**(Last updated: 12/13/2024)**

**1. Terminology/Core Concepts**

* TEI:
  + Definition: Text encoding initiative
  + The TEI guidelines are expressed in XML.
  + The TEI may be used as an interchange format, permitting projects to share resources even when their local encoding schemes differ. If there are n encoding formats, to provide mappings between each possible pair requires n X (n-1) translations. With an interchange format, only 2 X n such mappings are needed.
    - How does this work? Here might be the idea: Suppose that we have the following four formats: A, B, C, and D. To map each of them to each other one, we would need 12 (4 X (4-1)) mappings. That is, we would need the following mappings: (1) A -> B, (2) A -> C, (3) A -> D, (4) B -> A, (5) B -> C, (6) B -> D, (7) C -> A, (8) C -> B, (9) C -> D, (10) D -> A, (11) D -> B, and (12) D -> C. However, if we use an interchange format, we only need to map each format to the interchange format, and then the interchange format to each format. So, in this case we only need 8 (2 X 4) mappings. That is, we only need the following mappings: (1) A -> I (interchange format), (2) B -> I, (3) C -> I, (4) D -> I, (5) I -> A, (6) I -> B, (7) I -> C, and (8) I -> D.
* XML:
  + Definition: Extensible markup language.
  + Technically, XML is a meta-language, which is used to describe markup languages.
* Markup/encoding:
  + Definition #1: The process of inserting explicit markers for implicit textual features.
  + Definition #2: Annotation or other marks within a text intended to instruct a compositor or typist how a particular passage should be printed or laid out.
  + Definition #3: Any means of making explicit an interpretation of a text.
* Markup language/encoding scheme:
  + Definition #1: The complete set of rules associated with the use of markup in a given context.
  + Definition #2: A set of markup conventions used together for encoding texts.
  + Note on how XML and TEI work together: A markup language must specify (1) how markup is to be distinguished from text, (2) what markup is allowed, (3) what markup is required, and (4) what the markup means. XML does 1-3, while the TEI guidelines do 4.
* Markup vocabulary:
  + Definition: The specific set of markers or named distinctions employed by a given encoding scheme.
* Extensible:
  + Definition: Does not consist of a fixed set of tags.
* Element:
  + Definition: The technical term used in XML for a textual unit, viewed as a structural component.
  + Within a marked-up text, each element must be explicitly marked or tagged in some way. This is done by inserting a tag at the beginning of the element (a start-tag) and another at its end (an end-tag).
  + The material between the start-tag and the end-tag is known as the content of the element.
* Root-element:
  + Definition: A single element that encloses the whole document.
* Attribute:
  + Definition: Information that is in some sense descriptive of a specific element but not regarded as part of its content.
* Namespace:
  + Definition: Something that allows us to represent the fact that a name belongs to a group of names.
  + If the markup of a document is thought of as an expression in some language, then a namespace may be thought of as analogous to the lexicon of that language.
  + An XML document need not specify any namespace; it is then said to use the “null” namespace. Alternatively, the root element of the document may supply a default namespace, understood to apply to all elements which have no namespace prefix.
* Validation:
  + Definition: The performance of a check to ensure that the rules that apply to a document—e.g., that certain pieces of text should be given certain tags—have been followed.
* Schema:
  + Definition: A document that formally states the criteria for successful validation.
  + Schema validation for XML is usually written in the RELAX NG language.
  + The difference between a schema and a namespace: A schema is a set of element definitions, whereas a namespace is really only a property of a collection of elements. The only tangible form a namespace takes in an XML document is its distinctive prefix and the identifying name associated with it.
  + A TEI schema can be specified using the schemaSpec element.
* ODD:
  + Definition: The specification for a TEI schema, which is itself a TEI document. Stands for “One Document Does it all.”
* Glyph:
  + Definition: A visual symbol on a page.
* Abstract symbol:
  + Definition: A letter etc. that a visual symbol refers to.
* Glyph:
  + Definition: A visual symbol on a page.
* Characters:
  + Definition: The smaller units that scripts are composed of.
* Character set:
  + Definition: A collection of abstract characters.
  + There are two distinct types of character sets:
    - Character sets of documents: Pure collections of abstract characters.
    - Character sets of computing devices: Sets of abstract characters which have been mapped in a well-defined way to a set of numbers or code points by which the device represents those abstract characters internally. IOW: A coded character set.
  + Unicode:
    - Definition: A coded character set, devised and actively maintained by an international public body, where each abstract character is identified by a unique name and assigned a distinctive code point.
    - XML was designed with Unicode in mind as its way of representing abstract characters. It is possible to use other character encoding schemes with XML, but in general they are best avoided, as you run the risk of encountering compatibility issues with different XML processors, as well as potential difficulties with rendering their output.
    - Normalization: The process of ensuring that a given abstract character is represented in one way only in a given Unicode document or document collection.
* Prose:
  + Definition: A series of paragraphs, each marked as a p element.
* Phrase:
  + Definition: Implies character data, interspersed as need be with phrase-level elements, but not organized into paragraphs.
* Module:
  + Definition: A convenient way of grouping together a number of associated element declarations.
  + The relation between modules and a TEI schema: In a simple case, a TEI schema is made by combining together a small number of modules.
* Class:
  + Definition: Used to express commonality among elements.
  + There are two types of classes:
    - Attribute class:
      * Definition: A class composed of members that share attributes.
      * Attribute classes are given names composed of the prefix att., often followed by an adjective.
    - Modal class:
      * Definition: A class composed of members that appear in the same locations.
  + An element is said to inherit properties from any class of which it is a member.
  + Nested classes: Classes can be nested within each other. If A is a member of B and A and B are both classes, then A is a subclass of B and B is a superclass of A. Also, if any element which is a member of A will inherit the properties defined by both A and B.
* Element classification:
  + Definition: Describes the structural level of an element.
  + Three element classifications:
    - Chunk-level:
      * Definition: Paragraphs and other elements which have similar structural properties.
    - Phrase-level:
      * Definition: Elements which must be entirely contained within a paragraph or other chunk-level structure.
      * Includes quoted phrases, names, dates, etc.
    - Inter-level:
      * Definition: Elements which can appear either within a paragraph or between paragraphs.
      * Includes bibliographic citations, notes, and lists.

**2. The TEI Header**

* Definition: A set of descriptions and declarations that provides an electronic analogue to the title page attached to a printed work.
* Tag: <teiHeader>
* The five major parts of the TEI header:
  1. File description:
     + Tag: <fileDesc>
     + Definition: This element is intended to contain a full bibliographic description of the computer file itself, from which a user of the text could derive a proper bibliographic citation, or which a librarian or archivist could use in creating a catalogue entry recording its presence within a library or archive. (The term computer file here is to be understood as referring to the whole entity or document described by the header, even when this is stored in several distinct operating system files).
     + This is the only element required in all TEI headers.
     + The seven elements within the the fileDesc element, listed in the order in which they must be given:
       1. Title statement:
          - Tag: <titleStmt>
          - Definition: Groups information about the title of a work and those responsible for its content.
          - Status: Mandatory.
          - Title element:

Tag: <title>

Definition: Groups information about the title of a work and those responsible for its content.

Where the electronic work is derived from an existing source text, it is strongly recommended that the title for the former should be derived from the latter, but clearly distinguishable from it, for example by the addition of a phrase such as “an electronic transcription” or “a digital edition.” This will distinguish the electronic work from the source text in citations and in catalogues with contain both types of material.

* + - * + Author element:

Tag: <author>

Definition: In a bibliographic reference, contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

* + - * + Editor element:

Tag: <editor>

Definition: Contains a secondary statement of responsibility for a bibliographic item, for example the name of an individual, institution, or organization, (or of several such) acting as editor, compiler, translator, etc.

* + - * + Sponsor element:

Tag: <sponsor>

Definition: Specifies the name of a sponsoring organization or institution.

* + - * + Funding body element:

Tag: <funder>

Definition: Specifies the name of an individual, institution, or organization responsible for the funding of a project or text.

* + - * + Principal researcher element:

Tag: <principal>

Definition: Supplies the name of the principal researcher responsible for the creation of an electronic text.

* + - * + Responsibility statement:

Tag: <respStmt>

Definition: Supplies a statement of responsibility for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply. May also be used to encode information about individuals or organizations which have played a role in the production or distribution of a bibliographic work.

Responsibility element:

Tag: <resp>

Definition: Contains a phrase describing the nature of a person’s intellectual responsibility, or an organization’s role in the production or distribution of a work.

Name element:

Tag: <name>

Definition: Contains a proper noun or noun phrase.

* + - 1. Edition statement:
         * Tag: <editionStmt>
         * Definition: Groups information relating to one edition of a text.
         * Status: Optional.
      2. Extent:
         * Tag: <extent>
         * Definition: Describes the approximate size of a text stored on some carrier medium or of some other object, digital or non-digital, specified in any convenient units.
         * Status: Optional.
      3. Publication statement:
         * Tag: <publicationStmt>
         * Definition: Groups information concerning the publication or distribution of an electronic or other text.
         * Status: Mandatory.
      4. Series statement:
         * Tag: <seriesStmt>
         * Definition: Groups information about the series, if any, to which a publication belongs.
         * Status: Optional.
      5. Notes statement:
         * Tag: <notesStmt>
         * Definition: Collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.
         * Status: Optional.
      6. Source description:
         * Tag: <sourceDesc>
         * Definition: Describes the source from which an electronic text was derived or generated, typically a bibliographic description in the case of a digitized text, or a phrase such as “born digital” for a text that has no previous existence.
         * Status: Mandatory.
  1. Encoding description:
     + Tag: <encodingDesc>
     + Definition: Describes the relationship between an electronic text and its source or sources.
  2. Text profile:
     + Tag: <profileDesc>
     + Definition: Contains classificatory and contextual information about the text, such as its subject matter, the situation in which it was produced, the individuals described by or participating in producing it, and so forth.
  3. Container element:
     + Tag: <xenoData>
     + Definition: Allows easy inclusion of metadata from non-TEI schemes (i.e., other than elements in the TEI namespace).
  4. Revision history:
     + Tag: <revisionDesc>
     + Definition: Allows the encoder to provide a history of changes made during the development of the electronic text.
* Types of content in the TEI header:
  + Free prose.
  + Grouping elements: Elements whose names end with the suffix Stmt and the xenoData element.
  + Declaration elements: Elements whose names end with the suffix Decl. These enclose information about specific encoding practices applied in the electronic text.
  + Descriptions: Elements whose names end with the suffix Desc. These contain a prose description.

**3. Relevant Elements**

* Paragraph:
  + Tag: <p>
  + Definition: A mid-level unit of sense, a coherent grouping of sentences.
* Line beginning:
  + Tag: <lb>
  + Definition: Starts a new line.
* Page beginning:
  + Tag: <pb>
  + Definition: Starts a new page.
* Highlighted:
  + Tag: <hi>
  + Definition: Marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.
* Foreign:
  + Tag: <foreign>
  + Definition: Identifies a word or phrase as belonging to some language other than that of the surrounding text.
  + This element should not be used to represent foreign words which are *mentioned* within the text.
* Emphasized:
  + Tag: <emph>
  + Definition: Marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.
* Distinct:
  + Tag: <distinct>
  + Definition: Identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectical, non-preferred, etc., or as forming part of a sublanguage.
* Quoted:
  + Tag: <q>
  + Definition: Contains material which is distinguished from the surrounding text using quotation marks or a similar method, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used.
* Said:
  + Tag: <said>
  + Definition: Indicates passages thought or spoken aloud, whether explicitly indicated in the source or not, whether directly or indirectly report, whether by real people or fictional characters.
* Quotation:
  + Tag: <quote>
  + Definition: Contains a phrase or passage attributed by the narrator or author to some agency external to the text.
* Cited quotation:
  + Tag: <cit>
  + Definition: Contains a quotation from some other document, together with a bibliographic reference to its source.
  + It is common for quotations to be separated from their bibliographic reference by intervening text. In such cases, the use of cit is impractical; the quotation can instead by linked to a bibliographical reference using the source attribute.
* Mentioned:
  + Tag: <mentioned>
  + Definition: Marks words or phrases mentioned, not used.
* So called:
  + Tag: <soCalled>
  + Definiition: Contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.
* Term:
  + Tag: <term>
  + Definition: Contains a single-word, multi-word, or symbolic designation which is regarded as a technical term.
* Gloss:
  + Tag: <gloss>
  + Definition: Identifies a phrase or word used to provide a gloss or definition for some other word or phrase.
  + A gloss can be linked to the term it is glossing by means of its target attribute. To establish such a link, the encoder should give an xml:id value to the term or mentioned element and provide that value, prefixed by a sharp sign, as the value of the target attribute on the gloss term.
  + In the absence of a gloss in the source text, a term can also be associated with a gloss or definition by means of its ref attribute.
* Sic:
  + Tag: <sic>
  + Definition: Contains text reproduced although apparently incorrect or inaccurate.
* Correction:
  + Tag: <corr>
  + Definition: Contains the correct form of a passage apparently erroneous in the copy text.
* Choice:
  + Tag: <choice>
  + Definition: Groups a number of alternative encodings for the same point in a text.
* Regularization:
  + Tag: <reg>
  + Definition: Contains a reading which has been regularized or normalized in some sense.
* Original form:
  + Tag: <orig>
  + Definition: Contains a reading which is marked as following the original, rather than being normalized or corrected.
* Gap:
  + Tag: <gap>
  + Definition: Indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible.
* Deliberately marked omission:
  + Tag: <ellipsis>
  + Definition: Indicates a purposeful marking in the source document signaling that content has been omitted, and may also supply or describe the omitted content.
  + Within ellipsis, a required metamark is used to represent the characters or markings in the source that indicate the presence of omitted material.
* Unclear:
  + Tag: <unclear>
  + Definition: Contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible.
* Addition:
  + Tag: <add>
  + Definition: Contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.
* Deletion:
  + Tag: <del>
  + Definition: Contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector.
* Referencing string:
  + Tag: <rs>
  + Definition: Contains a general purpose name or referring string.
* Name:
  + Tag: <name>
  + Definition: Contains a proper noun or noun phrase.
* Address:
  + Tag: <address>
  + Definition: Contains a postal address, for example of a publisher, an organization, or an individual.
* Address line:
  + Tag: <addrLine>
  + Definition: Contains one line of a postal address.
* Email:
  + Tag: <email>
  + Definition: Contains an email address identifying a location to which email messages can be delivered.
* Number:
  + Tag: <num>
  + Definition: Contains a number, written in any form.
* Measure:
  + Tag: <measure>
  + Definition: Contains a word or phrase referring to some quantity of an object or commodity, usually comprising a number, a unit, and a commodity name.
* Unit:
  + Tag: <unit>
  + Definition: Contains a symbol, a word or phrase referring to a unit of measurement in any kind of formal or informal system.
* Date:
  + Tag: <date>
  + Definition: Contains a date in any format.
* Time:
  + Tag: <time>
  + Definition: Contains a phrase defining a time of day in any format.

**4. Relevant Attributes**

* xml:lang:
  + Definition: Identifies the language and writing system used.
  + Defined for all TEI elements.
* xml:id:
  + Definition: Provides a unique identifier for the element bearing the attribute.
  + Defined for all TEI elements.
  + The value supplied for this attribute must be a legal name.
* n:
  + Definition: Gives a number (or other label) for an element, which is not necessarily unique within the document.
  + Defined for all TEI elements.
* rend:
  + Definition: Indicates how the element in question was rendered or presented in the source text.
  + Defined for all TEI elements.
* style:
  + Definition: Contains an expression in some formal style definition language which defines the rendering or presentation used for this element in the source text.
  + Defined for all TEI elements.
* rendition:
  + Definition: Points to a description of the rendering or presentation used for this element in the source text.
  + Defined for all TEI elements.
* xml:base:
  + Definition: Provides a base URI reference with which applications can resolve relative URI references into absolute URI references.
  + Defined for all TEI elements.
* xml:space:
  + Definition: Signals an intention about how white space should be managed by applications.
  + Defined for all TEI elements.
* source:
  + Definition: Specifies the source from which some aspect of this element is drawn.
  + Defined for all TEI elements.
* cert:
  + Definition: Signifies the degree of certainty associated with the intervention or interpretation.
  + Defined for all TEI elements.
* resp:
  + Definition: Indicates the agency responsible for the intervention or interpretation (for example, an editor or transcriber).
  + Defined for all TEI elements.
* ref:
  + Definition: Provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URIs.
  + Comes from the att.canonical class.
* key:
  + Definition: Provides an externally-defined means of identifying the entity (or entities) being named, using a coded value of some kind.
  + Comes from the att.canonical class.
  + The value of the key attribute may be an unexpanded code with no particular significance. More usually, however, it will be an externally defined code of some kind, as provided by a standard reference source. The standard reference source should be document, for example using a taxonomy element in the TEI header.

**5. Relevant Modules**

* The tei module:
  + Definition: Defines classes, macros, and datatypes which are used by all other modules.
* The core module:
  + Definition: Contains declarations for elements and attributes which are likely to be needed in almost any kind of document, and is therefore recommended for global use.
* The header module:
  + Definition: Provides metadata elements and attributes constituting the TEI header, a component which is required for TEI conformance.
* The textstructure module:
  + Definition: Declares basic structural elements needed for the encoding of most book-like objects.

**6. Relevant Classes**

* att.global
  + Definition: Provides attributes common to all elements in the TEI encoding scheme.
* model.highlighted
  + Definition: A phrase level class which provides elements used for highlighting.
  + This class is further broken up into the following two classes:
    - model.emphLike:
      * Definition: Used to indicate that a passage is linguistically emphasized for some purpose.
    - model.hiLike:
      * Definition: Used to indicate that a passage is typographically distinct in some way.
* att.canonical:
  + Definition: Provides attributes that can be used to associate a representation such as a name or title with canonical information about the object being named or reference.
* att.measurement:
  + Definition: Provides attributes to represent a regularized or normalized measurement.
* att.ranging:
  + Definition: Provides attributes for describing numerical ranges.
* model.dateLike:
  + Definition: Contains elements used to represent dates and times.