The ACH-ACL-ALLC Text Encoding Initiative: An Overview

Susan Hockey
Chair, TEI Steering Committee

Document Number: TEI J16

June 1991 (Rev. Feb. 28, 1993)

Draft February 28, 1993 (15:10:59)

The Text Encoding Initiative (TEI) is a major international project, sponsored jointly by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). Its task is to develop and disseminate guidelines for the interchange of machine-readable texts among researchers, and to make recommendations for the encoding of new texts. Funding of approximately $1,000,000 has been provided by the US National Endowment for the Humanities, DG XIII of the Commission of the European Communities, and the Andrew W. Mellon Foundation. The TEI has also received substantial indirect support from the host institutions of participants in the project.

1

Background

1.1 The Need for a Common Text-Encoding Scheme

Textual data has been analysed and otherwise manipulated by computer for over thirty years, but until now there has not been any common encoding format for scholarly machine-readable texts. Scholars have developed many different encoding schemes for representing non-standard characters, footnotes, marginalia, text-critical apparatus, for marking the logical divisions of the text (e.g. book, chapter, verse) or the analytic or interpretive information relevant to the text (e.g. syntactic, morphological, or semantic analysis) as well as for documenting the source of an encoded text and the nature of the recording.

None of the existing encoding schemes has been able to gain acceptance as a standard. Most reflect the research interests of their originators and are applicable in only one subject area. Some were created to serve the needs of large-scale projects such as the Thesaurus Linguae Graecae at Irvine or the Responsa Project at Bar-Ilan (Israel) or as specifications for input to text-analysis software such as OCP or WatCon. Others were devised by single individuals for their own projects. None is sufficiently flexible or generalizable to apply to the encoding of textual materials across the full spectrum of applications and research interests. The practice followed in encoding them is often poorly documented and much time has been wasted writing software to translate from one format to another.

A common text encoding scheme developed for the needs of scholarly research would eliminate or minimize many of these problems and would also provide a text format to be used as a starting point by developers of new software.

1.2 Planning meeting

The TEI arose out of a planning conference convened by ACH at Vassar College, Poughkeepsie, New York on 12-13 November 1987. Prior to that there had been a number of meetings at which the problem was noted but no firm commitment given to finding any solution. At this meeting thirty-one experts from universities, learned societies, and text archives from North America, Europe, Israel, and Japan discussed the desirability, feasibility, and basic principles of a common set of guidelines for machine-readable text encoding. The firm consensus at this meeting was that such a common framework was necessary and feasible, and the group agreed on several basic principles to govern the scope and the organization of a set of guidelines for encoding textual materials. These principles are set out in Appendix A and have become known as the “Poughkeepsie Principles”.

Three key factors contributed to this consensus:

- More is known now about the problems of text encoding than at the time of previous attempts.
- Even though funding limitations and a compressed timetable made it impossible to invite everyone with a major potential contribution to the effort, the Vassar conference succeeded in bringing together more representatives of key organizations and active research centres than had ever met in one place before to discuss these problems.
- The recently developed Standard Generalized Markup Language (SGML), defined by the international standard ISO 8879, appears to provide an invaluable tool for developing a simple, flexible, extensible encoding scheme capable of satisfying the widely varying needs of textual researchers.

The newly achieved consensus also reflected the growing urgency of the need as more and more machine-readable texts are being created and exchanged. At Vassar, the status quo was several times described as ‘chaos'.

2
Objectives of the TEI

At Vassar three overall objectives were defined for the project:

1. To specify a common interchange format for machine readable texts.
2. To provide a set of recommendations for encoding new textual materials. The recommendations would specify both what features are to be encoded and how those features are to be represented.
3. To document the major existing encoding schemes, and develop a metalanguage in which to describe them.

3

Organization of the Project

3.1 Steering Committee

Following the Vassar conference ACL and ALLC joined with ACH to form a Steering Committee for the project with two representatives from each association. The Steering Committee is responsible for raising funds and overseeing the project. It meets approximately every three months, using electronic mail to conduct its business between meetings. The position of Chair of the Steering Committee rotates between the associations.

The first task of the Steering Committee was to establish a structure for the project and raise funds to carry it out. Initial funding was obtained from the NEH for two years beginning in June 1988 (first cycle), with the anticipation of a further two-year funding period. The project structure outlined below was in place by the end of 1988.

3.2 Participating Organizations — Advisory Board

Fifteen scholarly organisations participate in the TEI by representation on the project's Advisory Board. Members of the Advisory Board act as links between their organizations and the working committees and circulate progress reports and interim drafts within their organizations. The Advisory Board met in February 1989 to approve the project workplan and will meet again in late spring 1992 to review the guidelines.

3.3 Editors

Two editors co-ordinate the work and are responsible for drafting the TEI guidelines. They ensure the compatibility of the activities of the TEI committees, and review and edit the final document, integrating the products of the various committees into a single coherent whole, as well as dealing with day-to-day enquiries and acting as a TEI secretariat. The editors are based in Chicago and Oxford.

3.4 Working Committees

Following a distinction which was first discussed at the planning meeting,[1] four working committees were set up, with membership drawn from a broad cross section of the scholarly community.

3.4.1 Committee on Text Documentation

This committee, with expertise in librarianship and archive management, was charged with dealing with the problems of labeling a text with in-file encoding of cataloguing information about the electronic text itself, its source and the relationship between the two.

[1] See Appendix A item 6

3.4.2 Committee on Text Representation

This committee deals with the problems of representing in machine-readable form: character sets, the logical structure of a text, layout and other features physically represented in the source material. It first produced recommendations for ways of dealing with character sets and laid ground rules for encoding textual features common to most types of prose text.

3.4.3 Committee on Text Analysis and Interpretation

This committee was asked to provide discipline-specific sets of tags appropriate to the analytic procedure favoured by that discipline, but in such a way as to permit their extension and generalization to other disciplines. Its initial focus was on linguistic analysis and it has developed a number of powerful theory-independent mechanisms for encoding analytic features. Encoding of dictionaries is also dealt with by this committee with a subgroup which was able to build on previous work.

3.4.4 Committee for Metalanguage and Syntax Issues

This committee with expertise in formal language theory rapidly decided that SGML was the best choice of an encoding language for the TEI. It has produced recommendations about how SGML should best be used, suggested modifications in SGML that might be needed to accommodate the TEI, and has addressed the problems of conversion between the TEI and other encoding schemes.

4

Development of the First Draft of the TEI Guidelines

The work of the first cycle was to develop the first draft of the TEI Guidelines with input from all the Working Committees. The editors first prepared documents for the committees outlining their brief in more detail. Most of the committee work was then done by electronic mail with some face-to-face meetings to finalize the recommendations. At the end of the first development cycle the editors were able to put together the first draft of the TEI Guidelines (TEI P1, hereinafter "P1"),[2] a document of almost 300 pages that incorporates the recommendations of the Working Committees. This document was made available publicly in summer 1990 at the end of the first cycle. Section 7 describes the rationale behind the Guidelines in more detail. The Executive Summary of P1 is reproduced as Appendix B.

---

[2] C.M. Sperberg-McQueen and Lou Burnard, ed., *Guidelines For the Encoding and Interchange of Machine-Readable Texts* (Chicago and Oxford: Text Encoding Initiative, 1990).

5

Second Development Cycle

Substantial funding was obtained for the second cycle, beginning in June 1990 and this enabled the Steering Committee to consider more options on how to organize the work for this period. In the first draft of the Guidelines some topics are only sketched out; others need revision in the light of comments, and some are omitted completely. The major objectives during the second funding cycle are to test the guidelines and extend their scope and coverage in the light of user comments. A final document is scheduled for June 1992 at the end of the second funding cycle.

5.1 Work Groups

The work of the first cycle concentrated on the core requirements of the TEI Guidelines, that is encoding those features which are found in most text types. In the second cycle specific areas are being addressed in more detail by a number of small but tightly-focused work groups which the TEI has set up. The work groups will make recommendations in these areas, either directly where an area is already well-defined, or indirectly by sketching out a problem domain and proposing other work groups which need to be set up within it. Each work group has been given a specific charge and is working to a specified deadline. So far, about a dozen such groups have been set up, including: character sets, text criticism, hypertext and hypermedia, mathematical formulae and tables, language corpora, physical description of manuscripts, analytic bibliography, general linguistics, spoken texts, literary studies, historical studies, machine-readable dictionaries and computational lexica.

Each group is formally assigned to one of the two major working committees of the TEI, depending on whether its work is primarily concerned with Text Representation or Text Analysis and Interpretation. These two committees will then review and endorse the findings of each work group, though we expect that for some areas we will also seek expert outside reviewers, perhaps with the assistance of the Advisory Board.

The work of the Metalanguage Committee is continuing. Only a small amount of work remains to be done in the Documentation Committee.

5.2 Affiliated Projects

The TEI has set up arrangements with a number of projects which are engaged in the creation or maintenance of machine-readable texts. The projects have chosen to be representative of different discipline areas. They are encoding samples of their own data according to the Guidelines and reporting back to the TEI on the results of these tests. Each project has been assigned a TEI consultant to advise on the encoding and to assist in the preparation of the feedback report. Workshops for the affiliated projects were held in July 1991. The workshops provided an opportunity for affiliated project staff to learn more about the TEI guidelines and for consultants to become familiar with the work of the projects to which they are assigned. Each project will provide a formal report to the TEI on the problems encountered and how they have been overcome.

6

Syntax of the Scheme — SGML

No final decision about the syntactic basis for the new text encoding scheme was made at the planning conference. All present agreed that if possible the syntax should be borrowed from some existing scheme. The syntax must be relatively simple, capable of expressing the fine distinctions and occasionally complex overlapping hierarchical structures required by textual material, and must allow for user-defined extensions to the pre-defined set of tags. The most obvious candidate was the Standard Generalized Markup Language (SGML). It was decided to begin work on the syntax with SGML as a model, and abandon SGML only if it proved inadequate for the needs of research. SGML was soon shown to meet the requirements of the TEI, the one problem area being the need to handle multiple conflicting hierarchies for which some possible solutions do now exist.

SGML is not itself an encoding scheme but a framework within which encoding schemes (tag sets) may be developed. Because multiple tag sets may be used in the same texts, SGML-based encoding schemes can easily support a diversity of opinions about the basic text features to be tagged. SGML is device-independent and is supported by software products from an increasing number of vendors. It can handle any natural language for which an electronic representation exists and is by design application-independent. This means that it enables the same text base to be used for e.g. word processing and research-oriented analytic functions.

7

The TEI Guidelines

The function of the Guidelines is to specify both what features are to be encoded and how those features are to be represented. Because of diversity of the texts, no simple set of absolute requirements can be applicable to all texts and purposes, but the encoding of a certain minimum number of features is highly recommended. In many types of textual analysis, the textual features studied vary widely depending upon the theoretical orientation of the researcher. Research is needed to define sets of textual features relevant to specific disciplines or text types (e.g. lexical analysis, textual criticism, thematic study, etc.), and to provide techniques for representing them. The TEI approach to dealing with a wide variety of text types is to attempt to define a comparatively small number of features which all texts share, and to allow for these to be used in combination with user-definable sets of more specialized features.

The TEI has followed the Poughkeepsie Principles in the development of the Guidelines. These have been amplified in TEI ED P1, "Design Principles for Text Encoding Guidelines", the main points of which are summarized here.

The Guidelines should meet the following design goals

1. suffice to represent the textual features needed for research
2. be simple, clear and concrete
3. be easy for researchers to use without special-purpose software
4. allow the rigorous definition and efficient processing of texts
5. provide for user-extensions
6. conform to existing and emergent standards.

The Guidelines should contain explicit recommendations for

1. the encoding of new texts (which textual features should be captured, as a minimum, and how they should be marked)
2. the addition of new information or corrections to existing encodings
3. the interchange of existing encodings
4. archival documentation of encodings
5. text and encoding documentation for purposes of bibliographic control
6. documentation of selected markup schemes in terms of the recommended scheme.

The Guidelines should support work in any discipline based on textual material, concentrating initially on encoding the textual features of interest to those disciplines most commonly using computers: lexicography; thematic, metrical and stylistic studies; historical editing, computational branches of linguistics, and content analysis. Discipline-related work has concentrated first on linguistic issues and is now being extended to literary and historical studies and the June 1992 version of the Guidelines will provide explicit guidance for the text types most commonly encountered in textual research and for the handling of corpora.

Following the basic philosophy of SGML and the need for reusability, the Guidelines are based on descriptive rather than procedural markup. Where possible the encoding tags describe structural or other fundamental text features independently of their representation. However it is recognized that in some disciplines the physical appearance of the original text is important either because it is the primary object of interest or because there is no consensus on the meaning of its appearance. Therefore tags for appearance features are also provided, with the recommendation that the primary purpose of the markup should not be viewed simply as a faithful reproduction of the appearance of the original.

It should be stressed that the first draft of the Guidelines is very much a discussion document and is far from being complete or definitive. At the time of writing this we are half way through the second cycle and there is time for further comment and revision.

Further detail on the content of P1 can be found in the Executive Summary (Appendix B).

8

Information about the TEI

The TEI maintains an electronic bulletin board (TEI-L) on which news about the TEI is posted and which provides a forum for detailed discussion of the TEI Guidelines. To subscribe, send an electronic mail message containing only the line SUBSCRIBE TEI-L to LISTSERV@UICVM (bitnet) or LISTSERV@UICVM.UIC.EDU (internet).

The TEI also keeps a document register at Chicago, which holds information about all TEI-related working papers, reports and publications. Wherever possible, we try to make sure that finalized reports of general interest are made available publicly. Documents may be obtained using the fileserver mechanism of the LISTSERV. To find out what is currently available, send a message to LISTSERV@UICVM (bitnet) or LISTSERV@UICVM.UIC.EDU (internet) containing the line GET TEI-L FILELIST.

9

The Future

Once the technical specification of the TEI Guidelines has been established, the focus of the TEI will turn more to usability, user education and general promotion of the Guidelines.

9.1 Publications

At the time of writing this overview, the TEI Steering Committee is considering what the final deliverables will be when the current funding finishes in June 1992. It seems that a single large report (a revised and extended version of the current document TEI P1) will not be enough. There is a need for a brief introductory guide, setting out the basic TEI framework and philosophy, as well as for tutorial material, and for examples of TEI-encoded texts.

9.2 Feedback

The Guidelines were made available in draft in an admittedly incomplete state in the hope of stimulating the widest possible discussion of their proposals. The TEI is committed to respond to and summarize all comments on the proposals.

9.3 Maintenance

The Sponsoring Organizations have agreed to provide a mechanism for revising and extending the Guidelines after their first publication, based on experience with the Guidelines in practice. This joint maintenance mechanism will be responsible for updates and revisions. It may prove appropriate, after approval and publication of the Guidelines, for the Sponsoring Organizations to seek adoption of the Guidelines as national or international standards.

9.4 Training Materials

The TEI has held several training workshops. These workshops have provided the basis of information to be used as training materials and also have acted as another means of obtaining feedback from users. As time passes the content of these training materials will stabilize. It is hoped that there will be more people available to do the training and that some self-teaching material will be available.

It has already become apparent that a different approach to training may be required for the different types of users. What is easy for the computer scientist to understand is not necessarily as easy for the humanities scholar and vice versa. The TEI Guidelines are important for both and so the method of putting across the information may need to be tailored to the type of user.

9.5 Software

The need for software to handle TEI SGML is becoming more pressing. The requirement seems to be for systems which behave like well-known packages, but can also take advantage of the new capabilities offered by SGML. As a first step, we need programs to translate from the TEI encoding scheme to those required by the application programs we use, and back in the other direction. For writing one's own software, the community needs generally available routines which can read and understand TEI documents and which can be built into software individuals or projects develop for themselves or others (TEI parsers). Equally important for the usability of the encoding scheme in the community at large will be TEI-aware data-entry software, e.g. editors and word processors which can exploit the rich text structure provided by SGML, simple routines to allow TEI tags to be entered into a text with a keystroke or two, and other tools to help make new texts in the form recommended by the TEI.

Approximations to some of these are already available, but there is still a real need to provide such tools cheaply on the machines which humanities scholars most often use, i.e. IBM PCs and Macintoshes, and in such a way that they can be incorporated into software which is already in use.

9.6 Further Documents

The Metalanguage Committee has also prepared a more explicit definition of the notion 'TEI conformance', which has been requested by those who would like to develop TEI-conformant software. Some of the TEI working papers are also suitable for publication.

9.7 Conversion of Existing Archives

It is unreasonable to expect existing archives to use the new scheme for new texts, which they will need to keep consistent with their existing holdings, but the text archivists at the Vassar conference nevertheless saw a pressing need for a common format for data interchange. Such a format would allow each archive to write software with which to convert its present data format into the common interchange format before sending texts to other users, and convert incoming texts from the common interchange format into the local format.

At some later time, once the Guidelines are more widely used, it is likely that some archives will want to convert to the TEI scheme. The work of the Metalanguage Committee in providing a formal specification of existing schemes and the of TEI scheme should provide the groundwork for this.

10

The Ultimate Goal: User Acceptance

No standard can be imposed. The success of the TEI will depend on the ultimate acceptance of the Guidelines by the community they are intended to serve. Involving users as much as possible in their development has been the first stage in this effort. The TEI Steering Committee is now formulating plans for the future, once the June 1992 version is in place. The activities outlined in the previous section are some of the things we will consider. Suggestions for other will be very welcome.

11

Acknowledgements

Much of the material in this overview has been taken from other TEI documents, most of which were originally written by the editors Michael Sperberg-McQueen and Lou Burnard.

This work has been made possible with the generous financial support provided by the US National Endowment for the Humanities, DG XIII of the Commission of the European Communities and the Andrew W. Mellon Foundation.

The TEI is also grateful for the substantial indirect support from the host institutions of participants in the project.

Appendix A

Closing Statement of the Vassar Planning Conference

The Preparation of Text Encoding Guidelines

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
   a. define a recommended syntax for the format,
   b. define a metalanguage for the description of text-encoding schemes,
   c. 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
   a. text documentation
   b. text representation
   c. text interpretation and analysis
   d. 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.

Appendix B

Executive Summary of Draft 1 of the TEI Guidelines July 1990

This document summarizes the first public draft of the Text Encoding Initiative's Guidelines for the Encoding and Interchange of Machine-Readable Texts. It describes briefly the background and goals of the project and then summarizes the content of the Guidelines with a view to stimulating informed public comment and discussion of its draft proposals.

The Text Encoding Initiative (TEI) arose out of a planning conference convened by the Association for Computers and the Humanities (ACH) at Poughkeepsie, New York, in November 1987. It is an international research project, sponsored jointly by the ACH, the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC), with the further participation of several other organizations and learned societies. The associations concerned, whose representatives constitute an Advisory Board, are listed in the preface to the Guidelines.

Chapter 1 of the draft Guidelines describes their purpose and form. The goal of the TEI is to develop and disseminate a clearly defined format for the interchange of machine-readable texts among researchers, so as to allow easier and more efficient sharing of resources for textual computing and natural language processing. This interchange format is intended to specify how texts should be encoded or marked up so that they can be shared by different research projects for different purposes. The use of specific types of delimiters to distinguish markup from text, the use of specific delimiter characters, and the use of specific tags to express specific information, are all specified by this interchange format, which is closely based on the international standard ISO 8879, Standard Generalized Markup Language (SGML), on which see more below.

In addition, the TEI has taken on the task of making recommendations for the encoding of new texts using the interchange format. These recommendations give specific advice on what textual features to encode, when encoding texts from scratch, to help ensure that the text, once encoded, will be useful not only for the original encoder, but also for other researchers with other purposes. In these Guidelines, the requirements of the interchange format are expressed as imperatives; other recommendations, which are not binding for TEI-conformant texts, are marked as such by the use of such phrases as should or where possible.

The current Guidelines thus have two closely related goals: to define a format for text interchange and to recommend specific practices in the encoding of new texts. The format and recommendations may be used for interchange, for local processing of any texts, or for creation of new texts. Their use in local processing or text creation is always a local decision and it is expected that formats used in local processing may vary from the specifications presented here, even if the Guidelines are used as the basis for local work. A somewhat stricter definition is provided of TEI-conformant texts to ensure the utility of the Guidelines in interchange. Many of the details of the current draft will change in the coming development cycle, though the commitment to SGML will continue, as will the basic approach to the problem of representing natural language texts in computer-processable form.

Chapter 2 of the Guidelines provides a tutorial introduction to SGML, a formal language for the definition of markup schemes which consistently distinguishes the content of the text being encoded from the markup used to mark its structure or other features of interest. This chapter also contains a technical discussion of the SGML features used in these Guidelines and an overview of the structure of TEI-conformant documents in SGML terms.

Chapter 3 discusses the problems of character encoding and character set translation, and outlines some recommendations:

- In interchange, the character stream of a document should be restricted to a subset of ISO 646 (the ISO 7-bit character code corresponding to ASCII in the U.S.); which contains no national-use characters and no exclamation point. This temporary restriction allows transmission over current networks without data loss; it will become unnecessary when interchange no longer takes place over 7-bit channels.
- Other characters should be represented either with SGML entity references or by transliteration. (These mechanisms are described in more detail in chapters 2 and 3 of the document.)
- Transliteration schemes should be documented using a Writing System Declaration, that is, a formal specification defining a transliteration scheme in terms of some existing internationally defined character set. A formal SGML definition of the Writing System Declaration is provided so that such declarations can be used locally to document the character set in use. Ways of formalizing the use of such declarations in interchange will be the subject of further work within the TEI.

Chapter 4 discusses the TEI header, a mechanism for providing full in-file documentation of the text encoded, including its source or provenance, those responsible for its production and details of the encoding conventions it employs. The TEI header is designed to provide information adequate to the needs of researchers, data archivists and library cataloguers cataloguing the file. To that end, it attempts to address the requirements of the relevant national and international standards, notably ISBD(CF).

Tagging for features common to many text types is described in some detail in chapter 5. First an overall structure for conventionally-structured (prose) texts is provided, with tags for marking parts, chapters, sections, etc. (under whatever name), together with their headings and closings. Front-matter tags are provided for marking title pages, forewords, acknowledgements, abstracts, dedications, frontispieces, and the like. Back-matter tags are similarly provided for appendixes, glossaries, notes, etc. Within the component parts of these structural features, normal prose allows for the appearance of a wide variety of non-structural features, which we call phrase-level tags; these include such things as lists, notes, names, abbreviations, numbers, dates etc. Index entry tags are also provide, to serve either in book production or in information retrieval.

Chapter 5 also contains specific recommendations for the encoding of the following features:

1. bibliographic citations and references to them
2. the traditional or canonical reference systems used to structure many literary and historical texts
3. cross references and hypertextual links
4. different ways of segmenting running text, for example the division of paragraph-level units into orthographic sentences
5. non-textual items such as formulas, tables, and figures
6. ways of representing textual variation and traditional critical apparatus
7. ways of representing some aspects of the visual presentation of a text, in particular its typographic rendition, where (as in older texts) this is of analytic importance.

Chapter 6 addresses the problem of defining a general-purpose formalism for describing the many different kinds of linguistic analysis, in particular the most widely-used types of linguistic and morphological analyses. Some simple theory-independent mechanisms are proposed, together with a few special-purpose tag sets for particular areas. The general tools include:

1. tags for named, value-bearing feature structures which may be nested and grouped into sets or lists
2. ways of representing tree structures composed of such feature structures

3. ways of grouping feature structures containing logical operators
4. ways of representing multiple analyses for the same word or textual feature
5. ways of representing how such multiple analyses are to be aligned

This chapter also includes some simple examples of the application of the proposed mechanisms to problems in syntax and morphology.

Chapter 7 considers in more detail particular aspects of specific types of text. The text-types discussed in this draft are: language corpora and collections; verse, drama, and narrative; dictionaries; and office documents. In each case, an overview of the problems specific to these types of discourse is given, with some preliminary proposals for tags appropriate to them.

An important design goal of the TEI scheme is that it should be easily extended and modified: extensions and modifications of the Guidelines are both expected and intended, and the final form of the Guidelines will be designed to facilitate them. Chapter 8 describes the techniques initially proposed for carrying out such modifications, largely by introducing levels of indirection into the formal SGML specifications. The methods proposed have not yet been fully implemented, but are intended to be both simple and, so far as possible, self-documenting.

A number of technical appendixes are planned for the next draft of the Guidelines, preliminary or partial versions of which are provided in the current draft. These will include the following items:

1. an alphabetical reference section with a summary of each tag, its attributes, its usage, and an example of its use. This will be made available before the next draft.
2. annotated examples, illustrating the application of the TEI encoding scheme to the markup of a wide range of texts. Several examples are included in this draft, but these should not be regarded as definitive or complete.
3. formal SGML document type declarations (DTDs) for all the tags and groups of tags defined in the TEI scheme. Again, some preliminary DTDs are included in the current draft (including one for the core tag set defined in chapter 5), but these should not be regarded as definitive or complete.
4. code pages and sample Writing System Declarations for commonly used character sets. Several code pages have been included in the present draft, but WSDs are not yet included.

As work progresses on these and other missing parts of the Guidelines, it is probable that some of them will be distributed as separate documents. All those noted above will however be incorporated into the next full draft of these Guidelines. Information on the availability of this and other TEI publications is available from the addresses given in the preface, and will also be posted on the TEI's electronic bulletin board, TEI-L@UICVM (on BITNET; TEI-L@uivcm.uic.edu on the Internet).

Appendix C

TEI Addresses

Steering Committee

Susan Armstrong-Warwick (ACL)
ISSCO University of Geneva
54 route des Acacias
CH-1227 Geneve SWITZERLAND
susan@divsun.unige.ch
(+41 22) 705-7113
(+41 22) 300-1086 fax

Susan Hockey (ALLC), Chair
Center for Electronic Texts in the Humanities
169 College Avenue
New Brunswick, New Jersey 08903
Hockey@Zodiac.rutgers.edu
(+1 908) 932-1384
(+1 908) 932-1386 Fax

Nancy M. Ide (ACH)
AFTER 1 JUNE
Vassar College
Department of Computer Science,
Box 520
Poughkeepsie, NY 12601, USA
ide@vaxsar.vassar.edu
(+1 914) 437-5988
(+1 914) 437-7187 Fax
UNTIL 31 MAY
Rue de la Gardy
F-13480 Cabries, FRANCE
ide@grtc.cnrs-mrs.fr
+33-42-229516

C. M. Sperberg-McQueen (ACH)
Computer Center (M/C 135),
University of Illinois at Chicago,
1940 W. Taylor St.-Rm. 124
Chicago, IL 60612-352, USA
u35395@uicvm.uic.edu
u35395@uicvm.bitnet
(+1 312) 413-0317
(+1 312) 996-6834 Fax

Donald E. Walker (ACL)
Bellcore, MRE 2A379,
445 South Street,
Box 1910,
Morristown, NJ 07960-1910, USA
walker@bellcore.com
(+1 201) 829-4312
(+1 201) 829-5981 Fax

Antonio Zampolli (ALLC)
Istituto di Linguistica Computazionale,
Via Della Faggiola 32,
I-56100 Pisa, Italy
glottolo@icnucevm.cnuce.cnr.it
glottolo@icnucevm.bitnet
(+39 50) 560481
(+39 50) 589055 Fax

Editors

In Europe:

Lou Burnard,
Oxford University Computing Service,
13 Banbury Rd,
Oxford OX2 6NN,
England
lou@vax.ox.ac.uk
(+44 865) 273238, 273200
(+44 865) 273275 Fax

Elsewhere:

C. M. Sperberg-McQueen,
Computer Center (M/C 135),
University of Illinois at Chicago,
1940 W. Taylor St. - Rm. 124
Chicago, IL 60612-7352, USA
u35395@uicvm.uic.edu
u35395@uicvm.bitnet
(+1 312) 413-0317
(+1 312) 996-6834 fax