deletions and Other Markings	466

18.1.7	Text Omitted from or Supplied in the Transcription	467
18.2	Non-Linguistic Phenomena in the Source	469
18.2.1	Document Hands	469
18.2.2	Hand, Responsibility, and Certainty Attributes	471
18.2.3	Damage, Illegibility, and Supplied Text	473
18.2.4	Use of the Gap, Del, Damage, Unclear and Supplied Tags in Combination	475
18.2.5	Space	476
18.2.6	Lines	477
18.3	Headers, Footers, and Similar Matter	478
18.4	Other Primary Source Features not Covered in These Guidelines	479
19	Critical Apparatus	481
19.1	The Apparatus Entry, Readings, and Witnesses	482
19.1.1	The Apparatus Entry	482
19.1.2	Readings	483
19.1.3	Indicating Subvariation in Apparatus Entries	486
19.1.4	Witness Information	488
19.1.4.1	Witness Detail Information	488
19.1.4.2	Witness Information in the Source	489
19.1.4.3	The Witness List	490
19.1.5	Fragmentary Witnesses	492
19.2	Linking the Apparatus to the Text	493
19.2.1	The Location-referenced Method	493
19.2.2	The Double End-Point Attachment Method	494
19.2.3	The Parallel Segmentation Method	496
19.3	Using Apparatus Elements in Transcriptions	497
20	Names and Dates	499
20.1	Personal Names	500
20.2	Place Names	504
20.2.1	Geo-political Place Names	506
20.2.2	Geographic Names	506
20.2.3	Relative Place Names	506
20.3	Organization names	508
20.4	Dates and Time	511
20.4.1	Absolute Dates and Times	512
20.4.2	Relative Dates and Times	513
21	Graphs, Networks, and Trees	517
21.1	Graphs and Digraphs	517
21.1.1	Transition Networks	521
21.1.2	Family Trees	523
21.1.3	Historical Interpretation	524
21.2	Trees	525
21.3	Another Tree Notation	529
22	Tables, Formulae, and Graphics	535
22.1	Tables	535
22.1.1	The TEI Table DTD	536
22.1.2	Other Table DTDs	538
22.2	Formulae and Mathematical Expressions	539
22.3	Specific Elements for Graphic Images	543
22.4	Overview of Basic Graphics Concepts	546
22.5	Graphic Image Formats	547
22.5.1	Vector Graphic Formats	547
22.5.2	Raster Graphic Formats	548

22.5.3	Photographic and Motion Video Formats	549
23	Language Corpora	551
23.1	Varieties of Composite Text	552
23.2	Contextual Information	554
23.2.1	The Text Description	555
23.2.2	The Participants Description	559
23.2.3	The Setting Description	563
23.3	Associating Contextual Information with a Text	565
23.3.1	Combining Corpus and Text Headers	565
23.3.2	Declarable Elements	566
23.3.3	Summary	569
23.4	Linguistic Annotation of Corpora	569
23.4.1	Levels of Analysis	570
23.5	Recommendations for the Encoding of Large Corpora	570
V	Auxiliary Document Types	573
24	The Independent Header	575
24.1	Definition and Principles for Encoders	575
24.2	Required and Recommended Tags	576
24.3	Header Elements and their Relationship to the MARC Record	580
24.4	MARC Fields for the File Description	580
24.5	MARC Fields for the Encoding Description	583
24.6	MARC Fields for the Profile Description	583
24.7	MARC fields for the Revision Description	583
24.8	Structure of the DTD for Independent Headers	584
25	Writing System Declaration	587
25.1	Overall Structure of Writing System Declaration	587
25.2	Identifying the Language	589
25.3	Describing the Writing System	590
25.4	Documenting the Character Set and Its Encoding	591
25.4.1	Base Components of the WSD	591
25.4.2	Exceptions in the WSD	593
25.4.3	Documenting Coded Character Sets and Entity Sets	597
25.4.4	Documenting Transliteration Schemes	597
25.5	Notes in the WSD	597
25.6	Linkage between WSD and Main Document	597
25.7	Predefined TEI WSDs	598
25.8	Details of WSD Semantics	598
25.8.1	WSD Semantics: General Principles	598
25.8.2	Semantics of WSD Base Components	599
25.8.3	Multiple Base Components	600
25.8.4	Semantics of Exceptions	600
25.8.4.1	Case 1: replacement	600
25.8.4.2	Case 2: merger	601
25.8.4.3	Case 3: expansion	602
25.8.5	Merger of Form and Character Elements	603
26	Feature System Declaration	605
26.1	Linking a TEI Text to Feature System Declarations	605
26.2	The Overall Structure of a Feature System Declaration	607
26.3	Feature Declarations	609
26.4	Feature Structure Constraints	612
26.5	A Complete Example	614
27	Tag Set Documentation	617

27.1	The TagDoc Documentation Element	619
27.1.1	The AttList Documentation Element	621
27.2	Element Classes	623
27.3	Entity Documentation	624
VI	Technical Topics	627
28	Conformance	629
28.1	Definitions of Terms	629
28.1.1	TEI-Conformant Document	629
28.1.2	TEI Local Processing Format	629
28.1.3	TEI Interchange Format	629
28.1.4	TEI Packed Interchange format	630
28.1.5	TEI Recommended Practice	630
28.1.6	TEI Abstract Model	630
28.2	Modifications to TEI SGML Declaration	631
28.3	Modifications to TEI Document Type Declarations	631
28.4	TEI Processing Model	632
28.4.1	Document Capture and Reclamation	632
28.4.2	Local Storage Format and Application Software	632
28.4.3	Enrichment and Other Processing	633
28.4.4	Data Export	633
28.4.5	Data Import	633
28.4.6	TEI Conformance in the Processing Model	633
28.5	Aspects of Conformance and Document Description	634
28.5.1	Character Sets	634
28.5.2	SGML Declaration	634
28.5.3	Document Type Declaration	634
28.5.4	Tag Usage and Feature Marking	635
28.5.5	Non-SGML, non-XML Markup	635
29	Modifying and Customizing the TEI DTD	637
29.1	Kinds of Modification	639
29.1.1	Suppressing Elements	639
29.1.2	Renaming Elements	640
29.1.3	Class Extension	641
29.1.4	New content models	641
29.2	Documenting the Modifications	642
29.3	TEI Lite: an example Customization	643
30	Rules for Interchange	647
30.1	Negotiated Interchange	647
30.2	Some Simple Examples	648
30.3	Non-Negotiated Interchange	649
30.4	Notes for Implementors	650
31	Multiple Hierarchies	653
31.1	Concurrent Markup of Multiple Hierarchies	654
31.2	Boundary Marking with Milestone Elements	654
31.3	Fragmentation of Elements	655
31.4	Reconstitution of Virtual Elements	655
31.5	Multiple Encodings of the Same Information	656
31.6	Concurrent Markup for Pages and Lines	657
32	Algorithm for Recognizing Canonical References	661
VII	Alphabetical Reference Lists of Classes, Entities, and Elements	665
33	Element Classes	667
34	Entities	703

35	Elements	719
VIII	Reference Material	1005
36	Obtaining the TEI DTD	1007
37	Obtaining TEI WSDs	1011
38	Sample Tag Set Documentation	1013
38.1	Tag Documentation for the TEI p Element	1013
38.2	Tag Documentation for the TEI head Element	1013
38.3	Tag Documentation for the TEI div Element	1014
38.4	Class Documentation for the TEI Divn Class	1015
39	Formal Grammar for the TEI-Interchange-Format Subset of SGML	1017
39.1	Notation	1017
39.2	Grammar for SGML Document (Overview)	1017
39.3	Grammar for SGML Declaration	1017
39.4	Grammar for DTD	1020
39.5	Grammar for Document Instance	1022
39.6	Common Syntactic Constructs	1024
39.7	Lexical Scanner	1025
39.8	Differences from ISO 8879	1027
Appendix A	Bibliography	1029
Appendix B	Index	1035
Appendix C	Prefatory Notes	1059
Appendix C1.	Introductory Note (November 2001)	1059
Appendix C2.	Introductory Note (June 2001)	1059
Appendix C3.	Introductory Note (May 1999)	1060
Appendix C3.1.	Typographic corrections made	1061
Appendix C3.2.	Specific changes in the DTD	1061
Appendix C3.3.	Outstanding errors	1062
Appendix C4.	Preface (April 1994)	1062
Appendix C5.	Acknowledgments	1063
Appendix C5.1.	TEI Working Committees (1990-1993)	1064
Appendix C5.2.	Advisory Board	1065
Appendix C5.3.	Steering Committee Membership	1066
Appendix D	Colophon	1067

Revision History: March 2002

© Text Encoding Initiative Consortium

**In memoriam
Donald E. Walker
22 November 1928 – 26 November 1993**

Introductory Note (March 2002)

The primary goal of this revision has been to make available a new and corrected version of the TEI Guidelines which:

- is expressed in XML and conforms to a TEI-conformant XML DTD;
- generates a set of DTD fragments that can be combined together to form either SGML or XML document type definitions;
- corrects blatant errors, typographical mishaps, and other egregious editorial oversights;
- 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.

In most respects, the TEI Guidelines have stood the test of time remarkably well. The present edition makes no substantial attempt to rewrite those few parts of them which have now been rendered obsolete by changes since their first publication, though an indication is given in the text of where such rewriting is now considered necessary. Neither does the present version attempt to address any of the many possible new areas of digital activity in which the TEI approach to standardization may have something to offer. Both these tasks require the existence of an informed and active TEI Council to direct and validate such extension and maintenance work, in response to the changing needs and priorities of the TEI user community.

Two exceptions to the above principles may be cited: firstly, the chapter which originally provided a ‘Gentle Introduction’ to SGML has been completely rewritten to provide a similarly gentle introduction to XML; secondly the chapter on character sets has been completely revised in light of the close connexion between Unicode and XML. The editors gratefully acknowledge the assistance of the *ad hoc* workgroup chaired by Christian Wittern, which undertook to provide expert advice and correction at very short notice, in the latter task.

The preparation of this new version relied extensively on preliminary work carried out by the former North American editor of the TEI Guidelines, C.M. Sperberg-McQueen. In a TEI working paper written in 1999¹ he sketched out a precise blueprint for the conversion of the TEI from SGML to XML, which we have implemented, with only slight modification.

The Editors would also like to express thanks to the team of volunteers from the TEI community who helped us with the task of proof reading the first draft during the summer of 2001; and to Sebastian Rahtz of Oxford University Computing Services, without whose skill and enthusiasm this new edition would not have been possible.

A substantial proportion of the work of preparing this new edition was funded with the assistance of a grant from the US National Endowment for the Humanities, whose continued support of the TEI has also been crucial to the effort of setting up the TEI Consortium.

Finally, we would like to thank all our colleagues on the interim management board of the TEI Consortium, in particular its Chairman John Unsworth, for their continued support of the TEI’s work, and their willingness to devote effort to the difficult task of overseeing its transition to a new organizational infrastructure.

Summary details of the changes made in the present and previous editions are given in their Prefatory Notes, all of which are now reproduced in an Appendix to the present edition: see Appendix C *Prefatory Notes*.

Lou Burnard and Syd Bauman (TEI Editors) Oxford and Providence, March 2002.

¹ *TEI ED W69*, available from the TEI website at <http://www.tei-c.org/Vault/ED/edw69.htm>.

