

Edition-specific TEI encoding guidelines

Nachwuchsgruppe Berliner Intellektuelle 1800–1830

Digital Edition group*

December 23rd, 2011

Contents

1	Introduction	2
1.1	The project “Berlin Intellectuals 1800–1830” and its Digital Edition	2
1.2	About this document	2
2	Basic encoding decisions	2
3	Manuscript encoding	3
3.1	The header	4
3.2	The body	13
3.2.1	Body elements of all manuscripts	14
3.2.2	Specific elements for encoding letters	24
3.2.3	Specific elements for encoding dramas	25
3.3	Back matter	28
4	Index encoding	28
4.1	Index of persons	30
4.2	Index of organizations	33
4.3	Index of places	34
4.4	Index of works	35
4.5	Index of contributors	37
	List of TEI elements	37

*For comments regarding the edition and its contents, please contact Anne Baillot (anne.baillot@hu-berlin.de). For comments regarding this documentation, please contact Alexander Meyer (alexander.meyer@student.hu-berlin.de).

1 Introduction

1.1 The project “Berlin Intellectuals 1800 – 1830” and its Digital Edition

The Digital Edition described here was conceived and is being realized by the Junior Research Group “Berlin Intellectuals 1800–1830” at the Institute for German Literature of the Humboldt-University in Berlin (Germany). The Research Group is interested in various corpora that will be partly edited in print and partly edited online, the overlap between both publications forms being very thin. This documentation describes the encoding guidelines for the digital edition. This edition is based on XML and follows the TEI-P5 encoding guidelines.

The digital edition contains following corpora: letters (60 %), work manuscripts (novel, drama: 15 %), lecture notes (15 %), manuscripts from the University Archive documenting the history of the philosophy faculty between 1810 and 1830 (10 %). All of them are handwritten documents, in the vast majority original documents (only occasionally handwritten copies), written mostly in German or, in some cases, in French. They were produced in Berlin or in connection with intellectuals from Berlin in the late 18th and early 19th century.

1.2 About this document

This document describes the encoding guidelines and decisions that were made specifically for our project and our edition. It is not a replacement for the original TEI documentation. Rather, it focuses on the way we use TEI elements in order to fit every phenomenon we need to encode. For this reason, only elements that we actually use are described together with exact details on how we use them.

2 Basic encoding decisions

The TEI encoding must not be adapted to the letters exclusively. It has to be more universal in order to treat other types of text, too. This is one of the reasons why we decided not to use special elements like in SIG Correspondence, though it is appropriate and functional for letters. Yet we considered creating new, non-standard elements as no sensitive approach because there are consequences on the exchange of data and their durability.

For each manuscript, a single XML document is created. Section 3 describes elements appearing in those documents including header and body. Apart from that, there are



several XML documents called indexes. Each of them contains one of the central indexes of the edition: Persons, organizations, places, works, and of course the index of contributors to the edition. The index documents are described in section 4.

3 Manuscript encoding

A single TEI document comprises the full text of the manuscript and the metadata attached to it:

- The <teiHeader> element contains all metadata on the digital file.
- The <text> element contains the textual content.
- The transcription itself is contained in the <body> element in a <div> with the attribution @type="transcription".
- Currently there aren't any <front> and <back> elements outside of the <body> except for dramas where <back> is used for a cast list (see section 3.3).

This is the skeleton of the TEI document (some elements are unlisted—see the attachment for the complete example):

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
  <titleStmt>
    [...]
  </titleStmt>
  <publicationStmt>
    <availability>
      <p><licence target="http://creativecommons.org/licenses/by/3.0/de/deed.en">Attribution 3.0 Germany (CC BY 3.0)</licence></p>
    </availability>
    <date when-iso="[...]">[...]</date>
  </publicationStmt>
  <seriesStmt>
    <title type="main">Berliner Intellektuelle 1800-1830</title>
    [...]
  </seriesStmt>
  <sourceDesc>
    <msDesc>
      [...]
    </msDesc>
    <correspDesc>
```



`<teiHeader> → <fileDesc>`

```

    [...]
  </correspDesc>
</sourceDesc>
</fileDesc>
<encodingDesc>
  [...]
</encodingDesc>
<profileDesc>
  <langUsage> [...] </langUsage>
</profileDesc>
<revisionDesc>
  <change when="[...]" who="#[...]">[...]</change>
</revisionDesc>
</teiHeader>
<text>
  <body>
    <div type="transcription">
      <pb n="1" facs="[...].jpg" />
      <opener>[...]</opener>
      <p>[...]</p>
      <closer>[...]</closer>
      <!-- opener and closer only used for letters -->
    </div>
  </body>
  <back> [...] </back>
  <!-- back only used for dramas -->
</text>
</TEI>

```

3.1 The header

`<teiHeader>`

As usual, the header comprises the `<teiHeader>` element.

`<teiHeader>`

Following the TEI guidelines, the `<fileDesc>` (the main characteristics of the document), `<encodingDesc>` (editorial rules), `<profileDesc>` (information about the content) and `<revisionDesc>` (the history of the encoding) elements are used in `<teiHeader>`.

`<teiHeader> → <fileDesc>`

Within `<fileDesc>`, there are the following elements:

`<fileDesc>`

`<teiHeader> → <fileDesc> → <titleStmt>`

- `<titleStmt>`
- `<publicationStmt>`
- `<seriesStmt>`
- `<sourceDesc>`

`<teiHeader> → <fileDesc> → <titleStmt>`

In the mandatory element `<titleStmt>`, we always encode the title of the document (`<title>`), authors and editors (*not* for letters), and information on its digital version: the funder (`<funder>`), the head of the project (`<principal>`), and, inside as many `<respStmt>` elements as needed, all other staff members and contributors who have worked on the digital version of the particular document. `<titleStmt>`

In case of letters, the `<title>` is structured as follows: “Brief von + *name of the sender* + an + *name of the addressee* + (place, date)”. `<title>`

For example:

```
<title>Brief von Ludwig Tieck an Friedrich von Raumer (Dresden,
6. Oktober 1823)</title>
```

In case of other manuscripts, the `<title>` element contains the title as written on the manuscript or given in another form.

For letters, we *do not* use the `<author>` element. Instead, the sender and the addressee are encoded in `<correspDesc>` inside `<sourceDesc>` (see page 11). For other types of texts, `<author>` and `<editor>` may be used. We have decided to use `<editor>` only for roles belonging to the original text, i. e. roles that have led to the original text as it is. For roles of digital edition project members, we use `<respStmt>` (see below). `<author>`
`<editor>`

These are examples of our intended use of the `<author>` and `<editor>` elements:

```
<author>
  <persName ref="#p0001">Ludwig Tieck</persName>
</author>
<editor role="annotator">
  <persName ref="#p0002">Wilhelm Heinrich Wackenroder</persName>
</editor>
<editor role="literary_executor">
  <persName ref="#p0003">Rudolf Köpke</persName>
</editor>
```

The `<principal>` element always contains the name of Anne Baillot, the “Nachwuchsgruppenleiterin”, and her affiliation. The indication is very detailed (forename, surname, and the complete address of Humboldt University). `<principal>`

`<teiHeader>` → `<fileDesc>` → `<seriesStmt>`

```
<principal>
  <persName>
    <forename>Anne</forename>
    <surname>Baillot</surname>
  </persName>
  <affiliation>
    <orgName>Humboldt-Universität zu Berlin</orgName>
    <address>
      <street>Unter den Linden 6</street>
      <postCode>10099</postCode>
      <settlement>Berlin</settlement>
      <country key="DE">Deutschland</country>
    </address>
  </affiliation>
</principal>
```

The next item `<respStmt>` comprises the declaration of all persons who took part in the research and edition of the encoded document. We have selected 5 different responsibilities: transcription, encoding, annotations, corrections, digitization of the original manuscript. Please note that this list cannot be exhaustive.

`<respStmt>`

Inside the `<respStmt>`, the name of the person is given inside a `<persName>` element. Using the attribute `@ref`, we link it with the corresponding entry in the index of contributors.

For example:

```
<respStmt>
<resp>Transcription by</resp>
<persName ref="#sabine.seifert">
  <forename>Sabine</forename>
  <surname>Seifert</surname>
</persName>
</respStmt>
```

For a more detailed revision history, we use `<revisionDesc>` (see page 13).

`<teiHeader>` → `<fileDesc>` → `<publicationStmt>`

The `<publicationStmt>` element typically contains all information about the publisher, the publication's place and date and the restrictions of its use. In this project, we only use `<availability>` and `<date>`.

`<publicationStmt>`

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <msIdentifier>
```

<teiHeader> → <fileDesc> → <seriesStmt>

The `<seriesStmt>` element groups information about the series which the document belongs to. This is in all cases the project “Berliner Intellektuelle 1800–1830”, but we decided to specify to which sub-project the single documents are attached. To do so, we use a `@type` attribute with the value “main” for the name of the project and the value “sub” for each sub-project. Project titles are given in `<title>` elements. `<seriesStmt>`

For example:

```
<seriesStmt>
  <title type="main">Berliner Intellektuelle 1800-1830</title>
  <title type="sub">Tieckiana</title>
</seriesStmt>
```

<teiHeader> → <fileDesc> → <sourceDesc>

The following element, `<sourceDesc>`, is used for all metadata describing the sources of the digital edition. This element is repeatable, so one could describe different sources if one TEI document for the digital edition is based on several manuscripts. `<sourceDesc>`

The documents we are encoding are handwritten. For this reason, the `<sourceDesc>` element always contains the sub-element `<msDesc>`, which is intended for the description of manuscripts. If the manuscript is a letter, it also contains `<correspDesc>` to identify the sender and addressee.

<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc>

The `<msDesc>` element contains the following four elements: `<msDesc>`

- `<msIdentifier>` (information identifying the manuscript)
- `<msContents>` (describes the intellectual content of a manuscript)
- `<physDesc>` (physical description of the manuscript)
- `<history>` (the history of the manuscript)

<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <msIdentifier>

In the `<msIdentifier>` element, all information concerning the institution, the manuscript’s collection and its identifier (`<idno>`) is placed. A `@type` attribute for `<idno>` can be used to be more specific about who gave this identifier. `<msIdentifier>`
`<idno>`

It’s necessary to differentiate between the `<repository>` and the `<institution>` element. Institution describes according to the TEI guidelines an “organization such as a university or library, with which a manuscript is identified, generally its holding institution”, whereas repository is the place “within which manuscripts are stored”, thus the `<repository>`
`<institution>`

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <physDesc>
```

place where the document is preserved. The @key attribute for <country> is mandatory.

For example:

```
<msIdentifier>
  <country key="DE">Deutschland</country>
  <settlement>Berlin</settlement>
  <institution>Staatsbibliothek zu Berlin-PK</institution>
  <repository>Handschriftenabteilung</repository>
  <collection>Nachlaß Raumer</collection>
  <idno type="StaBi">Blatt 224-225</idno>
</msIdentifier>
```

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <msContents>
```

In <msContents>, the <docDate> (document date) is indicated within <msItem>. <docDate> is the date originally written on the letter. It may be identified as incorrect and thus differ from the <origDate> (origin date) in the <history> section, where the actual date (which can be a supposition) is declared.

For example:

```
<msContents>
  <msItem>
    <docDate when="1823-10-06">6. Oktober 1823</docDate>
  </msItem>
</msContents>
```

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <physDesc>
```

<physDesc> comprises the physical description of the manuscript. In our case, <physDesc> may contain four or five elements: <objectDesc>, <handDesc>, <accMat>, <sealDesc> and <decoDesc>.

In the <objectDesc> element, the material, the dimensions, the state of conservation etc. of the manuscript are specified. The statements are grouped inside <supportDesc>.

For example:

```
<objectDesc>
  <supportDesc>
    <support>
      <material>Welches Papier, Qualität usw.</material>
    </support>
  <extent>
    <measure type="leaf">4 Blätter</measure>
```

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <physDesc>
```

```

<dimensions unit="cm">
  <height>die Höhe</height>
  <width>die Breite</width>
</dimensions>
</extent>
<foliation corresp="#major_hand">Beschreibung einer
Follierung</foliation>
<foliation corresp="#red_ink">Beschreibung einer anderen
Follierung</foliation>
<foliation corresp="#sabine.seifert">unsere Follierung</foliation>
<condition>wohl erhalten zum Beispiel</condition>
</supportDesc>
</objectDesc>

```

`<foliation>` contains descriptions of foliations. The `@corresp` attribute links each `<foliation>` foliation with the hand who made it (as described by a `<handNote>`) or – only if it is made by us – with the corresponding person in the index of project members.¹

`<objectDesc>` *could* contain the `<layoutDesc>` element besides `<supportDesc>`. `<layoutDesc>` deals with the layout of the manuscript, but as we stated that the documents generally don't have a specific layout, we decided not to use this element. Therefore, `<supportDesc>` is the only child element of `<objectDesc>`.

The `<handDesc>` element is also part of `<physDesc>`. Here, the different hands found in the manuscript can be identified and described with as many `<handNote>` elements as needed. More than one `<handNote>` element can belong to the same person: if an author completed a text after a long time, two different hands could be identified for him. `<handDesc>`
`<handNote>`

`<handNote>` is also used to describe which scripture the hand uses predominantly (in general, this would be *Kurrentschrift*).

The following attributes are used:

- `@xml:id`: An identifier of the hand.
- `@scope`: The hand's preponderance in this manuscript.
- `@medium`: Which instrument has the hand been using?
- `@scribeRef`: Points to the person responsible for the hand, i. e. to the respective `<person>` element in the index of persons. This attribute is always given, even if the person is the author/sender. If the hand is of an archivist that cannot be identified as a person, then this attribute points to the corresponding archive in the index of organizations.
- `@scribe`: The function that the person has in this particular text.

¹In the text, we use `<note type="foliation">` to give foliation numbers. (See page 23.)

```
<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <physDesc>
```

- @corresp: Only used with scribe="publisher" (see list below).

In the full text, any @hand attribute points to a <handNote> element in the TEI header where the particular hand is described.

In the TEI header:

```
<handDesc>
  <handNote xml:id="major_hand" scope="major" medium="black_ink"
    scribe="author" scribeRef="#p0023">
    Hand des Absenders, Briefftext in schwarzer Tinte,
    vorwiegend in Kurrentschrift
  </handNote>
  <handNote xml:id="red_ink" scope="minor" medium="red_ink"
    scribe="publisher" scribeRef="#p0042" corresp="#w0221">
    Hand von ..., wenige Anmerkungen in roter Tinte,
    vorwiegend in Kurrentschrift
  </handNote>
</handDesc>
```

Corresponding @hand in the body:

```
<add place="margin-right" hand="#red_ink">Tieck an Raumer</add>
```

There may only be one hand with scope="major". This is defined to be the *default hand* of the document, i. e. the hand that is active in parts of the body where no @hand attribute is given. All other hands have scope="minor".

When there is only one hand in the whole manuscript, scope="sole" is used instead of scope="major". Then this hand is the default hand, of course.

The following values can be used for @scribe:

- author: The hand is of the author/sender.
- publisher: The hand is of somebody who played a role in publishing the text. In this case, the attribute @corresp can be used to point to the entry of the corresponding publication in the index of works (see example above).
- archivist: The hand is of an archivist.
- addressee: In case of letters, the hand is of the addressee (i. e. the person given in the <addressee> element of <correspDesc>).
- unknown: The person responsible for the hand is unknown.
- other: The hand is of a person whose role does not fit in the other categories but who is not unknown.

`<teiHeader> → <fileDesc> → <sourceDesc> → <correspDesc>`

Furthermore, `<physDesc>` may contain `<sealDesc>` where seals used to close the letter (`<sealDesc>` wax etc.) are specified.

Any material accompanying the letter—for example an envelope, a picture, a drawing etc.—is given inside `<physDesc>` using the `<accMat>` element. If an envelope is labelled, `<accMat>` the text can be added here:

```
<accMat>
  <p>Briefumschlag, beschrieben mit: <quote>Für meinen lieben Freund
  ...</quote></p>
</accMat>
```

`<teiHeader> → <fileDesc> → <sourceDesc> → <msDesc> → <history>`

After the `<physDesc>` element, we can now take a look at the `<history>` element inside `<msDesc>`. The full history of a manuscript is described here: from its creation until today (going through the place and date of creation, the provenance, the acquisition ...). For letters, we always use `<origin>` inside which we encode the place of creation in `<origPlace>` and the date of the letter—that is to say, the date which is written on the letter if it is identified as correct, otherwise the date of origin assumed by the editor—in `<origDate>`:

For example:

```
<history>
  <origin>Der Brief wurde in <origPlace ref="#l1234">Dresden</origPlace>
  am <origDate when-iso="1823-10-06">6. Oktober 1823</origDate>
  verfasst</origin>
  <acquisition />
</history>
```

Sometimes it might be impossible to give an exact date, but one can see from the contents that for example, a letter cannot be written before a certain date because it refers to something that happened on that date. In those cases, we use the `@notBefore-iso` and `@notAfter-iso` attributes.

`<teiHeader> → <fileDesc> → <sourceDesc> → <correspDesc>`

For letters, the second element inside of `<sourceDesc>` is `<correspDesc>`.² Here, the sender and addressee are identified using the sub-elements `<sender>` and `<addressee>`. As stated before, we *do not* use the `<author>` element for letters. `<correspDesc>`

For example:

²`<correspDesc>` is currently a proposal and not a part of the TEI guidelines yet. However, because it fits our case very well, we have decided the use it already.

`<teiHeader> → <encodingDesc>`

```
<correspDesc>
  <sender>
    <persName ref="#p1204">Ludwig Tieck</persName>
  </sender>
  <addressee>
    <persName ref="#p3420">Friedrich von Raumer</persName>
  </addressee>
</correspDesc>
```

In order to keep the TEI header of a letter as similar as possible to the headers of other manuscripts, the place where the letter was written and its date are described elsewhere and not inside `<correspDesc>`. (See `<msItem>`, `<history>` and `<dateline>`.)

`<teiHeader> → <encodingDesc>`

After `<fileDesc>`, the second top-level portion of the header is `<encodingDesc>`. It comprises `<projectDesc>` (which simply names the respective project) and `<editorialDecl>`.

The `<editorialDecl>` element is used for documenting decisions concerning `<correction>`, `<normalization>` and `<hyphenation>` in the particular manuscript.

For example:

```
<encodingDesc>
  <projectDesc>
    <p>Projekt: "Berliner Intellektuelle 1800-1830"</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p>Spezifische Entscheidungen, die Korrekturen
        in diesem Dokument betreffend</p>
    </correction>
    <normalization>
      <p>Spezifische Entscheidungen, die Normalisierung
        in diesem Dokument betreffend</p>
    </normalization>
    <hyphenation eol="hard">
      <p>Spezifische Entscheidungen, die Worttrennung
        in diesem Dokument betreffend</p>
    </hyphenation>
  </editorialDecl>
```

Apart from very few cases (see page 18), we do not do any correction or normalization.

`<hyphenation>`, though, always has the attribute `eol="hard"` which describes the way

`<teiHeader> → <revisionDesc>`

we deal with end-of-line hyphenation (see page 14). If no other particular decisions apart from that were made for the particular document, then `<hyphenation>` may be empty apart from the `@eol` attribute:

```
<hyphenation eol="hard" />
```

`<teiHeader> → <profileDesc>`

In `<profileDesc>`, we specify the languages of the manuscript. The only sub-element of `<profileDesc>` is `<langUsage>`.

`<profileDesc>`
`<langUsage>`

Example of `<langUsage>`:

```
<langUsage>
  <language ident="de" usage="80">Deutsch</language>
  <language ident="fr" usage="20">Französisch</language>
</langUsage>
```

Each language is identified in a `<language>` element. The approximate percentage of language usage is given in the attribute `@usage`. When there is only one language, usage becomes 100, of course. In those cases, `@usage` may be left out.

`<language>``<teiHeader> → <revisionDesc>`

`<revisionDesc>` is the last element directly bei `<teiHeader>`. It summarizes the revision history for a file. Each revision is recorded in a `<change>` element with a `@who` attribute which refers to the `@xml:id` attribute of the corresponding `<person>` entry within the index of contributors.

`<revisionDesc>`
`<change>`

```
<revisionDesc>
  <change when="2011-06-15" who="#anna.busch">Added annotations and
  corrections</change>
  <change when="2011-06-11" who="#solenne.coutagne">Added
  <att>hand</att></change>
  <change when="2011-06-05" who="#solenne.coutagne">Encoding</change>
  <change when="2011-05-02" who="#sabine.seifert">Transcription</change>
</revisionDesc>
```

3.2 The body

After `<teiHeader>`, the element `<text>` follows. It always contains the `<body>` of the text and in case of dramas `<back>` where we decided to put cast lists. For other manuscripts types apart from dramas, `<body>` is the only sub-element of `<text>`.

`<text>`

Of `<body>`, the only direct sub-element is a `<div>`: `<div type="transcription">`. In-

`<body>`
`<div>`

side of it, the actual body, i. e. the text of the manuscript appears.

This section describes elements appearing inside the text. It is split in three parts: Section 3.2.1 contains elements that may appear in any kind of manuscript, section 3.2.2 deals with elements that are specific to letters, section 3.2.3 deals with specific body elements for dramas.

3.2.1 Body elements of all manuscripts

Page breaks and facsimilies: <pb>

Until the image link is more precise, the only element which refers to the picture is the page break element <pb>. Each <pb> has a @n attribute which indicates the page number. For technical reasons it is required to always start with page number 1. The differing page number of an archivist etc. must be marked up as a <note> (see page 23). <pb>

The @fac attribute is used to link to the picture on the server. We will have to review this solution if we want to have a closer link to the picture (and maybe envisage using the <graphic> and <facsimile> elements).

The first child element of the <div type="transcription"> must be <pb>, because of the structure of the XSLT stylesheet.

For example:

```
<body>
  <div type="transcription">
    <pb n="1" facs="00000356.jpg"/>
    <opener>
```

Line breaks: <lb>

All line breaks are tagged with the <lb> element.³ <lb>

When a word is being split at a line break, we *leave out* the hyphenation character. The text

Werd ich denn hier nun end-
lich etwas finden?

thus becomes:

```
Werd ich denn hier nun end<lb break="no"/>lich
etwas finden?
```

³<l> is only used for verses, see section 3.2.3.

As you can see, we use the attribute `break="no"` to denote that there is no new word starting there. There must be *no spaces and no line breaks* between the `<lb>` tag and the surrounding letters!

In case the hyphenation character is not a standard single hyphen, the attribute `@type` may be used in `<lb>` with the following values:

- sh: single hyphen ("-"). This is the default if `@type` is not given.
- dh: double hyphen ("=").⁴
- none: no hyphenation character at all in the original text.

Thus, the text

Werd ich denn hier nun end=
lich etwas finden?

becomes:

Werd ich denn hier nun end<lb break="no" type="dh"/>lich
etwas finden?

However, a hyphenation character is retained at the end of a line if it belongs in the word. Thus, the text

Köln-
Ehrenfeld

becomes

Köln-<lb break="no"/>Ehrenfeld

(`break="no"` is still used, of course, because there is no beginning of a new word.)

The way we encode end-of-line hyphenation is described in the `<hyphenation>` element in the header.

Logical text divisions: <div>, <head>, <trailer>

For logical text divisions (sections, subsections and so on), we used nested `<div>` elements. We *do not* use the numbered division elements `<div1>`, `<div2>` and so on. `<div>`

When appropriate, `<div>` may have the attributes `@type` specifying the type of division and `@n` specifying the number of the division. Here is an example from a drama:

⁴Inside a line, double hyphens are treated differently. See page 18.



Changes in the manuscript: `<subst>`, `/<add>`

```
<div type="act" n="2">
  [...]
  <div type="scene" n="1">
    [...]
  </div>
</div>
```

Heading and trailing elements of the section as appearing in the manuscript are given inside `<head>` and `<trailer>`, respectively. These elements always belong *only* to the `<div>` of which they are direct sub-elements:

`<head>`
`<trailer>`

```
<div type="act" n="2">
  <head>II. Akt</head>
  [...]
  <div type="scene" n="1">
    <head>I. Scene</head>
    [...]
  </div>
</div>
```

Sometimes, an obvious logical unit of a manuscript can be identified but it has no heading given in the text. In these cases, we can (but do not need to) make up our own heading text. It can be included in `<head>`, but needs to be inside a `<note>` with `@resp` (see pages 23 and 24 for explanations of our uses of `<note>` with `@resp`):

```
<head><note type="structure" resp="#johanna.preusse">2. Akt</note></head>
```

Changes in the manuscript: `<subst>`, `/<add>`

Our documents are handwritten manuscripts, so they all contain abbreviations, deletions, additions, returns, mistakes, and corrections (which are sometimes hard to decipher) etc. We therefore decided to use the element couples `/<add>`, `<abbr>/<expn>`, and `<sic>/<corr>`.⁵

For the `@rend` attribute of `<add>`, the following values are recommended (other values `<add>` may be used if appropriate):

- strikethrough
- overwritten
- erasure
- none (e. g. if the deletion is only marked by writing the correct word alongside)

⁵Sometimes, it is not clear whether an addition should be encoded with `<add>` or with `<note>`. See page 23 for a discussion of this problem and an explanation of our solution. Of course, when an addition appears together with a deletion belonging to it, then `<add>` und `` should always be used.

For the `@place` attribute of ``, the following values are recommended (other values `` may be used if appropriate):

- `across` (used in combination with `<del rend="overwritten">`)
- `above`
- `below`
- `bottom`
- `top`
- `margin` (*may*—but does not need to—be more specific: `margin-left`, `margin-right`, `margin-top`, `margin-bottom` denote which margin it is, and combining these values with `...-horizontal` and `...-vertical` denotes whether the text is written horizontally or vertically (`margin-left-horizontal`, `margin-left-vertical`, `margin-right-horizontal` and so on).
- `opposite`
- `overleaf`
- `end`
- `inline` (if the text is simply written in the line, usually when some text immediately before has been struck through)
- `dots` (if the text was first deleted and then marked to be included again by writing dots below it; see below for examples of nesting of additions and deletions)

The `<subst>` element groups a combination of deletion and addition in case they definitely belong together: `<subst>`

```
<subst>
  <del rend="strikethrough">allem</del>
  <add place="above">beiden</add>
</subst>
```

Sometimes, additions and deletions are nested in themselves: Some text had been added that was (partly) deleted afterwards, or some text had been deleted that was marked for inclusion again afterwards (for example by writing dots below it). For these cases, we use two basic rules:⁶

1. What is meant to be final, goes outside:

- `A <add>B</add> C` means that “A C” was written first, then “B” was added in its middle, then everything was deleted.

⁶For simplicity reasons, `@rend` and `@place` are missing in the following examples but should always be used, of course.

Non-standard hyphenation marks: `<pc>`

- `<add>A</add>` means that “A” was once written as an addition but then, the whole addition was deleted.
- `<add>A</add>` means that “A” was first deleted but then marked as to be included again.

2. If rule 1 cannot be obeyed, then the attribute `@seq` is used to denote the sequence of the additions and deletions:

`<add seq="1">A <del seq="2">B C</add>` means that “A B C” was once written as one large addition but then, the “B” part was deleted (and not the other way round as it would be if there was no `@seq` attribute).

`@seq` may use higher numbers in case more than two additions and deletions are nested.

Sometimes, additions and deletions may also be nested with (parts of) other elements. If, however, an addition or deletion goes over such (parts of) other elements in a way that `<add>` and `` cannot be used with proper XML nesting, `<addSpan>` and `<delSpan>` are used with the same attributes and with `<anchor>`.

`<addSpan>`
`<delSpan>`
`<anchor>`

Normalization made by us: `<choice>`, `<abbr>/<expan>`, `<sic>/<corr>`

We have decided not to normalize spelling and diction. However, in a few cases, we do change the original text (but also keep the original dictions).

We use the `<choice>` element for abbreviations and their expansions (`<abbr>` and `<expan>`) as well as for corrections of obvious spelling mistakes (`<sic>` and `<corr>`).

`<choice>`
`<abbr>`
`<expan>`
`<sic>`
`<corr>`

For example:

```
<choice>
  <abbr>phil.</abbr>
  <expan>philosophische</expan>
</choice>
```

```
<choice>
  <sic>prangest</sic>
  <corr>pranget</corr>
</choice>
```

Non-standard hyphenation marks: `<pc>`

We say that the single hyphen “-” is the standard hyphenation mark in our edition. Thus, hyphenation done with this hyphen is not marked in any way. However, in some texts, the double hyphen “=” appears for hyphenation. In these cases, we do not transform it into a single hyphen. Instead, we use the ASCII equal sign but include it in a `<pc>`

`<pc>`

element to denote that is not a mathematical sign but a punctuation character. For example:

`Köln<pc>=</pc>Ehrenfeld`

There must be *no space and no line break* between the letters and the tags.

Please note that this *only* applies to hyphenation marks inside a line of text. Hyphenation marks at the end of the line are treated differently (see page 14).



Dashes and hyphens

We differentiate between dashes and hyphens as the standard character entity references in XML allow it:

- - is the single hyphen (“-”).
- `–` is the en-dash (“–”, German “Gedankenstrich”).
- `—` is the em-dash (“—”).
- `−` is the mathematical minus sign.

Hard-to-read passages: `<unclear>`, `<gap>`

The `<unclear>` element with the attributes `@cert`, `@reason` and `@extent` can be used in case of passages that are difficult to read: `<unclear>`

```
<unclear cert="medium" extent="one word" reason="stain">nur</unclear>
```

Sometimes, it happens that a hard-to-read passage could be either this or that. In those cases, `<unclear>` can be used with `<choice>`:

```
der de<choice><unclear extent="one char" reason="illegible"
cert="high">n</unclear><unclear extent="one char" reason="illegible"
cert="low">m</unclear></choice> Derwisch
```

Here, it is recommended to use `@cert` to give one possible version a high and the other one a low certainty. `cert="high"` here means, however, that there is a high likelihood for the passage to *be* the version described—and *not* that the certainty of unclarity was high. The same goes for `cert="low"`, respectively. When the editor could not decide to give one version a higher possibility than the other one, `cert="medium"` may be used twice, but this is not recommended.

For illegible passages or when a piece of text is missing, the `<gap>` element is recommended. It contains the same attributes as the `<unclear>` element, but is an empty element: `<gap>`

```
<gap extent="two chars" reason="illegible" />
```

The following values are recommended for @reason, both in <unclear> and <gap>:

- illegible
- stain
- wax
- damage

Highlighted text: <hi>

We only use the <hi> element, which, according to the TEI guidelines, “marks a word or phrase as graphically distinct from the surrounding text”. A change of scripture is indicated with <hi>, too (for example if one word is written in “Latin” and the rest of the text in “Kurrent” scripture). We add the @rend attribute which can adopt several values.

For example:

```
<hi rend="underline" n="2">Erstens</hi>
```

The following values are recommended for @rend:

- underline (with @n indicating how often the word is underlined; if n=1, then it may be left out)
- superscript
- subscript
- small
- large
- quoted
- majuscule
- latin (if the respective text is written in “Latin” scripture as opposed to “Kurrent” scripture surrounding it)
- kurrent (if the respective text is written in “Kurrent” scripture as opposed to “Latin” scripture surrounding it)
- mark (if there is a vertical line to mark a part of the text)

<hi> may be used with @hand to denote the hand which has done the *highlighting* if it is not the hand that has written the highlighted *text*.⁷ @hand corresponds with <handNote> in header as it does in the <add> and <note> elements.

⁷The TEI guidelines currently do not allow to use @hand with <hi> but we have decided to use it anyway.

Lists: <list>

Lists should be encoded with the `<list>` element:

`<list>`

```
<list>
<label>1)</label>
<item>die Rechtslehre,</item>
<lb/>
<label>2)</label>
<item>die philosophischen Briefe,</item>
<lb/>
<label>3)</label>
<item>über die Mythologie,</item>
</list>
```

The labels given inside the `<label>` element must really exist in the document. The `<label>` element must not be used if they don't appear in the source.

Dates: <date>

We place dates in a `<date>` element and we add the `@when-iso` attribute in order to encode the date in a standardized form according to ISO 8601. That is to say the following format: yyyy-mm-dd. For the interval, we use a slash "/".

For example:

```
<date when-iso="1823-10-06">den 6t October. 23</date>
<date when-iso="1824/1825">1824-1825</date>
```

Sometimes it might be impossible to give an exact date because there is none written on the letter or manuscript, but one can see from the contents that for example, a letter cannot be written before a certain date because it refers to something that happened on that date. In those cases, we use the `@notBefore-iso` and `@notAfter-iso` attributes.

Dates can also refer to a different calendar, mainly to the French Revolutionary calendar. The value of the `@when-iso` attribute has always to be in the form of the Gregorian calendar. The `@calendar` attribute can be used to precise that the date refers to another calendar.

Named entitites: Persons, organizations, places, works

Every time, a name of a person, an organization, a places or a work appears in a text, it gets marked using `<persName>`, `<orgName>`, `<placeName>` or `<title>`, respectively. For every such entity, an entry in the corresponding index (see section 4) is created including an `@xml:id` that has a specific format. From inside the text, the name is

```
<persName>
<orgName>
<placeName>
<title>
```

linked with the index entry using the @ref attribute which points to the corresponding @xml:id.

For example:

```
<placeName ref="#l1234">Dresden</placeName>
<orgName ref="#o0001">Humboldt-Universität zu Berlin</orgName>
<persName ref="#p0003">Sofokles</persName>
<title ref="#w0010">Nachgelassene Schriften und Briefwechsel</title>
```

However, it sometimes happens that the author writes about a person, an organization, a place or a work without quoting their name. We can't use the elements <persName>, <orgName>, <placeName> or <title> in those cases, because there is neither a name nor a title. Instead, the <rs> ("referencing string") element can be applied:

<rs>

```
<rs ref="#p0002">Bruder</rs>
```

Parts a in different language: <foreign>

The majority of the manuscripts is written in German or French. We specify the document's languages in the <profileDesc> as part of the TEI header.

To denote the actual parts text of the written in another language than its main language, we add an @xml:lang attribute to the correspondent element (<div>, <p>, <hi>, ...). If there is no element directly corresponding with the change of language, we use the <foreign> element with @xml:lang as its attribute.

<foreign>

As defined in the XML standard, value of the attribute is the code of the language according to IETF's BCP 47. In our case, it usually has only one part—the code of the language according to ISO 639.

For example:

```
<foreign xml:lang="la">malus</foreign>
```

Notes: <note>

We distinguish two types of notes—but use the <note> element for both of them.

<note>

On the one hand, there are *notes that appear on the manuscript page*. They can be written by the author/sender himself, by a publisher, by an archivist, and so on. We use the attribute @hand to link them to a hand that is described in a <handNote> element.⁸ The @place attribute can adopt the same values as it can in the <add> element (see page 16).

For example, an author has written annotations in his own list:

⁸The TEI guidelines currently do not allow @hand to appear with <note>. We have decided to use it anyway because it makes much sense in our case.

```
<list>
  <note place="inline" hand="#black_ink">
    Erste Periode der
    <persName ref="#p0001">
      <choice><abbr>Shaksp.</abbr><expan>Shakespeare</expan></choice>
    </persName>
    Werke von 1584-92.
  </note>
  [...]
</list>
```

In case of footnotes, we use the @n attribute to denote the note number.

Signatures of archives are marked using the @type attribute:

```
<note type="signature" hand="#red_ink">2b1830</note>
```

On the other hand, there are *notes that we write ourselves*. They do not appear on the manuscript page, of course. For these notes, we do not use @hand, but @resp to link the note with the person that has written it, i. e. with their entry in the index of contributors.

For example:

```
<note resp="#sabine.seifert">Die Nibelungensage ist ein im deutschen
und skandinavischen Mittelalter weitverbreiteter heldenepischer
Stoff, der über Jahrhunderte in zahlreichen voneinander abweichenden
 Fassungen überliefert ist.</note>
```

Both the <note> and the <add> element may describe additions. We have discussed extensively inside the group on how to distinguish between cases where <note> should be used and cases where <add> should be used.⁹ We have decided that this cannot be told for the whole project and thus settled on making this decision dependent on the habits of each author.

Foliation numbers: <note>

We also use the <note> element to denote foliation numbers found on the manuscript page or made by us (e. g. if the foliation on the manuscript is wrong). They are marked <note>

⁹On the one hand, it had been proposed to differentiate between the two elements according to the position of a note on the manuscript page, and on the other hand, to differentiate between them according to the content of a note. Both possibilities, however, would not lead to consistent decisions between all texts in the edition and were thus judged inappropriate as a rule for the whole project.

Opener: <opener>

using type="foliation". As with other notes, they are linked with @hand or @resp, depending on who made them.¹⁰

For example:

```
<note type="foliation" hand="#red_ink">Foliation number on the
manuscript page</note>
<note type="foliation" hand="#pencil">Other foliation number on the
manuscript page</note>
<note type="foliation" resp="#johanna.preusse">A foliation made
by us</note>
```

Logical text structure: <note>

Furthermore, the <note> element may be used with type="structure" to denote numbers or other means of text structuring in a logical way, for example the number of an act in a drama (see also page 16). <note>

```
<head><note type="structure" resp="#johanna.preusse">2. Akt</note></head>
```

Stamps: <stamp>

The wording of stamps on the document can be marked up with the <stamp> element in the body. <stamp>

For example:

```
<stamp>Litteraturarchiv Berlin.</stamp>
```

3.2.2 Specific elements for encoding letters

In every letter, we use specific elements to encode its opening and its closing parts.

Opener: <opener>

According to the TEI guidelines, “<opener> groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter”. We use the <dateline> element to encode the place and the date of the letter, <salute> for the salutation. <opener>
<dateline>
<salute>

For example:

¹⁰Foliation numbers in the text are *not* linked directly to the corresponding <foliation> element in the header. The link between <foliation> and the responsible hand or person is done using the @corresp attribute with <foliation>. (See page 9.)

```

<opener>
  <dateline>
    <settlement>
      <placeName ref="#l1234">Dresden</placeName>
    </settlement>
    <date when-iso="1823-10-06">den 6t Octobr. 23</date>
  </dateline>
  <salute>Mein geliebter Freund,</salute>
</opener>

```

Closer: <closer>

Similarly, the TEI guidelines say: “<closer> groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter”. <salute> may also be used for the final salutations, <signed> for the signature. It is also possible to have a <dateline> at the end of a letter. Postscripts are encoded within the <postscript> element. We only integrate the specified elements in the <closer>, primarily the last salutations. Different (preceding) content in the last paragraph is not part of the <closer>, but of the running text.

<closer>
<salute>
<signed>
<dateline>
<postscript>

For example:

```

<p>Die Hohenstaufen habe ich jetzt erhalten, zu <lb/>meiner grossen
Freude.</p>
<closer><salute>Meine Grüsse Ihrer Frau und Kindern, <lb/>
mein herzlichen Dank.<lb/></salute>
<signed>Ihr L. Tieck</signed>
</closer>

```

3.2.3 Specific elements for encoding dramas**Speech: <sp>, <speaker>**

<sp> is used for speech with an attribute @who linking its speaker with the corresponding entry of the cast list in the back (see page 28). The name of the speaker or the speakers as given in the manuscript text are included in a <speaker> element.

<sp>
<speaker>

See this example:

```

<sp who="#Achmet"><speaker>Achmet.</speaker>...</sp>

```

Stage directions: <stage>

Stage directors are given in <stage> with the attribute @type. We use the values of @type as suggested in the TEI guidelines. However, sometimes it is hard to decide which value to use when more than one is possible. Here are some further explanations on how we interpret the meanings of the values and some examples of our uses of them:

- **setting:** We use this typically at the beginning of a scene when the figures on the stage are listed. It can also be used for other descriptions of stage setting as long as they don't relate to the location (use `location` instead) or any movement (use `business` instead).

```
<stage type="setting">Vorige. Mustapha.</stage>
<stage type="setting">(Sklaven stehn im Hintergrunde zur
Bedienung.)</stage>
```

- **entrance:**

```
<stage type="entrance">(Schahriar kommt über die Mauer
zurück.)</stage>
```

- **exit:**

```
<stage type="exit">(<hi rend="underline">sie gehn ab.</hi>)</stage>
```

- **business:** describes actions on stage. It is sometimes hard to decide when to use `business` and when to use `delivery` because both may describe actions (see below).

```
<stage type="business">(<hi rend="underline">Die Sklaven bereiten
eine kostbare Tafel.</hi>)</stage>
```

- **delivery:** describes how a figure speaks which may thus include what the figure does while speaking. Clear use cases are directions like "loud", "crying" or "embracing him". However, use may be extended to include everything the figure does during speaking. In these cases, it is left to the encoder to decide between `delivery` and `business`.

```
<sp who="#Solimann"><speaker>Solimann</speaker>
<stage type="delivery">(umarmt ihn<lb/> heftig)</stage><l>Nein!
&ndash; Das kann das Schicksal nicht verlangen!</l>
```

- **modifier:** is used to describe a figure in detail and also for non-moving, i. e. "passive" actions like "standing and thinking".

```
<sp who="#Ali">Ali,</speaker><stage type="modifier"> ein Greis
von<lb/> neunzig Jahren, sitzt in seiner<lb/> Höle, sein Bart
reicht<lb/> fast bis auf die Knie.</stage>
```

```
<l>Das Gift &ndash; &ndash; </l><stage type="modifier">(sie
sinkt)</stage>
```

- location: is typically used to describe the location at the beginning of a scene.

```
<stage type="location">(<hi rend="underline">Garten des
Raschid.</hi>)</stage>
```

- mixed:

```
<stage type="mixed">(er öffnet die Thür der Höle,<lb/>
Achmet tritt herein, der ehr<lb break="no"/>würdige Anblick
des Greises<lb/> macht daß er einige Augen<lb break="no"/>blikke
zurückstarrt, dann<lb/> tritt er näher.)</stage>
```

Verses: <l>, <lg>

Single lines of verses are encoded in <l>:

<l>

```
<l>Die Sonne glänzt am Himmel</l>
<l>Geschmückt mit tausend Strahlen</l>
```

```
<l>Werd ich denn hier nun end<lb break="no"/>lich
etwas finden? &ndash;</l>
```

<lg> may be used to group several <l> elements to form a logical unit, for example a song, a stanza or a verse paragraph. @type may be used to specify the type of line group:

<lg>

```
<sp who="#Musikerinnen">
  <lg type="song">
    <l>Die Sonne glänzt am Himmel</l>
    <l>Geschmückt mit tausend Strahlen</l>
    <l>der Mond beglänzt den Himmel</l>
    <l>Mit sanftem Silberscheine,</l>
    <l>So glänzt der edle Sultan,</l>
    <l>So seine schöne Gattinn.</l>
  </lg>
```

When a verse or a line group goes over more than one speech or if it is, for instance, broken by a page break, we use the @part attribute to denote this. Values of @part are "I" (initial), "M" (medial) and "F" (final). For example:

`<castList>`

```

<sp who="#Solimann"><speaker>Solimann</speaker>
<l part="I">Was ist das?</l></sp>
<sp who="#Mustapha"><speaker>Mustapha.</speaker>
<l part="M">Aufruhr!</l></sp>
<sp who="#Solimann"><speaker>Solimann.</speaker>
<l part="F">Ha, wer wagt so frevelhaft &ndash; &ndash;</l></sp>

```

3.3 Back matter

Currently, the only type of text where a `<back>` appears in our edition is drama. Here, we use it to group a cast list which gives an `@xml:id` to each role. *Please note that the cast list is compiled by us, i. e. by a contributor to the digital edition, and thus does not appear on the manuscript.*

`<back>`

`<castList>`

The whole cast list is included in the `<castList>` element, a single role is a `<castItem>` described by `<role>` and `<roleDesc>`. See this example:

```

<castList>
<castItem>
<role>
<roleDesc>

```

```

</body>
<back>
  <castList>
    <castItem>
      <role xml:id="Roxane">Roxane</role>
      <roleDesc>Sultanin</roleDesc>
    </castItem>
    <castItem>
      <role xml:id="Achmet">Achmet</role>
      <roleDesc>[...]</roleDesc>
    </castItem>
  </castList>
</back>

```

4 Index encoding

The digital edition includes five indexes:

- index of persons
- index of organizations
- index of places

- index of works
- index of contributors (to the digital edition)

One XML document is created for each of the indexes. This is the skeleton of each of the index documents:

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Index [...]</title>
      <funder>Deutsche Forschungsgemeinschaft:
      Emmy Noether-Programm</funder>
      <principal>
        <persName>
          <forename>Anne</forename>
          <surname>Baillot</surname>
        </persName>
        <affiliation>
          <orgName>Humboldt-Universität zu Berlin</orgName>
          <address>
            <street>Unter den Linden 6</street>
            <postCode>10099</postCode>
            <settlement>Berlin</settlement>
            <country key="DE">Deutschland</country>
          </address>
        </affiliation>
      </principal>
      <respStmt>[...]</respStmt>
    </titleStmt>
    <publicationStmt>
      <availability>
        <p><licence target="http://creativecommons.org/licenses/by/3.0/de/deed.en">Attribution 3.0 Germany (CC BY 3.0)</licence></p>
      </availability>
      <date when-iso=" [...]">[...]</date>
    </publicationStmt>
    <seriesStmt>
      <title type="main">Berliner Intellektuelle 1800-1830</title>
    </seriesStmt>
    <sourceDesc>
      <p>Born digital</p>
    </sourceDesc>
  </fileDesc>
```

`<listPerson>`

```

<revisionDesc>
  <change when="[...]" who="#vorname.nachname">[...]</change>
</revisionDesc>
</teiHeader>
<text>
  <body>[...]</body>
</text>

```

The TEI header is almost identical for all indexes. Please note that the `<sourceDesc>` `<sourceDesc>` element contains only one line: “born digital”, because it originates in a digital form.

The indexes are included inside the `<body>` element, i.e. the elements `<listPerson>`, `<body>`, `<listOrg>` and so on are direct sub-elements of `<body>`.

4.1 Index of persons

The index of persons comprises any person appearing in our texts, whether as an author, as an addressee of a letter, otherwise involved or simply mentioned in a manuscript. This index is to be distinguished from the index of contributors listing persons who have worked on this edition (see page 37 for the latter).

`<listPerson>`

The index consists of a `<listPerson>` element which contains one `<person>` element for each person described and one `<relationGrp>` element for describing all relationships between persons. This is the basic structure whose elements are described in more detail below: `<listPerson>`

```

<listPerson>
<person xml:id="[...]">
  <persName>
    <forename>[...]</forename>
    <surname>[...]</surname>
    [...]
  </persName>
  <birth>
    <date when-iso="[...]">[...]</date>
    <placeName>[...]</placeName>
  </birth>
  <death>
    <date when-iso="[...]">[...]</date>
    <placeName>[...]</placeName>
  </death>
  <sex value="1" />

```

`<person>`

```

<occupation>[...]</occupation>
<education>[...]</education>
<affiliation>[...]</affiliation>
<idno type="PND">...</idno>
</person>
<person xml:id="[...]">
  [...]
</person>
<relationGrp>
  <relation name="..." active="... .." passive="... .." />
  <relation name="..." mutual="... .." />
  [...]
</relationGrp>
</listPerson>

```

<person>

Every `<person>` element has an `@xml:id` attribute used for referencing the person from the manuscript texts and other places whose value is composed by “p” plus four different letters:

`<person>`

```

<person xml:id="p0023">
  <persName>
    <forename>Friedrich Heinrich</forename>
    <nameLink>von der</nameLink>
    <surname>Hagen</surname>
  </persName>
</person>

```

The following information can be given inside of `<persName>`:

`<persName>`

- `<forename>` is used for the first name. It can be completed by a `@type` attribute in order to precise whether this is the name usually used, the second name etc. `<forename>`
- `<surname>` is used for the family name. It can be completed by a `@type` attribute in order to precise whether this is a maiden or married name (`type="maiden"` or `type="married"`). `<surname>`
- `<nameLink>` is used for the link in aristocratic names. `<nameLink>`
- `<addName>` is used for a nickname or alias. `<addName>`
- `<roleName>` with `type="honorific"` is used to indicate academic titles that are “Dr.” or on a higher level than that.¹¹ `<roleName>`

¹¹We include any such academic titles here, regardless of whether they are a part of the name according to German law. We do not, however, include academic titles that are on a level below “Dr.”, for example “Magister” and “Diplom”.

`<person>`

- `<roleName>` with other values of @type is used indicate functions such as king, pope, princess ... `<roleName>`
- `<genName>` is used in order to distinguish between identical names belonging to different persons (in case of a difference of generation etc.), for example if there is a number after the first name of kings or the words “junior” and “senior”. `<genName>`



For example:

```
<persName>
  <forename>Stanislaus</forename>
  <genName>II.</genName>
  <addName>Augustus</addName>
  <surname>Poniatowsky</surname>
  <roleName type="nobility">König von Polen<date>1764-1798</date></
    roleName>
  <roleName type="nobility">Großfürst von Litauen</roleName>
</persName>
```

Apart from the name, we encode the following information on persons inside a `<person>` element:

- information on birth** using `<birth>` with `<date>` and `<placeName>` inside. `<birth>`
- information on death** using `<death>` with `<date>` and `<placeName>` inside. `<death>`
- sex** using `<sex>` with `<sex value="1" />` for a male person and `<sex value="2" />` for a female person. `<sex>`
- occupations** using the `<occupation>` element. Any kind of occupation or profession can be encoded here. Thus, `<occupation>` may appear more than once. `<occupation>`
- affiliations** using the `<affiliation>` element. `<orgName>` can be used inside this element in order to denote the name of an organization and to provide a pointer to the corresponding entry in the index of organizations. `<placeName>` may also be used if appropriate. `<affiliation>` can appear more than once. `<affiliation>`
- educational experiences** using the `<education>` element. Inside, `<placeName>` and `<orgName>` can be used, too, when appropriate. `<education>` can appear more than once. `<education>`
- PND number** using `<idno>`. The German-language *Personennamendatei* (PND) is a powerful authority file of people that is maintained by the German National Library. It is used by libraries, the Allgemeine Deutsche Biographie (ADB, a large German biographical resource) and several other parties. Thus, we decided to include the PND number of every person inside an `<idno>` element. For example: `<idno type="PND">118829130</idno>`.¹² Our web interface then provides links to the various other resources. `<idno>`

¹²Currently, the TEI guidelines do not allow `<idno>` inside `<person>` but we have decided to use it anyway because it fits our case very well.

For every element, we can use the attribute @cert to indicate the degree of certainty associated with its content. We use the values high, medium and low. When no @cert attribute is given, this means that the content is definitely certain.

<relationGrp>, <relation>

Relationships between persons are encoded using <relation> elements that appear inside one <relationGrp> which is included in <listPerson> after all <person> elements. <relation>
<relationGrp>

Here are some examples:

```
<relation name="parent" active="#p0001 #p0002" passive="#p0003 #p0004" />
<relation name="spouse" mutual="#p0001 #p0002" />
<relation name="lover" mutual="#p0001 #p0006" />
<relation name="teacher" active="#p0006 #p0004" />
```

These examples indicate the following: Persons 1 and 2 are parents of persons 3 and 4. Persons 1 and 2 are also married. Person 1, however, has a lover which is person 6. Person 6 is also a teacher of person 4.

Names of relations (in the @name attribute) can be chosen as appropriate. For now, we have decided on the following names describing basic relationships that appear often:

- parent: describes a parent-child relationship with the parents encoded in @active and the children encoded in @passive.
- spouse: describes a marriage relationship with both partners encoded in @mutual.
- lover: describes a non-marriage romantic relationship with both partners encoded in @mutual.
- sibling: describes a sibling relationship with the siblings encoded in @mutual. Only used if the parents are not included in the index of persons. (Otherwise the parent relationship is used.)
- teacher: describes an educational teacher-student relationship with the teacher encoded in @active and the student encoded in @passive.

Other names may be used as appropriate. For reasons of uniformity, however, we do not use the inverse relationships "child" (use parent instead) and "student" (use teacher instead).

4.2 Index of organizations

The format of this index still needs to be specified.



`<place>`

4.3 Index of places

`<listPlace>`

The places, which are described in this index, are surrounded by a `<listPlace>` element. Each place has an `@xml:id` attribute whose value is composed by "l" (location) plus four different letters.

Skeleton of the places index:

```
<listPlace>
  <place xml:id="l1234">
    <placeName>Dresden</placeName>
    <location>
      <region>Sachsen</region>
      <country key="DE">Deutschland</country>
    </location>
    <desc><ref target="https://maps.google.com/maps/api/staticmap? ...">See
the map</ref></desc>
  </place>
  <place xml:id="[...]">
    [...]
  </place>
</listPlace>
```

`<listPlace>`

`<place>`

There are as many `<place>` elements as needed. The name of the place, the region and the country (with a standardized `@key` attribute defined by the norm ISO 3166) are indicated for each item. The `<desc>` element contains a link to a map. We use the googlemaps API web services.

`<place>`

The address of the link always starts according to this form: `https://maps.google.com/maps/api/staticmap?`. The following parameters have to be added:

- center: name of the place
- zoom: level of the zoom
- size: size of the picture
- sensor: gets the value "false" in any case.

We can add a marker with the parameter "markers". All parameters are separated by the sign "&".

For example:

`<listBibl>`

```
<desc><ref target="https://maps.google.com/maps/api/staticmap?center=
Dresden&zoom=6&size=512x512&sensor=false">See the map</ref></desc>
```

4.4 Index of works

`<listBibl>`

For the moment, this index gathers works, monographs, articles, journals . . . , shortly all objects that can be described by bibliographical information. Each title has an `@xml:id` attribute whose value is composed by “w” (work) plus four different letters.



A `<biblStruct>` element may be factorized in three parts, which correspond to different levels of bibliographical information. We only use the level(s) that we need.

`<biblStruct>`

- `<analytic>` “contains bibliographic elements describing an item (e.g. an article or poem) published within a monograph or journal and not as an independent publication”.
- `<monogr>` “contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item”.
- `<series>` “contains information about the series in which a book or other bibliographic item has appeared”.

`<analytic>``<monogr>``<series>`

The `<analytic>` element contains information about author (`<author>`), title (`<title>`), editor (`<editor>`), and, if necessary, other responsible persons (we use the same syntax as in the `<respStmt>` element in the TEI header).

`<analytic>``<author>``<title>``<editor>``<monogr>`

The `<monogr>` element contains information on a superior level, about publication and distribution (`<imprint>`). Within this element there are the sub-elements `<pubPlace>` (the publication’s place), `<publisher>` (the publisher’s name), `<date>` (the publication’s date) and `<biblScope>` (this element is mainly used to precise the number of pages).

`<imprint>``<pubPlace>``<publisher>``<biblScope>`

This is the skeleton of the list of works:

```
<listBibl>
  <biblStruct>
    <analytic>
      <author>[...]</author>
      <title xml:id="w0023" level="a">[...]</title>
      <editor>[...]</editor>
    </analytic>
    <monogr>
      <title xml:id="w0024" level="j">[...]</title>
      <imprint>
```

```

    <biblScope type="vol">5</biblScope>
    <biblScope type="pp">11-14</biblScope>
    <publisher>[...]</publisher>
    <pubPlace>[...]</pubPlace>
    <date type="article" when-iso="1826">1826</date>
    <date type="journal" when-iso="1825/1827">1825 bis 1827</date>
  </imprint>
</monogr>
</biblStruct>
</listBibl>

```

We have to take a closer look at the three elements <title>, <biblScope> and <date>:

Title

We have all types of bibliographic elements in the work index. In this case, the <title> element requires a @level attribute which specifies the kind of object. The values for @level are:

- “a” for analytic title (article, poem ...)
- “m” for monographic title
- “j” for journal title
- “s” for series title
- “u” for unpublished material

BiblScope

<biblScope> can obtain several attributes to describe the encoded information.

- @type with several values. The most common values are “vol” (volume number, also used for the issue number of a journal) and “pp” (the number of pages or the extent).
- @from “specifies the starting point of the range of units”.
- @to “specifies the end-point of the range of units”.

Date

The <date> element sometimes needs a @type attribute in order to specify whether it tags the publication period of the journal (type="journal") or the publication date of the article (type="article"). These dates are encoded in a standardized form like described above.

4.5 Index of contributors

The format of this index still needs to be specified.



List of TEI elements (with page numbers of their descriptions)

<abbr>, 18
<accMat>, 11
<add>, 16
<addName>, 31
<addSpan>, 18
<affiliation>, 32
<analytic>, 35
<anchor>, 18
<author>, 5, 35

<back>, 28
<biblScope>, 35
<biblStruct>, 35
<birth>, 32
<body>, 13, 30

<castItem>, 28
<castList>, 28
<change>, 13
<choice>, 18
<closer>, 25
<corr>, 18
<correction>, 12
<correspDesc>, 11

<date>, 21
<dateline>, 24, 25
<death>, 32
, 17
<delSpan>, 18
<div>, 13, 15
<docDate>, 8

<editor>, 5, 35
<editorialDecl>, 12
<education>, 32
<encodingDesc>, 12
<expan>, 18

<fileDesc>, 4
<foliation>, 9

<foreign>, 22
<forename>, 31

<gap>, 19
<genName>, 32

<handDesc>, 9
<handNote>, 9
<head>, 16
<hi>, 20
<history>, 11
<hyphenation>, 12

<idno>, 7, 32
<imprint>, 35
<institution>, 7

<l>, 27
<label>, 21
<language>, 13
<langUsage>, 13
<lb>, 14
<lg>, 27
<list>, 21
<listPerson>, 30
<listPlace>, 34

<monogr>, 35
<msContents>, 8
<msDesc>, 7
<msIdentifier>, 7
<msItem>, 8

<nameLink>, 31
<normalization>, 12
<note>, 22–24

<objectDesc>, 8
<occupation>, 32
<opener>, 24
<orgName>, 21
<origDate>, 11

List of TEI elements (with page numbers of their descriptions)

<origin>, 11
<origPlace>, 11

<pb>, 14
<pc>, 18
<persName>, 21, 31
<person>, 31
<physDesc>, 8
<place>, 34
<placeName>, 21
<postscript>, 25
<principal>, 5
<profileDesc>, 13
<projectDesc>, 12
<publicationStmt>, 6
<publisher>, 35
<pubPlace>, 35

<relation>, 33
<relationGrp>, 33
<repository>, 7
<respStmt>, 6
<revisionDesc>, 13
<role>, 28
<roleDesc>, 28
<roleName>, 31, 32
<rs>, 22

<salute>, 24, 25
<sealDesc>, 11
<series>, 35
<seriesStmt>, 7
<sex>, 32
<sic>, 18
<signed>, 25
<sourceDesc>, 7, 30
<sp>, 25
<speaker>, 25
<stage>, 26
<stamp>, 24
<subst>, 17
<supportDesc>, 8
<surname>, 31

<teiHeader>, 4

<text>, 13
<title>, 5, 21, 35
<titleStmt>, 5
<trailer>, 16

<unclear>, 19