The Association for Computers and the Humanities (ACH)

The Association for Computational Linguistics (ACL)

The Association for Literary and Linguistic Computing (ALLC)

Guidelines for Electronic Text Encoding and Interchange

Edited by

C. M. Sperberg-McQueen and Lou Burnard

TEI P2, Chapter 26 SH
Independent Header
Text Encoding Initiative
Chicago, Oxford

September 16, 1993

Draft Version 2, September 16, 1993

Note

This is the current draft (as of September 16, 1993) of one section of the *Guidelines for Electronic Text Encoding and Interchange*, version 2 (also known as TEI P2), produced by the Text Encoding Initiative. This fascicle contains:

- a table of contents for TEI P2, indicating the status of each section as of the release date of this fascicle
- a draft for one section of TEI P2
- entries derived from this section for the reference list of tags of TEI P2 (if any)
- a User Response and Comment Form

Please feel free to send us your comments on the contents of this draft and their applicability to your needs. Comments may be sent to the TEI-L list server for general discussion, or directly to either of the TEI editors (adresses on the User Response and Comment form). Bear in mind in reading, however, that TEI P2 is a reference manual, not a tutorial in the use of electronic text, and unavoidably assumes either a certain amount of prior knowledge or patience with unfamiliar technical material. Note also that some matters which may appear obscure in one section may well be treated in more detail elsewhere in the full text of TEI P2.

TEI P2, like P1, is a working draft. The observant will find several typographic and presentational flaws in the current text: for example, cross-references are not properly resolved in the LaTeX versions and some parts of the reference material are not yet complete. We hope to rectify as many as possible of these flaws in the coming months, but have decided not to allow them to delay any further the release of the draft. We hope that those interested in the Guidelines will find the current state of the text readable and usable enough for the purpose of substantive comment. As far as possible, all comments received on this section will be taken into account in the preparation of the next version of the Guidelines, which is due by the end of 1993.

Updates on the status of the draft as a whole will be distributed automatically to all subscribers of the Listserv list TEI-L. To subscribe to TEI-L, send electronic mail to the address LISTSERV@UICVM (or Listserv@uicvm.uic.edu) containing the single line

```
subscribe tei-l J. Smith
```

(substituting your name for “J. Smith”). TEI-L is also the appropriate place to pose questions or offer public comments on the TEI guidelines and other relevant issues. The TEI-L file server contains all the sections of TEI P2 thus far released, as well as other TEI materials. Its contents are shadowed at various other sites around the world; for more information consult one of the following documents:

- TEI ED J8, “Obtaining the Second Version of the TEI Guidelines (TEI P2)” (describes how to retrieve electronic copies of TEI P2 and the various formats they are available in)
- TEI ED J9, “Obtaining Paper Copies of the Second Version of the TEI Guidelines (TEI P2)” (describes how to request paper copies of TEI P2, for those without electronic mail access)

These documents are available from the TEI-L file server or through the editors at the addresses given on the User Response and Comment Form.

Overall Table of Contents for TEI P2

The following list shows the expected contents of TEI P2 as a whole and indicates the current status of each chapter.

Contents

Part 5

Auxiliary Document Types

Chapter 26

The Independent Header

Many libraries, text repositories, research sites and related institutions collect bibliographic and documentary information about machine readable texts without necessarily collecting the texts themselves. Such institutions may thus want access to the header of a TEI document without its attached text in order to build catalogues, indexes and databases that can be used by people to locate relevant texts at remote locations, obtain full documentation about those texts, and learn how to obtain them. This chapter of the Guidelines describes a set of practices by which the headers of TEI documents can be extracted from those documents and exchanged as freestanding independent TEI documents to remote locations. Headers exchanged independently of the documents they describe are called *independent headers*.[1]

This chapter outlines practices recommended for encoders (especially those responsible for the documentation of text) when creating independent headers to be distributed, and specifies the set of recommended elements that should be included in the independent header. Of interest to librarian cataloguers who may receive independent headers from remote sites, it also discusses the relationship between the elements of TEI headers and MARC tags, in order to facilitate the cataloguing of these headers or the loading of independent headers into local MARC-based bibliographic databases. This chapter does *not* describe how to create a header. Guidance on the creation of headers and descriptions of each element in the header can be found in chapter 5, "The TEI Header," [in separate fascicle].

26.1 Definition and Principles for Encoders

A *independent header* is a header extracted from a TEI text that can be exchanged as an independent document between libraries, archives, collections, projects, and individuals. The file description of the independent header (enclosed by the **<fileDesc>** element) can be used to generate bibliographic records. The profile description, encoding description, and revision history (encoded by the **<profileDesc>**, **<encodingDesc>**, and **<revisionDesc>** elements) can form part of a bibliographic description or, more appropriately, be used as an attached "codebook" for full documentation of the analysis of the text, and how it was encoded. Thus, the independent header can serve as the primary means by which libraries, archives, related repositories, research projects, and individual researchers can obtain bibliographic, descriptive, and full documentary information on machine-readable texts that reside in remote locations.

The structure of an independent header is exactly the same as that of a **<teiHeader>** attached to a document, and can therefore be validated using the same document type definition (DTD). In practice, this means that a **<teiHeader>** and its DTD can be extracted from a TEI document and shipped to a receiving institution with little or no change. However, some fields that are listed as "optional" in the

---

[1] The responsibility for this chapter was borne by the Working Committee for Text Documentation, whose members are listed in the front matter. The chapter was drafted by Richard Giordano.

header are listed as "recommended" for the independent header. For this reason, this chapter should be consulted in connection with any plan to send headers as independent documents to remote locations.

When deciding which information to include in the independent header, and the format or structure of that information, the following should be kept in mind:

The independent header should provide full bibliographic information on the encoded text, its source, where the text can be located, and any restrictions governing its use.

The header should contain useful information about the encoding of the text itself. In this regard, it is highly recommended that the encoding description be as complete as possible. The Guidelines do not require that the encoding description be included in the header (since some simple transcriptions of small items may not require it), but in practice the use of a header without an encoding description would be severely limited.

The independent header should be amenable to automatic processing, particularly for loading into databases and for the creation of publications, indexes, and finding aids, without undue editorial intervention on the part of the receiving institution. For this reason, two recommendations are made regarding the format or structure of the header: first, where there is a choice between a prose content model and one that contains a formal series of specialized elements, *wherever possible and appropriate the specialized elements should be preferred to unstructured prose*. For instance, the source description can contain either a free-prose citation (tagged **<bibl>** or even **<p>**) or else a **<biblStruct>** element, which provides a more rigorous structure for the bibliographic information (see examples in section 6.10, "Bibliographic Citations and References," [in separate fascicle]). The more structured **<biblStruct>** element is more suitable for automatic processing, and is therefore recommended over the less structured alternatives whenever the header is to be exchanged as an independent header. Second, with respect to corpora, information about each of the texts within a corpus should be included in the overall corpus-level **<teiHeader>**. That is, source information, editorial practices, encoding descriptions, and the like should be included in the relevant sections of the corpus **<teiHeader>**, with pointers to them from the headers of the individual texts included in the corpus. There are three reasons for this recommendation: first, the corpus-level header will contain the full array of bibliographic and documentary information for each of the texts in a corpus, and thus be of great benefit to remote users searching for particular texts, who may have access only to the independent header; second, such a layout is easier for the coder to maintain than searching for information throughout a text; and third, generally speaking, this practice results in greater overall consistency, especially with respect to bibliographic citations.

26.2 Required and Recommended Tags

The richness and size of the header reflect the diversity of uses to which electronic texts conforming to these Guidelines will be put. It is not intended, however, that all of the elements recommended in this chapter be present in every header. As described in section 5.6, "Minimal and Recommended Headers," [in separate fascicle], the TEI header allows for the provision of a very large amount of information concerning the text itself, its source, encodings, and revisions as well as detailed descriptive information that can be used by researchers in analysing the text. The amount of encoding will depend on the nature and intended use of the text. At one extreme, an encoder may expect that the header will be needed only to provide bibliographic information of the text adequate to local needs. At the other, wishing to ensure that their texts can be used for the widest range of applications, encoders will want to document as explicitly as possible both bibliographic and descriptive information in such a way that no prior or ancillary knowledge about the text is needed in order to process it. The header, in the latter case, will be very full, approximating to the kind of documentation often supplied in the form of a manual. Most texts will lie somewhere between these extremes; textual corpora in particular will tend toward the latter extreme.

There follows a list of the components of the header, in the order in which they are presented in chapter 5, "The TEI Header," [in separate fascicle], together with an indication of their importance in constructing an independent header.

- **<fileDesc>** required. Some subelements are required, others optional or recommended:
  - **<titleStmt>** required; subelements are required or optional:
    - **<title>** required
    - **<author>** required, if known
    - **<sponsor>** optional
    - **<funder>** optional
    - **<principal>** required, if known
    - **<resp>** required, if known
    - **<role>** and **<name>** required, if known, when the responsibility is not an author, sponsor, funding body, or principal researcher. Details may be found in section 5.2.1, "The Title Statement," [in separate fascicle].
  - **<editionStmt>** recommended
    - **<edition>** recommended
    - **<resp>** recommended
    - **<role>** and **<name>** recommended primarily to distinguish editions.
  - **<extent>** optional
  - **<publicationStmt>** required
    - **<date>** recommended
    - **<publisher>**, **<distributor>**, or **<authority>** required
    - **<city>** recommended
    - **<address>** recommended: prose is sufficient
    - **<idno>** recommended
    - **<availability>** recommended
  - **<seriesStmt>** optional
    - **<title>** required
    - **<idno>** recommended
    - **<resp>** and **<name>** optional
  - **<notesStmt>** recommended
- **<sourceDesc>** required. As much information as possible should be provided to identify the source. The following tags are either required or recommended, but other tags not listed here should be used wherever applicable in order to provide an accurate identification of the source. In some instances, the **<biblFull>** tag is preferable to the **<biblStruct>** tag.

— **<biblStruct>** recommended. For a full discussion of **<biblStruct>**, see section 6.10, “Bibliographic Citations and References,” [in separate fascicle].
— **<analytic>** required when the citation describes an item within a larger collection, such as an essay within a collection or an article in a journal, and is not an independent publication. If used, it should contain the following elements in this order:
  -- **<author>** required, if known.
  -- **<title>** required.
  -- **<editor>** recommended.
— **<monogr>** mandatory when applicable; this element should contain the following elements in this order:
  -- **<author>** required, if known.
  -- **<title>** required. The *level* attribute must be used to indicate whether this is the title of a book, journal, or series. All tags that indicate. It is highly recommended that the *type* attribute be used to distinguish the main title from subordinate, parallel, or other titles. All elements that indicate intellectual responsibility for a work, such as **<editor>**, are required, if known.
  -- **<imprint>** required.
  -- **<city>** required, if known.
  -- **<pubPlace>** recommended when **<city>** is unknown.
  -- **<org>** recommended.
  -- **<date>** required. If the date is unknown, “n.d.” may be used.
  -- **<idno>** recommended.
  -- **<series>** required, if the item is part of a series.
  -- **<title>** required, but *type* attribute is optional.
— **<scriptStmt>** required for transcribed speech. See section 5.2.9, “Computer Files Composed of Transcribed Speech,” [in separate fascicle].
— **<recordingStmt>** mandatory when applicable:
  -- **<resp>** and **<name>** recommended
  -- **<recording>** recommended
  -- **<equipment>** recommended
  -- **<broadcast>** recommended
  -- **<comment>** optional

• **<encodingDesc>** very highly recommended, especially for projects, collections, or corpora. If the **<encodingDesc>** element is used, it is recommended that it contain one or more of the following elements, rather than a prose description. See section 5.3, “The Encoding Description,” [in separate fascicle].
— **<projectDesc>** optional
— **<samplingDecl>** optional
— **<editorialDecl>** recommended; it is also recommended that the editorial declaration make use of the specialized elements defined in section 5.3.3, “The Editorial Practices Declaration,” [in separate fascicle], rather than only consisting of prose paragraphs. Prose may of course be used in addition to these elements for material otherwise not handled.
— **<tagsDecl>** recommended
— **<refsDecl>** optional in general, but recommended if a standard referencing system is built into the encoded works. Section 5.3.4, “The Ref-

erence System Declaration," [in separate fascicle] describes three different methods for documenting the referencing system: the prose method, the stepwise method, and the milestone method. No preference is expressed for one type of method over another, since this depends on the convenience of the coder and the likely efficiency of the particular software applications envisaged for the text. Only one method can be used within a single **<refsDecl>** element. If a text uses both hierarchical and milestone tagging, this can only be described in prose.

— **<classDecl>** required where the *scheme* attribute has been used to identify the classification scheme or taxonomy used by any of the elements **<keywords>**, **<classcode>**, **<occupation>** or **<sococstatus>**. Even where this is not done, this element may usefully document the classification employed, either explicitly as a series of **<taxonomy>** elements, or implicitly by means of bibliographic citation.

- **<profileDesc>** recommended
  - — **<langUsage>** recommended
  - — **<language>** recommended
- **<textDesc>** optional in most instances, but recommended when the encoder wants to provide a full description of the situation within which a text was produced or experienced, characterize it in a relatively continuous manner (in contrast to discrete categories based on type or topic), and believes that this characterization of the text will be helpful to the understanding, analysis, or retrieval of this text by remote users. If a collection or corpus uses a pre-existing descriptive typology as its organizing principle, it is recommended that its components be re-expressed in terms of the parameters listed here. If the encoder believes that pre-existing text categories (such as a standard classification scheme) are sufficient, then it is recommended that the **<textClass>** element be used instead. See section 25.2.1, "The Text Description," [in separate fascicle] for details and guidance.
  - — **<channel>** required
  - — **<constitution>** required
  - — **<derivation>** required
  - — **<domain>** required
  - — **<factuality>** required
  - — **<interaction>** required
  - — **<preparedness>** required
  - — **<purposes>** required
  - — **<purpose>** required
- **<textClass>** optional in most instances; this element may may be used as an alternative or in addition to the **<textDesc>** element. **<textClass>** is recommended in the following situations:
  - — a standard text category, such as the Library of Congress List of Subject Headings or a Dewey Decimal Classification category, clearly describes the text
  - — situational parameters (or the demographic elements of the **<particDesc>** element) are used and a text category can be constructed by the encoder based on a recurring set of values for those parameters.

  See section 5.4.3, "The Text Classification," [in separate fascicle] for details and guidance. One or more of the following sub-elements can be used.
  - — **<keywords>** recommended only if using a standard thesaurus such as the Library of Congress List of Subject Headings, a discipline-

specific thesaurus, or a thesaurus defined explicitly in the header. In each case, the source should be indicated by the *scheme* attribute and defined in the **<classDecl>** element.

- — **<classCode>** recommended only if the text is categorized by an internationally accepted classification scheme, such as the Dewey Decimal or Universal Decimal classification schemes. The scheme should be indicated by the *scheme* attribute and defined in the **<classDecl>** element.
- — **<catRef>** optional in most instances, but recommended when a user-defined classification is in use. The scheme should be indicated by the *scheme* attribute and defined in the **<classDecl>** element.

- **<particDesc>** optional, but recommended for spoken text when the encoder judges that such information is useful to remote users in the analysis of that text, and for both written and spoken text if such information is useful in the analysis of language usage. For details and guidance, see section 25.2.2, “The Participants Description,” [in separate fascicle].
- **<participant>** or **<particGroup>** recommended. Though the substructure of both the **<participant>** and **<particGroup>** elements can be prose, in independent headers one or more of the following sub-elements providing more specific details should be used in preference. Users of these Guidelines are free to extend the set of headings listed below.
  - — **<name>** recommended when the information is available
  - — **<birthDate>** recommended when the information is available
  - — **<birthPlace>** recommended when the information is available
  - — **<firstLang>** recommended when the information is available
  - — **<langKnown>** recommended when the information is available
  - — **<residence>** recommended when the information is available
  - — **<education>** recommended when the information is available
  - — **<affiliation>** recommended when the information is available
  - — **<occupation>** it is recommended that, where possible, the classification of the trade, occupation, or profession be derived from a standard classification or taxonomy, and that the source taxonomy be identified in the *scheme* attribute.
  - — **<socecstatus>** it is recommended that, where possible, the encoding of social and economic status be derived from a standard classification or taxonomy, and that the source taxonomy be identified in the *scheme* attribute.
- **<particRelations>** optional, but recommended where it is judged by the encoder that such information is important to the analysis of the text. If the **<particRelations>** tag is used, it is recommended that the special purpose **<relation>** element be used. See section 25.2.2, “The Participants Description,” [in separate fascicle].
- **<settingDesc>** optional, but recommended when the encoder judges that this information is useful in the analysis of the text, particular in the analysis of language usage.
- **<revisionDesc>** required in the independent header when available. It is recommended that the **<revisionDesc>** be encoded with a series of special purpose **<change>** elements with grouped **<name>**, **<what>** and **<date>** tags.

26.3 Header Elements and their Relationship to the MARC Record

This section offers some guidance to both cataloguers and bibliographic analysts who want to load TEI independent headers into a MARC-based retrieval system. Because there are variations in cataloguing practice across local sites, among bibliographic utilities (such as OCLC and RLIN), and differences in MARC usage in different countries, only tentative advice is possible. Note that the following examples are based on USMARC, *not* UNIMARC.[2] UNIMARC offers cataloguers in different countries the opportunity to combine different national practices in a single MARC format, and is the preferred "flavour" of MARC records for distribution across national boundaries. The implementation of UNIMARC, however, will be affected by local practice and by guidelines offered by the bibliographic utilities. Though UNIMARC is a stable format, the guidelines for its implementation are not sufficiently known or stabilized to be included in this chapter.

There are some major differences between the MARC record and the TEI header that will cause problems for librarians trying to map from the TEI independent header to the MARC record. The most important difference between the MARC record and the TEI header is the function of each. Despite the efforts and claims of some members of the library community, the MARC record remains fundamentally an electronic version of the catalogue card, with the limitations of its model.[3] The catalogue card is a unitary record for a physical object containing complex *bibliographic data* of varying sorts. The catalogue card points to the physical object. The TEI header provides full bibliographic information (as would a card), as well as documentary non-bibliographic information that supports the analysis, either by humans or machines, of the electronic text documented by header. Most of this analytical information, which is found in profile description, encoding description, and revision history, has little direct provision for it in the MARC record, and if retained must be recorded as unstructured notes (55XX) fields. Notes fields usually do not have the structure to support machine retrieval and analysis, while properly formatted profile, encoding, and revision descriptions lend themselves to retrieval, can support machine processing (including analysis) and point directly to the electronic text attached to the header. Moreover, the electronic text points back to the relevant elements in the header.

---

[2] For more information on UNIMARC, see Brian P. Holt, *UNIMARC Manual* (London, U.K.: IFLA Universal Bibliographic Control and International MARC Programme, British Library, 1987). For USMARC, see Walt Crawford, *MARC for library use: understanding USMARC* (Boston: G.K. Hall, 1989), *USMARC format for bibliographic data, including content designation* (Washington, D.C.: Library of Congress, 1987), and Deborah J. Byrne, *MARC manual : understanding and using MARC records* (Englewood, Colo.: Libraries Unlimited, Inc., 1991).

[3] The primary function of the MARC record when it was first designed in the mid-1960s was to allow for the electronic distribution of cataloguing records in support of card production. See Henriette Avram, *The MARC Pilot Project* (Washington D.C.: Library of Congress, 1968), p. 3. For discussion of the relationship between the MARC record and the catalogue card, see Michael Gorman, "After AACR2R: The Future of the Anglo-American Cataloging Rules," in Richard Smiraglia, ed., *Origins, Content and Future of AACR2 Revised* (Chicago: American Library Association, 1992).

Though this chapter offers some advice on where the profile, encoding, and revision descriptions might go in a MARC record, for practical reasons a repository might want create a codebook from these divisions of the header, and create a MARC record from the file description only. The MARC record should contain a reference to the codebook.

Subfields (or delimiters) are indicated by the dollar sign ($).

26.4 MARC fields for the File Description

Note that there is no provision for the “Main Entry” (or USMARC 1XX fields) in the TEI header. The main entry should be constructed, using appropriate name authority control, by the cataloguer from information derived from the header that indicates who is primarily responsible for the intellectual content of the work. There is an **<author>** tag, but the form of the name will have to be checked by a cataloguer before the main entry is constructed.

- **<titleStmt>** corresponds to title and statement of responsibility fields in MARC, typically 240 (for uniform title) and 245 (for title proper).
- **<title>** 240 $a (for uniform titles) or 245 $a fields. Put any subtitles in 24X $b. Insert the constant, [computer file] in the 24X $h gmd subfield.

The following elements belong in the 245 $c subfield: statement of responsibility.

- **<sponsor>**
- **<funder>**
- **<principal>**

Example:

```
<titleStmt>
  <title>Two stories by Edgar Allen Poe: electronic
         version</title>
  <author>Poe, Edgar Allen (1809-1849)</author>
  <respStmt><resp>compiled by</resp>
  <name>James D. Benson</name></respStmt>
```

This might be tagged in MARC as:

```
245 Two stories by Edgar Allen Poe :$belectronic version ;
    compiled by $cJames D. Benson.
```

- **<edition>** 250 $a
- **<name>** 250 $b

Example:

```
<editionStmt><edition>
    Student's edition,
    <date>June 1987</date>
</edition><respStmt>
    <resp>New annotation by</resp>
```

```
    <name>George Brown</name>
</respStmt></editionStmt>
```

This might be tagged in MARC as:

```
250  $aStudent's edition, June, 1987, new annotation by
     $bGeorge Brown.
```

- **<extent>**. The extent is analogous to the “Physical Description” MARC field. Fields 256 or 3XX, depending on local practice are appropriate.
- **<date>** 260 $c, and appropriate fixed fields.
- **<publisher>**, **<distributor>**, or **<authority>** 260 $b
- **<city>** 260 $a

Example:

```
<publicationStmt>
  <publisher>Columbia University Press</publisher>
  <pubPlace>New York</pubPlace>
  <date>1993</date>
</publicationStmt>
```

This may be tagged in MARC as:

```
260 $aNew York :$bColumbia University Press, $c1993.
```

Local practice will determine appropriate MARC fields for **<address>**, **<idno>**, and **<availability>**. Restrictions on access should normally be placed in the 506 field, while the place where an item may be ordered will be located in a local notes (590) field. If local practice warrants it, the address of the publisher should be indicated in the 260 field.

The series **<title>** and the **<idno>** should be placed in the appropriate 490 fields (series untraced), if series authority checking needs to be done. Further, because the TEI tags do not differentiate between name, conference, or title series, there is no simple mechanical method for determining which MARC tag (410, 411, etc.) should be used. Safe practice would be to load any series statements into 490 fields, and then to conduct authority work on those fields.

- **<notesStmt>** These are usually reserved for general notes (500) fields.

The **<sourceDesc>** can be mapped to be a “source of data” note (537 in RLIN MDF format) with the print constant “Transcribed from:” at the beginning of the note. The **<biblStruct>** itself can be mapped onto a 581 field (note on primary publication) using the ISBD format to separate each data element.

The **<scriptStmt>**, **<recordingStmt>**, **<recording>**, **<equipment>**, and **<broadcast>** elements do not easily map on to existing MARC fields, and should be put into a local notes field (590) treating the TEI tag introducing each component as a print constant at the head of the field in order to facilitate future local processing and retrieval. Example:

```
<scriptStmt id=CNN12>
  <bibl><author>CNN Network News</author>
    <title>News Headlines</title>
    <date>12 Jun 1991</date>
  </bibl>
</scriptStmt>
```

This may be tagged in MARC thus:

```
590  <scriptStmt id=CNN12>
     <bibl><author>CNN Network News</author>
     <title>News Headlines</title>
     <date> 12 Jun 1991></date></bibl>
     </scriptStmt>
```

Example:

```
<recordingStmt>
   <recording type=video dur="10 mins">
   <equipment><p>Recorded from FM radio to chrome
       tape</p></equipment>
   <broadcast>
      <bibl>
         <title>Britain's pleasure parade</title>
         <author>BBC Radio 4 FM</author>
         <editor role=interviewer>Robin Day</editor>
         <editor role=interviewee>Margaret Thatcher</editor>
         <series><title>The World Tonight</></series>
         <date>27 Nov 89</date>
       </bibl>
    </broadcast>
    </recording>
</recordingStmt>
```

This can be tagged in MARC as:

```
590 <recordingStmt>
    <recording type=video dur="10 mins">
    <equipment><p>Recorded from FM radio to chrome
         tape</p></equipment><broadcast>
    <bibl><title>Britain's pleasure parade</title>
    <author>BBC Radio 4 FM</author>
    <editor role=interviewer>Robin Day</editor>
    <editor role=interviewee>Margaret Thatcher</editor>
    <series><title>The World Tonight</></series>
    <date>27 Nov 89</date>
    </bibl>
    </broadcast>
    </recording>
    </recordingStmt>
```

26.5 MARC Fields for the Encoding Description

The **<encodingDesc>** element provides useful information documenting the relationship between an electronic text and the source or sources from which it was derived. The **<projectDesc>**, **<samplingDecl>**, **<editorialDecl>**, and **<refsDecl>** elements provide details of decisions and rationales used about the process and purposes of the project, how text was sampled, principles of editorial practice, and how canonical references are constructed. The 567 field (notes on methodology) appears to be the most appropriate for this sort of information, though this field is normally intended for methodologies characterizing the social sciences. Practically, it would be wise to transcribe the **<projectDesc>**, **<editorialDecl>**, **<refsDecl>**, and **<classDecl>** elements directly as one or more 567 fields without intervention, with the element name at the beginning of each field, and any TEI tags left intact. This may facilitate any locally-developed retrieval software.

Example:

```
<encodingDesc>
  <projectDesc><p>Texts were collected to illustrate the
    full range of twentieth-century spoken and written Swedish,
    written by native Swedish authors.</projectDesc>
  <samplingDecl><p>Sample of 2000 words taken from the
    beginning of the text.</p></samplingDecl>
  <editorialDecl>
       <interpretation><p>Errors in transcription controlled
         by using the SUC spell checker, v.2.4</p></interpretation>
  </editorialDecl>
</encodingDesc>
```

This may be tagged in MARC as:

```
567  <projectDesc><p>Texts were collected to illustrate the
      full range of twentieth-century spoken and written
      Swedish, written by native Swedish authors.</p>
567  <samplingDecl><p>Sample of 2000 words taken from the
     beginning of the text.</p>
567  <editorialDecl>
     <interpretation><p>Errors in transcription controlled
      by using the SUC spell checker, v. 2.4</p>
     </interpretation>
     </editorialDecl>
```

26.6 MARC Fields for the Profile Description

The profile description is the most problematical element in the TEI header for librarian cataloguers, because it provides a detailed description of the *non-bibliographic* aspects of the text, specifically the languages and sublanguages used, the situation in which it was produced, and the participants and their setting. This information can be used for retrieval purposes, or in machine-supported analysis of the text. The information can be loaded into a separate “codebook” and referenced by the MARC record. Little guidance can be offered on the appropriate MARC location for the elements that make up the profile description, except to suggest that if a site wants to load the profile description into a MARC record for archival and possibly retrieval purposes, then the contents of the profile description may be mapped into a locally-defined notes field (59X) with its TEI tags intact, as in the examples above.

26.7 MARC fields for the Revision Description

The revision history (**<revisionDesc>**) logs all changes to a machine readable file whether or not these constitute a new edition of the file. Aside from the edition area of the MARC record, there are no MARC fields that deal specifically with changes of this sort. This information might be best included in a "codebook", rather than a MARC record. As before, the simplest way of approaching this problem is to include the material with its TEI tags intact as a locally-defined note (59X) in order to support future local processing.

26.8 Structure of the DTD for Independent Headers

The following document type definition is provided in file *teishd2.dtd* and constitutes the auxiliary DTD for independent headers, as described in this chapter.

```
<!-- 26.8: File teishd2.dtd: Auxiliary DTD for Independent    -->
<!-- Header                                                   -->
<!-- Text Encoding Initiative: Guidelines for Electronic      -->
<!-- Text Encoding and Interchange. DRAFT Version 2. 1992-93. -->

<!-- Copyright (c) 1990, 1992, 1993 ACH, ACL, ALLC.           -->
<!-- Permission to copy in any form is granted for use with   -->
<!-- TEI-aware systems and applications, provided this        -->
<!-- notice is included in all copies.                        -->

<!-- These materials may not be altered; modifications to     -->
<!-- these DTDs should be performed as specified in chapter   -->
<!-- MD of the Guidelines.                                    -->

<!-- These materials subject to revision. Current versions    -->
<!-- available from the Text Encoding Initiative.             -->


<!-- Embed entities for TEI generic identifiers.              -->

<!ENTITY % TEI.elementNames system 'teigis2.ent' >
%TEI.elementNames;

<!-- Embed entities for TEI keywords.                         -->

<!ENTITY % TEI.keywords.ent system 'teikey2.ent' >
%TEI.keywords.ent;

<!-- Define element classes for content models, shared        -->
<!-- attributes for element classes, and global attributes.   -->
<!-- (This all happens within the file teiclas2.ent.)         -->
```

```
<!ENTITY % TEI.elementClasses system 'teiclas2.ent' >
%TEI.elementClasses;

<!-- Now declare the IHS element.                           -->

<!ELEMENT ihs           - O  (teiHeader+)                    >
<!ATTLIST ihs                %a.global;                      >

<!-- Finally, embed the TEI header and core tag sets.       -->

<!ENTITY % TEI.header.dtd system 'teihdr2.dtd' >
%TEI.header.dtd;
<!ENTITY % TEI.core.dtd system 'teicore2.dtd' >
%TEI.core.dtd;
```

The overall structure of a set of independent headers, encoded for interchange as a group, is thus:

```
<!DOCTYPE ihs system 'teishd2.dtd'>
<ihs>
<teiHeader>
  <fileDesc> ...     </fileDesc>
  <encodingDesc> ... </encodingDesc>
  <profileDesc> ...  </profileDesc>
  <revisionDesc> ... </revisionDesc>
</teiHeader>
<teiHeader>
  <fileDesc> ...     </fileDesc>
  <encodingDesc> ... </encodingDesc>
  <profileDesc> ...  </profileDesc>
  <revisionDesc> ... </revisionDesc>
</teiHeader>
<teiHeader> ... </teiHeader>
<!-- ... etc. -->
</ihs>
```

In practice, headers might be stored in separate operating system files, to reduce redundant storage requirements; in this case, the top-level file for a typical document might have the following structure:

```
<!DOCTYPE tei system 'tei2.dtd' [
  <!ENTITY txt01 system 'text01.tei' >
  <!ENTITY hdr01 system 'text01.hdr' >
]>
<tei.2>
&hdr01
&txt01
</tei.2>
```

while that for a set of independent headers might have this structure:

```
<!DOCTYPE ihs system 'teishd2.dtd' [
  <!ENTITY hdr01 system 'text01.hdr' >
  <!ENTITY hdr02 system 'text02.hdr' >
  <!ENTITY hdr03 system 'text03.hdr' >
  <!-- ... etc. -->
]>
<ihs>
&hdr01
&hdr02
&hdr03
<!-- etc. -->
</ihs>
```

Part 7

Alphabetic Reference List of Tags

Class Name: **analysis**

Description: defines additional analytic attributes specifying alternate structural arrangements for the components of a text.

Attributes:

equiv (i.e. equivalent): specifies the identifiers of elements considered structurally equivalent to the one bearing the attribute.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

exclude: specifies the identifiers of elements that are considered excluded from the structure when the element bearing this attribute is selected.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

select: specifies which of several mutually exclusive elements actually occurs in a structure.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:

Discussed in 17, “Additional Tag Set for Simple Analysis,” [in separate fascicle].

Class Name: **certainty**
Description: defines additional attributes which are enabled for all elements when the additional tag set for certainty and uncertainty is selected.
Attributes: [None: global and inherited attributes only.]
Part: additional tag set for.
Member of classes:
DTD file:
Discussed in 19, “CE: Indications of Certainty and Responsibility,” [in separate fascicle].

Class Name: **fs** (i.e. feature-structures)
Description: defines additional attributes for all elements, which are enabled when the additional tag set for feature-structure notation is selected.
Attributes: [None: global and inherited attributes only.]
Remarks: [No attributes declared yet.]
Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:
Discussed in 18, “Additional Tag Set for Feature Structures,” [in separate fascicle].

Class Name: **global**
Description: defines attributes common to all elements in the TEI encoding scheme.
Attributes:

id (i.e. identifier): provides a unique identifier for the element bearing the ID value.
Data type: ID
Value: any valid SGML name (in the reference concrete syntax, a name must begin with a letter and contain only letters, numeric digits, hyphen, and full stop).
Default value: #IMPLIED
Example:

```
<p id=names>Paragraph with the ID <mentioned>names</>....
<p id=dates>Paragraph with the ID <mentioned>dates</>....
```

Remarks: The *id* attribute may be used to specify a canonical reference for an element; see section 6.9, “Reference Systems,” [in separate fascicle].

This attribute is optional.

n (i.e. number, name, etc.): gives a number (or other label) for an element, which is not necessarily unique within the document.
Data type: CDATA
Value: any string of characters; often, but not necessarily, numeric.
Default value: #IMPLIED
Example:

Remarks: The *n* attribute may be used to specify the numbering of chapters, sections, list items, etc.; it may also be used in the specification of a standard reference system for the text.
This attribute is optional.

lang (i.e. language): indicates the language of the element content, usually using a two- or three-letter code from ISO 639.
Data type: IDREF
Value: The value must be the identifier specified for a writing system declaration declared in the TEI header, as described in section “REFID= not defined” target=HD>.
Default value: %INHERITED
Example:

```
<p lang=EN>The only surviving work by <person>Ari</>
(died 1148) is the ten-page <title
lang=IS>&Iacute;slendingab&oacute;k</> (<title
lang=LA>Libellus Islandorum</>), written in the early
twelfth century.
```

Remarks: If no value is specified for *lang*, the *lang* value for the immediately enclosing element is inherited; for this reason, a value should always be specified on the outermost element (**<tei.2>**).
This attribute is optional.

rend (i.e. rendition or presentation): indicates how the element in question was rendered or presented in the source text.
Data type: CDATA
Value: any string of characters; if the typographic rendition of a text is to be systematically recorded, a systematic set of values for the *rend* attribute should be defined.
Default value: #IMPLIED
Example:

Remarks: These Guidelines make no binding recommendations for the values of the *rend* attribute; the characteristics of visual presentation vary too much from text to text and the decision to record or ignore individual characteristics varies too much from project to project. Some potentially useful conventions are noted from time to time at appropriate points in the Guidelines.

This attribute is optional.

Remarks: The global attributes described here are made part of the attribute definition list declaration of each element by including the string "%a.global" in each such declaration. Some global attributes are made available when certain base or additional tag sets are selected; these are incorporated into the global attributes by references to the appropriate parameter entities. When the tag sets in question have not been selected, the parameter entities in question expand to the empty string.
Part: base tag set for common core features.
Member of classes: clterm cllink clana clfs clcert clspoken
DTD file:
Discussed in 3.5, "Global Attributes," [in separate fascicle].

<ihs> (i.e. independent header set)

Description: contains a set of TEI Headers exchanged independently of their associated text or group elements.
Attributes: [None: global and inherited attributes only.]
Example:

```
<ihs>
<teiHeader> ... </teiHeader>
<teiHeader> ... </teiHeader>
<teiHeader> ... </teiHeader>
<!-- ... -->
</ihs>
```

Remarks: This element is required.
Part: auxiliary tag set for TEI headers.
Member of classes:
DTD file:
Data description: Contains an arbitrary sequence of **<teiHeader>** elements only.
May occur within:
May contain: teiHeader
Declarations:

```
<!ELEMENT ihs           - O  (teiHeader+)                      >
<!ATTLIST ihs                %a.global;                        >
```

Discussed in 26, "The Independent Header," on page 3.
This tag is required.

Class Name: **linking**

Description: defines additional attributes for hypertext and other linking, which are enabled for all elements when the additional tag set for linking is selected.
Attributes:

corresp (i.e. correspondents): specifies the identifiers of corresponding elements.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

link (i.e. link): specifies the identifiers of elements that may be linked with this one to form an aggregate.
Data type: IDREFS
Value: one or more valid SGML identifiers for elements in the current document separated by white space.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

partof: specifies the identifier of a **<join>** element in the current document representing an aggregate of which this element is a component part.
Data type: IDREF
Value: a valid SGML identifier attached to a **<join>** element.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

prev (i.e. previous): specifies the identifier of an element which may be linked with the current element to form an aggregate, and which immediately precedes it in the aggregate.
Data type: IDREF
Value: a valid SGML identifier for an element in the current document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

next (i.e. next): specifies the identifier of an element which may be linked with the current element to form an aggregate, and which immediately follows it in the aggregate.
Data type: IDREF
Value: a valid SGML identifier for an element in the current document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Part: additional tag set for analysis and interpretation.
Member of classes:
DTD file:
Discussed in 16, "Segmentation and Alignment," [in separate fascicle].

Class Name: # spoken

Description: defines additional attributes which are enabled for all elements when the base tag set for transcription of spoken texts is selected.
Attributes: [None: global and inherited attributes only.]
Part: additional tag set for spoken materials.
Member of classes:
DTD file:
Discussed in 11, "Base Tag Set for Transcriptions of Spoken Texts," [in separate fascicle].

Entity Name: # TEI.core.dtd (i.e. TEI core)

Description: identifies the file containing element and attribute list declarations for the TEI core elements.
Remarks: This entity is included in all TEI DTDs.
Expansion: system 'teicore2.dtd'
Discussed in 3.2, "Core, Base, and Additional Tag Sets," [in separate fascicle].

Entity Name: # TEI.elementClasses

Description: file containing the parameter entity declarations which define element classes for content models, attributes shared among elements of a class, and global attributes.
Remarks:
Expansion: system 'teiclas2.ent'
Discussed in 3.6, "The TEI2.DTD File," [in separate fascicle].

Discussed in 3.7, “Element Classes,” [in separate fascicle].

Entity Name: **TEI.elementNames**

Description: file containing parameter entity declarations for all generic identifiers of the encoding scheme.

Remarks: The parameter entities in this file all take the same form as the two shown below:

```
<!ENTITY % n.div0 'div0' >
<!ENTITY % n.div1 'div1' >
```

Element and attribute-list declarations in the DTDs refer to the parameter entity *n.div1*, not directly to the generic identifier *div1*. As a result, the declarations will function as desired even if a new generic identifier is substituted (e.g. “caput” for “div1” and “liber” for “div0”:

```
<!ENTITY % n.div0 'liber' >
<!ENTITY % n.div1 'caput' >
```

This allows generic identifiers to be renamed conveniently, to provide names in languages other than English, or to provide shorter names than those documented here. See further chapter 31, “Modifying TEI DTDs,” [in separate fascicle].

Expansion: system 'teigis2.ent'

Discussed in 3.8.2, “Parameter Entities for Element Generic Identifiers,” [in separate fascicle].

Entity Name: **TEI.header.dtd** (i.e. TEI header)

Description: identifies the file containing element and attribute list declarations for the TEI header.

Remarks: This entity is included in all TEI DTDs.

Expansion: system 'teihdr2.dtd'

Discussed in 3.2, “Core, Base, and Additional Tag Sets,” [in separate fascicle].

Entity Name: **TEI.keywords.ent**

Description: file containing the parameter entity declarations for TEI keywords.

Remarks: The keywords defined in this file define non-SGML data types (e.g. ISO dates) and control the selection of base and additional tag sets.
Expansion: system 'teikey2.ent'
Discussed in 3.8.3, “Parameter Entities for TEI Keywords,” [in separate fascicle].

Class Name: # terminology
(i.e. global attributes for terminological data.)
Description: defines attributes for all elements in documents which use the base tag set for terminological data.
Attributes:

group: indicates the group (term and related elements) to which this element should be associated by specifying a string matching the *n* attribute value on an appropriate element.
Data type: CDATA
Value: any string matching the *n* attribute value on the **<term>** element to which the group is attached.
Default value: #IMPLIED
Example:

Remarks: The *group* attribute provides a specialized pointing mechanism for use within **<termEntry>** elements.
This attribute is optional.

grpPtr: indicates the group (term and related elements) to which this element should be associated by specifying its unique identifier, where this is available.
Data type: IDREF
Value: the value specified must match a value supplied as the value for an *id* attribute on some **<term>** element in the current SGML document.
Default value: #IMPLIED
Example:

Remarks: The *group* attribute provides a specialized pointing mechanism for use within **<termEntry>** elements.
This attribute is optional.

depend: indicates the parent element to which this element should be associated by specifying a string matching the *n* attribute value on an appropriate element.
Data type: CDATA
Value: any string matching the *n* attribute value on the element to which the dependent element is attached.
Default value: #IMPLIED
Example:

This attribute is optional.

depPtr: indicates the parent element to which this element should be associated by specifying its unique identifier, where this is available.
Data type: IDREF
Value: the value specified must match a value supplied as the value for an *id* attribute on some **<term>** element in the current SGML document.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Remarks: The attributes shared by this element class are used for linking elements, possibly not adjacent in the record, which are related (e.g. a grammatical annotation and the term it describes). If no attribute is specified, the element is assumed to relate to the most recently specified **<term>** or **<otherForm>** element.
Part: base tag set for terminological data.
Member of classes:
DTD file:
Discussed in 14.2, "Tags for Terminological Data," [in separate fascicle].
Discussed in 14.3.3, "Flat Term Entries Using Group and Depend Attributes," [in separate fascicle].
Discussed in 14.4, "Overall Structure of Terminological Documents," [in separate fascicle].

Usability Report and Comment

If you have comments of any kind on the current draft of the TEI Guidelines, please complete this form and return it to one of the following addresses:

C. M. Sperberg-McQueen, Computer Center (M/C 135), University of Illinois at Chicago, 1940 W. Taylor Room 124, Chicago IL 60612-7352, USA. Bitnet: U35395@UICVM, Internet: U35395@uicvm.uic.edu

Lou Burnard, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, UK. Fax: +44 (865) 273275, Janet: LOU@UK.AC.OXFORD.VAX Internet: LOU@vax.oxford.ac.uk

Further copies of this form may be downloaded from LISTSERV @ UICVM (or listserv @ uicvm.uic.edu). Send a mail message containing the line 'GET TEIURC MEMO TEI-L'

Your name:
Your postal address:

Your e-mail address (if any):
Your occupation:
Your academic background:
Your immediate reactions to the Guidelines (please tick or cross one only):

Relevance to your present concerns or interests: high medium low
Importance to your future research plans: high medium low
Comprehensibility/usability of current draft: high medium low

Detailed technical comment: [Please feel free to comment on any part of the Guidelines or in general. Comments relating to specific parts of the text are most helpful if they include a section or chapter number]

Thank you!

Temporary Index