

**DRAFT Guidelines for the EpiDoc Header<sup>1</sup>**  
**version 0.1**

An EpiDoc text represents an edition of a particular inscription, not the inscription itself. This means that information about the inscription is placed alongside the inscription itself in the body of the document in typed `<div>s`. For this reason, some of the usual metadata functions of the TEI header will be served by typed `<div>s` in the document body. Thus there will tend to be some overlap of the categories of information described in the EpiDoc header and body. The editor may therefore be faced with the question of where to put information, in the header or the body. In general, these questions may be resolved by considering the nature of TEI header elements. There are places where persons and places relevant to the inscription may be described in the header, and it is perfectly reasonable to do so and then to refer to those declarations in the body of the document. The best way to make this decision is to think of the header as data which will be globally available. If a person or place name occurs repeatedly in the document, placing a full record of that entity in the header will allow it to be pointed to by all subsequent references. Since header information is more highly structured, it may also be easy to copy or import header definitions from one document to another.

In cases where there is true overlap, the primary distinction should be understood as between data-oriented and document-oriented (or narrative-oriented) information. Data-oriented information is highly structured and tends to be less readable for humans. Elements in this type of markup tend to have content which is either text (often of a particular data type) or other elements, but not both, while document-oriented markup will tend to mix text and other elements. The TEI header elements typically allow free prose descriptions, but this is not recommended for EpiDoc documents. If a type of information already placed in the header must also be discussed in the body, the representation in the body should be in the form of marked-up prose, while the header

---

<sup>1</sup> For the current version of the EpiDoc Guidelines (0.4), see <http://www.unc.edu/awmc/epidoc/EpiDocGuidelines04.pdf>. This document, once it has been properly formatted and accepted by the EpiDoc Collaborative, will be included within a future version of the Guidelines.

information should be of a more data-oriented variety.

The discussion below follows closely the TEI Guidelines for the document header and is intended to be read in conjunction with that document. Section numbers are taken from the P4 Guidelines. Elements which are not likely to be useful in EpiDoc are marked as deprecated.

## 5.2 The File Description (<fileDesc>)

The File Description is the only mandatory header element in the TEI Guidelines. In EpiDoc, however, a minimal header should also contain an Encoding Description (5.3) containing a <classDecl> which declares the division types to be used in the document, and a Profile Description (5.4) which declares the languages used in the document and their encodings.

### 5.2.1 <titleStmt>

In Epidocs, the <title> element will frequently be simply a number or identifying string associated with the inscription. If this type of notation is used, the collection to which the inscription belongs should be noted in the <seriesStmt>.

The usage of other titleStmt child elements (<author>, <sponsor>, <funder>, <principal>, <respStmt>, <resp>, and <name>) in EpiDoc follows the discussion in the TEI Guidelines.

### 5.2.2 <editionStmt>

The Edition Statement is optional but recommended, following the TEI Guidelines. Its child elements are <edition> and <respStmt>, which in turn contains a <name> (the person responsible) and a <resp> element (indicating the nature of the responsibility).

### 5.2.3 <extent>

The extent element is repeatable and may contain statements about the size of the document. We recommend that file size be given in kilobytes (Kb), rather than bytes, since the precise size of the file may be changed by factors such as added whitespace and header information.

### 5.2.4 <publicationStmt>

As in standard TEI, the Publication Statement is mandatory and should contain information about the entity responsible for the electronic publication of the document. The Publication Statement should contain sufficient information for anyone reading an EpiDoc document to contact the organization or person named therein. “Publication” in this discussion should be understood broadly as any distribution of the document beyond the environment in which it was created so that the publisher is no longer in direct control of the document. Publication can include the placement of the document on a web site, emailing the document to anyone with permission to distribute it further, or any other action which might result in the document's dissemination.

The child elements of <publicationStmt> are <publisher>, <distributor>, <authority>, <pubPlace>, <address>, <idno>, <availability>, and <date>. The EpiDoc guidelines follow the TEI recommendations.

#### 5.2.5 <seriesStmt>

The Series Statement should be used for individual EpiDocs that are part of a larger collection, especially when the edition title is merely a numerical or string identifier within that collection. The child elements are <title>, <idno>, and <respStmt>. Their usage in EpiDoc follows the TEI Guidelines.

#### 5.2.6 <notesStmt>

The Notes Statement may contain one or more <note> elements which may be used to record additional pieces of information which are important but which appear nowhere else in the document.

#### 5.2.7 <sourceDesc>

The Source Description is a mandatory element and should be used for bibliographic citation(s) of the source or sources from which the EpiDoc is derived. The Source Description's content should not overlap with the contents of the bibliography <div> in the body. If there is a need to refer to the source in the bibliography, this should be handled by cross-references.

### 5.3 The Encoding Description <encodingDesc>

The Encoding Description deals with “the methods and editorial principles which governed the transcription or encoding of the text in hand.”<sup>2</sup> Not all of the available features of the Encoding Description are discussed in detail here, but future versions of this document may deal with them. In general, elements in this section should be used when the practices of a particular project differ from those outlined in the EpiDoc guidelines.

#### 5.3.1 <projectDesc>

Prose description (in <p> tags) of the project responsible for the creation of the document. This will typically be used more often in corpora than individual EpiDocs.

#### 5.3.2 <samplingDecl>

Unlikely to be used in EpiDoc texts.

#### 5.3.3 <editorialDecl>

Editorial practices for the publication of inscriptions are highly regularized, but in cases where those practices deviate from the norm, they should be noted here. The Editorial Declaration may contain a prose description or one or more of the elements discussed below.

<correction> should not be needed unless there is some sort of important variation from common practice in supplying missing text or expanding abbreviations. In general, none of these actions will be performed silently.

<normalization> should not normally be used. Typically epigraphic texts will not be normalized.

<quotation> is deprecated.

<hyphenation> is deprecated.

<segmentation> should be used when <seg> elements are used to describe fragmentary inscriptions and should contain a description of how segments are marked up.

<stdVals> should be used if date values are given according to a system other than the proleptic Julian Calendar, which is the standard

---

2 <http://www.tei-c.org/P4X/HD.html#HD5>.

method for recording ancient dates, or the Gregorian calendar for modern dates.

`<interpretation>` Typically, interpretation will be provided in the form of commentary in a typed `<div>`, but this is standard practice and so no discussion of it will generally be necessary.

#### 5.3.4 `<tagsDecl>`

A Tag Declaration need only be used when the usage pattern differs from the recommendations specified in these guidelines.

#### 5.3.5 `<refsDecl>`

A Reference System Declaration won't typically be needed for a single EpiDoc, given the small size of most inscriptions, and the fact that most of them are not naturally subdivided, though it might be useful for a corpus.

#### 5.3.6 `<classDecl>`

The Classification Declaration's uses in EpiDoc differ somewhat from those recommended in the TEI Guidelines. The `<classDecl>` in EpiDoc is used to group together definitions or sources for descriptive classification schemes used in the document.<sup>3</sup> `<classDecl>` contains one or more `<taxonomy>` elements, which in turn contain either a bibliographic citation or one or more `<category>` elements. The primary use of the Classification Declaration in EpiDoc is to describe the structure of the system of typed `<div>`s. It achieves this by means of a `<taxonomy>` that declares the division types and subtypes employed in the EpiDoc document.

The division types and subtypes define categories of information represented in the document.<sup>4</sup> Subtypes may appear in the "n" attribute of a typed `div`, or on the "type" attribute of the `rs` element. The value of the attribute in either case should be set to the id of the category element and the value of the `<catDesc>` may be used in representations derived from the document as (e.g.) headers for the information inside the `<div>` or `<rs>` elements. Taxonomic categories for types of inscription (to be referenced in

---

<sup>3</sup> In the TEI Guidelines, the contents of the `<classDecl>` apply only to other elements in the document header. In EpiDoc, they may be referred to by elements in the body also.

<sup>4</sup> EpiDoc Guidelines, pp. 9-12.

the <textClass> element of the Profile Description (see below) may also be defined here.

The division taxonomy for an EpiDoc that uses all of the basic division types<sup>5</sup> with description subtypes of “monument,” “text,” and “date” might look like this:

```
<taxonomy id="divs">
  <category id="description">
    <catDesc>Description</catDesc>
    <category id="monument">
      <catDesc>Description of the Monument</catDesc>
    </category>
    <category id="text">
      <catDesc>Description of the Text</catDesc>
    </category>
    <category id="date">
      <catDesc>Date</catDesc>
    </category>
  </category>
  <category id="edition">
    <catDesc>Edition</catDesc>
  </category>
  <category id="translation">
    <catDesc>Translation</catDesc>
  </category>
  <category id="commentary">
    <catDesc>Commentary</catDesc>
  </category>
  <category id="history">
    <catDesc>History of the Inscription</catDesc>
  </category>
  <category id="bibliography">
    <catDesc>Bibliography</catDesc>
  </category>
</taxonomy>
```

Projects using EpiDoc can use this method to declare variations in the division taxonomy. The associated descriptions may be used as section headers when the document is rendered.

A third use of the Classification Declaration in EpiDoc is as a basis for linking to external registries which contain further information about persons or places mentioned in the document. Further discussion of this usage may be found below in the “Proposal for Linking to External Registries” appendix.

### 5.3.7 <fsdDecl>

The use of Feature Systems is outside the scope of the current

---

5 EpiDoc Guidelines, p. 10.

guidelines, but discussion of their use may appear in future versions.

#### 5.3.8 <metDecl>

The Metrical Declaration may be used to allow metrical analysis to be performed on the text in cases where the inscription is metrical in nature. Its usage should follow the TEI Guidelines.<sup>6</sup>

#### 5.3.9 <variantEncoding>

Since EpiDoc will use XPath notation for links and cross-references, typically a location-referenced apparatus (if there is one) will be used. The apparatus should be placed in its own div, therefore the location attribute should be set to “external” if the <variantEncoding> header is used.

---

<sup>6</sup> See also section 9.4 of the TEI Guidelines (<http://www.tei-c.org/P4X/VE.html#VEME>).

## 5.4 The Profile Description <profileDesc>

The Profile Description is mandatory in EpiDoc headers, since the languages used in the document must be specified in the <langUsage> element. There are also six optional elements that may appear.

### 5.4.1 <creation>

Since the primary focus of an EpiDoc is the *edition* of an inscription, rather than just the inscription itself, the creation element, if it is used, refers to the creation of the edition, not the inscription itself. Information about the creation of the inscription should be placed in a <div> in the body, of type “description”.

### 5.4.2 <langUsage>

The Language Usage element is mandatory in EpiDoc. It will contain a <language> element for each language used in the text. The id attribute on the <language> element should be drawn from the three-letter codes specified by ISO 639-2.<sup>7</sup> The usage attribute is optional, as it is unlikely to be a very useful metric for EpiDocs. The <language> element in the TEI Guidelines also allows a “wsd” attribute which indicates an unparsed entity referring to a Writing System Declaration.<sup>8</sup> We do not recommend using the wsd attribute. Instead the EpiDoc-specific “encoding” attribute should be applied to <language> tags specifying languages that use an encoding scheme different than the encoding of the document (that is, the value of the encoding attribute on the XML declaration at the top of the document). For most documents, we recommend the use of Unicode, usually UTF-8, which is the default encoding for XML documents. But, for example, when Greek texts are encoded using Beta Code, the value of the encoding attribute should indicate this usage (i.e. <language id="grc" encoding="BetaCode" />).

If a proprietary encoding for Ancient Greek (such as a GreekKeys-compatible font) is used for Greek sections in the document, the document encoding must be compatible with this encoding. GreekKeys, for example,

---

<sup>7</sup> See <http://lcweb.loc.gov/standards/iso639-2/langcodes.html>.

<sup>8</sup> See <http://www.tei-c.org/P4X/WD.html>.

uses almost all of the 256 codepoints available to handle characters with diacriticals. This makes it, strictly-speaking, incompatible with Unicode and ISO-8859-x, which use codepoints 128 to 159 as control characters for dumb terminals. Because characters in this range will be understood as control characters by any parsing software reading the XML document as Unicode (which, again, is the default), they may be garbled during parsing.<sup>9</sup> Therefore, documents which use GreekKeys-compatible fonts, or similar encodings to record Greek must set the document encoding to Windows Cp1252 or a similar encoding. Beta Code, because it uses only the lower 127 characters of ASCII, is fully compatible with both UTF-8 and ISO-8859-x encodings. Because of these issues, we strongly recommend that legacy methods of encoding Greek other than Beta Code not be used in EpiDoc documents.

#### 5.4.3 <textClass>

The Text Classification, if used, should follow the documentation in the TEI Guidelines.

The following four elements (<textDesc>, <particDesc>, <settingDesc>, and <handList>) are optional, and may duplicate information in the body of the document. They may be used for purposes of automatic categorization of documents or other types of analysis. All four are extensions to the Profile Description and are discussed in other sections of the TEI Guidelines. Their section numbers in this document correspond to the sections in the TEI Guidelines where they are discussed.

#### 23.2.1 <textDesc>

The Text Description may be used to categorize the text according to various situational parameters. It provides an alternative to the classificatory mechanisms described in the Text Classification and the Classification Declaration. Editors should follow the TEI Guidelines for its use.

The particDesc and settingDesc elements are very likely to overlap with information presented in the body. It is therefore recommended that if these elements are used in the header, their id attributes be referenced via a cross-reference from an element

---

<sup>9</sup> This problem is language-dependent. Java, for example, will garble characters in this range if they are read as Unicode, but other languages do not.

of the names class when the persons or places they refer to appear in the document body. It may be desirable, especially in the case of corpora, to define persons or places related to the inscription once in the corpus header, and thereafter refer to them when they appear in the text.

#### 23.2.2 <particDesc>

The Participants Description may be used to provide information about persons mentioned in the text, if that information is not provided in the text itself.

#### 23.2.3 <settingDesc>

The Setting Description may be used to provide information about the setting of the inscription (findspot, original location, current location, etc.) if that information is not provided in the text itself.

#### 18.2.1 <handList>

The <handList> element too is likely to overlap somewhat with information presented in the discussion of the inscription in the body of the document. As is the case with the <particDesc> and <settingDesc> elements, its contents (two or more hand elements) may be referred to via cross references when those hands are discussed in the description or the commentary.

## **5.5 The Revision Description <revisionDesc>**

The Revision Description is optional in EpiDoc, as in regular TEI, but it is a useful and recommended feature. The <revisionDesc> element may consist of one or more <change> elements, which comprise a log of the changes made to the document during its development. Maintenance of the Revision Description during development will require some self-discipline, since each change made to the document should be documented as a <change>. Projects that do not wish to do this may still benefit from the use of versioning systems, such as CVS or SourceSafe, which provide a framework for documenting changes and preserving older document versions.

## A minimal EpiDoc header

At the bare minimum, an EpiDoc header should contain a <fileDesc> and a <profileDesc> with a <langUsage> child. An <encodingDesc> element with at least a <classDecl> describing the logical organization of the <div> types used and references to any external registries or classificatory systems is also highly recommended. The example below is fuller than the bare minimum and includes the <encodingDesc> element. The <taxonomy> therein follows the divisions and subdivisions used in the EPAPP project.<sup>10</sup>

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>13</title>
      <editor>Charlotte Roueché</editor>
      <respStmt>
        <resp>converted to XML by</resp>
        <name>Gabriel Bodard</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <distributor>The Centre for Computing in the Humanities,
King's College</distributor>
      <pubPlace>London</pubPlace>
    </publicationStmt>
    <seriesStmt>
      <title>Electronic Aphrodisias in Late Antiquity</title>
    </seriesStmt>
    <sourceDesc>
      <bibl>Roueché, Charlotte, Aphrodisias in Late Antiquity,
Society for the Promotion of Roman Studies, London, 1989</bibl>
    </sourceDesc>
  </fileDesc>
  <encodingDesc>
    <taxonomy id="divs">
      <category id="description">
        <catDesc>Description</catDesc>
        <category id="monument">
          <catDesc>Description of the Monument</catDesc>
        </category>
        <category id="text">
          <catDesc>Description of the Text</catDesc>
        </category>
        <category id="letters">
          <catDesc>Description of the Letters</catDesc>
        </category>
        <category id="date">
          <catDesc>Date</catDesc>
        </category>
      </category>
      <category id="edition">
        <catDesc>Edition</catDesc>
        <category id="text">
```

---

<sup>10</sup> <http://www.kcl.ac.uk/humanities/cch/epapp/>.

```

        <catDesc>Text</catDesc>
    </category>
    <category id="text-constituted-from">
        <catDesc>Text Constituted From</catDesc>
    </category>
</category>
<category id="translation">
    <catDesc>Translation</catDesc>
</category>
<category id="commentary">
    <catDesc>Commentary</catDesc>
</category>
<category id="history">
    <catDesc>History of the Inscription</catDesc>
    <category id="found">
        <catDesc>Found</catDesc>
    </category>
    <category id="original-location">
        <catDesc>Original Location</catDesc>
    </category>
    <category id="last-recorded-location">
        <catDesc>Last Recorded Location</catDesc>
    </category>
    <category id="record">
        <catDesc>History of Recording</catDesc>
    </category>
</category>
<category id="bibliography">
    <catDesc>Bibliography</catDesc>
</category>
<category id="metadata">
    <catDesc>Metadata</catDesc>
    <category id="category-text">
        <catDesc>Category of Text</catDesc>
    </category>
    <category id="category-monument">
        <catDesc>Category of Monument</catDesc>
    </category>
</category>
</taxonomy>
</encodingDesc>
<profileDesc>
    <langUsage>
        <language id="eng">English</language>
        <language id="fre">French</language>
        <language id="grc">Ancient Greek</language>
        <language id="lat">Latin</language> </langUsage>
    </profileDesc>
</teiHeader>

```

## Appendix: A Proposal for Linking EpiDoc Texts to External Registries

One of the basic information needs of EpiDoc texts is a way to link to external data registries. These registries contain canonical records pertaining to categories of data like persons and places. This essay will sketch out a method for producing preliminary linkages to external data registries using RDF (the Resource Description Framework).<sup>11</sup> The construction of these linkages will make it possible to tie features of individual EpiDocs to external registries even if those registries are not yet digitized. It should be noted that RDF in this proposal could be replaced by another linking /description mechanism, such as a database.

There are several reasons why such linking mechanisms are a desideratum. One major reason is disambiguation: it will be important to identify which of the dozen-or-so Alexandrias one is referring to in the markup, for example. A second important reason is the ability to refer to additional information concerning a person or place which is out of scope for the current document. For example, given a reference to an individual, it would be nice to be able to link to external information about that individual, such as data about his career, other places where he is mentioned, etc. For various reasons, it is impractical to embed all of this information in any document that mentions the person, so an automated link to the external information is a better solution.

There is little scope at the present time for linking directly to person and placename registries because these resources are not yet available in digital form for the most part (though placenames are in a much better state, digitally speaking, than persons).<sup>12</sup> There are, however, various standard printed works which supply catalogs of people and places. The goal then, is a standard way of referring to individual references in these works and, if they are in digital form, to enable querying of those resources. RDF(S) seems to provide just such functionality. An RDF application that supports this functionality would be capable of providing unique identifiers to refer to catalogs and catalog entries. The great advantage of RDF is that each resource can be referred to as a URI. This suggests a method for easily embedding references to external resources (even non-digital ones) in EpiDoc documents: all that will be necessary in order properly to

---

<sup>11</sup> <http://www.w3.org/RDF/>.

<sup>12</sup> See, for example, the Getty Thesaurus of Geographic Names (<http://www.getty.edu/research/tools/vocabulary/tgn/>).

identify a person, place, etc. will be to attach to the reference the URI of its registry and the unique identifier that registry applies to the data item. TEI provides a 'key' attribute on the "names" class of elements, which may serve as the identifier for the catalog entry.<sup>13</sup> Available registries can be described in an online, queryable format, such as a database, or in this example, an RDF document.<sup>14</sup> The registry can then be referenced in a <persName> tag:

```
<persName key="PIR2:A 0008">L. Aelius Aurelius Septimius Abgarus</persName>.
```

The key attribute should be composed of the id of the registry from the classDecl header (see below), followed by a colon, and then an identifying string drawn from the registry itself.

Two additional elements are still needed in order to enable automatic linking: the URI of the registry needs to be referenced in the document somewhere, and (if electronic querying of the registry is possible) a method for accessing the entry itself must be described in the RDF document. The <classDecl> header can be used to construct a TEI version of the reference to the registry:<sup>15</sup>

```
<classDecl>
  <taxonomy id="LGPN">
    <bibl>
      <editor>P.M. Fraser and E. Matthews</editor>
      <title>A Lexicon of Greek Personal Names</title>
      <!-- etc. -->
    </bibl>
    <xref
url="http://www.unc.edu/awmc/epidoc/RDF/personNameRegistry.rdf#">
  </taxonomy>
</classDecl>
```

The second requirement can be satisfied by the addition of some more elements to the RDF document. The following examples will enable a functioning link to a registry entry to be constructed from a place name reference in an EpiDoc file:

#### 1. The registry listing:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:epidoc="http://www.epidoc.org/RDF/epidoc.rdfs#"
xmlns:teiBibl="http://www.epidoc.org/RDF/BiblStruct.rdfs#">
  <rdf:Description rdf:ID="TGN">
    <teiBibl:title>Getty Thesaurus of Geographic
```

---

13 See <http://www.tei-c.org/P4X/ref-NAMES.html>.

14 A draft example of such a document may be found at:  
<http://www.unc.edu/projects/awmc/epidoc/RDF/personNameRegistry.rdf>.

15 Note that this will require the extension of the TEI DTD to allow an "xref" element inside the "taxonomy" element.

```

Names</teiBibl:title>
  <epidoc:website>http://www.getty.edu/research/tools/vocabulary/tgn/
</epidoc:website>
  <epidoc:queryBase>http://vocab.pub.getty.edu/cgi-
bin/tgn_browser/tgn.spl?searchtype=record&file=/tgn_browser/index.html&key=
</epidoc:queryBase>
</rdf:Description>
</rdf:RDF>

```

## 2. The classDecl from the TEI header:

```

<classDecl>
  <taxonomy id="TGN">
    <bibl>
      <title>Getty Thesaurus of Geographic Names</title>
      <xref
url="http://www.unc.edu/awmc/epidoc/RDF/placeNameRegistry.rdf#"/>
      <xref
url="http://www.getty.edu/research/tools/vocabulary/tgn/"
n="website"/>
      <xref url=http://vocab.pub.getty.edu/cgi-
bin/tgn_browser/tgn.spl?searchtype=record&file=/tgn_browser/index.html&key=*
n="queryBase"/>
    </bibl>
  </taxonomy>
</classDecl>

```

## 3. The place name reference.

```
<placeName key="TGN:7000874">Rome</placeName>
```

Given this information, the data necessary to query the online TGN database can be assembled by replacing the wildcard character ("\*") in the URL of the xref element with n="queryBase" in the appropriate classDecl with the key component following the colon. This will produce the string "http://vocab.pub.getty.edu/cgi-bin/tgn\_browser/tgn.spl?searchtype=record&file=/tgn\_browser/index.html&key=7000874", which when dereferenced returns information on the place in question. If the registry has a website, but not an online querying mechanism, a similar query could be formulated to return the site URI, or whatever bibliographic information is contained in the RDF registry. The means by which this information is to be rendered for a user viewing an EpiDoc text will depend on the capabilities of the server delivering the document and/or those of the client application viewing the document. Having the linking mechanism described in the external registry means that if it is changed, it should be possible to update the EpiDocs that utilize that registry.

The mechanism outlined in this proposal should allow for linking between references to data such as personal and place names in EpiDoc texts and information about the registries in which those data are cataloged. There are still a number of issues

to be clarified, such as what namespaces should be used in these registries and where the registry files should reside. If these issues can be resolved, this method will provide the creators of EpiDoc texts with a means to identify data items according to their representation in standard reference works and simultaneously to link to those reference works if they have an online presence.