Appendix D - Specification of transformation of scans to TEI-XML

This appendix specifies how scans should be converted to TEI-XML.
The requirements and examples are numbered.

Example source texts are shown in sans serif font Calibri.
Example precoded texts are shown in handwriting font *Segoe Script* or Arial.
The actual precoded scans may contain handwritten digital coding or typed coding. All coding is red. In special situations the colour of the precoding can differ if that makes it more legible. Regardless, we aim to make it easily distinguishable from the original text.
Literal TEI-XML is shown in monospaced font `Courier`. An ellipsis (…) represents arbitrary TEI-XML.

1 General requirements 2
2 The overall structure of the TEI-XML file 3
3 Chapters and headings 4
4 Paragraphs 8
5 Colophon 10
6 Interp 12
7 Poems 14
8 Plays 18
9 Line beginning tags 20
10 Pages and page begin tags 22
11 Lists 24
12 Tables 26
13 Emphasis 31
14 Hyphenation 32
15 Notes 33
16 Images 35
17 Spaces in abbreviations 39
18 Character-level rules 40
19 White spaces 44
20 Cross-references and links 46

1 General requirements

1. The rules for the transformation of scans to TEI-XML described in this document may be overruled by instructions written in the instruction document or in the precoded scan.

2. The character encoding of the TEI-XML file should be UTF-8.

3. The TEI-XML file should adhere to the following DTD: https://www.dbnl.org/xml/dtd/teixlite.dtd

4. The TEI-XML file should be a valid XML version 1.0 document, as specified by W3C (https://www.w3.org/TR/REC-xml).

5. There should not be any white space characters at the start of a line.

6. Every XML tag specified in chapters 2 to 8 (including) should appear at the start of a line, except for end-tags (i.e. tags that start with `</`).

End-tags and tags specified in other chapters may appear anywhere, including at the start of a line, unless otherwise specified.

The examples in this document show TEI-XML-code divided over separate lines. It is not required to produce lines of code in exactly the same way, as long as the requirements stated above are satisfied.

7. If a piece of source text or precode is not readable, or it is otherwise not clear how to transform a piece of source text to TEI-XML, then `{problem}` should be inserted at the position of the problem and the problem should be described in a problem document.

8. If a cross is drawn through a piece of source material on the precoded scan, then this piece of source material should be skipped.

9. If a certain pattern occurs many times in the source material, only a few instances will be precoded and it will be indicated that the same precodes apply to all other occurences of the pattern as well. In that case, the source material should be converted to TEI-XML as if the precodes were present on all occurrences of the pattern.

Boer: ← speaker
'Hoe heet uw kat eigenlijk?'
Psycholoog: ← speaker, etc.
'Waarom vraagt u dat?'
Boer:
'Nou, ik zag die kat van u langs mij schieten.'
Psycholoog:
'Het gaat om u, niet om de kat.'
Boer:
'Het is toch uw kat? Hoe heet-ie nou?'
	`<sp>`
	`<speaker>Boer:</speaker>`
	`<p>'Hoe heet uw kat eigenlijk?'</p></sp>`
	`<sp>`
	`<speaker>Psycholoog:</speaker>`
	`<p>'Waarom vraagt u dat?'</p></sp>`
	`<sp>`
	`<speaker>Boer:</speaker>`
	`<p>'Nou, ik zag die kat van u langs mij schieten.'</p></sp>`
	`<sp>`
	`<speaker>Psycholoog:</speaker>`
	`<p>'Het gaat om u, niet om de kat.'</p></sp>`
	`<sp>`
	`<speaker>Boer:</speaker>`
	`<p>'Het is toch uw kat? Hoe heet-ie nou?'</p></sp>`

Example 1

2 The overall structure of the TEI-XML file

2.1 Top-level tags

1. The overall structure of the TEI-XML file should be as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI.2 PUBLIC "-//DBNL//DTD TEI.2 XML//NL" "http://www.dbnl.org/xml/dtd/teixlite.d
<TEI.2>
<teiHeader>
       ≪the TEI header≫
</teiHeader>
<text>
       ≪optional first page interp group≫
<body>
       ≪the TEI body≫
</body>
</text>
</TEI.2>
```

Example 2

2. The TEI header is mandatory, but its content is not important. It will be replaced completely by KB. The TEI header may contain any data, as long as it adheres to the teixlite.dtd.

3. Interps are described in chapter 6. Interps written on the first page of the precoded scan should be inserted as an interp group between `<text>` and `<body>`.

2.2 The TEI body

1. The TEI body consists of one or more chapter divisions.
2. A chapter division is represented by the tag `<div type="chapter"> … </div>`. In case of a play, the type attribute has value `"act"` (see chapter 8).
3. A chapter or a sequence of chapters may be enclosed within `<cf>` and `</cf>`, as explained in chapter 5.
4. All other tags, described in the rest of this document, are placed inside a chapter division.

For more information about chapters, see chapter 3.

```
       <body>
       <div type="chapter">
              ≪content≫
       </div>
       <cf>
       <div type="chapter">
              ≪content≫
       </div>
       </cf>
       <div type="chapter">
              ≪content≫
       </div>
       </body>
```

Example 3

3 Chapters and headings

3.1 Default chapter, section and heading tags

If the source material has a clear division in chapters, then the precoded scan or the instruction document usually won't give any specific instructions on how to tag the chapters. In that case, the following rules apply.

1. Each chapter in the book should be represented by a chapter division, as specified in section 2.2.

2. The chapter title (if present) should be enclosed within `<head rend="h3">` and `</head>`.

3. Parts of the source material outside of the book chapters, like a title page or a table of contents, should be represented by a chapter division as well.

4. If a chapter is divided into sections, as identified by their headings, then each section should be enclosed within `<div type="section">` and `</div>` and its heading should be enclosed within `<head rend="h4">` and `</head>`.

5. Any subsections, subsubsections etc. should be represented by nested `<div type="section">` tags. Their headings should be enclosed within `<head>` and `</head>`.

Hoofdstuk 12
 	
We gaan nu nader in op het verschijnsel taal.
 	
12.1 Het taalbegrip
 	
Al vroeg in hun leven beginnen kinderen iets te leren over de taal.
 	
Het eerste levensjaar
…
	`<div type="chapter">`
	`<head rend="h3">Hoofdstuk 12</head>`
	`<p>We gaan nu nader in op het verschijnsel taal.</p>`
	`<div type="section">`
	`<head rend="h4">12.1 Het taalbegrip</head>`
	`<p>Al vroeg in hun leven beginnen kinderen iets te leren over de taal.</p>`
	`<div type="section">`
	`<head>Het eerste levensjaar</head>`
	`...`
	`</div></div></div>`

Example 4

6. If a heading is immediately preceded by another heading, then the inner div tags may be omitted. 'Immediately' means that there is only white space or a page break in between.

Hoofdstuk 12
 	
12.1 Het taalbegrip
Het eerste levensjaar
…
	`<div type="chapter">`
	`<head rend="h3">Hoofdstuk 12</head>`
	`<head rend="h4">12.1 Het taalbegrip</head>`
	`<head>Het eerste levensjaar</head>`
	`...`
	`</div>`

Example 5

7. Chapter divisions should not be nested.

8. Multiple sections on the same level should not be nested.

9. If there are two chapter levels, e.g. Part I, containing chapters 1 to 7 and Part II containing chapters 8 to 14, then the following rules are applicable:

   9.1. Both levels should be represented by chapter divisions. See Example 6.

9.2. The heading of the outer level chapter is enclosed within `<head rend="h2">` and `</head>` and the head of the inner level within `<head rend="h3">` and `</head>`. See Example 6.

9.3. The exception to rule 9.1 is the first inner level chapter in each outer level chapter of the book (in the above-mentioned example: chapters 1 and 8). If there is *no* content between the heading of the outer level chapter and the heading of the first inner level chapter of the book, then rule 6 is applicable, which means that there should not be a div-tag for the first inner level chapter. See Example 6.

9.4. If there is content between the heading of the outer level chapter and the heading of the first inner level chapter of the book, then precodes will indicate how this should be captured. See Example 7 in which the first inner level chapter is indicated with precodes to be part of the outer level chapter of the book. Hence, the first inner level chapter becomes a section.

9.5. In any of the cases mentioned the outer level chapter division should be closed before the second inner level chapter division begins: chapters may not be nested.

Part I h2
Chapter 1 h3
…
Chapter 2 h3
…
	`<div type="chapter">`
	`<head rend="h2">Part I</head>`
	`<head rend="h3">Chapter 1</head>`
	…
	`</div>`
	`<div type="chapter">`
	`<head rend="h3">Chapter 2</head>`
	…
	`</div>`

Example 6

Same chapter
Part I h2
Lorem ipsum dolor sit amet.
Chapter 1 h3
…
Chapter 2 h3
…
	`<div type="chapter">`
	`<head rend="h2">Part I</head>`
	`<p>Lorem ipsum dolor sit amet.</p>`
	`<div type="section">`
	`<head rend="h3">Chapter 1</head>`
	…
	`</div></div>`
	`<div type="chapter">`
	`<head rend="h3">Chapter 2</head>`
	…
	`</div>`

Example 7

3.2 Precoded heads

1. Headings are indicated in the precoded scans as *h2*, *h3*, *h4* or *head*. The latter can also be shortened to *h*. Capture these headings using the start tag `<head rend="h2">`, `<head rend="h3">`, `<head rend="h4">` or `<head>` respectively and the end-tag `</head>`.

2. The h2 and h3 headings should be considered as the start of chapters and the other headings should be considered as the start of sections. The rules specified in 3.1 are applicable.

h2 Deel I h3 Hoofdstuk 1 h3 Eerste bedrijf …	 <head rend="h2">Deel I</head> <head rend="h3">Hoofdstuk 1</head> <head rend="h3">Eerste bedrijf</head> …
h4 **2. De achttiende eeuw** head 2.1 Napoleon Napoleon werd geboren. head 2.2 Staatsgreep Hij greep de macht.	 <head rend="h4">2. De achttiende eeuw</head> <head>2.1 Napoleon</head> Napoleon werd geboren. <head>2.2 Staatsgreep</head> Hij greep de macht.

Example 8

3.3 Chapter starts before its heading

Sometimes a chapter starts before its heading, because some text or an illustration is present above the heading, or even on the previous page. In that case, there is an indication on the precoded scan or in the instruction document about where the chapter starts. The chapter head should then be included in a section.

chapter starts here Lorem ipsum dolor sit amet. h3 Lorem ipsum …	 Lorem ipsum dolor sit amet. <head rend="h3">Lorem ipsum</head> …

Example 9

3.4 Headings to be added to the original

If the precoded scan or the instruction document instructs to add headings, for example to chapters or poems that don't have a heading, these should be typed between square brackets.

please add h3 [Lorem ipsum] Lorem ipsum dolor sit amet.	 <head rend="h3">[Lorem ipsum]</head> Lorem ipsum dolor sit amet.

Example 10

3.5 Surtitle, Subtitle and Captions

Additionally, chapters and sections headings with h2 and h3 might be provided with precodes to indicate surtitles, subtitles and captions.

1. The precode sur or surtitle should be captured between `<head rend="surtitle">` and `</head>`
2. The precode sub or subtitle should be captured between `<head rend="subtitle">` and `</head>`
3. The precode cap or caption should be captured between `<head rend="caption">` and `</head>`

This is illustrated in Example 11.

sur Hoofdstuk 1 h3 **Een voetreis naar Rome** sub *Het leven van Bertus Aafjes* cap Door Marcel Janssens Lorem ipsum dolor sit amet.	 <head rend="surtitle">Hoofdstuk 1</head> <head rend="h3">Een voetreis naar Rome</head> <head rend="subtitle">Het leven van Bertus Aafjes</head> <head rend="caption">Door Marcel Janssens</head> Lorem ipsum dolor sit amet.

Example 11

3.6 Uppercase and lowercase

1. If a heading is completely printed in uppercase *or* completely in small capitals, the precoded scan indicates with a double underline which letters are to remain capitalized. The letters that are not underlined should be captured in lowercase.

2. If a heading consists of uppercase *and* lowercase, or uppercase *and* small capitals, the casing should be captured as how it is shown in the source material.

h3 OP WEG NAAR PARIJS	<head rend="h3">Op weg naar Parijs</head>
h3 OP WEG NAAR PARIJS	<head rend="h3">op weg naar parijs</head>
h3 OP WEG NAAR PARIJS	<head rend="h3">Op weg naar Parijs</head>
h3 Op weg naar Parijs	<head rend="h3">Op weg naar Parijs</head>

Example 12

4 Paragraphs

1. Paragraphs should be enclosed within and . Whether the first line is indented or not is ignored.

2. White lines between paragraphs (i.e. empty space, more than the regular paragraph spacing used in the text) should be coded as   . See chapter 19 for more rules on white lines.

3. Simple decorative paragraph separators are ignored. See chapter 16 for more information about decorations.

Ik keek een beetje schuins achter mij en zag rijen boeken in dezelfde band: Jung, Freud en verder veel ander geleerd Duits. 'Hoe heet uw kat eigenlijk?' vroeg ik. 'Waarom vraagt u dat?' 'Nou, ik zag die kat van u langs mij schieten – hij is toch van u?'  De therapie zou jaren in beslag nemen. Soms fietste ik wel vijf keer per week naar de stille laan.	Ik keek een beetje schuins achter mij en zag rijen boeken in dezelfde band: Jung, Freud en verder veel ander geleerd Duits. 'Hoe heet uw kat eigenlijk?' vroeg ik. 'Waarom vraagt u dat?' 'Nou, ik zag die kat van u langs mij schieten – hij is toch van u?' De therapie zou jaren in beslag nemen. Soms fietste ik wel vijf keer per week naar de stille laan.

Example 13

4. If a paragraph continues on the next page, a page begin tag should be inserted at the right position (somewhere between and). See chapter 10, rule 7 for an example. The paragraph should not be closed at the page break.

5. Precode *speaker* may be written next to a paragraph. In that case, the marked text should be placed between <speaker> tags before the subsequent paragraph with tags. Both the <speaker> and element are then placed between <sp> tags. This only applies in plays, see chapter 8.

Boer: ← *speaker* 'Hoe heet uw kat eigenlijk?' Psycholoog: ← *speaker* 'Waarom vraagt u dat?' Boer: ← *speaker* 'Nou, ik zag die kat van u langs mij schieten.'	<sp> <speaker>Boer:</speaker> 'Hoe heet uw kat eigenlijk?' </sp> <sp> <speaker>Psycholoog:</speaker> 'Waarom vraagt u dat?' </sp> <sp> <speaker>Boer:</speaker> 'Nou, ik zag die kat van u langs mij schieten.' </sp>

Example 14

6. An indented paragraph is marked by the precode *bq*. This means 'block quote'. In that case, the following rules apply:
   6.1. The tag pair … should be replaced by <q rend="bq">…</q>.
   6.2. White lines before or after the block quote should be ignored.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit.
bq Aenean commodo ligula eget dolor tanuem massa.
Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.
	`<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit.</p>`
	`<q rend="bq">Aenean commodo ligula eget dolor tanuem massa.</q>`
	`<p>Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.</p>`

Example 15

See also chapter 9, rule 3 regarding new and empty lines in block quotes.

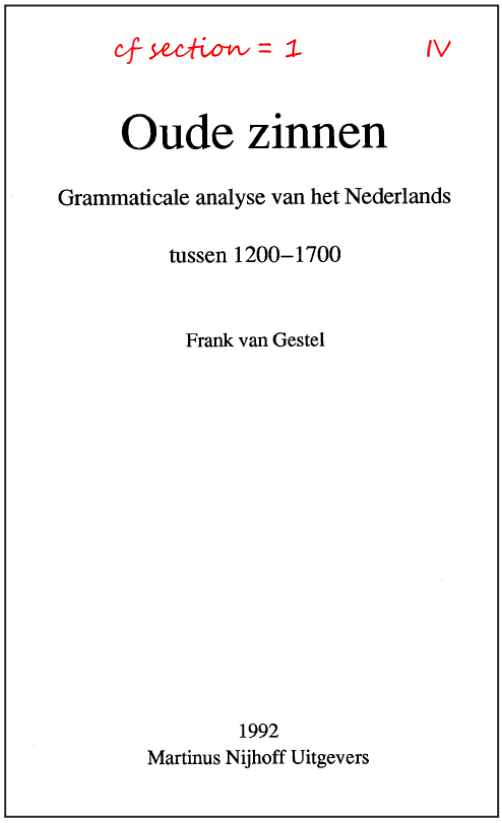
5 Colophon

1. Complete pages or individual paragraphs marked by *cf* should be placed between colophon tags, i.e. between `<cf>` and `</cf>`.

2. If precode *cf* is written at the top of a page, then the whole page should be placed between colophon tags, including tags like `<pb>` (this tag is described in chapter 10).

3. If precode *cf section = n* is written then the following XML tag should be used: `<cf section="n">`.

4. If all text that belongs to a `<div>…</div>` is marked by *cf*, then the `<div>` and `</div>` tags should also be placed between colophon tags.

<table>
<tr>
<td>
</td>
<td><pre>
<cf section="1">
<div type="chapter">
<pb n="IV"/>
<p>Oude zinnen</p>
<p>Grammaticale analyse van het
Nederlands</p>
<p>tussen 1200-1700</p>
<p>Frank van Gestel</p>
<p>1992</p>
<p>Martinus Nijhoff Uitgevers</p>
</div>
</cf>
</pre></td>
</tr>
<tr>
<td>Lorem ipsum dolor sit amet, consectetuer adipiscing elit.

Aenean con ligula eget dolor.</td>
<td><pre>
<p>Lorem ipsum dolor sit amet, consectetuer
adipiscing elit.</p>
<cf>
<p>Aenean con ligula eget dolor.</p>
</cf>
</pre></td>
</tr>
</table>

Example 16

Often the first and last printed page(s) of a scan contain(s) the front, back and or inside cover(s) of the book (in Dutch: ‘voorplat’ and ‘achterplat’). This is often captured as a colophon. If this is the case, it will be indicated with precodes.

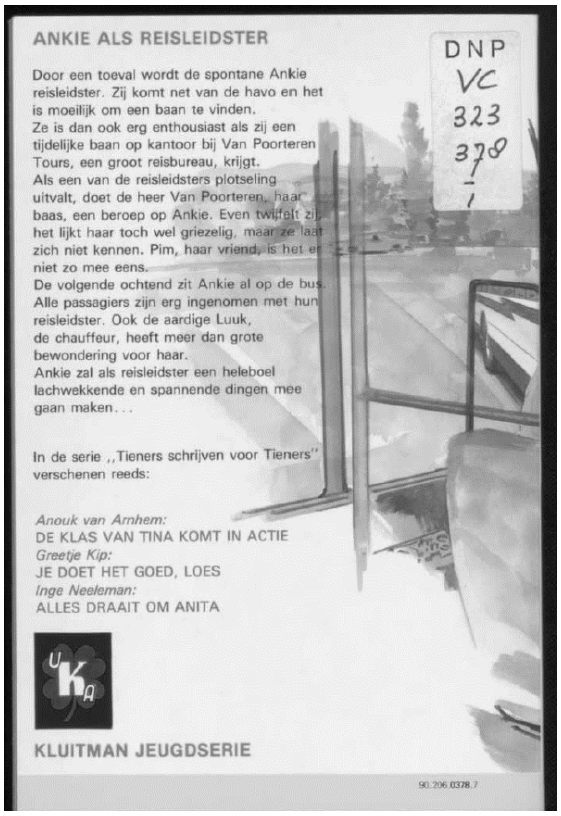
Part of the text on (one of) these pages can be precoded as ‘blurb’. This part should be tagged as `<cf section=”blurb”>`.

The following rules apply:

1. Do not add library-stickers to the xml-file.
2. If a sticker covers (part of) the text, put {problem} in the file at the position of the unreadable text.
3. If a sticker covers (part of) an image, ignore this and do not make an image.
4. In case of a fold-out page, put the inside-part alongside the voorplat or achterplat.

cf achterplat

blurb




```
<cf section="blurb">
<div type="chapter">
<pb n="achterplat"/>
<p>ANKIE ALS REISLEIDSTER</p>
<p>Door een toeval wordt de
spontane Ankie reisleidster.
Zij komt net van de havo en het
is moeilijk om een baan te
vinden. Ze is dan ook erg
enthousiast als zij een
tijdelijke baan op kantoor bij
Van Poorteren Tours, een groot
reisbureau, krijgt. ...</p>
</blurb>
<p>In de serie “Tieners
schrijven voor Tieners”
verschenen reeds</p>
<p>Anouk van Arnhem</p>
<p>...</p>
<p>...</p>
<figure><p>
<xptr to="«image file
name»"/></p></figure>
<p>KLUITMAN JEUGDSERIE</p>
</div>
</cf>
```

Example 17

6 Interp

The tag <interp> ('interpretation') is used to specify that a publication, a part of a publication or an illustration pertains to a particular entity, which can be a person, an organization, a location, a time or a publication. The tag always has two attributes: type and value. The type specifies the relationship of entity with the text and the value is the identifier of the entity. For example, <interp type="secundair" value="19830400"> means that the text describes events that happened in April, 1983.

Insertion of <interp> tags is always requested explicitly by precode *int* or *interp*. Such a precode can be on the first page, near a head of type h2, h3 or h4, or near an illustration. The type and the value are written behind this precode, e.g. *int sec 19830400*.

1. If the precode is "*int «type» «value»*" or "*interp «type» «value»*", then the tag should be <interp type="*«type»*" value="*«value»*"/>. Sometimes abbreviations will be used for the <<*type*>>. The abbreviations listed below, should always be translated as stated. If there are abbreviations used for other types, then there will be specific instructions in the precode scan on how to translate these to XML. The full list of possible types is listed in Table 2.

Table 1

Type in precode	Type in TEI-XML
pri	primair
sec	secundair
bio	biografie
ill	illustrator

Table 2

primair
vertaler
zender
biografie
illustrator
secundair
editeur
ontvanger
beeld
taalkunde

2. Every interp tag, even if there is only one, should be part of an interp group. An interp group contains a set of interps that all pertain to the publication as a whole, or all pertain to the same part of the publication or all pertain to the same illustration. The interp tags should be enclosed within <interpGrp> and </interpGrp>.

3. If the precode *int* or *interp* is written on the first page of the scan, then the interp pertains to the publication as a whole. The interp group that contains such interps should be inserted between <text> and <body>. See chapter 2, rule 3.

4. If the precode *int* or *interp* is written near a head of type h2 or h3, then the interp tag should be included in an interp group right after the tag of the corresponding chapter.

5. Interps with the type "zender" or "ontvanger", *or* interps with the type "primair" or "secundair" *and* a value consisting of 8 numbers, can also be written near a head of type h4. In that case, the interp group should be placed right after the tag of the corresponding section. Other interps (except the one in rule 6 below)

should always be treated as explained above in rule 4 (even if they seem to be written near a head of type h4).

6. If the precode *int* or *interp* is written near an illustration, then the interp tag should be included in an interp group right after the `<figure>` tag. This only applies to interps of the type “beeld”.

Hoofdstuk 13 h3
 	
int pri deco030
int pri abic001
 	
Lorem ipsum dolor sit amet.
	`<div type="chapter">`
	`<interpGrp>`
	`<interp type="primair" value="deco030"/>`
	`<interp type="primair" value="abic001"/>`
	`</interpGrp>`
	`<head rend="h3">Hoofdstuk 13</head>`
	`<p>Lorem ipsum dolor sit amet.</p>`

 interp beeld gou_001	`<figure>`
	`<interpGrp>`
	`<interp type="beeld" value="gou_001"/>`
	`</interpGrp>`
	`<p><xptr to="`*«image file name»*`"/>`
	`</p>`
	`</figure>`

Example 18

7. Interp types and values are always in lowercase and contain no spaces. Values can only consist of letters, numbers or underscores, no other characters are allowed.

7 Poems

1. Poems are marked in the precoded scan as *poem*.

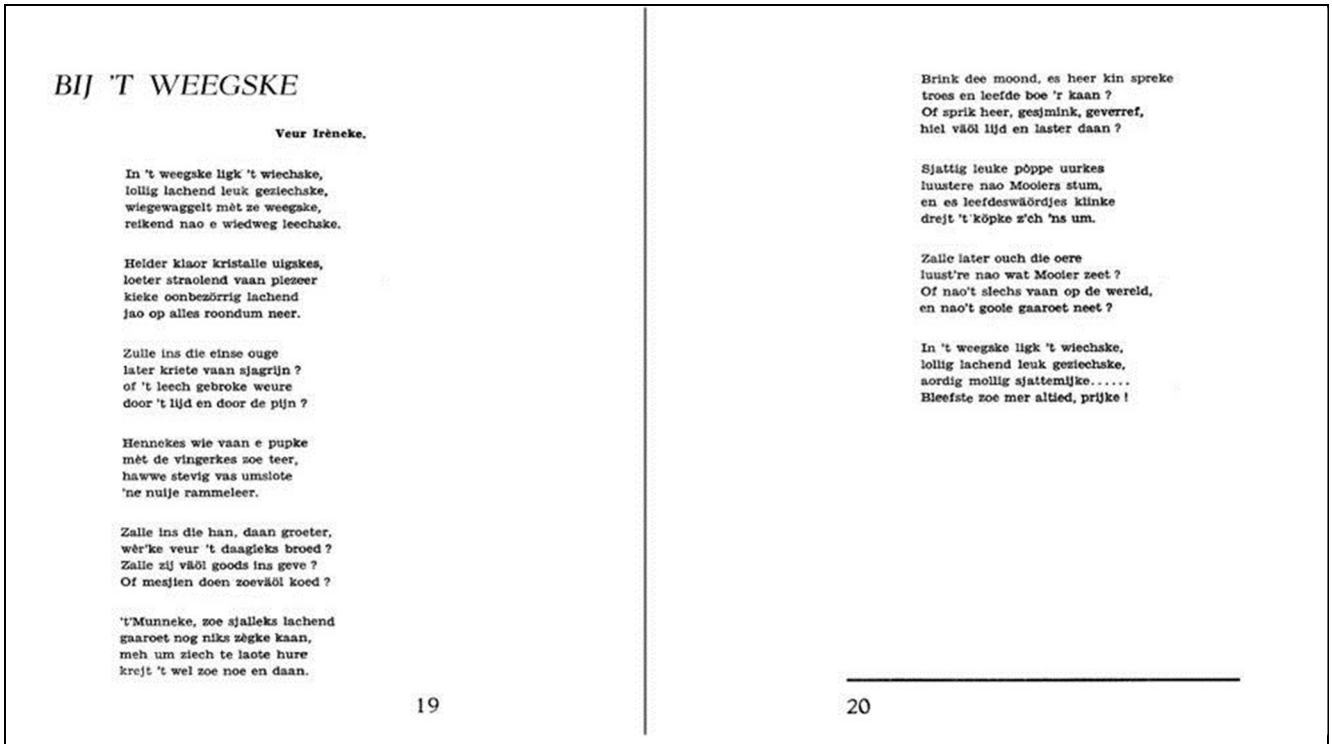
2. Poems should be enclosed within `<lg type="poem">` and `</lg>`.

3. A poem can contain a `<head>`. In that case, the tag `<head>` should appear after the tag `<lg type="poem">`, without a `<div>` tag. This only applies if the head belongs to the poem, not to an overarching section that begins with the poem. If the head belongs to the poem this will be indicated by precodes as shown in Example 20. Furthermore, the capitalization rules defined in paragraph 3.5 on page 6 are applicable.

4. Each poem line should be enclosed within `<l>` and `</l>`.

5. White lines within a poem should be captured as `<l> </l>`.

6. White lines should be added if the end of a verse coincides with the end of a page in the original. In this case it should be clear from the structure of a poem that a new verse starts on the next page. In Example 19 it is clear from the structure of the poem that each verse consists of four lines, so a new verse starts on page 20 and a white line should be added accordingly. If the length of the verses varies, and it is therefore not clear from the structure whether the new page starts with a new verse, no white line has to be added.



BIJ 'T WEEGSKE

Veur Irèneke.

In 't weegske ligk 't wiechske,
lollig lachend leuk geziechske,
wiegewaggelt mèt ze weegske,
reikend nao e wiedweg leechske.

Helder klaor kristalle uigskes,
loeter straolend vaan piezeer
kieke oonbezörrig lachend
jao op alles roondum neer.

Zulle ins die einse ouge
later kriete vaan sjagrijn ?
of 't leech gebroke weure
door 't lijd en door de pijn ?

Hennekes wie vaan e pupke
mèt de vingerkes zoe teer,
hawwe stevig vas umslote
'ne nuije rammeleer.

Zalle ins die han, daan groeter,
wèr'ke veur 't daagleks broed ?
Zalle zij vääl goods ins geve ?
Of mesjien doen zoevääl koed ?

't'Munneke, zoe sjalleks lachend
gaaroet nog niks zègke kaan,
meh um ziech te laote hure
krejt 't wel zoe noe en daan.

19

Brink dee moond, es heer kin spreke
troes en leefde boe 'r kaan ?
Of sprik heer, gesjmink, geverref,
hiel vääl lijd en laster daan ?

Sjattig leuke pöppe uurkes
luustere nao Mooiers stum,
en es leefdeswäördjes klinke
drejt 't köpke z'ch 'ns um.

Zalle later ouch die oere
luust're nao wat Mooier zeet ?
Of nao't slechs vaan op de wereld,
en nao't goole gaaroet neet ?

In 't weegske ligk 't wiechske,
lollig lachend leuk geziechske,
aordig mollig sjattemijke.....
Bleefste zoe mer altied, prijke !

20

```
<lg type=”poem”>
…
<l>meh um ziech te laote hure</l>
<l>krejt 't wel zoe noe en daan</l>
<l> </l>
<pb n=”20”/>
<l>Brink dee moond, es heer kin spreke</l>
<l>troes en leefde boe 'r kaan?</l>
…
</lg>
```

Example 19

7. White lines that are part of the head or outer structure of the poem are to be ignored. Thus the following rules apply:

   7.1. No white line directly before or after `<lg type="poem">`
   7.2. No white line directly before or after `</lg>`
   7.3. No white line directly before or after `<head>` or `</head>`

head **Lorem ipsum**
poem Lorem ipsum dolor sit
Cum sociis natoque it
Aenean commodo ligula eget dolor.
	`<div type="section">`
	`<head>Lorem ipsum</head>`
	`<lg type="poem">`
	`<l rend="; 1tab">Lorem ipsum dolor sit</l>`
	`<l rend="; 1tab">Cum sociis natoque it</l>`
	`</lg>`
	`<p>Aenean commodo ligula eget dolor.</p>`
	`</div>`

Lorem ipsum head
poem Lorem ipsum dolor sit
Cum sociis natoque it
Aenean commodo ligula eget dolor.
	`<lg type="poem">`
	`<head>Lorem ipsum</head>`
	`<l rend="; 1tab">Lorem ipsum dolor sit</l>`
	`<l rend="; 1tab">Cum sociis natoque it</l>`
	`</lg>`
	`<p>Aenean commodo ligula eget dolor.</p>`

Example 20

8. Poem lines may have line numbers. The line number should be represented by the value of attribute `n` of the line tag, e.g. `<l n="5">`. Line numbers should only be captured where they appear in the original text. See Example 21. Note that this is different for proze, see chapter 9 rule 4.

Verkapt chanson Head
poem 1. Bonsoir mon amour - de rozen
op je kleed doen mij onwillekeurig
denken aan het ziltgeurig
onderkleed en het schaamteloze
5. van toen je onlangs dronken
lachende stond te pronken
met hetgeen aan struikgewas
onder het kleed met die rozen was.
	`<div type="section">`
	`<head>Verkapt chanson</head>`
	`<lg type="poem">`
	`<l n="1">Bonsoir mon amour - de rozen</l>`
	`<l>op je kleed doen mij onwillekeurig</l>`
	`<l>denken aan het ziltgeurig</l>`
	`<l>onderkleed en het schaamteloze</l>`
	`<l> </l>`
	`<l n="5">van toen je onlangs dronken</l>`
	`<l>lachende stond te pronken</l>`
	`<l>met hetgeen aan struikgewas</l>`
	`<l>onder het kleed met die rozen was.</l>`
	`</lg>`
	`</div>`

Example 21

9. If a poem line is indented, attribute `rend` should be added to the line tag. The value of the attribute is `"; 1tab"`. If the poem lines have different indentation levels, use `"; 1tab"` for the least indented line, `"; 2tab"` for the next level, etc. The maximum level is 9, i.e. `"; 10tab"` is not permitted.

10. The source text sometimes shows one or more additional lines below the poem, for example the name of the author or a reference to the source of the poem.

    10.1. These lines should be enclosed within `<l>` and `</l>`. In the source text, these lines are usually right aligned, but in the TEI-XML, the indentation should be ignored.

10.2. These lines should be separated from the poem by a white line: `<l> </l>`.

poem

'k Sal als Diana u dit luck gansch niet benijen,
Maer mij van harten daer volkomen in verblijen;
Ick heb al over langh de deughden van uw Bruyt
Gehoort, doch korts gekent; doe pickte ik haer uyt,
En gunde haer yet meer als d' andre Jonckvrouwen,
En siet nu sal se één, die ick haer gunde, trouwen.
Dan 't spyt mij, dat ick niet kan ieder maken kondt,
Hoe veel geluck en heyl dat U mijn herte gont.

U beyder Jonstige Vriendin,
ANNA ROEMERS.

```
<lg type="poem">
<l>'k Sal als Diana u dit luck gansch niet benijen,</l>
<l>Maer mij van harten daer volkomen in verblijen;</l>
<l rend="; 1tab">Ick heb al over langh de deughden van uw Bruyt</l>
<l rend="; 1tab">Gehoort, doch korts gekent; doe pickte ik haer uyt,</l>
<l>En gunde haer yet meer als d' andre Jonckvrouwen,</l>
<l>En siet nu sal se één, die ick haer gunde, trouwen.</l>
<l rend="; 1tab">Dan 't spyt mij, dat ick niet kan ieder maken kondt,</l>
<l rend="; 1tab">Hoe veel geluck en heyl dat U mijn herte gont.</l>
<l> </l>
<l>U beyder Jonstige Vriendin,</l>
<l>ANNA ROEMERS.</l>
</lg>
```

Example 22

11. A poem may have lines marked by the precode *speaker*. For these lines, attribute `rend="speaker"` should be added to the `<l>` tag.

poem

BOER ← speaker
Goedenavond en een goet Jaer.
GAEUW-DIEF ← speaker
Mag ick hier slapen?
BOER ← speaker
Ja, gy myn vaar.

```
<lg type="poem">
<l rend="speaker">BOER</l>
<l rend="; 1tab">Goedenavond en een goet
Jaer.</l>
<l rend="speaker">GAEUW-DIEF</l>
<l rend="; 1tab">Mag ick hier slapen?</l>
<l rend="speaker">BOER</l>
<l rend="; 2tab">Ja, gy myn vaar.</l>
</lg>
```

Example 23

12. If a line of a poem continues on the next line because it is too long for the page, then this should be captured in the XML as one line. This is often indicated by an opening bracket "[" or closing bracket "]", which can be ignored in the XML. See Example 24.

poem	<lg type="poem"> <l>Soms roept een kind:</l> <l>Moeder, daar gaat de duivel met zijn staart!</l> <l>Maar moeder antwoordt bleek: Kind, daar gaat Vader!</l> </lg>
Soms roept een kind: Moeder, daar gaat de duivel met zijn staart! Maar moeder antwoordt bleek: Kind, daar [gaat Vader!	

Example 24

8 Plays

1. Plays are marked in the precoded scan as *This is a play*, or defined in the instruction document. Plays consist of a number of acts and each act consists of a number of scenes.

2. The highest level head within a play indicates the beginning of an act. The `type` attribute in the corresponding `<div>` should get the value `"act"`. Apart from the value of the `type` attribute, an act should be treated as a chapter.

3. The next level head indicates the beginning of a scene. The `type` attribute in the corresponding `<div>` should get the value `"scene"`. Apart from the value of the `type` attribute, a scene should be treated as a section.

4. Scenes contain speeches. A speech is a piece of text that starts with a speaker description, followed by his or her speech text. Either the instruction document explains how to distinguish the speaker description from the rest of the text, or the precoded scan marks a few examples of speaker descriptions by using the precode *speaker*. For each speech, the following TEI-XML structure should be used:

```
<sp>
<speaker>«speaker description»</speaker>
<p>«speech text»</p>                                    «repeat for each paragraph»
</sp>
```

Example 25

5. White lines before or after a speech and white lines before or after a speaker description should be ignored.

6. Scenes may contain stage directions. Either the instruction document or the precoding explains how to identify these, and/or the precoded scan marks them by using the precode *stage*. Stage directions may appear inside or outside speech. Stage directions should be enclosed within the tags `<stage>` and `</stage>`. See Example 26.

This is a play	
Eerste bedrijf h3 **Scene 1** h4 Boer en gaeuw-dief komen op. stage BOER ← speaker Goedenavond en een goet Jaer. GAEUW-DIEF ← speaker Mag ick hier slapen? **Scene 2** h4 Boer en gaeuw-dief lopen verder. stage BOER ← speaker Nee, dat mag niet . . .	 <head rend="h3">Eerste bedrijf</head> <head rend="h4">Scene 1</head> <stage>Boer en gaeuw-dief komen op.</stage> <sp> <speaker>BOER</speaker> Goedenavond en een goet Jaer. </sp> <sp> <speaker>GAEUW-DIEF</speaker> Mag ick hier slapen? </sp> <head rend="h4">Scene 2</head> <stage>Boer en gaeuw-dief lopen verder.</stage> <sp> <speaker>BOER</speaker> Nee, dat mag niet </sp> . . .

Example 26

7. Other precodes, e.g. *poem*, may also occur in plays and should be treated as specified in chapter 7 Poems.

8. A play may have lines instead of paragraphs. This will be indicated with the precode *lines* or <l> at the start of the scan. In this case the following rules apply:

   8.1. Each line throughout the play will be placed between <l> tags.
   8.2. Contrary to the lines of a poem, lines of a play are not placed within <lg> tags but directly within the or element.
   8.3. It is possible that within a play with lines also a poem occurs. This will be indicated with *poem* and should be treated as a normal poem as described in chapter 7. This means that for these lines the <lg type="poem"> tag is to be used.
   8.4. Lines may also be indicating a speaker. This is indicated with the precode *speaker*. If the *speaker* precode occurs within a play with lines, use <l rend="speaker"> tags to capture these lines.

9. If the text of a play has line numbers, then add attribute n="*«line number»*" to the corresponding <l> or <lb> tags.
   9.1. If the precode *lines* is applicable, use <l n="number">
   9.2. If the precode *lines* is *not* stated at the beginning of the scan, use <lb n="number"/> (see chapter 9, rule 4)

9 Line beginning tags

1. Line beginnings *within a heading* should only be captured if explicitly indicated by the precode Ub/. See Example 27.

1. Inleiding en overzicht van de inhoud
H3 met een voorwoord van dr. Janssen
 	
Het is lang geleden dat ik mij voornam om een boek te schrijven over dit onderwerp.
	`<div type="chapter">`
	`<head rend="h3">1. Inleiding en overzicht van de inhoud met een voorwoord van dr. Janssen</head>`
	`<p>Het is lang geleden dat ik mij voornam om een boek te schrijven over dit onderwerp.</p>`
	`</div>`

1. Inleiding en overzicht van de inhoud Ub/
H3 Door dr. Janssen
 	
Het is lang geleden dat ik mij voornam om een boek te schrijven over dit onderwerp.
	`<div type="chapter">`
	`<head rend="h3">1. Inleiding en overzicht van de inhoud <lb/>Door dr. Janssen</head>`
	`<p>Het is lang geleden dat ik mij voornam om een boek te schrijven over dit onderwerp.</p>`
	`</div>`

Example 27

2. Line beginnings *within a paragraph* (`<p>` element) should only be captured if explicitly indicated with precodes.

   2.1. Precode Ub/ ('line beginning') should be represented by the `<lb/>` tag. See Example 28.

Nu volgt een lied over de lente, ook al is het nog winter.
Ub/ Daarna bezingen we de zomer.
Ub/ Ten slotte de herfst.
	`<p>Nu volgt een lied over de lente, ook al is het nog winter.`
	`<lb/>Daarna bezingen we de zomer.`
	`<lb/>Ten slotte de herfst.</p>`

Example 28

   2.2. Line beginnings in a paragraph, not explicitly precoded by Ub/, should be represented by starting a separate paragraph. See Example 29.

Nu volgt een lied over de lente, ook al is het nog winter.
Daarna bezingen we de zomer.
Ten slotte de herfst.
	`<p>Nu volgt een lied over de lente, ook al is het nog winter.</p>`
	`<p>Daarna bezingen we de zomer.</p>`
	`<p>Ten slotte de herfst.</p>`

Example 29

3. Line beginnings *outside a paragraph* (for example, in a block quote or a note) should be represented by the `<lb/>` tag, even if not explicitly precoded. See Example 30.

   3.1. Empty lines in a block quote or note should be represented by two times `<lb/>`.

Lorem ipsum. bq Aenean commodo ligula eget dolor tanuem massa. Cum sociis natoque penatibus et magnis dis parturient montes.	Lorem ipsum. <q rend="bq">Aenean commodo ligula eget dolor tanuem massa. <lb/> <lb/>Cum sociis natoque penatibus et magnis dis parturient montes.</q>

Example 30

4. If the source text is a *paragraph with line numbers*, then these line numbers must be inserted and every line will have an `<lb n=…/>` tag, even if the line numbers are not repeated in the original text. This will be precoded by "*prose with line numbers*". The value of attribute n is equal to the line number. See Example 31. This rule does not apply to poems (see chapter 7, rule 8).

prose with line numbers : Lorem ipsum dolor sit amet, consec tetuer adipiscing elit. Aenean con ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nasce tur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, retium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu.	 <lb n="1"/>Lorem ipsum dolor sit amet, consec <lb n="2"/>tetuer adipiscing elit. Aenean con <lb n="3"/>ligula eget dolor. Aenean massa. <lb n="4"/>Cum sociis natoque penatibus et <lb n="5"/>magnis dis parturient montes, nasce <lb n="6"/>tur ridiculus mus. Donec quam felis, <lb n="7"/>ultricies nec, pellentesque eu, retium <lb n="8"/>quis, sem. Nulla consequat massa <lb n="9"/>quis enim. Donec pede justo, fringilla <lb n="10"/>vel, aliquet nec, vulputate eget, arcu.

Example 31

5. Line beginning tags may only appear at the start of an XML line, except for line number tags in hyphenated words (as specified in chapter 14, rule 4).

6. The `<lb/>` tag should always be preceded by a space, unless otherwise specified in the rules above.

10 Pages and page begin tags

1. All pages should have a page begin tag `<pb n="…"/>`, where the value of attribute n is equal to the page number. The page number should be derived as follows:
   1.1. If the page has a printed page number, then that number should be taken.
   1.2. Else, if a page number is written in the top right corner of the page, then that number should be taken.
   1.3. Else, the page number should be equal to the number of the previous page plus one.
   1.4. Do not number the blank pages between the cover pages and the first or last numbered page. The first and last numbered page of the book will be either printed in the book or precoded.
   1.5. If the page number cannot be determined, use the following tag: `<pb n="{problem}"/>`.

Note: Page numbers are not always numeric. Roman numbers, like `<pb n="XIV"/>`, are common. If written on the precoded scan, special characters like a hyphen, pi sign (π) or an asterisk may be used and should appear also in the TEI-XML, e.g. `<pb n="12-14"/>`, `<pb n="π1"/>` or `<pb n="*1"/>`.

Sometimes a page number consists of one or two words, e.g. `<pb n="binnenkant voorplat"/>`. This occurs frequently for unnumbered cover pages in precoded titles. If present, they appear in the following order:

- voorkant stofomslag
- voorplat
- binnenkant voorplat

[…]

- binnenkant achterplat
- achterplat
- achterkant stofomslag

If the words mentioned above are used in place of a page number, please also add the attribute ed followed by an empty value in the pb-tag, e.g. `<pb n="binnenkant voorplat" ed=""/>`. If a pb-tag already contains the attribute rend, please do not add the attribute ed.

2. If a page has precode ~~pto~~ followed by a page number, then the tag should be `<pb n="t.o. ≪page number≫"/>`.

~~pto~~ IX Lorem ipsum	`<pb n="t.o. IX"/>` `<p>Lorem ipsum</p>`

Example 32

The abbreviation "t.o." means "opposite to" and is used for unnumbered pages opposite to a numbered page.

3. Page begin tags should be placed at the start of the page (i.e. to indicate start of new page) even if page numbers appear at the bottom of the pages in the source material.

4. If the page indicates the start of a chapter or a section, the page begin tag should appear after `<div>` or `<interpGrp>` and before `<head>`.

5. If there is no other content on the page except for the page begin tag, attribute `rend` should be added to the page begin tag. The value of the attribute depends on the reason why there is no other content, as specified below. Example: `<pb rend="blank" n="17"/>`.

   5.1. If the page is blank, the value is `"blank"`.

5.2. If the page only contains advertisements, as indicated in the instruction document or by the precode '*ad*', the value is `"ad"`. Frequently the image/text belonging to the advertisement will also be crossed-out as to emphasize that it can be ignored.
5.3. If the page only contains notes, as defined in chapter 15, or only notes and advertisements, the value is `"notes"`.
5.4. If there is any other reason why there is no XML for the page, except for the page begin tag itself, the value is `"misc"`.
5.5. If the blank page occurs directly after `<pb n="voorkant stofomslag"/>` or directly before `<pb n="achterkant stofomslag"/>`, or the page occurs between the cover pages and the first or last numbered page and has no page number printed or precoded, then the pb-code should be: `<pb rend="hidden"/>`. These pages should *not* get an n attribute in the pb-tag.

6. Page begin tags should appear at the start of an XML line, except for page begin tags in hyphenated words (as specified in chapter 14, rule 4).

7. Page headers and footers as well as standard decorations should be ignored.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque eco
 	
© ABC 1922
 	
~ 344 ~
 	
penatibus et magnis dis parturient montes, nascetur ridiculus mus.
	`<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque eco`
	`<pb n="344"/>`
	`penatibus et magnis dis parturient montes, nascetur ridiculus mus.`
	`</p>`

Example 33

8. A page begin tag should never occur between `<l>` and `</l>`, nor between `<note>` and `</note>`, nor between `<stage>` and `</stage>` nor between `<table>` and `</table>`.

This is incorrect:
```
<l> . . . . . .
<pb n="233"/></l>
<l> . . . . . .</l>
```

This is correct:
```
<l> . . . . . . </l>
<pb n="233"/>
<l> . . . . . .</l>
```

11 Lists

Definition:
A list is an enumeration of items, where each item begins on a new line, sometimes labeled by numbers or special characters, like bullets.

Exceptions: The enumeration is not a list, but a table in the following cases:

- If at least one of the labels consists of more than three characters.
- If, apart from the labels, the items themselves are divided in multiple columns.
- If it is the Table of Contents (in Dutch: Inhoud or Inhoudsopgave).

1. Where the source text contains a list, as defined above, or the precoded scan or instruction document shows the precode *list*, the TEI-XML should be as follows:
   1.1. If none of the items have a label:

```
<list>
<item>«item text»</item>                    repeat for every item
</list>
```

   1.2. If all items have a label:

```
<list>
<label>«label»</label><item>«item text»</item>    repeat for every item
</list>
```

   1.3. If some but not all items have a label, then the missing labels should be coded as `<label/>`.

2. Lists should be nested, as indicated by the source text.

3. Empty space between items (more than the regular item spacing used in the text) should be coded as `<item/>` if there are no labels, or `<label/><item/>` if there are labels.

Example 34 shows a combination of the application of rules 1 to 3. Note that the empty line before item 1 is coded differently than the empty line after item 1.

Lorem ipsum dolor:
 	
1) Julius Caesar
 	
2) Augustus (Gaius Octavius Thurinus)
a) Clodia minor
Clodia major
b) Scribonia
	`<p>Lorem ipsum dolor:</p>`
	`<p> </p>`
	`<list>`
	`<label>1)</label><item>Julius Caesar</item>`
	`<label/><item/>`
	`<label>2)</label><item>Augustus (Gaius Octavius Thurinus)`
	`<list>`
	`<label>a)</label><item>Clodia minor</item>`
	`<label/><item>Clodia major</item>`
	`<label>b)</label><item>Scribonia</item></list></item></list>`

Example 34

4. A list may have a head, usually explicitly precoded as *head*. The `<head>` tag should be positioned right after the `<list>` tag, without a `<div>` tag. The capitalization rules defined in paragraph 3.5 on page 6 are applicable.

IMPERATORI ROMANA head = = 1) Julius Caesar 2) Augustus list	<list> <head>Imperatori Romana</head> <label>1)</label><item>Julius Caesar</item> <label>2)</label><item>Augustus</item> </list>

Example 35

An item may indicate that a part of a previous item is repeated. This can be indicated in two ways:

- By means of the character „ beneath every word of the previous item that should be repeated;
- By means of an underline beneath the part of the previous item that should be repeated.

5. If an item contains an indication that a part of a previous item is repeated, then that part should be repeated in the TEI-XML. See Example 36

– Julius Caesar – „ Augustus	<list> <label>-</label><item>Julius Caesar</item> <label>-</label><item>Julius Augustus</item> </list>
Laatste keizer van het Perzische Rijk Eerste _____ Romeinse ___	<list> <item>Laatste keizer van het Perzische Rijk</item> <item>Eerste keizer van het Romeinse Rijk</item> </list>

Example 36

12 Tables

Definition:
A table is a part of a page on which text is aligned over multiple columns.

1. Where the source text shows a table, or the precoded scan or instruction document shows the precode *table*, the following TEI-XML structure should be used:

   ```
   <p>
   <table>
   <row>
   <cell>«cell contents»</cell>
   </row>
   </table>
   </p>
   ```

   repeat for every column / *repeat for every row*

   See Example 37.

2. If the first row(s) of the table contain(s) headers, then attribute `role="header"` should be added to each of the cells of the first row(s). See Example 38.

3. An empty cell should be represented by `<cell/>`.

4. If a cell contains a horizontal line, continuous or dashed, then the cell content should be represented by five underscore characters: `<cell>_____</cell>`. See Example 39.

5. Whenever text spans over more than one column in a row, `<cell cols="`*«number of columns»*`">` should be used.

6. Each row in a table should contain the same number of cells, where a cell with attribute `cols="`*«number of columns»*" is counted as *«number of columns»* cells.

7. An empty row should be represented by `<row><cell cols="`*«number of columns»*`"/></row>` or, if there is just one column: `<row><cell/></row>`.

8. A page begin tag in a table is not allowed. A new table should be started on the next page.

9. White space before and after a table should be ignored (i.e. omit the `<p> </p>` code).

10. A table may have a head, usually explicitly precoded as *head*. The `<head>` tag should be positioned right after the `<table>` tag, without a `<div>` tag. The capitalization rules defined in paragraph 3.5 on page 6 are applicable. See Example 38.

11. If a cell contains an indication that a part of a previous row is repeated, then that part should be repeated in the TEI-XML (see chapter 11, rule 5).

12. Series of dots or dashes within a table used to connect two columns should be ignored. See Example 40.

N.B.: Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor.	 <table> <row> <cell>N.B.:</cell> <cell>Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor.</cell> </row> </table>

Example 37

Tabel 3. HAVO/VWO-statistiek head

		HAVO	VWO
1.	Culturele vorming	28%	38%
2.	Leesplezier	27%	13%

```
<p>
<table>
<head>Tabel 3. HAVO/VWO-
statistiek</head>
<row>
<cell cols="2" role="header"/>
<cell role="header">HAVO</cell>
<cell role="header">VWO</cell>
</row>
<row>
<cell>1.</cell>
<cell>Culturele vorming</cell>
<cell>28%</cell>
<cell>38%</cell>
</row>
<row>
<cell>2.</cell>
<cell>Leesplezier</cell>
<cell>27%</cell>
<cell>13%</cell>
</row>
</table>
</p>
```

Example 38

NAAM _____

Adres _____

Woonplaats _____

```
<p>
<table>
<row>
<cell>NAAM</cell>
<cell>_____</cell>
</row>
<row><cell cols="2"/></row>
<row>
<cell>Adres</cell>
<cell>_____</cell>
</row>
<row>
<cell>Woonplaats</cell>
<cell>_____</cell>
</row>
</table>
</p>
```

Example 39

Inleiding . . . . . 2 Gedachten 5	 <table> <row> <cell>Inleiding</cell> <cell>2</cell> </row> <row> <cell>Gedachten</cell> <cell>5</cell> </row> </table>

Example 40

13. If columns are just a means to avoid long lines, as in a newspaper, then these columns should be ignored and tables should not be used.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Cum sociis et natoque penatibus et magnis dis parturient montes, nascetur riculus mus. Donec quam felis.	Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Cum sociis et natoque penatibus et magnis dis parturient montes, nascetur riculus mus. Donec quam felis.

Example 41

See chapter 14 for hyphenation rules.

14. If a table has more than 8 columns or contains a large amount of text, the table should be split into two (or more) parts. The precode "*split table*" will be used, and arrows will indicate where the split is to be made. If there is no further instruction, this means that this table should be split into two (or more) tables at the indicated location(s), where the cells of the very first column should be repeated in the following table(s). There can be an added instruction to repeat more or no columns, for example "*split table, repeat first two columns*" or "*split table and don't repeat first column*".

See below for an example of a "*split table*" precode followed by the actual XML (with the left side of the table omitted because of the length). The repeated column is marked yellow:

split table and repeat first column

	LEIDEN	FRANEKER	GRONINGEN	AMSTERDAM	UTRECHT	HARDERWIJK
Jaar van stichting der academie:	Febr. 1575	Juli 1585	Aug. 1614	Jan. 1632	Maart 1636	April 1648
	1775	1785	1814		1786	
Theologanten	zwart	wit	blauw		rozenrood	1772
Juristen	rood	rood	wit		donkerrood	rood
Medici	groen	groen	rood			1807
Phil. en Litt.	wit	licht blauw	groen		hemelsblauw	lichtblauw
						1768
					korenblauw	blauw
Jaar van stichting van het Corps:	Maart 1839		Sen. V.A.P. Febr. 1815	Mei 1851	Sen. Vet. Febr. 1814	
Godgeleerdheid	zwart	opgeheven decreet 1811 K.B. 1843	blauw	wit	rose	opgeheven decreet 1811 K.B. 1818
Rechtsgeleerdheid	rood		wit	rood	rood	
Geneeskunde	groen		rood	geel	groen	
Wis- en Natuurk.	blauw		geel	paars	geel	
Lett. en Wijsbeg.	wit		purper	blauw	donkerblauw	

(naar Mr. H. Waller)

```
<table>
<row>
<cell role="header"/>
<cell role="header">AMSTERDAM</cell>
<cell role="header">UTRECHT</cell>
<cell role="header">HARDERWIJK</cell>
</row>
<row>
<cell>Jaar van stichting der academie:</cell>
<cell>Jan. 1632</cell>
<cell>Maart 1636</cell>
<cell>April 1648</cell>
</row>
<row>
<cell/>
<cell/>
<cell/>
<cell/>
</row>
<row>
<cell/>
<cell/>
<cell>1786</cell>
<cell/>
</row>
<row>
<cell>Theologanten</cell>
<cell/>
<cell>rozenrood</cell>
<cell>1772</cell>
</row>
<row>
<cell>Juristen</cell>
<cell/>
<cell>donkerrood</cell>
<cell>rood
<lb/>
1807</cell>
</row>
<row>
<cell>Medici</cell>
<cell/>
<cell>hemelsblauw</cell>
<cell>lichtblauw
<lb/>
1768</cell>
</row>
<row>
<cell>Phil. en Litt.</cell>
<cell/>
<cell>korenblauw</cell>
<cell>blauw</cell>
</row>
```

<row>
<cell/>
<cell/>
<cell/>
<cell/>
</row>
<row>
<cell>Jaar van stichting van het Corps:</cell>
<cell>Mei 1851</cell>
<cell>Sen. Vet. Febr. 1814</cell>
<cell/>
</row>
<row>
<cell/>
<cell/>
<cell/>
<cell/>
</row>
<row>
<cell>Godgeleerdheid</cell>
<cell>wit</cell>
<cell>rose</cell>
<cell>opgeheven</cell>
</row>
<row>
<cell>Rechtsgeleerdheid</cell>
<cell>rood</cell>
<cell>rood</cell>
<cell>decreet 1811 K.B. 1818</cell>
</row>
<row>
<cell>Geneeskunde</cell>
<cell>geel</cell>
<cell>groen</cell>
<cell/>
</row>
<row>
<cell>Wis- en Natuurk.</cell>
<cell>paars</cell>
<cell>geel</cell>
<cell/>
</row>
<row>
<cell>Lett. en Wijsbeg.</cell>
<cell>blauw</cell>
<cell>donkerblauw</cell>
<cell/>
</row>
</table>

13 Emphasis

13.1 Forms of emphasis

The various forms of typographical emphasis should be captured as indicated in the following table.

Table 3

Description	Example text	TEI-XML
Italic	*example text*	`<hi rend="i">example text</hi>`
Bold	**example text**	`<hi rend="b">example text</hi>`
Underlined	example text	`<hi rend="u">example text</hi>`
Double underlined	example text	`<hi rend="uu">example text</hi>`
Superscript	example text	`example <hi rend="sup">text</hi>`
Subscript	example text	`example <hi rend="sub">text</hi>`
Spaced text	e x a m p l e t e x t	`<hi rend="spat">example text</hi>`
Small capitals	EXAMPLE TEXT EXAMPLE TEXT	`<hi rend="sc">example text</hi>` `<hi rend="sc">Example Text</hi>`
Strike through	~~example~~	`<hi rend="strikethrough">example text</hi>`

Note: Punctuation characters may have a different emphasis than the surrounding text. See Example 42.

En zie, het was *zeer goed*!	`En zie, het was <hi rend="i">zeer goed</hi>!`
En zie, het was *zeer goed!*	`En zie, het was <hi rend="i">zeer goed!</hi>`

Example 42

13.2 Emphasis in heads

Only italic, superscript and subscript emphasis are allowed within heads. If the heads in the source material are presented with other emphasis, like small capital letters or bold or spaced text, omit all of these.

Examples:

h3 **Example text**	`<head rend="h3">Example text</head>`
h3 Example **text**	`<head rend="h3">Example text</head>`
h3 *Example* **text**	`<head rend="h3"><hi rend="i">Example</hi> text</head`

Example 43

14 Hyphenation

1. Hyphenation, used to indicate that a word continues on the next line or the next column (see chapter 12, rule 13), should be eliminated (but there are exceptions, see next rules).

Lorem ipsum dolor sit amet, consec tetuer adipiscing elit. Aenean contra eget dolor.	`<p>Lorem ipsum dolor sit amet, consec tetuer adipiscing elit. Aenean contra eget dolor.</p>`

Example 44

2. If a word is hyphenated and the two parts of the word appear on different pages, then the hyphen should not be eliminated and the page number tag should be inserted into the word after the hyphen. See Example 45.

3. If a word is hyphenated while line number tags are required, then the hyphen should not be eliminated and the line number tag should be inserted into the word after the hyphen. See Example 45.

4. If a page number tag and/or a line number tag is inserted after a hyphen, then there should not be any spaces before or after the tag and the word including the tag should be on the same XML line.

prose with line numbers :
Lorem ipsum dolor sit amet, consec
tetuer adipiscing elit. Aenean con-
tra eget dolor. Aenean massa. Cum
donoris sociis natoque penatibus et
5 magnis dis parturient montes, nasce-
 	
233
 	
tur ridiculus mus. Donec quam felis,
ultricies nec, pellentesque eu, retium
quis, sem. Nulla consequat massa
quis enim. Donec pede justo, fringilla
10 vel, aliquet nec, vulputate eget, arcu.
	`<lb n="1"/>Lorem ipsum dolor sit amet, consec`
	`<lb n="2"/>tetuer adipiscing elit. Aenean`
	`con-<lb n="3"/>tra eget dolor. Aenean massa. Cum`
	`<lb n="4"/>donoris sociis natoque penatibus et`
	`<lb n="5"/>magnis dis parturient montes,`
	`nasce-<pb n="234"/><lb n="6"/>tur ridiculus mus.`
	`Donec quam felis,`
	`<lb n="7"/>ultricies nec, pellentesque eu, retium`
	`<lb n="8"/>quis, sem. Nulla consequat massa`
	`<lb n="9"/>quis enim. Donec pede justo, fringilla`
	`<lb n="10"/>vel, aliquet nec, vulputate eget, arcu.`

Example 45

5. If a period of time is notated as two years with a hyphenation in between, then the hyphenation should not be eliminated and the page number tag should be inserted into the word after the hyphen. See Example 46.

Dit gebeurde in België tijdens de eerste twee maanden van de Tweede Wereldoorlog (1940- 233 1945). De consequenties hiervan kunnen niet genoeg benadrukt worden.	`<p>Dit gebeurde in België tijdens de eerste twee maanden van de Tweede Wereldoorlog (1940<hy>-</hy><pb n="233"/>1945). De consequenties hiervan kunnen niet genoeg benadrukt worden.`

Example 46

15 Notes

Notes are short pieces of text that reference to other pieces of text, such as words or sentences. They can be put in any element that contains text such as `<p>`, `<l>`, `<q>`, `<item>`, `<cell>`, `<head>` and `<figDesc>` elements. Notes themselves may contain other tags, like `<figure>` or `<table>`.

15.1 Footnotes

Footnotes are usually printed at the bottom of a page. A footnote consists of a footnote identifier and a footnote text. The footnote text gives additional information related to a particular piece of the main text. There are two ways in which a footnote can be connected to that piece of main text:

a) By means of a footnote reference in the main text. A footnote reference consists of the footnote's identifier, usually printed in superscript.

b) By means of a line number, if the lines of the main text are numbered. In this case, there isn't any footnote reference, but the footnote identifier is equal to a line number.

1. A footnote should be coded as follows:

   <note n="≪footnote identifier≫" place="foot">≪footnote text≫</note>

2. In the TEI-XML, footnotes should not be included at the positions where they are printed, but

   a) at the positions of the footnote references, replacing the footnote references, or

   b) at the end of the corresponding line of text (see Example 48).

Lorem ipsum dolor sit amet, con sectetuer adipiscing[a)] elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque eco. ------ [a)] Nero 202 p34	`<p>Lorem ipsum dolor sit amet, con sectetuer adipiscing <note n="a)" place="foot">Nero 202 p34</note> elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque eco.</p>`

Example 47

15.2 Endnotes

Endnotes differ from footnotes in the sense that they are not printed at the bottom of the page, but at the end of the chapter or at the end of the book.

1. Endnotes should be treated just like footnotes, except that the `place` attribute should have value `end` instead of `foot`.

   `<note n="`*≪endnote identifier≫*`" place="end">`*≪endnote text≫*`</note>`

15.3 Margin notes

Margin notes are printed in the page margin. A margin note does not always have an identifier.

1. A margin note should be coded as follows:

   `<note place="margin">`*≪margin note text≫*`</note>`
   or
   `<note n="`*≪marginnote identifier≫*`" place="margin">`*≪margin note text≫*`</note>`

2. In the TEI-XML, margin notes should be inserted in the main text, near the positions where they are printed, or, if the precoded scan specifies the required position of the margin note: at the specified position.

15.4 Textcritical notes

A textcritical note is a special type of footnote, endnote or margin note. Which notes should be treated as textcritical is explicitly specified on the precoded scan using precode *text crit note*, or in the instruction document.

1. In the TEI-XML, the textcritical note should be inserted right after the `<pb>` tag of the page containing the text to which the note pertains.

2. A textcritical note should be coded as follows:

   `<note n="`*≪note identifier≫*`" place="krit">`*≪note text≫*`</note>`

Lorem ipsum dolor sit amet, con sec
tetuer adipiscing elit. Quorum con-
tra eget dolor. Aenean massa. Cum
donoris sociis natoque penatibus et
5 magnis dis parturient montes, nasce-
tur ridiculus mus. Donec quam felis,
ultricies nec, pellentesque eu, retium
quis, sem. Nulla consequat massa
quis enim. Donec pede justo, fringilla
10 vel, aliquet nec, vulputate eget, arcu.
 	
------ 	
3 Aenean = Enean *text crit note*
------ 	
3 Lorem ipsum:
Dolor sit amet *bq*
 	
233
	`<pb n="233"/><note n="3" place="krit">Aenean = Enean</note>`
	`<p>`
	`<lb n="1"/>Lorem ipsum dolor sit amet, con sec`
	`<lb n="2"/>tetuer adipiscing elit. Quorum`
	`con-<lb n="3"/>tra eget dolor. Aenean massa. Cum`
	`<note n="3" place="foot">Lorem ipsum:`
	`<q rend="bq">Dolor sit amet</q>`
	`</note>`
	`<lb n="4"/>donoris sociis natoque penatibus et`
	`<lb n="5"/>magnis dis parturient montes,`
	`nasce-<lb n="6"/>tur ridiculus mus. Donec quam felis,`
	`<lb n="7"/>ultricies nec, pellentesque eu, retium`
	`<lb n="8"/>quis, sem. Nulla consequat massa`
	`<lb n="9"/>quis enim. Donec pede justo, fringilla`
	`<lb n="10"/>vel, aliquet nec, vulputate eget, arcu.`
	`</p>`

Example 48

3. If the page begins with a poem line of a poem that started on the previous page, then the note should be enclosed within `<l>` and `</l>`:

   `<l><note n="`*≪note identifier≫*`" place="krit">`*≪note text≫*`</note></l>`

16 Images

Parts of the source material that cannot be captured in TEI-XML code are captured in image files (PNG). The file names of these image files are included in the TEI-XML in the right tags.

There are two types of such source material:

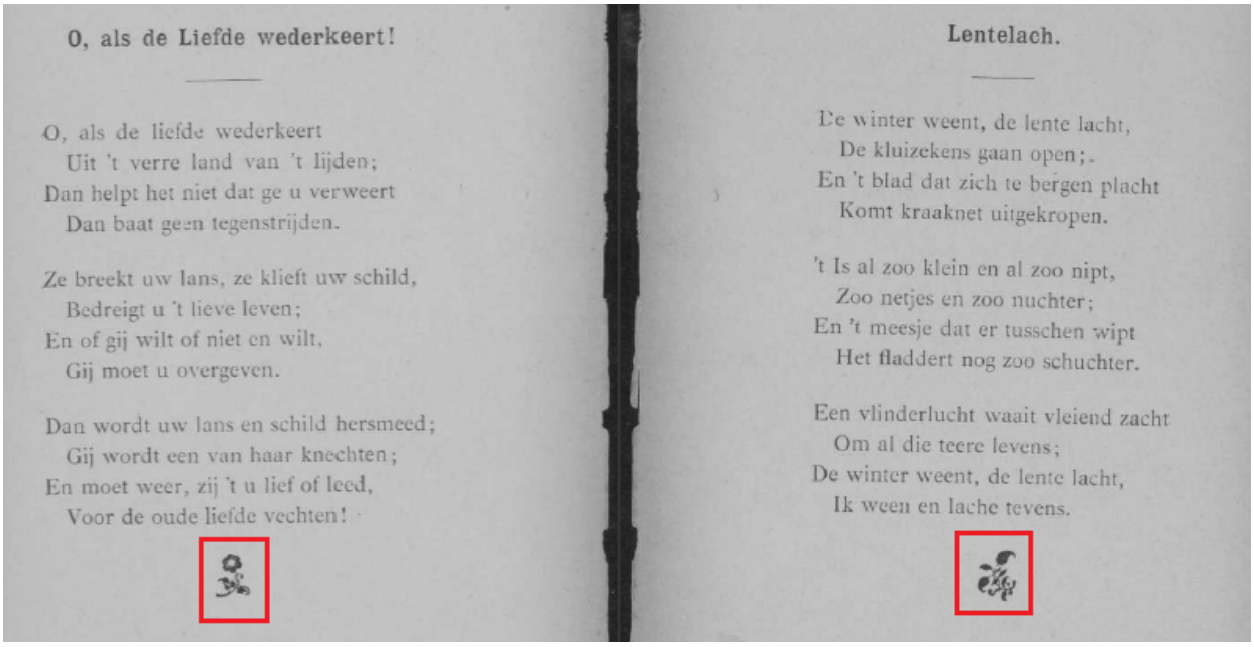
- Illustration: a graphical part of the source material.
- Irregular text: a textual part of the source material that cannot be captured in TEI-XML code (see paragraph 16.2).

16.1 Illustrations

The following types of source material should not be captured in image files:

- Simple, repeated page decorations or paragraph/chapter separators; these may be ignored.
- Decorated characters at the beginning of a chapter or paragraph; these should be treated as regular bold characters (see chapter 18, rule 12).

See chapter 4, rule 3 for an example of a decorative paragraph separator. The scan in Example 49 shows pages with decorations which should be captured as illustrations, because they are not ‘simple’ and not ‘repeated’.



O, als de Liefde wederkeert!

O, als de liefde wederkeert
Uit 't verre land van 't lijden;
Dan helpt het niet dat ge u verweert
Dan baat geen tegenstrijden.

Ze breekt uw lans, ze klieft uw schild,
Bedreigt u 't lieve leven;
En of gij wilt of niet en wilt,
Gij moet u overgeven.

Dan wordt uw lans en schild hersmeed;
Gij wordt een van haar knechten;
En moet weer, zij 't u lief of leed,
Voor de oude liefde vechten!

Lentelach.

De winter weent, de lente lacht,
De kluizekens gaan open;
En 't blad dat zich te bergen placht
Komt kraaknet uitgekropen.

't Is al zoo klein en al zoo nipt,
Zoo netjes en zoo nuchter;
En 't meesje dat er tusschen wipt
Het fladdert nog zoo schuchter.

Een vlinderlucht waait vleiend zacht
Om al die teere levens;
De winter weent, de lente lacht,
Ik ween en lache tevens.

Example 49

1. Illustrations should be captured in the TEI-XML with the following code:

```
<figure><p><xptr to="«image file name»"/></p></figure>
```

The location of this tag depends on the placement of the image in the original and how it interrupts the text. Rule 2 and 3 describe the two possible placements.

2. For an illustration that breaks a text element, such as a `<p>`, `<q>`, `<head>` and `<lg>`, rule 2.1 and 2.2 apply. This usually is the case when an illustration stretches the whole width of a page, or when it covers a full page.

   2.1. Always place the illustration between two text elements, using the text element break *closest* to the illustration in the original.

2.2. Illustrations can never be tagged inside an heading tag `<head>`. In the unusual case that an illustration breaks up an heading, it should be placed after the closing heading tag.

3. For an illustration that does not break a text element, rule 3.1, 3.2 and 3.3 apply. This usually is the case when an illustration is placed besides a body of text and the text continues around it (see Example 51) or because it is placed at the bottom or top of a page, after or before the text (see Example 49).

   3.1. These kind of illustrations should always be placed outside text elements such as `<p>`, `<q>`, `<head>` and `<lg>`, and they should be placed between the two text elements that are the closest in the original. See Example 51.

   3.2. If there is only one text element on the page, then the illustration is to be placed at the top (before the `<p>`) or at the bottom (after the `</p>`) of the page.

   3.3. If there is no break in the text element at all on the page, the illustration tag will be placed after the end of a sentence (so after the dot, exclamation mark, question mark, etc.) and before the next sentence starts, near the illustration in the original.

4. If the image is in a table (chapter 12) or a list (chapter 11) and the `<figure>` tag does not have the attribute `rend="inline"`, then the table or list should be closed (using the closing tags `</cell> </row> </table>` or `</item> </list>` respectively) before the `<p><figure>` tags and reopened after the `</figure></p>` tags.

5. Never move an illustration to another page, unless this is specifically indicated with precodes.

6. White lines before or after an illustration should be ignored.

7. The precode *ill* may be used to indicate that a decoration should be regarded as an illustration.

8. A figure with a figure description should be captured as follows:

   `<figure><p><xptr to="`*«image file name»*`"/></p><figDesc>`*«figure description»*`</figDesc></figure>`

9. The precode *fd* may be used to indicate which part of the source text is a figure description. If it is not clear which figure description belongs to which figure, the precode may have an arrow to the corresponding illustration or a clarification in text, e.g. *fd to figure on page 16*.

10. A figure description may contain tags, like `<lb/>` to indicate the beginning of a new line.

 *fd*
Leonard de Gou
© KbPhoto 1967
	`<figure><p><xptr to="`*«image file name»*`"/></p>`
	`<figDesc>Leonard de Gou <lb/>© KbPhoto 1967</figDesc>`
	`</figure>`

Example 50



JOHANNES BOSBOOM. 5

uitgezien, toen daar op 21 October 1653 Vondel werd gevierd en gekroond.

Dáar in Pulchri zat aan het hoofd van den disch ook een krachtig grijsaard, omstuwd door de meesters van de groote Hollandsche school van het heden, in het bijzijn van een reeks van vrienden en bewonderaars. Die man was Jan Bosboom. Pulchri vierde Bosboom's zeventigsten jaardag met al de artistieke praal, die het aan zijne feesten weet bij te zetten, met al de oprechte opgewondenheid, die een man als Bosboom in de harten van kunstenaars wekken kan. Daar werd getoast en gefeest; daar vertoonde zich ook eensklaps in de geopende deur een kerk, zooals Bosboom er zoovelen schilderde, terwijl een zacht koraal zich deed hooren. Het was eene diepgevoelde, eerbiedige hulde, die niet alleen den kunstenaar tranen in de oogen dreef. Daar sprak eenige oogenblikken later de geest van den heer v. Nieuwkoop zijn afscheidsgroet; en toen het uur kwam, dat Bosboom van scheiden sprak, besloot men hem naar huis te geleiden. Op, om en achter het rijtuig van den meester schaarden zich Pulchri's leden, en met flambouwen en chineesche lampions toog — een eenig schouwspel — de opgewonden schare door de hoofdstraten der stad naar Bosboom's woning.

Hoekje uit het atelier van den meester.

Die hulde, in één oogwenk gewekt en gebaard, was de eenvoudige maar roerende uiting van genegenheid, die elkeen, die Bosboom persoonlijk kent, bezielt voor den man met zijn groot en zijn goed hart. En welke Hagenaar, wien kunst ter harte gaat, kent hem niet?

Bosboom werd op 18 Februari 1817 te 's-Gravenhage geboren en heeft, bijna onafgebroken, daar geleefd, één der beroemde mannen, en wat voor den tijdgenoot meer zegt, een der meest sympathieke persoonlijkheden der oude Gravenstad. Hij heeft heel het leven van zijn volk medegeleefd gedurende vele jaren, en alles wat belangrijk, groot en schoon was heeft nagetrild in zijn hart. Hij heeft vorsten zien heerschen en zien sterven; hij heeft ontelbare ministers zien komen en zien gaan; hij heeft mannen zien bekend worden en hunne glorie zien tanen. En hij staat daar nog, bijna onverzwakt, met de innige overtuiging, dat hij allen overleven zal en dat zijn naam nog zal worden genoemd als de echo van de meesten der anderen zal zijn

```
<pb n="5"/> uitgezien, toen daar op 21
October 1653 Vondel werd gevierd en
gekroond.</p>
<figure><p><xptr to="≪image file
name≫ill0005.gif"/></p><figDesc>Hoekje uit
het atelier van den
meester.</figDesc></figure>
<p>Dáar in Pulchri zat aan het hoofd van den
disch ook een krachtig grijsaard, omstuwd
door de meesters van de
groote Hollandsche school van het heden, in
het bijzijn van een reeks van vrienden en
bewonderaars...</p>
```

Example 51

16.2 Irregular text

1. Textual parts of the source material that cannot be captured as text in XML, like mathematical formulas or non-UTF-8 characters should be captured in the same way as illustrations, but with attribute `rend="inline"` in tag `<figure>`.

    1.1. Greek and Hebrew characters, as well as initials (bigger or decorated characters at the beginning of a paragraph or chapter) should always be captured as UTF-8 characters. See chapter 18, rule 12 for how to treat initials.
    1.2. Non-latin (e.g. Arabic) characters should preferably be captured with UTF-8 but may also be treated as inline figures.
    1.3. Fractions with only numbers above and below the line should *not* be treated as illustrations. See Example 52 and see chapter 18 rule 1 for more information.

dit is gelijk aan $\frac{x+1}{y}$ minimaal	`dit is gelijk aan <figure rend="inline"><p>` `<xptr to="≪image file name≫"/></p></figure> minimaal`
het was $\frac{2}{7}$ van de lengte maal π	`het was 2/7 van de lengte maal π`

Example 52

2. If precode *inline* is used, then treat this content as irregular text and add the attribute `rend="inline" to` the `<figure> tag`.

3. If part of the text is precoded as an illustration, or cannot be captured as text in XML, but is too big for figure rend=inline, please leave the illustration in the exact location as in the original. This could for example apply to a formula:

gen op de hoogte is en niet onmiddellijk inziet dat de samenstelling *wandelstok* in die kringen met behulp van de formule  wordt afgeleid, kan bepaalde gedeelten beter overslaan. Gelukkig zijn de meeste polemische passages in afzonderlijke paragrafen ondergebracht zodat ieder gemakkelijk de rozijnen uit de pap kan vissen.	gen op de hoogte is en niet onmiddelijk inziet dat de samenstelling <hi rend="i">wandelstok</hi> in die kringen met behulp van de formule <figure> <xptr to="≪*image file name*≫"/> </figure> wordt afgeleid, kan bepaalde gedeelten beter overslaan. Gelukkig zijn de meeste polemische passages in afzonderlijke paragrafen ondergebracht zodat ieder gemakkelijk de rozijnen uit de pap kan vissen.

Example 53

17 Spaces in abbreviations

1. Spaces in abbreviations should be omitted. This applies to initials as well.

Table 4

Dit was m. m. v. het ministerie	`Dit was m.m.v. het ministerie`
K. N. A. W.	`K.N.A.W.`
J. H. C. de Jong	`J.H.C. de Jong`

18 Character-level rules

1. For each character in the source text, the most appropriate UTF-8 character should be selected, even if this is a rarely used character. Equivalent XML character entities (numeric entities, or named entities defined in teixlite.dtd) may be used as well, but this is not preferred. Table 5 gives an overview of common XML escape codes and UTF-8 characters. Other characters are possible.

Table 5

Source text	**Correct**			**Incorrect**
	Numeric entities	**Named entities**	**UTF-8 characters (preferred)**	
2×2	`2×2`	`2×2`	`2×2`	`2x2`
± 100	`± 100`	`± 100`	`± 100`	`+/- 100`
§4.5 • 3	`§4.5 • 3`	`§4.5 • 3`	`§4.5 • 3`	`$4.5 * 3`
‘a’	`‘a’`	`‘a’`	`‘a’`	`'a'`

In old Dutch literature, the long s character ‘ſ’ (ſ) is used instead of ‘s’.

2. All occurrences of ſ should be replaced by s. Characters ſ and f should be recognized as different characters. See Example 54.

ſelfs ofje ſchoon	`selfs ofje schoon`

Example 54

3. Capture fractions as follows:
   3.1. Use the corresponding UTF-8 fraction characters or character entities if possible (e.g. ¾ may be captured by ¾, `¾` or `¾`). See Example 55.
   3.2. Otherwise, if there are only numbers above and below the line, use a slash. If there is a number immediately preceding the fraction, add a space in between. See Example 55.
   3.3. Otherwise, if 3.1 and 3.2 are not possible, use an inline figure (see chapter 16, rule 1).

marge van $3\frac{3}{4}$ cm	`marge van 3¾ cm`
marge van 3 $\frac{3}{7}$ cm	`marge van 3 3/7 cm`
het was $\frac{3}{7}$ van de lengte	`het was 3/7 van de lengte`

Example 55

4. Ligatures should be captured as separate letters. Note that ‘ij’ is a very common ligature in Dutch. See Example 56. In linguistic texts, there may be an instruction in the precoding to keep the ligatures.

Het rijke œuvre van Stæl	`Het rijke oeuvre van Stael`

Example 56

5. Dashes should be captured as hyphens (UTF-8 code #002D).

6. Spaces around a dash in a numeric range should be omitted, but dashes within non-numeric text should be separated from the text by spaces.

Lorem ipsum 24 — 32	`Lorem ipsum 24-32`
Lorem ipsum—dolor sit amet	`Lorem ipsum - dolor sit amet`

Example 57

7. Lines of dots and lines of dashes used as paragraph separators should be regarded as a white line (`<p> </p>`).

8. Lines of dots and lines of dashes used in poems should not be ignored.

9. Series of dots within a text should not be ignored.

poem [Lorem ipsum dolor sit
— — — — — — —
Cum . . . . . sociis
	`<lg type="poem">`
	`<l>Lorem ipsum dolor sit</l>`
	`<l>- - - - - - -</l>`
	`<l>Cum . . . . . sociis</l>`
	`</lg>`

Example 58

10. Series of dots within a table used to connect two columns should be ignored. See Example 40..

11. Ellipses should always be captured as separate dot characters (. . . instead of …) without spaces.

Lorem ipsum…	`Lorem ipsum...`

Example 59

12. Bigger or decorated characters at the beginning of a paragraph or chapter should be captured as bold.

 aer ghy sult ouer de Iordaen gaen, ende in ’t Lant woonen	`<p><hi rend="b">M</hi>aer ghy sult ouer de Iordaen gaen, ende in ’t Lant woonen</p>`

Example 60

13. Quotation marks should be captured as single quotation marks, regardless of the form of the quotation mark in the original, which might sometimes also be << and >>. The only exception are quotation marks within a quotation, which should be captured as double quotation marks. See Example 61.

14. For quotation marks, the following UTF-8 characters should be used:
    Start single: ‘ `‘` `‘`
    End single: ’ `’` `’`
    Start double: “ `“` `“`
    End double: ” `”` `“`

Hij zei: „Ik lust geen ""gezond"" eten.”	`<p>Hij zei: ‘Ik lust geen “gezond” eten.’</p>`
Hans zei: <<Salade is gezonder dan patat!>>	`<p>Hans zei: ‘Salade is gezonder dan patat!’</p>`

Example 61

15. In old books, quotations sometimes have an opening quotation mark at each new line or paragraph. Only the first and last quotation mark should be captured.

Standard XML character entities (<, >, &) should obviously be used to obtain valid XML (see chapter 1, rule 4).

Als x < 2 dan V&D	`<p>Als x < 2 dan V&D</p>`

Example 62

16. A curly bracket spanning more than one line indicates these lines should be taken as a group. In many cases, the other side of the curly bracket points to another line, which we call the repeater. The purpose of this notation is to indicate that the repeater has a relationship with all of the grouped lines at the other side of the curly bracket.

    16.1. A curly bracket spanning more than one line should be captured as one curly bracket character on each line. It should be separated from adjacent text by at least one space character. An instance of repeated curly brackets should always be followed by {problem} as to allow for a manual check afterwards.

list
Lorem ipsum
Dolor sit amet
Consectetuer{2}
	`<list>`
	`<item>{ {problem} Lorem ipsum</item>`
	`<item>{ Dolor sit amet</item>`
	`<item>Consectetuer{2}</item>`
	`</list>`

Example 63

16.2. If a repeater is present, then the repeater should be repeated on each line.

poem Lorem ipsum Dolor sit amet Consectetuer{2} repeat	<lg type="poem"> <l>Lorem ipsum } {problem} repeat</l> <l>Dolor sit amet } repeat</l> <l>Consectetuer{2}</l> </lg>
table Lorem ipsum Dolor sit amet Consectetuer{2} repeat	 <table> <row> <cell>Lorem ipsum</cell> <cell>} {problem} repeat</cell> </row> <row> <cell>Dolor sit amet</cell> <cell>} repeat</cell> </row> <row> <cell>Consectetuer{2}</cell> <cell/> </row> </table>

Example 64

19 White spaces

19.1 Vertical white space (between lines)

1. White lines between paragraphs (i.e. empty space, more than the regular paragraph spacing used in the text) should be coded as `<p> </p>`.

2. White lines between poem lines (i.e. empty space, more than the regular poem line spacing) should be coded as `<l> </l>`.

3. Multiple white lines should be captured as one white line.

4. White lines on the title page should be ignored. See Example 16

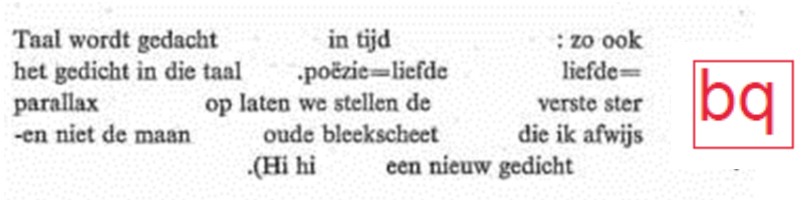
5. White lines before and after the following tags should be ignored:

   - `<head> </head>` (see chapter 3), except when used as a table column head
   - `<q> </q>` (see chapter 4)
   - `<lg> </lg>` (see chapter 7)
   - `<table> </table>` (see chapter 12)
   - `<figure> </figure>` (see chapter 16)

19.2 Horizontal white space (between columns or within a line)

1. White space in the middle of a line of text should be captured as 5 or 10 times ` `. See Example 65.

2. As an exception to the previous rule, if the name or initial letters of the author are written at the end of a line, this should be put in a separate paragraph or line. See Example 66.




```
<q rend="bq">Taal wordt gedacht
           in tijd
           : zo ook
…
</q>
```

Example 65

Groet Bep hartelijk van mij; naar haar vonnis zal ik niet minder gespannen uitzien. Een hart. hand van je Menno
Groet Bep hartelijk van mij; naar haar vonnis zal ik niet minder gespannen uitzien. Een hart. hand van je   Menno

Example 66

20 Cross-references and links

A cross-reference is a reference to a particular spot (the 'target') elsewhere in the text. The target is usually on a different page than the reference. The target is a tag containing an `id` attribute. The value of this attribute is the target identifier. A cross-reference is represented by `<xref>` or `<xptr>`, having an attribute called `to` with a value equal to the target identifier.

A link is a reference to something outside the text, e.g. to a web site. A link is represented by `<xref>`.

Target identifiers used by KB always consist of one letter followed by a number of digits.

20.1 Cross-references within the same document

1. Precode *xptr* followed by a target identifier should be represented by `<xptr to="»target identifier«"/>`. See Example 67 and a (repeated) snippet in Example 70.

Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Cum donoris sociis natoque penatibus. (vz. 189, 31) *xptr v189*	`<p>Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Cum donoris sociis natoque penatibus. (vz. 189, 31)<xptr to="v189"/></p>`

Example 67

2. If a piece of text is marked by precode *xref* followed by a target identifier, then this piece of text should be enclosed between `<xref to="»target identifier«">` and `</xref>`.

Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Cum donoris sociis natoque penatibus. (vz. 189, 31) *xref v189*	`<p>Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Cum donoris sociis natoque penatibus. <xref to="v189">(vz. 189, 31)</xref></p>`

Example 68

3. Precode *id=* followed by a target identifier should be represented by attribute `id="»target identifier«"`, in the nearest tag of type head, p, q, lg, l, table, cell (not row!), list, label, item, figure or figDesc.

id=v189 Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit.	`<p id="v189">Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit.</p>`

Example 69

p. 15-313: please add xptr's after each h3 to the id numbers of the corresponding text on p. 346-444
int pri 16000000
xptr to a001 → h3 *1600. Jubel-Jaar-Lied. Op: Simpelijk geloven en weldoen.*
h4 *Stem*: Zoo 't beghint
	`<div type="chapter">`
	`<interpGrp>`
	`<interp type="primair" value="16000000"/>`
	`<interpGrp>`
	`<head rend="h3">`
	`1600. Jubel-Jaar-Lied. Op: Simpelijk geloven en weldoen`
	`<xptr to="a001"/></head>`
	`<head rend="h4">`
	`Stem: Zoo 't beghint`
	`</head>`
	`...`
	`</div>`

Example 70

20.2 Cross-references to other documents

1. In addition to paragraph 20.1 rule 2, a reference may also point to a target in another document. In this case there will also be a "doc" attribute with the name of the other file in addition to the "to" attribute. The reference is thus captured between the `<xref doc="«`*file name*`»" to="«`*target identifier*`»">` and `</xref>` tag. The value of the "doc" attribute will be indicated with precodes as shown in Example 71.

Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. ~~Cum~~ donoris sociis natoque penatibus. (vz. 189, 31) ~~xref to~~=v189 doc=mult001idee03_01	`<p>Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Cum donoris sociis natoque penatibus. <xref doc="mult001idee03_01" to="v189">(vz. 189, 31)</xref></p>`

Example 71

Example 72 shows an example of a document linked to another document. Example 73 shows how this is captured in the second document. The examples also show how the precodes are indicated to be repeated throughout both documents.

h3 I EEN AMOREUS LIEDEKEN
add xref to a001
p. 9-491: please add below each h3: [Commentaar] and make it an <xref doc=_ant001antw09 to=...> Refer to an id number with the same number as in the h3.
	`<div type="chapter">`
	`<head rend="h3">Een amoreus liedeken</head>`
	`<p>`
	`<xref doc="_ant001antw09" to="a001">[Commentaar]</xref></p>`
	`...`
	`</div>`

h3 3 EEN NYEU LIEDEKEN
add xref to a003, etc.
	`<div type="chapter">`
	`<head rend="h3">Een nyeu liedeken</head>`
	`<p>`
	`<xref doc="_ant001antw09" to="a003">`
	`[Commentaar]</xref></p>`
	`...`
	`</div>`

Example 72

p. 50-493: please add an id-number to each h3, using the number the h3 starts with. id=a001 h3 1 ALS ALLE DIE CRUYDEKENS SPRUYTEN Dialooglied waarin een man en een vrouw ruziënd een punt achter hun relatie zetten. 1 (De vrouw is aan het woord)	 <head rend="h3" id="a001"> Als alle die cruydekens spruyten </head> ...
id=a003 etc. h3 3 ALLE MIJN GEPEYS DOET MI SO WEE Liefdesklacht. Een minnaar wordt door liefdesverdriet verteerd, omdat de vrouw die hij bemint zijn liefde niet langer beantwoordt. De oorzaak van hun verwijdering is het gestook van jaloerse roddelaars.	 <head rend="h3" id="a003"> Alle mijn gepeys doet mi so wee </head> ...

Example 73

20.3 Links

1. If a URL is marked by precode *xref*, then this URL should be enclosed between <`xref type="url" to="«URL»"`> and `</xref>`.

2. If a piece of text is marked by precode *xref* followed by a URL, then this piece of text should be enclosed between <`xref type="url" to="«URL»"`> and `</xref>`.

Zie ook http://www.dbnl.org ← xref	Zie ook <xref type="url" to="http://www.dbnl.org"> http://www.dbnl.org</xref>
Mail de redactie → xref mailto:info@dbnl.org	Mail <xref type="url" to="mailto:info@dbnl.org"> de redactie</xref>

Example 74