## TABLE OF CONTENTS

6. TEI: History........................................................................................................................1
   6.1 Poughkeepsie Principles...............................................................................................1
   6.2 TEI P1 and TEI P2......................................................................................................2
   6.3 TEI P3..........................................................................................................................3
   6.4 TEI P4..........................................................................................................................3
   6.5 TEI P5..........................................................................................................................4
6. TEI: History

After the concise overview of the most recent version of TEI (P5) in the preceding section, this section explains the historical development of the TEI Guidelines.

6.1 Poughkeepsie Principles

Shortly after the publication of the SGML specification as an ISO Standard, a diverse group of 32 humanities computing scholars gathered at Vassar College in Poughkeepsie, New York in a two-day meeting (11 & 12 November 1987) called for by the Association for Computers and the Humanities (ACH [http://www.ach.org]), funded by the National Endowment for the Humanities (NEH), and convened by Nancy Ide and Michael Sperber McQueen. The main topic of the meeting was the question how and whether an encoding standard for machine-readable texts intended for scholarly research should be developed. Amongst the delegates were representatives from the main European text archives and from important North American academic and commercial research centres. Contrary to the disappointing outcomes of other such meetings in San Diego in 1977 or in Pisa in 1980, this meeting did reach its goal with the formulation and the agreement on the following set of methodological principles – the so called Poughkeepsie Principles – for the preparation of text encoding guidelines for literary, linguistic, and historical research:\1\2\3:

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. 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

\begin{footnotes}
\footnote{1 Burnard, L. Report of Workshop on Text Encoding Guidelines. Literary and Linguistic Computing 1988, 3 (2): 131-133 (132-133).}
\end{footnotes}
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. We 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.

For the implementation of these principles the ACH was joined by the Association for Literary and Linguistic Computing (ALLC http://www.allc.org) and the Association for Computational Linguistics (ACL http://www.aclweb.org/). Together they established the Text Encoding Initiative (TEI) whose mission it was to develop the Poughkeepsie Principles into workable text-encoding guidelines. The Text Encoding Initiative very soon came to adopt SGML, published a year before as ISO standard, as its framework. Initial funding was provided by the US National Endowment for the Humanities, Directorate General XIII of the Commission of the European Communities, the Canadian Social Science and Humanities Research Council, and the Andrew W. Mellon Foundation.

6.2 TEI P1 and TEI P2

From the Poughkeepsie Principles the TEI concluded that the TEI 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;
- Allow the rigorous definition and efficient processing of texts;
- Provide for user-defined extensions;
- Be application independent;
- Be simple, clear, and concrete;
- Be simple for researchers to use without specialized software.

A Steering Committee consisting of representatives of the ACH, the ACL, and the ALLC appointed Michael Sperberg-McQueen as editor-in-chief and Lou Burnard as European editor of the Guidelines.

The first public proposal for the TEI Guidelines was published in July 1990 under the title Guidelines for the Encoding and Interchange of Machine-Readable Texts with the TEI document number TEI P1 (for Proposal 1). This version was reprinted with minor changes and
corrections, as version 1.1 in November 1990.  

Further development of the TEI Guidelines was done by four Working Committees (Text Documentation, Text Representation, Text Analysis and Interpretation, Metalanguage and Syntax) and a number of specialist Working Groups amongst which groups on character sets, text criticism, hypertext and hypermedia, formule, tables, figures, and graphics, language corpora, manuscripts and codicology, verse, drama and performance texts, literary prose, linguistic description, spoken text, literary studies, historical studies, print dictionaries, machine lexica, and terminological data. The extensions and revisions resulting from this work, together with extensive public comment resulted in the drafting of a revised version, TEI P2, that was released chapter by chapter between March 1992 and the end of 1993 and that included substantial amounts of new material.

6.3 TEI P3

The following step was the publication of the TEI P3 Guidelines for Electronic Text Encoding and Interchange in 1994 that presented a further revision of all chapters published under the document number TEI P2, and the addition of further chapters. A final revised edition of these P3 Guidelines correcting several typographic and other errors, and introducing one new element was published in 1999. The publication of this 1,292 page documentation of the definitive guidelines defining some 439 elements marked the conclusion of the initial development work. With this work, the Poughkeepsie Guidelines were met by providing a framework for the encoding of texts in any natural language, of any date, in any literary genre or text type, without restriction on form or content and treating both continuous materials (“running text”) and discontinuous materials such as dictionaries and linguistic corpora.

6.4 TEI P4

Recognizing the benefits for the TEI community, the P4 revision of the TEI Guidelines was published in 2002 by the newly formed TEI Consortium in order to provide equal support for XML and SGML applications using the TEI scheme. The chief objective of this revision was


8 Sperberg-McQueen, C.M.; Burnard, L. (eds.). TEI P4: Guidelines for Electronic Text Encoding and Interchange. XML-compatible edition. XML conversion by Syd Bauman, Lou Burnard, Steven DeRose,
to implement proper XML support in the Guidelines, while ensuring that documents produced to earlier TEI specifications remained usable with the new version. The XML support was realized by the expression of the TEI Guidelines in XML and the conformation to a TEI conformant XML DTD. The TEI P4 generated a set of DTD fragments that can be combined together to form either SGML or XML DTDs and thus achieved backwards compatibility with TEI P3 encoded texts. In other words, any document conforming to the TEI P3 SGML DTD was guaranteed to conform to the TEI P4 XML version of it. This ‘double awareness’ of the TEI P4 is the reason why this version was called an ‘XML-compatible edition’ rather than an ‘XML edition’. This was achieved by restricting the revisions needed to make the P4 version with its 441 elements to error correction only. During this process of revision, however, many possibilities for other, more fundamental changes have been identified. Which led to the current TEI P5 version of the Guidelines.

6.5 TEI P5

In 2003 the TEI Consortium asked their membership to convene Special Interest Groups (SIGs) whose aim could be to advise revision of certain chapters of the Guidelines and suggest changes and improvements in view of the P5. With the establishment of the new TEI Council, which superintends the technical work of the TEI Consortium, it became possible to agree on an agenda to enhance and modify the Guidelines more fundamentally which resulted in a full revision of the Guidelines published as TEI P5.9 TEI P5 contains a full XML expression of the TEI Guidelines and introduces new elements, revises content models, and reorganizes elements in a modular class system that facilitates flexible adaptations to users' needs. Contrary to its predecessor, TEI P5 does not offer backwards compatibility with previous versions of the TEI. The TEI Consortium has, however, maintained and corrected errors in the P4 Guidelines for 5 more years, up to the end of 2012. Since that date, the TEI Consortium has ceased official support for TEI P4, and deprecated it in favour of TEI P5.

......