

1 About These Guidelines

These Guidelines have been developed by the Text Encoding Initiative (TEI); see 1.3 *Historical Background*. They are addressed to anyone who works with any text in electronic form.

They provide means of representing those features of a text which need to be identified explicitly in order to facilitate processing of the text by computer programs. In particular, they specify a set of markers (or *tags*) which may be inserted in the electronic representation of the text, in order to mark the text structure and other textual features of interest. Without such explicit markers, many important features remain difficult to locate by mechanical means such as computer programs, and thus difficult to process effectively. The process of inserting such explicit markers for implicit textual features is often called ‘markup’, ‘encoding’, or ‘tagging’, and the term *encoding scheme* or *markup language* denotes the rules which govern the use of markup in a set of encodings.

The Guidelines formulated in this document are intended for use in interchange between individuals and research groups using different programs and computer systems over a broad range of applications. Since they contain an inventory of the features most often found useful for text processing, the Guidelines also provide help to those creating texts in electronic form. They can also be used for the local storage of text which is to be processed with multiple software packages requiring different input formats.

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. Though principally directed to the needs of the scholarly research community, the Guidelines are not restricted to esoteric academic applications. They should also be useful for librarians who maintain and document electronic materials, as well as for publishers and others creating or distributing electronic texts. Although they focus on problems of representing in electronic form texts which already exist in traditional media, these Guidelines should also be useful for the creation of electronic texts. They are adequate to, but not limited by, existing practices.

The rules and recommendations made in these Guidelines are designed to enable the creation of documents that conform to either the Standard Generalized Markup Language (SGML, defined by ISO 8879) or the Extensible Markup Language (XML, defined by the World Wide Web Consortium’s XML Recommendation). XML is a subset of SGML, and the modifications to these Guidelines to support XML are designed to maximize compatibility with both specifications. For more information on markup languages see chapter 2 *A Gentle Introduction to XML*.

These Guidelines also make reference to character encoding standards such as ISO 646, ISO 10646 and Unicode. ISO 646 defines a standard seven-bit character set in terms of which recommendations on character-level interchange are formulated; this is the most portable character set for broad interchange, but requires indirect encoding of many characters. Unicode provides a much larger character set appropriate for international use, and all XML implementations must support it; however, it is not as of this writing so widely portable as ISO 646.

This document provides the authoritative statement of the requirements and usage of the TEI encoding scheme. Although it includes numerous small examples, it must be stressed that it is intended as a reference manual and that readers unfamiliar with SGML, XML, or text markup in general will find it difficult to learn the encoding scheme by reading this document alone.

This document will be complemented by a series of tutorials in text encoding (document TEI U1 et seq.) and a case book of extended examples with discussion of the rationale for various markup choices (TEI T1).² Readers seeking an introduction to text markup and the use of the TEI encoding scheme in a specific area should consult an appropriate tutorial; those already familiar with the scheme and interested in seeing examples of its application should consult the case book.

² TEI documents bear identifying numbers which indicate the provenance of the document (here simply “TEI”, in other cases the TEI work group number, e.g. “TEI A15”), the type of document (here “U” and “T”, meaning *users’ guide* or *users’ manual* and *sample text(s)*), and a sequential number. The TEI document number of the document in hand is TEI P4 (for *TEI public proposal number 4*).

The remainder of this chapter comprises three sections. The first gives an overview of the structure and notational conventions used throughout the document. The second enumerates the design principles underlying the TEI scheme and the application environments in which it may be found useful. Finally, the third section gives a brief account of the origins and development of the Text Encoding Initiative itself.

1.1 Structure and Notational Conventions of this Document

1.1.1 Structure

Part I provides some relevant background information about the Guidelines themselves (in this chapter); a brief technical review of markup languages (chapter 2 *A Gentle Introduction to XML*); and a description of how the TEI document type definition (DTD) is organized (chapter 3 *Structure of the TEI Document Type Definition*).

Part II provides a systematic treatment of issues common to all text types: character representation (chapter 4 *Languages and Character Sets*); in-file documentation of the text (chapter 5 *The TEI Header*); tags for text features found in all sorts of text: lists, notes, emphasis, quotations, cross-references, technical terms, names, dates, numbers, etc. (chapter 6 *Elements Available in All TEI Documents*); and a definition for the default structure of all TEI documents (chapter 7 *Default Text Structure*).

Part III documents various *base tag sets*: these include specialized tags for prose, for verse, for drama and other performance materials, for spoken materials, as well as for letters and memoranda, printed dictionaries, and terminological data. Additional sections discuss user-defined and mixed base tag sets. An instance of the TEI DTD must use one and only one base tag set, unless one of the ‘mixed’ bases is used.

Part IV documents various *additional tag sets*, which may be included or excluded, as appropriate. Topics covered include a variety of approaches to the analysis and interpretation of texts, and include representations for hypertextual links and other non-hierarchical structures, as well as specialized tags for the encoding of critical editions and language corpora.

Part V defines certain specialized *auxiliary document types*, used to encode information about the way that texts have been encoded, specifically: the TEI header regarded as a distinct document; the TEI Writing System Declaration; the Feature System declaration; and the Tag Set Documentation.

Part VI contains a number of technical discussions of a more specialist interest. Topics covered include the notion of formal *conformance* to the TEI Guidelines; the controlled user-modification of the TEI DTD; practical aspects of the use of TEI markup both in local processing and in interchange; and the relationship of TEI markup to other markup standards.

Part VII consists of an alphabetical reference list of all elements and element classes defined in the TEI encoding scheme.

Part VIII provides further reference material: specifically, a description of how to obtain current versions of the full TEI DTDs and the set of standard Writing System Declarations, a sample Feature System Declaration for basic grammatical annotation, sample tag documentation, and a formal grammar for the subset of SGML used in the TEI interchange format. No formal subset has been defined for XML, since XML itself is a subset appropriate to these Guidelines.

In the back matter, a bibliography lists works cited in the text of the Guidelines. A mechanically generated index is also provided, which can serve, it is hoped, as a finding aid for the use of the Guidelines.

1.1.2 Notational Conventions

This section describes the typographic and stylistic conventions used throughout this document. The use of many terms and concepts which have not yet been defined is unavoidable in this section. All such terms and concepts will be explained in later chapters of Part I.

When SGML or XML elements are mentioned in the text, they take the form <name>, where “name” is the *generic identifier* of the element. Sample tags mentioned in the text are displayed in the form <name att='value' att2='value two'>. References to attributes take the form attname, where “attname”

is the name of the attribute. Where the elements and attributes thus mentioned are part of the TEI encoding scheme, they are included in the index.

These Guidelines distinguish encoding practices and elements as required, recommended, or optional. The phrases “must”, “is required to”, etc., mark practices and tags which are required for TEI conformance. The phrases “should”, “it is recommended that”, “it is preferable to ...”, etc., are used in describing practices which are recommended but not required for TEI conformance. Modal verbs like “may”, “might”, etc., mark practices which are strictly optional. Qualifying phrases like “if desired”, “where appropriate”, or “under some circumstances” are used when some tag or practice described may be desirable or acceptable under some circumstances and not under others.

In the reference section in Part VII, elements and their attributes are all classed as one of:

- required** unconditionally required in a TEI-conformant document
- mandatory when applicable** required under the appropriate conditions; may be omitted if not applicable
- recommended** recommended unless there are good reasons, in the given circumstances, against it
- recommended when applicable** recommended under some circumstances (which should be clear from context)
- optional** strictly optional

This reference section includes cross-references to the chapter or section of the main text within which each element is discussed. Most sections of the main text in which elements are defined begin with a descriptive list of the elements concerned in the following format:

- <tag>** short description of the element marked by <tag>. Where appropriate this is followed by a list of significant non-global attributes for the element as follows:
 - attribute** description of the attribute’s meaning or usage, optionally followed by a list of suggested or legal values:
 - value1** meaning of value1
 - value2** meaning of value2

Not all attributes are always included in these lists; those which are shared with other elements in a class are usually listed separately, and those of relatively specialized interest are usually listed only in the reference section. The values of the attribute are introduced with one of the following formulaic phrases:

- ‘Legal values include:’** The attribute cannot take values other than those given. Other values will cause parsing errors. (This is used relatively rarely in these Guidelines.)
- ‘Suggested values include:’** The values listed constitute a set which should suffice for most purposes, and they should be used where appropriate. Developers of TEI-aware software should ensure that their software can process these values appropriately. In some cases, however, it is conceivable that other values might be necessary, so the declaration for the attribute does not restrict legal values to those given. TEI-aware software should have reasonable fallback processing for values not in the list.
- ‘Sample values include:’** The attribute can take any value; those listed are provided simply as examples of the kind of value possible.

Each list of elements is followed by some discussion of its semantics and usage, followed by one or more examples, taken wherever possible from real texts, and presented in the following format:

```
<p>This paragraph contains an <hi rend="it">italicized phrase</hi></p>
```

All the examples are (or should be) legal SGML or XML, but because they are fragmentary they may not be parseable without additional context. They also frequently make liberal use of white space to exhibit the logical structure of the encoding more clearly. Although this does not affect the validity of the examples, some users will prefer not to follow it in practice, since not all processors will ignore the extra white space. Except where otherwise noted, examples do not use minimization not permitted by XML, though SGML users may wish to exercise SGML’s options to:

- use empty end-tags (of the form `</>`) to close the most recently opened element
- omit end-tags where they may legally be omitted (the TEI DTDs do not permit omission of any start-tags)

Attribute values are given indifferently in single quotes or double quotes. Unquoted attribute values are not permitted in XML, and so are not used except where otherwise noted, for example to emphasize a comparison between SGML and XML.

After the examples and usage notes, each section typically concludes with a DTD fragment containing the formal declarations for the elements described. Each DTD fragment is given a heading, and may contain element and attribute list declarations, entity declarations, parameter entity references, comments, and references to DTD fragments in other sections. The DTD fragments of a single chapter almost invariably belong to the same DTD file, the structure of which is typically described (with references to the included fragments) in one of the first or last sections of the chapter.

The DTD fragments are identical to the DTDs distributed with these Guidelines, with the following exceptions:

- In the text, the DTD fragments appear in an order dictated by organization of this document; the actual DTD files may re-order the material slightly. This is indicated in the text by references from one DTD fragment to another.
- The DTD fragments in the text show the generic identifiers of all elements using the standard English names assigned in this document; the actual DTD files use parameter entities for all generic identifiers, so that elements can be conveniently renamed, as described in chapter 29 *Modifying and Customizing the TEI DTD*.
- The actual DTD files include conditional marked sections surrounding the element and attribute list declaration for each element, to ensure that elements can conveniently be suppressed or redefined, as described in chapter 29 *Modifying and Customizing the TEI DTD*. The fragments in the text suppress the marked-section-open and marked-section-close markup.

Note that, in both text and DTD, the omissibility indicators which must appear within an SGML declaration (but which are illegal in XML) are always given in parameterized form, as in the following examples. This is to enable a single source to support both XML and SGML versions of the DTDs, as further discussed in section 3.8.4 *Generation of an XML DTD*.

What appears in the text, therefore, as:

```
<!ELEMENT blort %om.R0; (farble+)>
```

will appear thus in the actual DTD file:

```
<![ %blort; [  
<!ELEMENT %n.blort; %om.R0; ((%n.farble;)+)>  
]]>
```

For further discussion, see chapter 3 *Structure of the TEI Document Type Definition*, or chapter 29 *Modifying and Customizing the TEI DTD*.

1.2 Underlying Principles and Intended Use

1.2.1 Design Principles of the TEI Scheme

The planning conference held at Vassar College in November, 1987 (see section 1.3 *Historical Background*) agreed on a number of principles concerning the basic design goals of the Text Encoding Initiative. These principles are expounded in various documents of the TEI (notably TEI ED P1 and TEI ED P2) and the interested reader is directed to those documents for further discussion.

Because of its roots in the humanistic research community, the TEI scheme is driven by its original goal of serving the needs of research, and is therefore committed to providing a maximum of comprehensibility, flexibility, and extensibility. More specific design goals of the TEI have been that the Guidelines should:

- provide a standard format for data interchange

- provide guidance for encoding of texts in this format
- support the encoding of all kinds of features of all kinds of texts studied by researchers
- be application independent

This has led to a number of important design decisions, such as:

- the choice of SGML, XML, ISO 646, and Unicode
- the provision of a large predefined tag set
- a distinction between required, recommended, and optional encoding practices
- encodings for different views of text
- alternative encodings for the same text features
- mechanisms for user-defined extensions to the scheme

These goals and principles are expounded in more detail below.

The goals of creating a common interchange format which is application independent require the definition of a specific markup syntax as well as the definition of a large predefined tag set. The syntax of the recommendations made in this document conforms to the international standard ISO 8879, which defines the Standard Generalized Markup Language, and to the World Wide Web Consortium's XML Recommendation, which defines the Extensible Markup Language. Full document type declarations are provided for the scheme described in these Guidelines; they are constructed so that they can be easily converted to either language. Reference is also made to ISO 646, which defines a standard seven-bit character set; and to Unicode, which defines a larger character set supporting most modern languages.

The goal of providing guidance for text encoding requires that recommendations be made as to what textual features should be recorded in various situations. This mandate is fulfilled by the explicit specification, in the reference section for each tag, that the tag is *required*, *mandatory when applicable* but otherwise omissible, *recommended* generally, *recommended when applicable* but not always applicable, or *optional*.

However, the TEI Guidelines make (with relatively rare exceptions) no suggestions or restrictions as to the relative importance of textual features. The philosophy of the Guidelines is “if you want to encode this feature, do it this way” — but very few features are mandatory.

The Guidelines have been written largely with a focus on text capture (i.e. the representation in electronic form of an already existing copy text in another medium) rather than text creation (where no such copy text exists). Hence the frequent use of terms like ‘transcription’, ‘original’, ‘copy text’, etc. However, the Guidelines should be equally applicable to text creation, and the two terms *text creation* and *text capture* are often used interchangeably.

Concerning text capture the TEI Guidelines do not specify a particular approach to the problem of fidelity to the source text and recoverability of the original; such a choice is the responsibility of the text encoder. The current version of these Guidelines, however, provides a more fully elaborated set of tags for markup of rhetorical, linguistic, and simple typographic characteristics of the text than for detailed markup of page layout or for fine distinctions among type fonts or manuscript hands.

In these Guidelines, no hard and fast distinction is drawn between ‘objective’ and ‘subjective’ information or between ‘representation’ and ‘interpretation’. These distinctions, though widely made and often useful in narrow, well-defined contexts, are perhaps best interpreted as distinctions between issues on which there is a scholarly consensus and issues where no such consensus exists. Such consensus has been, and no doubt will be, subject to change. The TEI Guidelines do not make suggestions or restrictions as to which of these features should be encoded. The use of the terms *descriptive* and *interpretive* about different types of encoding in the Guidelines is not intended to support any particular view on these theoretical issues, but reflects a purely practical division of responsibility between the two committees called Committee on Text Representation and Committee on Text Interpretation and Analysis.

In general, the accuracy and the reliability of the encoding and the appropriateness of the interpretation is for the individual user of the text to determine. The Guidelines provide a means of documenting the

encoding in such a way that a user of the text can know the reasoning behind that encoding, and the general interpretive decisions on which it is based. It is strongly recommended that the TEI header be used to give an account of these aspects of the encoding. The TEI header is described in chapter 5 *The TEI Header*.

In many situations more than one view of a text is needed. No absolute recommendation to embody one specific view of text can apply to all texts and all approaches to them. The syntaxes of SGML and XML ensure that some encodings can be ignored for some purposes. To enable encoding multiple views, these Guidelines not only treat a variety of text features, but sometimes provide several alternative encodings for what appear to be identical textual phenomena. These Guidelines therefore offer the possibility of encoding many different views of the text, simultaneously if necessary.

However, the Guidelines are built on the assumption that there is a common core of textual features shared by virtually all texts and virtually all serious work on texts. This core set of tags is defined in Chapter 6 *Elements Available in All TEI Documents*. Beyond this core, many different elements can be encoded.

In brief, the TEI Guidelines define a general-purpose encoding scheme which makes it possible to encode different views of text, possibly intended for different applications, serving the majority of scholarly purposes of text studies in the humanities. However, no predefined encoding scheme can serve all research purposes. Therefore, the TEI also provides means of modifying and extending the encoding scheme defined by the Guidelines (see chapter 29 *Modifying and Customizing the TEI DTD*).

1.2.2 Intended Use

We envisage three primary functions for these Guidelines:

- guidance for individual or local practice in text creation and data capture;
- support of data interchange;
- support of application-independent local processing.

These three functions are so thoroughly interwoven in practice that it is hardly possible to address any one without addressing the others. However, the distinction provides a useful framework for discussing the possible role of the Guidelines in work with electronic texts.

1.2.2.1 Use in Text Capture and Text Creation

The description of textual features found in the chapters which follow should provide a useful checklist from which scholars planning to create electronic texts should select the subset of features suitable for their project.

Problems specific to text creation or text ‘capture’ have not been considered explicitly in this document. For purposes of the TEI interchange format and for use of markup languages, it does not matter how a text is created or captured: it can be typed by hand, scanned from a printed book or typescript, read from a typesetter’s tape, or acquired from another researcher who may have used another markup scheme (or no explicit markup at all).

We include here only some general points which are often raised about markup and the process of data capture.

XML, and even SGML, can appear distressingly verbose, particularly when (as in these Guidelines) the names of tags and attributes are chosen for clarity and not for brevity. Editor macros and keyboard shorthands can allow a typist to enter frequently used tags with single keystrokes. Special-purpose software may be purchased which scans word-processor or scanner data and inserts tags. Markup-aware software can help with maintaining the hierarchical structure of the document, and display the document with visual formatting rather than raw tags.

The techniques described in chapter 29 *Modifying and Customizing the TEI DTD* may be used to give shorter names to the tags being used most often. It should also be noted that the examples in this text are chosen to exhibit the markup compactly, and thus have denser markup than will be typical in many texts.

The SGML standard provides ways of abbreviating, omitting, or otherwise *minimizing* the amount of markup which need be explicitly provided in a text. They are all forbidden in the TEI interchange format

because their use complicates processing; this does not however preclude their use in local processing, where this is felt appropriate or desirable. The XML Working Group followed this guideline as well, and XML prohibits essentially the same minimization practices proposed by these Guidelines.

1.2.2.2 Use for Interchange

When the TEI Guidelines are used for interchange, it is expected that researchers using other encoding schemes in their work will translate outgoing data from such schemes into the scheme described by these Guidelines, and similarly translate incoming data from the scheme described here into those used internally. For such translations to be carried out without loss of information, the scheme proposed here must be as expressive (in a formal sense) as any encoding scheme now known to be in wide use for textual research. To ensure that this is the case, a set of extension techniques is provided (see chapter 29 *Modifying and Customizing the TEI DTD*) which makes possible the addition of extra tags, the renaming of existing tags, and certain kinds of redefinition. Although the intention is to minimize the need for recourse to such extensions, they may be used to accommodate the encoding of new or unanticipated textual features. To translate between any pair of encoding schemes implies:

1. identifying the sets of textual features distinguished by the two schemes;
2. determining where the two sets of features correspond;
3. creating a suitable set of mappings.

For example, to translate from encoding scheme X into the TEI scheme:

1. Make a list of all the textual features distinguished in X.
2. Identify the corresponding feature in the TEI scheme. There are three possibilities for each feature:
 1. the feature exists in both X and the TEI scheme;
 2. X has a feature which is absent from the TEI scheme;
 3. X has a feature which corresponds with more than one feature in the TEI scheme.

The first case is unproblematic. The second requires an extension to the TEI scheme, as described in chapter 29 *Modifying and Customizing the TEI DTD*. The third requires that a consistent choice be made. The algorithm used to make that choice should be documented in the TEI header.

3. Using the table of equivalences so generated, a simple translation can be carried out between X and the TEI.

The ease with which this translation can be carried out will of course depend on the clarity and explicitness with which scheme X represents the features it encodes.

Translating from the TEI into scheme X follows the same pattern, except that if a TEI feature has no equivalent in X, and X cannot be extended, information must be lost in translation.

Similar procedures may be followed where the TEI scheme is to be used as an interlanguage for interchange among several different sites or applications, although the degree of TEI-conformance may vary.

In the simplest case, where two sites or individuals exchanging texts know each other and know or can inquire what equipment the other is using, these Guidelines serve primarily as documentation for a file format, which can be referred to without actually being transmitted together with the file. In the general case, where sender and recipient cannot communicate such information, a stricter degree of *TEI conformance* will be required for loss-free interchange.

The rules defining such strict conformance to the Guidelines are given in some detail in chapter 28 *Conformance*. The *interchange format* defined there requires that an electronic text:

1. adhere to the SGML declaration defined in these Guidelines (when using SGML), or to the XML syntax rules (which imply a particular SGML declaration). These constructs are further discussed in chapter 2 *A Gentle Introduction to XML*.

1 About These Guidelines

2. conform to the document type declarations defined in these Guidelines, unless modified or extended as described in chapter 29 *Modifying and Customizing the TEI DTD*. These constructs are further discussed in chapter 2 *A Gentle Introduction to XML*.³
3. provide external documentation as described in chapter 27 *Tag Set Documentation* for all elements not defined in these Guidelines, specifying a formal name (generic identifier) and a corresponding full natural-language name, describing its meaning and usage, specifying its legal content and also any attributes it may use.
4. adhere to the requirements of the TEI header in providing bibliographic identification of the text and description of the encoding practices used (as described in chapter 5 *The TEI Header*).

Note that the interchange format makes no formal restriction on the character set to be used in interchange, as this will depend on the medium of interchange and the local character sets in use by sender and receiver. For further information, refer to chapter 30 *Rules for Interchange*.

1.2.2.3 Use for Local Processing

Machine-readable text can be manipulated in many ways; some users:

- edit texts (e.g. word processors, syntax-directed editors)
- edit, display, and link texts in hypertext systems
- format and print texts using desktop publishing systems, or batch-oriented formatting programs
- load texts into free-text retrieval databases or conventional databases
- unload texts from databases as search results or for export to other software
- search texts for words or phrases
- perform content analysis on texts
- collate texts for critical editions
- scan texts for automatic indexing or similar purposes
- parse texts linguistically
- analyze texts stylistically
- scan verse texts metrically
- link text and images

These applications cover a wide range of likely uses but are by no means exhaustive. The aim has been to make the TEI Guidelines useful for encoding the same texts for different purposes. We have avoided anything which would restrict the use of the text for other applications. We have also tried not to omit anything essential to any single application.

1.3 Historical Background

The Text Encoding Initiative grew out of a planning conference sponsored by the Association for Computers and the Humanities (ACH) and funded by the U.S. National Endowment for the Humanities (NEH), which was held at Vassar College in November 1987. At this conference some thirty representatives of text archives, scholarly societies, and research projects met to discuss the feasibility of a standard encoding scheme and to make recommendations for its scope, structure, content, and drafting. During the conference, the Association for Computational Linguistics and the Association for Literary and Linguistic Computing agreed to join ACH as sponsors of a project to develop the Guidelines. The outcome of the conference was this set of principles, which determined the further course of the project.

1. The guidelines are intended to provide a standard format for data interchange in humanities research.
2. The guidelines are also intended to suggest principles for the encoding of texts in the same format.

³ These guidelines do not provide any other schema (XML Schema, RELAX NG, etc.) corresponding to the DTDs, although such may be provided at a later time.

3. The guidelines should
 1. define a recommended syntax for the format,
 2. define a metalanguage for the description of text-encoding schemes,
 3. describe the new format and representative existing schemes both in that metalanguage and in prose.
4. The guidelines should propose sets of coding conventions suited for various applications.
5. The guidelines should include a minimal set of conventions for encoding new texts in the format.
6. The guidelines are to be drafted by committees on
 1. text documentation
 2. text representation
 3. text interpretation and analysis
 4. metalanguage definition and description of existing and proposed schemes,
 coordinated by a steering committee of representatives of the principal sponsoring organizations.
7. Compatibility with existing standards will be maintained as far as possible.
8. A number of large text archives have agreed in principle to support the guidelines in their function as an interchange format, and have (since the publication of the prior edition), actually done so. We continue to encourage funding agencies to support development of tools to facilitate this interchange.
9. Conversion of existing machine-readable texts to the new format involves the translation of their conventions into the syntax of the new format. No requirements will be made for the addition of information not already coded in the texts.

In the course of the work, some of these goals assumed greater, some lesser importance; some proved easier, some harder to achieve. The document in hand does define a standard form for the interchange of textual material, and adumbrate principles for the creation of new electronic texts. The only metalanguage used, however, is that common to XML and SGML, and no formal definitions are given for other encoding schemes. These Guidelines do define a minimal set of conventions for text encoding (i.e. those elements classed as recommended or required), though few researchers will be satisfied to encode *only* what is required or recommended here, since the set of required and recommended elements is rather small. This document does not, however, define — at least not explicitly — “sets of coding conventions suited for various applications”, since consensus on suitable conventions for different applications proved elusive; this remains a goal for future work.

1.3.1 Origin and Development of the TEI

The Text Encoding Initiative proper began in June 1988 with funding from the NEH, soon followed by further funding from the Commission of the European Communities, the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada. Four working committees, composed of distinguished scholars and researchers from both Europe and North America, were named to deal with problems of text documentation (resulting largely in chapter 5 *The TEI Header*), text representation, text analysis and interpretation (together responsible for most of what has become parts II, III, and IV), and metalanguage and syntax issues (largely responsible for part VI).

A first draft version (1.0) of the Guidelines was distributed in July 1990 under the title *Guidelines for the Encoding and Interchange of Machine-Readable Texts*, with the TEI document number TEI P1. With minor changes and corrections, this version was reprinted as version 1.1 in November 1990.

Extensive public comment and further work on areas not covered in version 1 resulted in the drafting of a revised version, TEI P2, distribution of which began in April 1992. This version includes substantial amounts of new material, resulting from work carried out by several specialist working groups, set up in 1990 and 1991 to propose extensions and revisions to the text of P1. The overall organization, both of

the draft itself and of the scheme it describes, was entirely revised and reorganized in response to public comment on the first draft.

In June, 1993, the Advisory Board of the Text Encoding Initiative met to review the current state of the Guidelines, and recommended the formal publication of the work done to that time. That version of the TEI Guidelines, TEI P3, represents a further revision of all chapters published under the document number TEI P2, and the addition of further chapters. Although subject to revision and amendment on the basis of practical experience and public discussion, that version of the Guidelines was published in May of 1994 without the label 'draft', and marks the conclusion of the initial development work.

In February of 1998 the World Wide Web Consortium issued a final Recommendation for the Extensible Markup Language, XML. XML was developed as a far simpler subset of SGML, for many of the same reasons as the TEI interchange subset, and taking a very similar approach. Several TEI participants contributed heavily to the development of XML, most notably XML's senior co-editor C. M. Sperberg-McQueen, who until recently served as the North American co-editor for these Guidelines.

Following the ratification of XML and its rapid adoption, many projects found need for an updated version of these Guidelines which supported XML unambiguously. For example, because SGML element names are normally case-insensitive while XML ones are not, a decision had to be made on the normative case for TEI element names in XML. The TEI editors, with abundant assistance from others who have developed and used TEI, developed an update plan, and made tentative decisions on relevant syntactic issues. With the formation of the TEI Consortium in 2001, and with generous funding from the National Endowment for the Humanities, a formal update was undertaken. The goals of this update were to revise both the text and the DTDs of the scheme in a way compatible with the use of either SGML or XML. The present edition is the first public draft of that update; the present editors hope that it maintains the quality and usefulness of P3, and solicit comments, suggestions, and other input wherever it does not.

1.3.2 Future Developments

Work on areas still not satisfactorily covered in this manual will continue, and resulting recommendations will be issued as supplements to the published Guidelines. Work is expected to continue in at least the following areas:

- linguistic description and grammatical annotation
- historical analysis and interpretation
- base tag sets for further document types
- manuscript analysis and physical description of text

The encoding recommended by this document may be used without fear that future versions of the TEI scheme will be inconsistent with it in fundamental ways. The TEI will be sensitive, in revising these Guidelines, to the possible problems which revision might pose for those who are already using this version of the Guidelines. Wherever consistent with the long-term goals of the project, consistency with this version will be preserved in future revisions.