

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

Version 1.1, last updated 30-05-2024

This appendix specifies how scans should be converted to TEI-XML and is subject to change.
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.....	4
3	Chapters and headings	5
4	Paragraphs.....	10
5	Colophon.....	12
6	Interp.....	15
7	Poems.....	17
8	Plays.....	22
9	Line beginning tags.....	25
10	Pages and page begin tags.....	27
11	Lists.....	30
12	Tables.....	32
13	Emphasis.....	37
14	Hyphenation.....	38
15	Notes	39
16	Images.....	43
17	Character-level rules	49
18	White spaces	54
19	Cross-references and links	56

1 General requirements

1. The rules for the conversion 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. All pages of the book or periodical should be included in the XML. The text on the cover should always be included in the XML. For periodicals there is a slightly different rule: all cover pages should be completely included in the XML, text and image.
6. All text of a title should be captured as TEI-XML code, unless indicated otherwise. Only the parts of the source material that cannot be captured in TEI-XML code are captured in image files (PNG). Further specifications can be found in chapter 16.
7. All content of a title should be captured on the same page as it is on in the original scan.
8. There should not be any white space characters at the start of a line.
9. Every XML tag specified in chapters 2 to 8, 11 and 12 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.

10. 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.
11. If a cross is drawn through a piece of source material on the precoded scan, then this piece of source material should be skipped.
12. 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 occurrences 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: <i>← speaker</i> 'Hoe heet uw kat eigenlijk?' Psycholoog: <i>← speaker, etc.</i> 'Waarom vraagt u dat?' Boer: '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>
--	--

Psycholoog: 'Het gaat om u, niet om de kat.' Boer: 'Het is toch uw kat? Hoe heet-ie nou?'	<pre> <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> </pre>
--	---

Example 1

13. Do not add library-stickers to the XML file.
14. If a sticker covers (part of) the text, put {problem} in the file at the position of the unreadable text.
15. If a sticker covers (part of) an illustration, only make an image if less than 10% on the border of the image is covered.

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>`. 'If "[interp] between text and body' is mentioned in the instruction file, this interp should also 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>`.
 - 2.1. Note that headings cannot be used within the element `<cf>`.
3. If there is no title present, the precoded scan or instruction file will indicate if a head should be added between square brackets.
4. 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.
5. 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>`.
6. 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. <u>Het eerste levensjaar</u> ...	<pre><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></pre>
--	--

Example 4

7. 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 <u>Het eerste levensjaar</u> ...	<pre><div type="chapter"> <head rend="h3">Hoofdstuk 12</head> <head rend="h4">12.1 Het taalbegrip</head> <head>Het eerste levensjaar</head> ... </div></pre>
--	--

Example 5

8. Chapter divisions should not be nested.

9. Multiple sections on the same level should not be nested.
10. 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:
 - 10.1. Both levels should be represented by chapter divisions. See Example 6.
 - 10.2. 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.
 - 10.3. The exception to rule 10.1 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 7 is applicable, which means that there should not be a div tag for the first inner level chapter. See Example 6.
 - 10.4. 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.
 - 10.5. 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 Chapter 1 ... Chapter 2 ...	<pre> <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> </pre>
---	--

Example 6

Part I Lorem ipsum dolor sit amet. Chapter 1 ... Chapter 2 ...	<pre> <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> </pre>
---	--

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>`.

- The h2 and h3 headings should be considered as the start of chapters and the other headings (h4 and head) should be considered as the start of sections. The rules specified in 3.1 are applicable. See for an exception section 3.3 Chapter starts before its heading.

Deel I Hoofdstuk 1 Eerste bedrijf ...	<pre><div type="chapter"> <head rend="h2">Deel I</head> <head rend="h3">Hoofdstuk 1</head> <head rend="h3">Eerste bedrijf</head> ... </div></pre>
2. De achttiende eeuw 2.1 Napoleon Napoleon werd geboren. 2.2 Staatsgreep Hij greep de macht.	<pre><div type="section"> <head rend="h4">2. De achttiende eeuw</head> <head>2.1 Napoleon</head> <p>Napoleon werd geboren.</p> <div type="section"> <head>2.2 Staatsgreep</head> <p>Hij greep de macht.</p> </div></div></pre>

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 a 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 following the instruction should then be included in a section.

chapter starts here Lorem ipsum dolor sit amet. Lorem ipsum ...	<pre><div type="chapter"> <p>Lorem ipsum dolor sit amet.</p> <div type="section"> <head rend="h3">Lorem ipsum</head> ... </div></div></pre>
--	---

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 new headings should always be captured between square brackets. See Example 10.

please add h3 [Lorem ipsum] Lorem ipsum dolor sit amet.	<pre><div type="chapter"> <head rend="h3">[Lorem ipsum]</head> <p>Lorem ipsum dolor sit amet.</p> </div></pre>
---	--

3.5 Div type= "lang"

1. When a book contains the same text in two languages on opposite pages, we indicate this in the XML. Each piece of text must be placed in a section and given a ‘lang’ attribute. This will always be indicated in the precoded scan or instruction document.
2. If it applies to the whole title this will be indicated on the first page of the precoded scan or on top of the instruction document with the following instruction: “For all even pages, use <div lang="XX">. For all odd pages, use <div lang="XX">”, see Example 11.
If it applies to certain pages only, the instruction will be given at the relevant pages.
3. Sometimes two different languages may appear on one page. If this is the case, the instruction will be: “Please use <div lang="fr-nl">”

4. If div type=lang is used in a text, the value of the lang attribute has to be declared in the TEI-header. Please add therefore the following in the TEI-header between </encodingDesc> en <revisionDesc>:

```
<profileDesc>
<langUsage>
<language id="XX">XXXXXX</language>
</langUsage>
</profileDesc>
```

Dag Het leven van Bertus Aafjes.	Day The life of Bertus Aafjes.	<div lang="nl" type="section"> <head rend="h4">Dag</head> <p> Het leven van Bertus Aafjes.</p> </div> </div>	<div lang="en" type="section"> <head rend="h4">Day</head> <p> The life of Bertus Aafjes.</p> </div> </div>
--	--	--	--

Example 11

3.6 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 *surttitle* should be captured between <head rend="surttitle"> and </head>
2. The precode *sub* or *subttitle* 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 12.

sur Hoofdstuk 1 h3 Een voetreis naar Rome sub Het leven van Bertus Aafjes cap Door Marcel Janssens Lorem ipsum dolor sit amet.	<div type="chapter"> <head rend="surttitle">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> <p>Lorem ipsum dolor sit amet.</p> </div> </div>
---	---

3.7 Uppercase and lowercase

- Capitalization in heads may be based on:
 - if the word is a proper noun such as a (place) name
 - if the head is listed in a table of contents
 - if the same word occurs elsewhere in the book or, in case of a periodical, article.
- If a heading is completely printed in uppercase *or* completely in small capitals, the precoded scan indicates with an underline which letters are to remain capitalized. The letters that are not underlined should be captured in lowercase.
Note that instructions in the precoding take precedence over rule 1.
- 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 13

- If a heading consists of uppercase, lowercase and small capitals, the following rules apply:
 - capture uppercase as uppercase
 - capture lowercase as lowercase
 - capture small capitals as lowercase when a word contains both uppercase and small capitals and capture small capitals as uppercase when a word contains only small capitals. See Example 14.

head	TABLE 4 Number of 1-2 and 2-1 orders for three different types of modal auxiliary – infinitive clusters in the syntactically annotated part of the CGN*					table	<table> <head>Table 4 Number of 1-2 and 2-1 orders for three different types of modal auxiliary – infinitive clusters in the syntactically annotated part of the CGN</head> [...]
		<i>kunnen</i> – infinitive	<i>moeten</i> – infinitive	<i>zullen</i> – infinitive	Total		
	1-2	951	848	787	2586		
	2-1	7	4	3	14		

Example 14

4 Paragraphs

1. Paragraphs should be enclosed within `<p>` and `</p>`. Whether the first line is indented or not is ignored.
2. Line beginnings in a paragraph, not explicitly precoded by `ℳ/`, should be represented by starting a separate paragraph. See also chapter 9, rule 2.2.
3. White lines between paragraphs (i.e. empty space, more than the regular paragraph spacing used in the text) should be coded as `<p> </p>`. See chapter 18 for more rules on white lines.
4. 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.	<pre><p>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.</p> <p>'Waarom vraagt u dat?''</p> <p>'Nou, ik zag die kat van u langs mij schieten - hij is toch van u?''</p> <p>&nbsp;</p> <p>De therapie zou jaren in beslag nemen. Soms fietste ik wel vijf keer per week naar de stille laan.</p></pre>
---	---

Example 15

5. If a paragraph continues on the next page, a page begin tag should be inserted at the right position (somewhere between `<p>` and `</p>`). See chapter 10, rule 7 for an example. The paragraph should not be closed at the page break.
6. 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 `<p>` tags. Both the `<speaker>` and `<p>` 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.'	<pre><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></pre>
--	--

Example 16

7. An indented paragraph is marked by the precode `bq`. This means 'block quote'. In that case, the following rules apply:
 - 7.1. The tag pair `<p>...</p>` should be replaced by `<q rend="bq">...</q>`.
 - 7.2. White lines before or after the block quote should be ignored.
 - 7.3. If a block quote breaks a text element or is located beside the text, capture the `bq` between two (text) elements, using the element break closest to the `bq` in the original.

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

Example 17

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).
 - 2.1. Note that headings cannot be used within the element `<cf>`.
3. It is also possible for only part of the page to be placed between colophon tags; if this is the case, it will be indicated in the precoded scan or instruction file.
 - 3.1. The *cf*-part of the page must not be captured in a separate chapter.
 - 3.2. The *pb* tag should not be captured within the *cf* tag.
4. If precode *cf* is indicated for multiple pages in a row these pages should be captured within the same *cf* tag and within the same chapter, even if there are blank pages in between.
5. If precode *cf section = n* is written then the following XML tag should be used: `<cf section="n">`.
6. 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.

<i>cf section = 1</i> IV Oude zinnen Grammaticale analyse van het Nederlands tussen 1200–1700 Frank van Gestel 1992 Martinus Nijhoff Uitgevers	<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>
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean con ligula eget dolor.	<pre> <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit.</p> <cf> <p>Aenean con ligula eget dolor.</p> </cf> </pre>

Example 18

7. If precode *cf* is indicated for parts outside the book chapters and these are on consecutive pages, these parts should be captured within the same chapter and cf tag, even if there are blank pages and/or hidden pages in between.
 - 7.1. If precode *cf* is indicated for parts outside the book chapters at the end of the book, these pages should not be captured in a separate chapter but as part of the previous one.
 - 7.2. Hidden pages directly before or after a cf tag should be captured within the preceding or following cf tags.
 - 7.3. Blank and hidden pages do not need to be captured between cf tags. If a blank or hidden page is directly before or after a cf tag, then the blank/hidden page must be captured within this cf tag unless this cf tag is a blurb. In case of a cf-blurb the blank/hidden page must be captured outside the cf tag.
8. 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’). These pages should always be captured as colophon, unless indicated otherwise in the precoded scan or instruction file.
 - 8.1. All text on the ‘voorplat’ is captured as it appears in the original. Each line of text is captured on a new line in the transcription. See Example 18.
 - 8.2. If a `<cf>` or `<cf section="blurb">` starts on top of the page, the pb-code should be placed directly at the beginning within those cf tags. See Example 19.
9. Part of the text on (one of) these pages can be precoded as ‘*blurb*’. This part should be tagged as `<cf section="blurb">`. See Example 19.

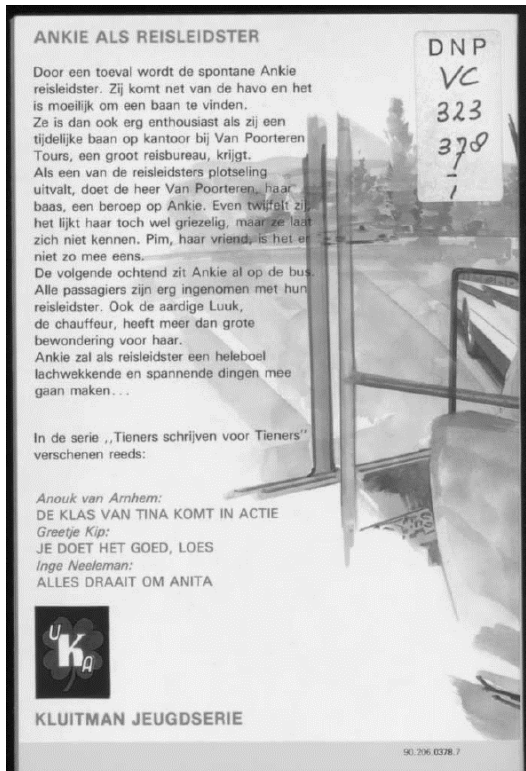
The following rules apply:

- 9.1. The tag `<cf section="blurb">` cannot be nested within other cf tags. Whenever this is the case, the first cf tag should be closed, after which the new cf-blurb tag can be opened.
- 9.2. Images may not be captured within a cf-blurb. Whenever this is the case, the image should be captured between regular colophon tags, i.e. `<cf>` and `</cf>`.
- 9.3. If all text that belongs to a `<div>...</div>` is cf AND cf-blurb, then do not place the `<div>` and `</div>` tags between colophon tags. Instead, merge the chapter with the previous or following chapter.

cf

achterplat

blurb



```

<div type="chapter">
<p>...</p>
<cf section="blurb">
<pb ed="" 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>
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>
[...]
```

Example 19

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*.

An interp should always be placed in an interp group directly under the div tag chapter, see for exceptions rule 6.3, 6.5 and 6.6.

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
<i>pri</i>	primair
<i>sec</i>	secundair
<i>bio</i>	biografie
<i>ill</i>	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, or if the instruction ‘[interp] between text and body’ is written in the instruction, 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 <div> tag of the corresponding chapter. This also applies if the chapter starts on the page before the head.
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 h4 or head. In that case, the interp group should be placed right after the <div> 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 or head in the precoded scan or instruction document).
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.	<pre><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></pre>
 interp beeld gou_001	<pre><figure> <interpGrp> <interp type="beeld" value="gou_001"/> </interpGrp> <p><xptr to="<<image file name>>" /> </p> </figure></pre>

Example 20

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 may have a `<head>`, usually explicitly precoded as head. The `<head>` tag should be positioned right after the `<lg type="poem">` tag, without a `<div>` tag. The capitalization rules defined in paragraph 3.7 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 21 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 weegake ligk 't wiechake, lollig lachend leuk geziechake, wiegewaggelt mèt ze weegake, reikend nao e wiedweg leechake. Heider klaor kristalle uigakes, loeter straolend vaan piezeer kieke oonbezörriг lachend jao op alles roondum neer. Zulle ins die einse ouge later krieke vaan sjagrijn ? of 't leech gebroke weure door 't lijd en door de pijn ? Hennekes wie vaan e pupke mèt de fingerkes zoe teer, hawwe stevig vas umsiote 'ne nulje rammeleer. Zalle ins die han, daan groeter, wèr'ke veur 't daagieks broed ? Zalle zij vāōl goods ins geve ? Of mesjen doen zoevāōl koed ? 'tMunneke, 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 vāōl lijd en laster daan ? Sjattig leuke pōppe uurkes luustere nao Moolers stum, en es leefdeswāōrdjes klinke drejt 't kōpke z'ch 'ns um. Zalle later ouch die oere luust're nao wat Mooler zeet ? Of nao't slechs vaan op de wereld, en nao't goole gaaroet neet ? In 't weegake ligk 't wiechake, lollig lachend leuk geziechake, aordig mollig sjattelijke..... Bleefste zoe mer altied, prijke ! 20
<pre><lg type="poem"> ... <l>meh um ziech te laote hure</l> <l>krejt 't wel zoe noe en daan</l> <l>&nbsp; </l> <pb n="20"/> <l>Brink dee moond, es heer kin spreke</l> <l>troes en leefde boe 'r kaan?</l> ... </lg></pre>	

Example 21

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>`

<i>head</i> Lorem ipsum <i>poem</i> Lorem ipsum dolor sit Cum sociis natoque it Aenean commodo ligula eget dolor.	<pre><div type="section"> <head>Lorem ipsum</head> <lg type="poem"> <l>Lorem ipsum dolor sit</l> <l>Cum sociis natoque it</l> </lg> <p>Aenean commodo ligula eget dolor.</p> </div></pre>
<i>poem</i> Lorem ipsum <i>head</i> Lorem ipsum dolor sit Cum sociis natoque it Aenean commodo ligula eget dolor.	<pre><lg type="poem"> <head>Lorem ipsum</head> <l>Lorem ipsum dolor sit</l> <l>Cum sociis natoque it</l> </lg> <p>Aenean commodo ligula eget dolor.</p></pre>

Example 22

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 23. Note that this is different for prose, see chapter 9 rule 4.

Verkapt chanson <i>head</i> <i>poem</i> 1. Bonsoir mon amour - de rozen op je kleet 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 kleet met die rozen was.	<pre><div type="section"> <head>Verkapt chanson</head> <lg type="poem"> <l n="1.">Bonsoir mon amour - de rozen</l> <l>op je kleet doen mij onwillekeurig</l> <l>denken aan het ziltgeurig</l> <l>onderkleed en het schaamteloze</l> <l>&nbsp;</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 kleet met die rozen was.</l> </lg> </div></pre>
---	--

Example 23

9. If a poem line is indented, attribute `rend` should be added to the line tag. Note that if all lines within one poem are equally indented, no indentation is needed. The amount of indentation within one poem does not depend on other poems, see Example 24.

<i>poem</i> De bloempjes moe gekust, Nu kwijnt het dartel windje, En lispelt op een trillend blaadje Zijn laatsten adem uit.	<pre><lg type="poem"> <l rend=""; 1tab">De bloempjes moe gekust,</l> <l>Nu kwijnt het dartel windje,</l> <l>En lispelt op een trillend blaadje</l> <l>Zijn laatsten adem uit.</l> </lg> <lg type="poem"> <l>De zonne stak zoo heerlijk</l> <l>Heur hoogtijdvlag in top.</l> <l>Wij ademden begeerlijk</l> <l rend=""; 1tab">De honinggeuren op.</l> </lg></pre>
---	---

Example 24

9.1. The value of the attribute is "; [n]tab". If the poem lines have different indentation levels, use no indentation for the least indented line, "; 1tab" for the next level, etc. The maximum level is 9, i.e. "; 10tab" is not permitted.

9.2. The indentation within one poem should be consistent and relative to the sheet mirror. Stanza's of the same poem but on different pages should be captured with the same amount of indentation compared to the other stanza's.

Note that in many books the even pages have a wider margin on the left and the odd pages have a wider margin on the right. This does not affect indentation.

9.3. If a poem is centered or right aligned on the page, do not use indentation. See Example 25.

<i>h3</i> JONGE BADERS <i>poem</i> Het was een schoone morgen Al op de purpren hei. We gingen zonder zorgen En zongen liedjes blij. De zonne stak zoo heerlijk Heur hoogtijdvlag in top. Wij ademden begeerlijk De honinggeuren op.	<pre><head rend="h3">Jonge baders</head> <lg type="poem"> <l>Het was een schoone morgen</l> <l>Al op de purpren hei.</l> <l>We gingen zonder zorgen</l> <l>En zongen liedjes blij.</l> <l>&nbsp;</l> <l>De zonne stak zoo heerlijk</l> <l>Heur hoogtijdvlag in top.</l> <l>Wij ademden begeerlijk</l> <l>De honinggeuren op.</l> </lg></pre>
--	--

Example 25

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>.

<i>poem</i> '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.	
<pre> <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=""; ltab">Ick heb al over langh de deughden van uw Bruyt</l> <l rend=""; ltab">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=""; ltab">Dan 't spyt mij, dat ick niet kan ieder maken kondt,</l> <l rend=""; ltab">Hoe veel geluck en heyl dat U mijn herte gont.</l> <l>&nbsp;</l> <l>U beyder Jonstige Vriendin,</l> <l>ANNA ROEMERS.</l> </lg> </pre>	

Example 26

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

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

Example 27

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 28.

<i>poem</i> Soms roept een kind: Moeder, daar gaat de duivel met zijn staart! Maar moeder antwoordt bleek: Kind, daar [gaat Vader!	<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>
---	---

Example 28

13. Paragraphs or stanza in poems are sometimes separated by asterisks. These asterisks should be captured in the XML.
 - 13.1. If asterisks in a poem are indented, the rules of indentation in poems apply, see rule above. The asterisks should be regarded as one of the lines of the poem.

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>
</sp>
```

<<repeat for each paragraph>>

Example 29

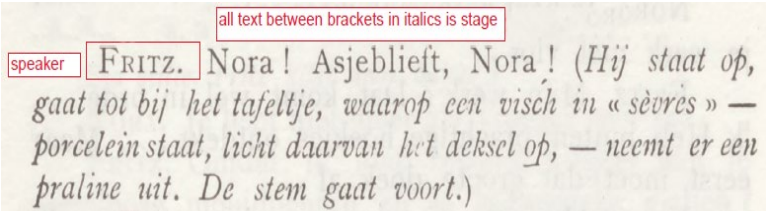
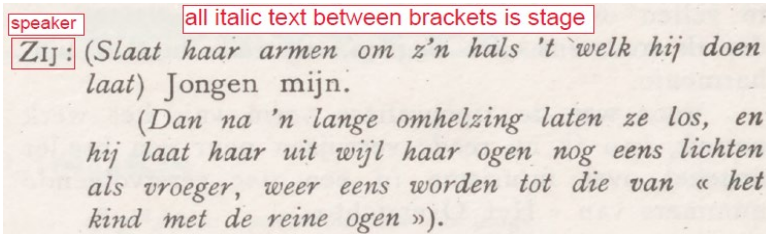
5. White lines before or after a speech and white lines before or after a speaker description should be ignored.
6. Scenes may contain stage directions. Either the instruction document or the precoding explains how to identify these, and/or the precoded scan marks them by using the precode *stage*. Stage directions may appear inside or outside speech. Stage directions should be enclosed within the tags `<stage>` and `</stage>`. See Example 30.

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 ...	<pre> <div type="act"> <head rend="h3">Eerste bedrijf</head> <head rend="h4">Scene 1</head> <stage>Boer en gaeuw-dief komen op.</stage> <sp> <speaker>BOER</speaker> <p>Goedenavond en een goet Jaer.</p> </sp> <sp> <speaker>GAEUW-DIEF</speaker> <p>Mag ick hier slapen?</p> </sp> <div type="scene"> <head rend="h4">Scene 2</head> <stage>Boer en gaeuw-dief lopen verder.</stage> <sp> <speaker>BOER</speaker> <p>Nee, dat mag niet</p> </sp> ... </pre>
--	---

Example 30

6.1. Stages can appear inside or outside a paragraph. See Example 31.

Note that all interpunction should be captured before the closing stage tag.

	<pre> <sp> <speaker>F<hi rend="sc">ritz.</hi></speaker> <p>Nora! Asjeblieft, Nora! <stage>(<hi rend="i">Hij staat op, gaat tot bij het tafeltje, waarop een visch in 'sèvrès' - porcelein staat, licht daarvan het deksel op, - neemt er een praline uit. De stem gaat voort.</hi></stage></p> </sp> </pre>
	<pre> <sp> <speaker>Z<hi rend="sc">ij</hi>:</speaker> <p><stage>(<hi rend="i">Slaat haar armen om z'n hals 't welk hij doen laat</hi></stage> Jongen mijn. </p> <stage>(<hi rend="i">Dan na 'n lange omhelzing laten ze los, en hij laat haar uit wijl haar ogen nog eens lichten als vroeger, weer eens worden tot die van 'het kind met de reine ogen'</hi>).</stage> </sp> </pre>

Example 31

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 *<div type="act">* or *<div type="scene">* element.
 - 8.3. Lines may also contain speakers. 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. The speaker and his or her speech text should also be placed between *<sp>* and *</sp>*.
 - 8.4. It is also possible that a poem occurs within a play with lines. 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 should be used.
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 *<p><lb n="number"/></p>* (see chapter 9, rule 4)

9 Line beginning tags

Line beginning tags are used to begin a new line *within* another code. The tag used for this is `<lb/>`. Line beginning tags may be used within all other codes that contain text. Note that in paragraphs, lines and heads the `<lb/>` tag should be explicitly precoded. See the relevant chapters for rules on line beginning tags within paragraphs, lines and heads.

1. Line beginnings *within a heading* should only be captured if explicitly indicated by the precode `Ub/`. There are two correct ways to capture a line beginning tag in a head. See Example 32. Only one of these two ways may be used within one XML file.

- 1.1. If an instruction document mentions ‘h3 with lb/’ it means the line beginning tag should be placed only in logical places such as after a full stop or comma, or before an evident subtitle.

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

Example 32

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 33.

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

Example 33

- 2.2. Line beginnings in a paragraph, not explicitly precoded by `Ub/`, should be represented by starting a separate paragraph. See Example 34.

Nu volgt een lied over de lente, ook al is het nog winter. Daarna bezingen we de zomer.	<pre><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></pre>
---	---

Ten slotte de herfst.	
-----------------------	--

Example 34

3. Line beginnings *outside a paragraph*, in a `<bq>`, `<cell>`, `<item>`, `<figcaption>`, `<head>`, `<note>` or `<stage>` should be represented by the `<lb/>` tag, even if not explicitly precoded. See Example 35.

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

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

Example 35

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 36. This rule does not apply to poems (see chapter 7, rule 8).

<i>prose with line numbers :</i> Lorem ipsum dolor sit amet, consec tetuer adipiscing elit. Aenean con ligula eget dolor. Aenean massa. Cum 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.	<pre> <p> <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. </p> </pre>
---	---

Example 36

5. Line beginning tags may appear at the start of or within an XML line.
6. The `<lb/>` tag should always be preceded by a space, unless it is placed at the start of a line.

10 Pages and page begin tags

1. All pages of the book or periodical should be included in the XML. 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 or name (see below) 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. The blank pages between the cover pages and the first or last numbered page should be captured as unnumbered pages. Use `<pb rend="hidden" />` for these unnumbered blank pages. 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}" />`.
2. 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="π1" />` or `<pb n="*1" />`.
 - 2.1. If for example '126-127' is written in the top right corner of a page or in the instruction document, this means that the two pages should be combined in the XML. Instead of adding two separate `<pb n="x" />`, you add one, in this case. `<pb n="126-127" />`. This is usually done when an illustration or table covers (part of) two pages.
 - 2.2. Sometimes a page number consists of one or two words, e.g. `<pb n="binnenkant voorplat" />`. If this is the case, this will be indicated in the precoded scan or instruction file:

<i>if present</i> : front of the dust jacket, including the flap (beginning of the book)	<code><pb n="voorkant stofomslag" ed=""/></code>
<i>if present</i> : white inside of dust jacket, including the flap	<code><pb rend="hidden" /></code> (use no <code>n</code> or <code>ed</code> attribute)
front cover	<code><pb n="voorplat" ed=""/></code>
inside front cover	<code><pb n="binnenkant voorplat" ed=""/></code>
inside back cover	<code><pb n="binnenkant achterplat" ed=""/></code>
back cover	<code><pb n="achterplat" ed=""/></code>
<i>if present</i> : white inside of dust jacket, including the flap	<code><pb rend="hidden" /></code> (use no <code>n</code> or <code>ed</code> attribute)
<i>if present</i> : back of the dust jacket, including the flap (end of the book)	<code><pb n="achterkant stofomslag" ed=""/></code>

- 2.3. If the words mentioned above are used in place of a page number, please always add the attribute `ed` followed by an empty value in the `pb` tag, e.g. `<pb n="binnenkant voorplat" ed=""/>`.
- 2.4. Voorplat and achterplat should always be captured as colophon, unless otherwise indicated in the precoded scan or instruction file.
- 2.5. If `<pb ed="" n="achterplat" />` is directly followed by `<cf>` or `<cf section="blurb">`, the `pb`-code should be placed directly at the beginning within those `cf` tags. See Example 19.
- 2.6. In case of a fold-out page, capture the inside-part on the dust jacket or cover (voorkant/achterkant stofomslag or voorplat/achterplat).

3. 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	<pre><pb n="t.o. IX"/> <p>Lorem ipsum</p></pre>
---	---

Example 37

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

4. 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.
5. 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>`.
6. 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"/>`.
 - 6.1. If the page is blank, the value is "blank".
 - 6.2. If the page only contains advertisements, as indicated in the instruction document or by the precode ‘*ad*’, the value is "ad". Frequently the illustration or text belonging to the advertisement will also be crossed-out as to emphasize that it can be ignored.
 - 6.3. If the page only contains notes, as defined in chapter 15, or only notes and advertisements, the value is "notes".
 - 6.4. 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".
 - 6.5. 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.
7. Page headers and footers as well as standard decorations should be ignored.

Lorem ipsum dolor sit amet, consectetur 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.	<pre><p>Lorem ipsum dolor sit amet, consectetur 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></pre>
--	---

Example 38

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.

Lists may be used in the elements <q>, <figDesc>, and <note>.

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/>.

1.4. If, in a list with labels, one line contains only an item, use <label/><item>. If a line contains only a label but not an item, use <label><<label>></label><item/>.

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> </item> if there are no labels, or <label/><item> </item> if there are labels.

Example 39 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	<pre> <p>Lorem ipsum dolor:</p> <list> <label>1)</label><item>Julius Caesar</item> <label/><item>&nbsp;</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> </pre>
---	--

Example 39

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 3.7 are applicable.

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

Example 40

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 41

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

Example 41

12 Tables

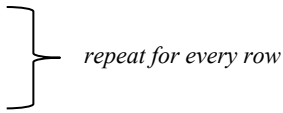
Definition:

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

Tables may be used in the elements `<q>`, `<figDesc>`, and `<note>`.

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>      repeat for every column
</row>
</table>
</p>
```



See Example 42.

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 43.
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 44.
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 3.7 are applicable. See Example 43.
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 45.

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

Example 42

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 43

NAAM	-----	 <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>
Adres	-----	
Woonplaats	-----	

Example 44

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

Example 45

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, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Cum sociis et nato-	que penatibus et magnis dis parturient montes, nascetur rculus mus. Donec quam felis.	<p>Lorem ipsum dolor sit amet, consectetur adipiscing elit.</p> <p>Aenean commodo ligula eget dolor. Cum sociis et natoque penatibus et magnis dis parturient montes, nascetur rculus mus. Donec quam felis.</p>
--	---	---

Example 46

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*”.

14.1. In case of a split table, always add {problem}.

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		hemelsblauw	1807
Phil. en Litt.	wit	licht blauw	groen		korenblauw	1768
						blauw
Jaar van stichting van het Corps:	Maart 1839		Sen. V.A.P. Febr. 1815	Mei 1851	Sen. Vet. Febr. 1814	
Godgeleerdheid	zwart	opgeheven	blauw	wit	rose	opgeheven
Rechtsgeleerdheid	rood	decreet 1811 K.B. 1843	wit	rood	rood	decreet 1811 K.B. 1818
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/>
<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>

```

Medici
hemelsblauw
lichtblauw
1768
Phil. en Litt.
korenblauw
blauw
Jaar van stichting van het Corps:
Mei 1851
Sen. Vet. Febr. 1814
Godgeleerdheid
wit
rose
opgeheven
Rechtsgeleerdheid
rood
rood
decreet 1811 K.B. 1818
Geneeskunde
geel
groen
Wis- en Natuurk.
paars
geel
Lett. en Wijsbeg.
blauw
donkerblauw

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	<i>example text</i>	<code><hi rend="i">example text</hi></code>
Bold	example text	<code><hi rend="b">example text</hi></code>
Underlined	<u>example text</u>	<code><hi rend="u">example text</hi></code>
Double underlined	<u><u>example text</u></u>	<code><hi rend="uu">example text</hi></code>
Superscript	example ^{text}	<code>example <hi rend="sup">text</hi></code>
Subscript	example _{text}	<code>example <hi rend="sub">text</hi></code>
Spaced text	e x a m p l e t e x t	<code><hi rend="spat">example text</hi></code>
Small capitals	EXAMPLE TEXT EXAMPLE TEXT	<code><hi rend="sc">example text</hi></code> <code><hi rend="sc">Example Text</hi></code>
Strike through	example	<code><hi rend="strikethrough">example text</hi></code>

Note: Punctuation characters may have a different emphasis than the surrounding text. See Example 47.

1. In case of a combination of spacing and any other form of emphasis, the spacing code should be captured inside the other code(s). See Example 47.

En zie, het was <i>zeer goed!</i>	En zie, het was <code><hi rend="i">zeer goed</hi>!</code>
En zie, het was <i>zeer goed!</i>	En zie, het was <code><hi rend="i">zeer goed!</hi></code>
En zie, het was <i>z e e r g o e d!</i>	En zie, het was <code><hi rend="i"><hi rend="spat">zeer goed!</hi></hi></code>

Example 47

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:

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

Example 48

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).

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

Example 49

2. If a word on the end of a line is hyphenated and the two parts of the word both begin with a capital, then the hyphen should not be eliminated.
3. 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 50.
4. 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 50.
5. 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.

<i>prose with line numbers:</i> 5 Lorem ipsum dolor sit amet, consec tetuer adipiscing elit. Aenean contra eget dolor. Aenean massa. Cum donoris sociis natoque penatibus et magnis dis parturient montes, nasce- 233	<pre><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.</pre>
10 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.	

Example 50

6. If a range of numbers or a period of time is notated as two numbers or two years with a hyphen in between, then the hyphen should not be eliminated and the page number tag should be inserted after the hyphen. See Example 51.

Dit gebeurde in België tijdens de eerste twee maanden van de Tweede Wereldoorlog (1940- 233	<pre><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.</pre>
1945). De consequenties hiervan kunnen niet genoeg benadrukt worden.	

Example 51

15 Notes

15.1 General

- 1. 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 the following tags: <lb/>, <figure>, <lg type="poem">, <q rend="bq">, <list> and <table>.
- 2. Notes should always be placed directly behind a character, without any spacing See Example 53.
- 3. If a note contains more than one paragraph, these paragraphs must be represented by the <lb/> tag. See also chapter 9, rule 3 regarding new and empty lines in block quotes. Other uses of <lb/> within notes are not allowed.
- 4. Notes should not be used within the cf tag. Instead use paragraph. See Example 52.

	<pre><p><table> <row> <cell>* Makandra</cell> <cell>(samen)</cell> <cell>4</cell> </row> <row> <cell>* Was yu tifi</cell> <cell>(tanden poetsen)</cell> <cell>6</cell> </row> <row> <cell>▪ Tan de</cell> <cell>(blijf daar)</cell> <cell>8</cell> </row> <row> <cell>▪ Piauw</cell> <cell>(piauw spelletje)</cell> <cell>10</cell> </row> </table></p> <p>* composities van Paul Middellijn</p> <p>▪ authentieke Surinaamse melodieën</p></pre>
--	--

Example 52

15.2 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 55.

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

Example 53

15.3 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>
```

2. If a chapter with endnotes contains an endnote on the last page, this is still an endnote and not a footnote. See Example 54.

<note n="<<note identifier>>" place="krit"><<note text>></note>

5 Lorem ipsum dolor sit amet, con sec tetuer adipiscing elit. Quorum con- tra eget dolor. Aenean massa. Cum donoris 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 10 vel, aliquet nec, vulputate eget, arcu. ----- 3 Aenean = Enean <i>text crit note</i> ----- 3 Lorem ipsum: Dolor sit amet <i>bq</i> 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 55

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 <1> and </1>:

<1><note n="<<note identifier>>" place="krit"><<note text>></note></1>

16 Images

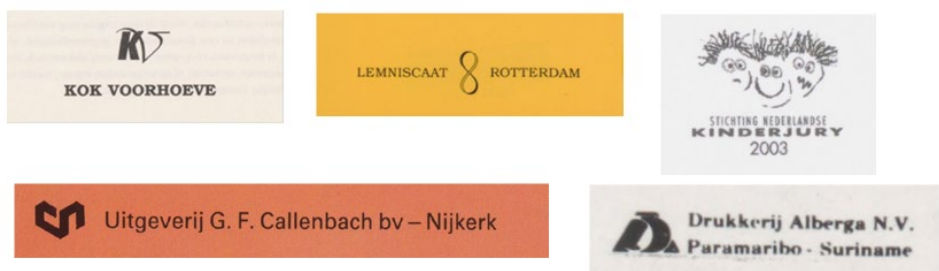
Parts of the source material that cannot be captured in TEI-XML code are captured in image files (PNG), unless indicated otherwise. The file names of these image files are included in the TEI-XML in the right tags, see Appendix E for the correct file names.

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 16.3).

Note:

- In case of logos, always capture the text both in the image and as text, see below for examples of logos.

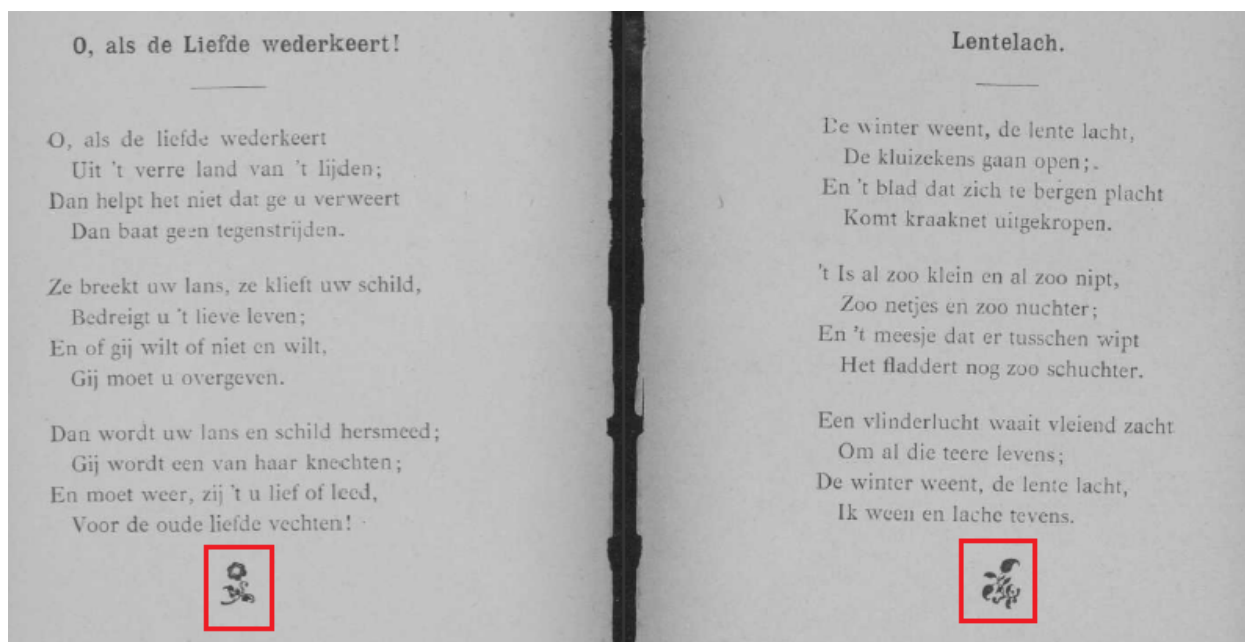


16.1 General

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 17, rule 13).
- Text overlapping an image should not be captured in the PNG file, unless indicated otherwise.

See chapter 4, rule 4 for an example of a decorative paragraph separator. Example 56 shows pages with decorations which should be captured as illustrations, because they are not 'simple' and not 'repeated' but figurative.



Example 56

1. Images 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. Text or images should never be captured on a different page than in the original scan, unless expressly indicated.

- In general, images should not be captured within another text element. See rule 1.4 for the exception.
- Rule 1 and 2 describe the two possible placements.
- Note: each image should be captured with a separate figure tag.

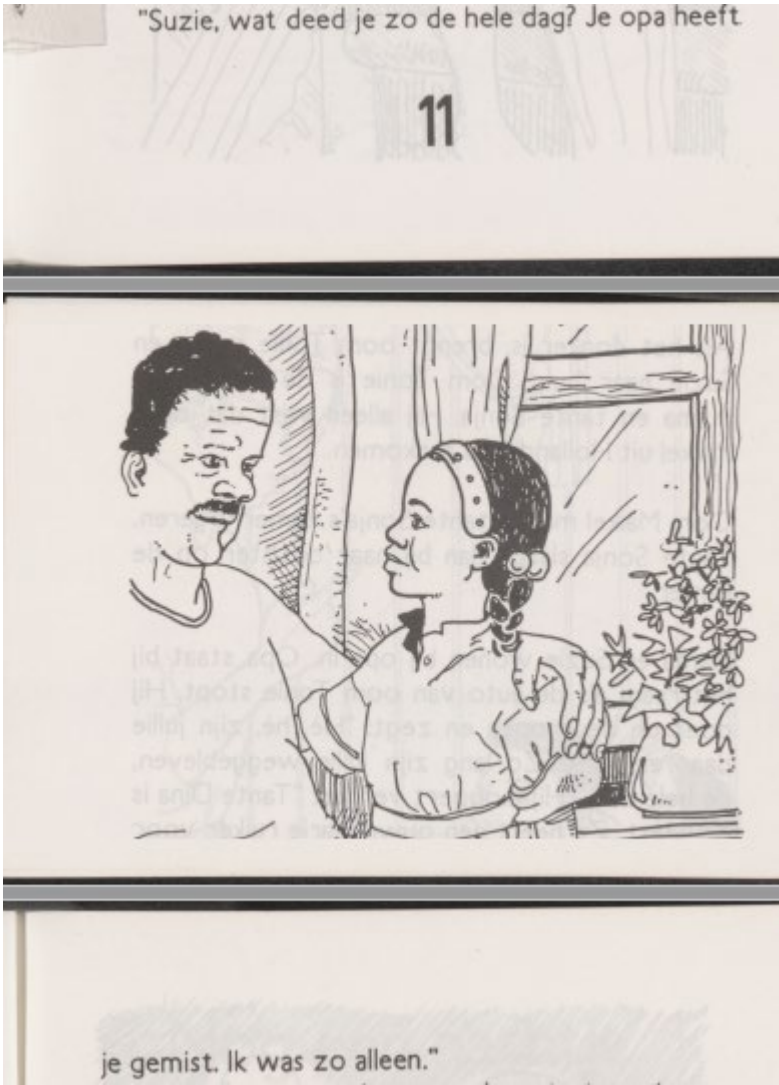
16.2 Illustrations

1. For an illustration that occurs within a text element, such as a <p>, <q>, <head>, <l> and <note> the following rules apply.

- 1.1. Always place the illustration between two text elements, using the text element break *closest* to the illustration in the original. See rule 2.3 for exceptions. If the illustration is placed between two text elements, the figure tag itself should be captured between paragraph tags:

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

- 1.2. If there is no break in the text element at all on the page, the illustration tag will be placed where the illustration is in the original.
- 1.3. Illustrations can never be tagged inside a heading tag <head>. In the unusual case that an illustration breaks up a heading, it should be placed after the closing heading tag. In case of inline images in headings, please contact us.
- 1.4. When the illustration breaks a text element but covers the whole of the (following) page, the text element should not be closed and opened around the figure tag. In the case of a full page illustration breaking a text element, the figure is captured within the text element (see Example 57).

	<pre> <p>'Suzie, wat deed je zo de hele dag? Je opa heeft <pb n="12"/> <p><figure><p><xptr to="<<image file name>>" /></p></figure></p> <pb n="13"/>je gemist. Ik was zo alleen.'</p> </pre>
--	--

Example 57

2. For an illustration that does not occur within a text element, rule 2.1, 2.2 and 2.3 apply. This usually is the case when an illustration is placed besides a body of text and the text continues around it (see Example 59) or because it is placed at the bottom or top of a page, after or before the text (see Example 57).
 - 2.1. 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 59.
 - 2.2. 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.
 - 2.3. 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.
3. In periodicals with more than one issue, the cover of every issue should be added to the XML file. See also Appendix E, 6.2.

4. If the illustration 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. When asked to provide extra large illustrations (for example 800 or 1200 pixels wide/high), the attribute `"rend="zoom""` should be added to the figure tag:

```
<figure rend="zoom"><p><xptr to="<<image file name>>" /></p></figure>
```

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

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

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

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

10. 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*.

11. A figure description may contain the following tags: `<figure>`, `<lb/>`, `<lg type="poem">`, `<list>`, `<note>`, `<table>`, `<xptr>`, `<xref>`.

 <i>fd</i> Leonard de Gou © KbPhoto 1967	<pre><figure><p><xptr to="<<image file name>>" /></p> <figDesc>Leonard de Gou <lb/>© KbPhoto 1967</figDesc> </figure></pre>
--	---

Example 58

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 zooveel 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.

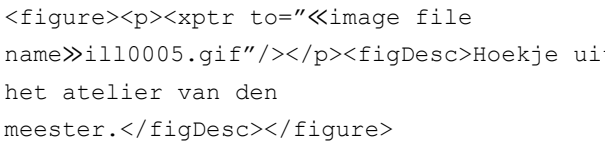


Hoekje uit het atelier van den meester.

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 persoonlijkheiden 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

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

Hoekje uit het atelier van den meester.

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...

Example 59

16.3 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 17, rule 13 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 60 and see chapter 17 rule 1 for more information.

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

Example 60

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 $\lambda z \{ \text{CAUS} [\text{SOTH}, \text{INSTR} (F, z)] \}$ $Q^1, \quad Q^2,$  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.	<pre><p>gen op de hoogte is en niet onmiddellijk inziet dat de samenstelling <hi rend="i">wandelstok</hi> in die kringen met behulp van de formule</p> <p><figure><p><xptr to="<<image file name>>" /></p></figure></p> <p>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.</p></pre>
--	---

Example 61

17 Character-level rules

- 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 4 gives an overview of common XML escape codes and UTF-8 characters. Other characters are possible.

- 1.1. In the case of Greek text/words, please do not use Greek symbols or rare characters, but the normal corresponding letter from the Greek alphabet. This does not apply to the use of Greek symbols in mathematical formulas and the like.

See Example 62. In this word, the **rho symbol** (unicode number U+03F1, HTML code ϱ) is used instead of the **rho letter** (unicode number U+03C1, HTML code ρ). Please capture with the rho letter, so ῥαριον instead of ῥαριον.

ῥαριον		ῥαριον
--------	--	--------

Example 62

Table 4

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’.

- All occurrences of ſ should be replaced by s. Characters ſ and f should be recognized as different characters. See Example 63.

felfs ofje fchoon	selfs ofje schoon
-------------------	-------------------

Example 63

- Sometimes, when indicating currency, an italic letter f is followed by numbers. Capture this character as an entity: Latin Small letter f with hook ƒ followed by a space. See Example 64.
In all other instances capture the italic f as <hi rend="i">f</hi>.

Los exemplaar f 0,60	Los exemplaar ƒ 0,60
----------------------	-----------------------------

Example 64

- Capture fractions as follows:

- 4.1. Use the corresponding UTF-8 fraction characters or character entities if possible (e.g. $\frac{3}{4}$ may be captured by `¾`, `¾` or `¾`). See Example 65.
- 4.2. 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 65.
- 4.3. Otherwise, if 4.1 and 4.2 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 65

5. Ligatures should be captured as separate letters. Note that ‘ij’ is a very common ligature in Dutch. See Example 66. This also applies to Greek ligatures. In linguistic texts, there may be an instruction in the precoding to keep the ligatures.
 - 5.1. Note that the ligature IJ should be captured as two separate capitals.
 - 5.2. ß is a ligature of ss. Always capture ß as two separate letters, see Example 66.

Het rijke œuvre van Stæl	Het rijke oeuvre van Stael
de eerste verordening van Seyß-Inquart	de eerste verordening van Seyss-Inquart

Example 66

6. Dashes should be captured as hyphens (UTF-8 code #002D).
7. 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. See Example 67.

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

Example 67

8. Lines of dots and lines of dashes used as paragraph separators should be regarded as a white line (`<p> </p>`).
9. Lines of dots and lines of dashes used in poems should not be ignored.
10. Series of dots within a text should not be ignored.

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

Example 68

11. Series of dots within a table used to connect two columns should be ignored. See Example 45.
12. Ellipses should always be captured as separate dot characters (. . . instead of ...) without spaces, regardless of the number of dots. Ellipses may be captured with or without a preceding space, depending on how it appears in the original.

Lorem ipsum...	Lorem ipsum...
----------------	----------------

Example 69

13. 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	<hi rend="b">M</hi>aer ghy sult ouer de Iordaen gaen, ende in 't Lant woonen
--	--

Example 70

14. 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 71.
Do NOT put {problem} on missing quotation marks. Capture the quotation marks that are present even w/out its corresponding pair.

15. 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.”	Hij zei: 'Ik lust geen "gezond" eten.'
Hans zei: <<Salade is gezonder dan patat!>>	Hans zei: 'Salade is gezonder dan patat!'

Example 71

16. 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.
- 16.1. In the case of a repeated opening quotation mark at the start of each line, the quotation will not always have a closing quotation mark. The end of the quotation will then only be indicated by the final opening quotation mark at the start of the last line of the quotation. In that case, capture the closing quotation mark at the end of the sentence in that line. See Example 72.

„Voor Coster bestaat, in welken vorm dan ook : een moraal. En „voor zijn tegenvoeters is iedere gedachte daaraan al een verschrikking, omdat zij den mensch liefhebben om zijn zelfs wil en zooals „hij reilt en zeilt. Zij beoordeelen het schilderij om den gloed en den „zwier der schildering, om zijn poëtische substantie ¹⁾ en niet om „wat het voorstelt.	<pre><p>'Voor Coster bestaat, in welken vorm dan ook: <hi rend="spat">een moraal</hi>. En [...] <hi rend="spat">voorstelt</hi>.'</pre>
---	--

Example 72

17. Apostrophes should be captured as ' or with the following UTF-8 character: '. They should not be captured as quotation mark.

17.1. If there is only one single quotation mark in the sentence, it should be captured as an apostrophe. Mostly this occurs when:

- the quotation mark is situated before or after a one-letter word, such as 't or 'k, often at the beginning of a sentence;
- the quotation mark is situated not at the end nor at the beginning of a sentence;
- the quotation mark is situated within a word, for example 'zo'n'.

18. 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.

18.1. 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.

<i>list</i> { Lorem ipsum Dolor sit amet Consectetuer{2}	<pre><list> <item>{ {problem} Lorem ipsum</item> <item>{ Dolor sit amet</item> <item>Consectetuer{2}</item> </list></pre>
--	---

Example 73

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

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

	<pre><cell>Dolor sit amet</cell> <cell> repeat</cell> </row> <row> <cell>Consectetuer{2}</cell> <cell/> </row> </table> </p></pre>
--	--

Example 74

19. Paragraphs or stanza in poems are sometimes separated by asterisks. These asterisks should be captured in the XML.

19.1. If asterisks in a poem are indented, the rules of indentation in poems apply (see chapter Poems, rule 9); the asterisks should be regarded as one of the lines of the poem.

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 76.

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

Example 76

18.3 Spaces in abbreviations

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

Table 5

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

2. Punctuation marks should always be captured without a preceding space, except for ellipses. See chapter 17, rule 12.

19 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.
- Only add cross-references or links if specifically instructed.

19.1 Cross-references within the same document

1. Precode `xptr` followed by a target identifier should be represented by `<xptr to="<<target identifier>>" />`. See Example 79 and a (repeated) snippet in Example 82.
 - 1.1. If the instruction document mentions the `xptr` should be added ‘on top of page’, the `xptr` code should be captured in a separate paragraph tag at the top of the page (see Example 78)
 - 1.2. If the paragraph was started on the previous page, close the paragraph on the previous page and open a new paragraph on top of the new page so the `xptr` may still be added in the separate paragraph tag on top of the page. See Example 79.

<i>Instruction:</i> "p. 3-563: Please add xptrs on top of each page to corresponding id numbers on p. 568-607 ("a003", "a007", "a017", etc.), if there is an id number with the same number on p. 568-607"	<pre>[...] <pb n="3"/> <p><xptr to="a003"/></p> [...]</pre>
--	---

Example 77

Please make xptrs on top of each page to corresponding id numbers on pdf p. 296-314, if there is an id number with the same number on pdf p. 296-314.	<pre><p>Tot 1830 van de bloemenbalsems, welke ze in hare vaart hebben opgenomen, zuchten van angstige beklemdeheid, bedaard in het lommer bij ruischende beeken, zuchten van lust en van liefde, gekoeld aan het kristal der</p> <pb n="10"/> <p><xptr to="a010"/></p> <p>fonteinen krampachtige trek.</p></pre>
---	--

Example 78

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

Example 79

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>`. See Example 80.

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

Example 80

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 or figure (not figDesc!). See Example 81 and Example 82.

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

Example 81

p. 15-313: please add xptr's after each h3 to the id numbers of the corresponding text on p. 346-444 <i>xptr to a001</i> → <i>h3</i> 1600. Jubel-Jaar-Lied. Op: Simpelijk geloven en weldoen. <i>h4</i> Stem: Zoo 't beghint <i>int pri 16000000</i>	<pre> <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> </pre>
--	--

Example 82

19.2 Cross-references to other documents

1. In addition to paragraph 19.1 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 83.

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

Example 83

Example 84 shows an example of a document linked to another document. Example 85 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 <u>EEN</u> AMOREUS LIEDEKEN add xref to a001 p. 9-491: please add below each h3: <p>[Commentaar]</p> and make it an <xref doc=_ant001antw09 to=...> Refer to an id number with the same number as in the h3.	<pre><div type="chapter"> <head rend="h3">Een amoreus liedeken</head> <p> <xref doc="_ant001antw09" to="a001">[Commentaar]</xref></p> ... </div></pre>
h3 3 <u>EEN</u> NYEU LIEDEKEN add xref to a003, etc.	<pre><div type="chapter"> <head rend="h3">Een nyeu liedeken</head> <p> <xref doc="_ant001antw09" to="a003"> [Commentaar]</xref></p> ... </div></pre>

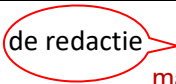
Example 84

p. 50-493: please add an id-number to each h3, using the number the h3 starts with. h3 I <u>ALS</u> ALLE DIE CRUYDEKENS SPRUYTEN id=a001 Dialogooglied waarin een man en een vrouw ruziënd een punt achter hun relatie zetten. I (De vrouw is aan het woord)	<pre><div type="chapter"> <head rend="h3" id="a001"> Als alle die cruydekens spruyten </head> <p>...</p> </div></pre>
id=a003 etc. h3 3 <u>ALLE</u> 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.	<pre><div type="chapter"> <head rend="h3" id="a003"> Alle mijn gepeys doet mi so wee </head> <p>...</p> </div></pre>

Example 85

19.3 Links

1. If a URL is marked by precode *xref*, then this URL should be enclosed between `<xref type="url" to="«URL»">` and `</xref>`. See Example 86.
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>`. See Example 86.

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

Example 86