Schenker Documents Online

Encoding Guidelines

```
Revision History
```

```
v. .09.8 September 2008. Author:tlopez.
v. .09.7 August 2008. Author: tlopez.
v. .09.2 June 2008. Author: tlopez.
v. .09 March, April 2008. Author: Eleonora Durban Litta Modignani Picozzi, revision: tlopez
In This Document:

Overview
The XML Schema
Records and Record Collections
```

Editorial Responsibility
Tracking Changes

Hand List

Describing Records

Describing the Record Collection

File Names and Directories

Structural Encoding Specifications

Content Type Structural Requirements

Diary Entries

Sub Entries

Symbols Marking Sub-entries

Translations

Lessonbook Entries

Correspondence

Opening, Closing and Postscripts

Other documents

Page breaks

Footnotes

References Between Records

Editorial Encoding Specifications

Doubtful Readings

Abbreviations

Alterations in original

Deletions (incl. one/two letter deletions)

Corrections

Omitted Text

Supplied Text

Script distinction

Typographical Emphasis

```
Foreign Words

Dates, Times, Temperatures

Notes

Marginal Marks (authorial)

Editorial Notes

Figures

Lists

Semantic Encoding Specifications

Titles

Names, places, institutions, terms
```

Overview

This section contains information about the use of XML in its broader aspects, including an overview of the XML Schema, the concepts of records and record collections, and the structural markup to be employed for diary, lessonbook and correspondence records.

The XML Schema

The XML schema for Schenker Documents Online [SDO] is a composite, drawing on elements from the TEI and Dublin Core that are combined within a custom tagset. To aid the encoder, tags will in every case need to be indicated using the namespace to which they belong:

TEI

Elements are prefixed with 'tei', e.g. <tei:pb>

XML

The attributes of some elements within the TEI, notably **lang** and **id** must use the namespace 'xml', e.g. xml:id="r0001" and xml:lang="eng".

Dublin Core

Elements are prefixed with 'dc' or 'dcterms', e.g. <dc:rights>, <dcterms:provenance>.

Schenker Documents Online

Custom elements designed to support project requirements prefixed with 'sdo', e.g. <sdo:record>

Here is the the basic structure for a Schenker Documents Online file, which shows how namespaces are declared and referred to when writing tags:

```
<sdo:recordCollection
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:sdo="http://www.cch.kcl.ac.uk/schenker"
xmlns:tei="http://www.tei-c.org/ns/1.0" xmlns:xml="http://www.w3.org/XML/1998/namespace"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
xsi:schemaLocation="http://www.cch.kcl.ac.uk/schenker schema/schenker.xsd">
<sdo:collectionDesc>
<dc:description>The collectionDesc element can be used to describe information about all
```

```
records in recordCollection. t will contain a mixture of Dublin Core and TEI elements.</dc:description>
</sdo:collectionDesc>
<sdo:record ID="r0001">
<sdo:record ID="r0001"

<sdo:record ID="r0001"
```

Namespaces in oXygen: to easily insert namespaces in oXygen XML, open a new tag by typing an angle bracket (<) to display a list of available tags and their namespaces in a select box. Using your up and down keys, navigate to the tag you desire and select 'Enter'. The tag will be inserted with the appropriate name space.

Records and Record Collections

Each XML file will contain one or more <sdo:record> elements. The <sdo:record> contains all of the information relating to a single diary entry, lessonbook entry or piece of correspondence and includes the German transcription, the English translation, notes, and descriptive metadata. A record must be given a unique ID within the XML file.

```
<sdo:record ID="r0001"> [...] </sdo:record>
```

record IDs in oXygen: A controlled list of record IDs can be selected in oXygen XML by typing <sdo:record **ID=** and the open quote ("). Upon this action, oXygen XML will display a list of valid record IDs for diary entries. A single XML file should not contain more than 1000 records

Records are collected together into a <sdo:recordCollection>. Each Record Collection (and therefore, each XML file) should contain records that correspond to a single shelfmark derived from the arrangement adopted by the holding institution. The shelfmark for a <sdo:recordCollection> is to be indicated as the first element within the <sdo:recordCollection>'s description:

```
<sdo:recordCollection>
<sdo:collectionDesc>
<sdo:source>
<sdo:diary shelfmark="OJ-01-02"/>
</sdo:source>
[...]
</sdo:collectionDesc>
```

In the above example, all records in this record collection will belong to the Special Collections and Archives Oswald Jonas Memorial Collection at the University of California Riverside and be located in Box 1, Folder 1.

shelfmarks in oXygen: A controlled list of shelfmarks can be selected in oXygen XML by typing shelfmark= and the open quote ("). Upon this action, oXygen XML will display a list of valid shelfmarks for diaries.

See Also: The HTML inventory which lists all archival information for the corpus, located here: http://curlew.cch.kcl.ac.uk/schenker/ref/content/planning/sdo-inventory.html

Describing the Record Collection

In addition to identifying the shelfmark and content type of the records in an XML file, the <sdo:collectionDesc> is used to track administrative information about all of the records it contains.

Hand List

The <sdo:collectionDesc> element is used to capture information about the hands that appear in the records collected in a single XML file. These are to be indicated using the element <tei:handNotes>. Each hand should be entered in it's own <tei:handNote> element, and given a unique ID based on the person's initials. The attribute **scope** may be used to roughly indicate the amount of text in each hand; the value of the note can include more detailed information about the appearance of the hand in the records.

```
<tei:handNotes>
<tei:handNote xml:id="HS" scope="minor">Heinrich Schenker's hand (corrections on pp. 45 and 50)</tei:handNote>
<tei:handNote xml:id="JS" scope="major">Jeanette Schenker's hand (dominant hand and some corrections) </tei:handNote>
</tei:handNotes>
```

Editorial Responsibility

In addition, the <sdo:collectionDesc> is used to identify the editors of the group of records collected in a single XML file. This is to be indicated using the required element <tei:respStmt>. Repeat this element for each person involved in editing; each element should have a <tei:resp> element that describes the role taken, and a <tei:persName element with the name of the person. The <tei:respStmt> should also be given a unique ID, formed using the person's initials.

```
<tei:respStmt xml:id="IB">
<tei:resp>Transcription, Translation, Commentary, Footnotes, and Summary</tei:resp>
<tei:persName > lan Bent</tei:persName>
</tei:respStmt>

<tei:respStmt xml:id="TL">
<tei:resp>XML Encoding</tei:resp>
<tei:persName > Tamara Lopez</tei:persName>
</tei:respStmt>
```

Tracking Changes

Major changes to a <sdo:recordCollection> should be tracked within the <tei:revisionDesc> element. Enter a new <tei:change> element for each major change, indicating in attributes **who** made the change, and **when** it was made. The value of **who** should reference and ID given in a corresponding <tei:respStmt>. Record comments about the work performed in the body of the element.

<tei:revisionDesc>
<tei:change when="2008-06-08" who="#MT">MovableType Migration</tei:change>
<tei:change when="2008-08-08" who="#TL">Hand encoding to demonstrate schema, fix Movable
Type</tei:change>
</tei:revisionDesc>

Describing Records

Descriptive information about individual records should be entered within the <itemDesc> element. This tag groups information about a single record, e.g. a single diary entry, lessonbook entry or piece of correspondence. The <itemDesc> is the first element to appear within an <sdo:record, and contains a series of Dublin Core elements.

<sdo:record ID="r0003"> <sdo:itemDesc>

More To Come: the exact specification of the Dublin Core elements that must, should and may be included to be specified and documented.

</sdo:itemDesc>

File Names and Directories

The filename should correspond to the shelfmark given in the <sdo:collectionDesc>/<sdo:source block.

✓ More To Come: the exact specification for directories should be given, with examples.

Structural Encoding Specifications

Content Type Structural Requirements

As indicated above, the textual components of all content types (diaries, lessonbooks, correspondence and other documents), are to be encoded into separate blocks within the <sdo:record>. The German transcription is to be placed in a <tei:transcription> block; the English translation is to be encoded within a <tei:translation> block. Specific content type requirements follow.

Diary Entries

Individual diary entries are to be encoded in a single <tei:record> with the German appearing in <tei:transcription> and the English translation appearing in <tei:translation>. All tags used in creating a diary entry structure will fall within the TEI namespace.

The main text of an entry is to be encoded within a single <tei:ab> (or anonymous block) element. If the entry

includes date, time of day, temperature or location information at the start of the entry, this should be indicated using the <tei:opener> element. Use a **rend** attribute to indicate placement of the opener, e.g. inline, left-margin.

```
<tei:transcription>
<tei:opener rend="inline">
<tei:date>1. II. 18</tei:date><tei:measure type="temperature" unit="celsius">-2°</measure>,
Glatteis.
</tei:opener>
<tei:ab> [...] </tei:ab>
</tei:transcription>
```

For more information on opener information, see relevant sections in the Editorial Encoding Sections below, for example: *Dates, Times and Temperatures*

Diary entries with multiple paragraph-like sections may contain multiple <tei:ab> blocks:

```
<tei:transcription>
<tei:ab> [...] </tei:ab>
<tei:ab> [...] </tei:ab>
</tei:transcription>
```

Sub Entries

Diary entries frequently encapsulate multiple pieces of information relating to different topics. These sub-entries are to be indicated using a combination of the <tei:seg> and <tei:ab> tags. If a sub-entry occurs within a paragraph-like block (which will have been encoded in a <tei:ab> tag), it should be wrapped within a <tei:seg> tag, and given a unique ID that indicates the record number, followed by a sub-entry number, e.g. *xml:id="r0003-se01*".

For example, if the entry includes a single block of text with smaller, sub-entries, use the <tei:seg> tag:

```
<tei:transcription>
<tei:ab>
<tei:seg type="sub-entry" xml:id="r0003-se01"> [...]</tei:seg>
</ab>
</tei:transcription>
```

If however the sub-entries are themselves paragraph-like blocks, the **type="sub-entry"** attribute should be directly added to each <tei:ab> tag. The <tei:ab> should be given a unique ID that indicates the record number, followed by a sub-entry number, e.g. **xml:id="r0003-se01"**.

```
<tei:transcription>
<tei:ab type="sub-entry" xml:id="r0003-se01">
[...]
</ab>
<tei:ab type="sub-entry" xml:id="r0003-se02">
[...]
```

```
</ab>
<tei:/transcription>
```

Symbols Marking Sub-entries

Sometimes, the start or end of a sub-entry is marked in the text with a symbol such as an em-dash or an asterisk. The symbol may appear within a line of text, or isolated on a line. These symbols should be indicated using the <tei:milestone> tag at the point within the text at which they occur, and within the sub-entry element to which they relate. Milestone is an empty tag, but should include a **unit** attribute indicating 'section'; the symbol used to indicate the sub-entry and it's placement are to be indicated using the **rend** attribute.

For example to show an em-dash marker that appears inline, use:

```
<tei:seg>
<tei:milestone unit="section" rend="inline em-dash" />
</tei:seg>
```

Possible values for rend include:

- rend="inline em-dash"
- rend="block em-dash"
- rend="inline asterisk"
- rend="block asterisk"
- others as they are encountered

Translations

Within the translation block, a corresponding <tei:ab> and <tei:seg> structure should be created to mirror the structure in the transcription. Each structural unit should reference the ID of the corresponding element in the transcription using the **corresp** attribute.

For sub-entries marked with <tei:seg>, references to the transcription will look like the following:

```
<tei:translation>
<tei:ab>
<tei:seg type="sub-entry" corresp="r0003-se01"> [...]</tei:seg>
</ab>
</tei:translation>
```

When sub-entries are indicated with <tei:ab>, create the following correspondence within the translation:

```
<tei:translation>
<tei:ab type="sub-entry" corresp="r0003-se01">
[...]
</ab>
<tei:ab type="sub-entry" corresp="r0003-se02">
```

```
[...]
</ab>
<tei:/transcription>
```

Lessonbook Entries

Individual lessonbook entries are to be encoded in a single <tei:record> with the German appearing in <tei:transcription> and the English translation appearing in <tei:translation>. All tags used in creating a lessonbook entry structure will fall within the TEI namespace.



To Be Written

Correspondence

Individual pieces of correspondence are to be encoded in a single <tei:record> with the German appearing in <tei:transcription> and the English translation appearing in <tei:translation>. All tags used in creating the correspondence structure will fall within the TEI namespace.

The main structural division within a letter is the paragraph; these are to be indicated using the <tei:p> element.

```
<tei:transcription>
<tei:p> Mit Dank sende Ihnen anbei den [...] </tei:p>
</tei:transcription>
```

Opening, Closing and Postscripts

Text appearing among the preliminary information at the start of the letter (e.g. the date, salutation) should be grouped together using <tei:opener> element.

```
<tei:transcription>
<tei:opener>
<tei:salute> Verehrter, liebster Meister! </tei:salute>
</tei:opener>
<tei:p> Mit Dank sende Ihnen anbei den [...] </tei:p>
[...]
</tei:transcription>
```

Text appearing at the close of the letter (e.g. salutations, signatures, dates) should be grouped together using the <tei:closer> element. This may be repeated as necessary, to indicate closing information that is interrupted by postscript text.

```
<tei:transcription>
[...]
<tei:closer>
<tei:salute>Bis dahin in aller Verehrung und Liebe Ihr</tei:salute>
<tei:signed>Hans.</tei:signed>
<tei:dateline>Wien, <tei:date>17/IX 14.</tei:date></tei:dateline>
</tei:closer>
```

</tei:transcription>

See Also: More information about the use of <tei:opener> and <tei:closer> may be read in the TEI guidelines, at: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSOC

Postscripts should be indicated with the <tei:postscript> element. Text within the postscript may be grouped into paragraphs.

```
<tei:transcription>
[...]
<tei:postscript>
<tei:p>Herzlichste Grüsse der lieben gnädigen Frau!</tei:p>
</tei:postscript>
</tei:transcription>
```

Other documents

Individual documents are to be encoded in a single <tei:record> with the German appearing in <tei:transcription> and the English translation appearing in <tei:translation>. All tags used to represent the structure of other document types will fall within the TEI namespace.

The main structural division within a piece may be the paragraph; in such cases, these are to be indicated using the <tei:p> element. In other cases, the main structural division may conceptually be more appropriately represented with <tei:ab>.



Page breaks

Page breaks are to be encoded using the empty element <tei:pb>, at the point in the transcription or translation at which they occur (even mid-word). The attribute **n** should be given a value that corresponds to the physical page number. In rendering, the number will be made bold, and displayed within parentheses, i.e. **{835}**.

```
<tei:pb n="835"/>
```

Footnotes

Notes should be indicated within the text at the point in the transcription or translation at which they occur, and given a unique identifier that references the record number, followed by a footnote number, e.g. xml:id="r0001-fn01". Indicate the **type** of the note as 'editorial' and include a **place** attribute with the value of "foot". Within the display, footnotes will be displayed at the foot of the page, with numbered, linked markers appearing at both points within the transcription and translation.

For example:

```
<tei:transcription>
<tei:ab>
Some text requiring a footnote.<tei:note xml:id="r0003-fn01" type="editorial" place="foot">Here is the note.</tei:note>
</tei:ab>
```

</tei:transcription>

To make reference to the same note within the translation, use a <tei:ptr> tag. For example:

<tei:translation>

<tei:ab>

Some text requiring a footnote.<tei:ptr corresp="#r0003-fn01" type="footnote"/>

</tei:ab>

</tei:translation>

References Between Records

Cross references between records can be made to any element that possesses a unique id, i.e. to records, to sub-entries (where they exist), or to footnotes.



More To Come

Editorial Encoding Specifications

Editorial encoding will be performed within the <tei:transcription> element, and all tags used will fall within the <tei> namespace.

Doubtful Readings

In the case of doubtful readings the element <tei:unclear> can be used with the @reason which will have the value of:

- 1. "reading" in cases in which there is uncertainty in the reading of the German.
- 2. "meaning" in cases in which there is uncertainty in the precise meaning of the word in the English.

For Example:

tat ich es erst in diesem <tei:unclear reason="reading">Augenblick</tei:unclear>

Abbreviations

When an abbreviation needs to be expanded, both abbreviation and expansion are encoded using a combination of two elements: <tei:abbr> and <tei:ex>. The <tei:abbr> indicates the boundary of the abbreviation, while the <tei:ex> tag indicates the letters of expansion. For example:

<tei:abbr>lhr erg<tei:ex>e</tei:ex>b<tei:ex>ener</tei:ex></tei:abbr>

Alterations in original

Alterations of the original must be reproduced and encoded within the element <tei:choice>. Within <tei:choice> there are different ways of encoding an alteration.

Deletions (incl. one/two letter deletions)

Deletions are to be encoded within the element <tei:del>:

```
ein<tei:del>e</tei:del>
mi<tei:del>ch</tei:del>r
```

This method can be used also for punctuation, for example:

```
<tei:del>,</tei:del>
```

Corrections

There are two primary types of corrections:

- 1. Authorial Corrections: those performed by Heinrich Schenker or Jeanette Schenker and others.
- 2. Editorial Corrections: those performed by the SDO editors
- 1. Authorial Corrections (in the hand of Heinrich or Jeanette)

Punctuation:

1. overstrike

```
<tei:subst> <tei:del type="overstrike" hand="#HS">;</tei:del> <tei:add place="inline" hand="#HS">;</tei:add> </tei:subst>
```

2. alterations:

```
<tei:subst> <tei:del type="altered" hand="#HS">;</tei:del> <tei:add place="superimposed">:</tei:add> </tei:subst>
```

Superscript additions:

```
<tei:subst> <tei:del type="overstrike" hand="#HS">beim</tei:del> <tei:add place="superscript" hand="#HS">zum</tei:add> </tei:subst>
```

2. Editorial Corrections

Where the uncorrected original is indecipherable:

```
tei:corr>Pontresina /tei:corr>
```

To indicate faults in the original:

```
<tei:choice>
<tei:sic>lhrer</tei:sic>
<tei:corr resp="#IB">lhrem
</tei:corr>
```

</tei:choice>

The responsibility for editorial corrections should be indicated with the **resp** attribute. The value of this attribute should correspond to a statement of responsibility in the Collection Description (see the section *Records and Record Collections* above for more information).

Omitted Text

When text is omitted from a transcription or translation for editorial reasons (usually because it is not of interest or relevance to the project at hand), use the empty <tei:gap> element. Add a **resp** attribute indicating the person taking responsibility for the omission, a **reason** attribute to indicate why the text is being ommitted. Optionally, use an **extent** attribute to indicate how much text is being omitted.

```
<tei:gap reason="irrelevant" resp="#IB" extent="paragraph"/>
```

This will be rendered as [paragraph].

```
<tei:gap reason="irrelevant" resp="#IB" />
```

This will be rendered as [...].

Supplied Text

Additional punctuation in case of difficult syntax in the transcription is to be indicated with the element <tei:supplied> which represents an addition supplied by the editor, as in the case shown below:

```
<tei:supplied resp="#IB">,</tei: supplied>
```

When text is to be added to a translation for editorial reasons (to improve readability, for example), use the <tei:supplied> element. Add a **resp** attribute indicating who is supplying the text, and a **reason** attribute indicating why the text is being supplied.

<tei:supplied reason="translation" resp="#IB" >What next?<tei:supplied>

Script distinction

The use of Lateinschrift, even if as part of a compound word, is encoded within the element **To Be** Specified.

Typographical Emphasis

In cases in which additional typographical distinction is required, e.g. bold, italics (that are NOT Lateinschrift), underlined, etc., words or pieces of text can be encoded as follows:

```
<tei:hi rend="bold">some text</tei:hi>
<tei:hi rend="italics"> some text</tei:hi>
<tei:hi rend="bold-italics"> some text</tei:hi>
<tei:hi rend="superscript"> some text</tei:hi>
<tei:hi rend="underline"> some text</tei:hi>
```

```
<tei:hi rend="double-underline"> some text</tei:hi> <tei:hi rend="triple-underline"> some text</tei:hi>
```

...Others as they are encountered and defined

Please note that "superscript" used here refers to typographical placement of text above the baseline. This usage is distinct from the placement of supra- or sub- text indicating corrections or insertions.

Foreign Words

Words that are foreign to the dominant language in the text can be marked up using the <tei:foreign> tag. The language should be indicated using the xml:lang attribute:

```
<tei:foreign xml:lang="fre">Courtoisie</tei:foreign>
```

<tei:foreign xml:lang="eng">consequent</tei:foreign>

The codes given to the xml:lang attribute should correspond to the three letter codes specified in ISO 639-2.

See Also: The Library of Congress ISO 639-2 for an alphabetical listing of these codes at: http://www.loc.gov/standards/iso639-2/php/English_list.php

Dates, Times, Temperatures

Dates appearing in the text are to be encoded within the <tei:date> element. Use the when attribute to indicate the date in a machine readable format, i.e. YYYY-MM-DD.

```
<tei:date when="1918-02-18">18</tei:date>
```

Materials which include a time of day, particularly in the openers associated with diary entries, can be indicated with the the <tei:time> element, for example:

```
<tei:date when="1918-02-18">18 <tei:time>Morgen</tei:time></tei:date>
```

Diary entry openers often also include some indication of the weather. These can be encoded using the <tei:measure> tag. For example:

```
<tei:measure type="temperature" unit="celsius">-2°</measure>
```

Date, temperature and time of day information that appears at the start of a diary or lessonbook entry should be encapsulated within the <tei:opener> tag. For more information about the use of this tag, see the *Structural Encoding* section above.

Notes

Marginal Marks (authorial)

Marginal marks are encoded differently according to whether they are marginal comments or glosses or whether they are additions to the text connected by a cue-mark in the text. Both elements have to be inserted at the point in the text to which the editor thinks the marginal note refers:

a.In the case of a marginal note:

<tei:note place="left-margin">some text</tei:note>

b.In case of a clear addition to the text, with or without a cue-mark:

<tei:add place="left-margin">some text</tei:add>

Editorial Notes

To make an editorial note in the text, insert a <tei:note> tag at the point at which you wish the note to appear. Add a type="editorial" attribute to the tag to distinguish the note from those made by the document authors.

<tei:note type="editorial">This letter was subsequently marked with a large 'X', indicating acknowledgement by Universal Edition.</tei:note>

Indication of certainty about translation should also be indicated using an editorial note.

<tei:note type="editorial">?</tei:note>

★ See Also: For a more comprehensive treatment of certainty in the TEI, see Chapter 21 in the TEI P5 Guidelines, *Certainty and Responsibility*, available here:http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CE.html

Figures

Inline diagrams are to be indicated at the point at which they appear in the text using the <tei:figure> tag. The figure must be given a unique ID, which references the record number followed by a figure number, e.g. xml:id="r0003-fg01". The *place" attribute may be used to indicate the position of the figure in the text. If needed, the figure may be given a caption, indicated with <tei:head> and a longer description, indicated with <tei:figDesc>.

To insert a figure, use:

```
<tei:figure xml:id="r0001-fg01"/>
```

To add a caption or other description information to the figure, use:

```
<tei:figure xml:id="r0001-fg02" place="margin-left">
<tei:figDesc>
<tei:caption> This caption will be displayed to the reader. </tei:caption>
Other text may be added here to describe or contextualize the figure.
</tei:figDesc>
```

oXygen Tip: opening the **place** attribute in oXygen will display a list of recommended values for this attribute.

Lists

Lists are to be encoded using the <tei:list> element. Use a type attribute to indicate the kind of list: bulleted,

gloss, ordered or simple. Individual items should be placed in <tei:item> elements.

```
<tei:list type="ordered">
<tei:item>Apples</tei:item>
<tei:item>Oranges</tei:item>
</tei:list>
```

Semantic Encoding Specifications

Titles

Titles (as of books, works) should be marked with the <tei:title> tag.

<tei:title>Die Fackel</tei:title>

Names, places, institutions, terms

Names, places, institutions and musical terms and works are to be indicated in the text.

a. Names:

<tei:persName>Beethoven</tei:persName>

b. Place Names

<tei:placeName>Wien</tei:placeName>

c. Institutions

<tei:orgName>Österreichische Nationalbibliothek</tei:orgName>

d. Musical Terms

<tei:term>Fuge</tei:term>

d. Works (musical, Schenkerian, and others)

To Be Specified

Each tag will also have a key attribute containing an ID that unambiguously identifies the entity to which the wrapped text refers. This key will be maintained and referenced within a separate system, the Entity Authority Tool Set (EATS).

This topic: XMLTeam > DocProjDigitalSchenker > DocProjDigitalSchenkerEncodingModel

History: r12 - 10 Sep 2008 - 10:27:57 - TamaraLopez