Encoding Guide

for Critical Editions

Release Version 1, 2023
Arlo Griffiths & Axelle Janiak

Version History

Author(s), Contributor(s)	Version	Changes	Date
Arlo Griffiths, Giulia Buriola, Axelle Janiak	0.1	Redaction of the first version	2020-2023
Arlo Griffiths, Axelle Janiak	1	Release of first version	2023

Table of Content

Preliminaries

This guide is primarily intended for use by scholars who are preparing a *digital* critical edition of a so-far unedited text.[1] The encoding strategies offered here also attempt to take into account the situation where a new edition of a previously edited text is envisaged, and will in most cases even be applicable for encoding a published critical (or even a not-so-critical) edition. However, this Guide is explicitly not intended for diplomatic editions based on single manuscripts. The method to be deployed for editing texts, especially texts written in vernacular languages, based on a *codex unicus* — whether to maintain two separate files with diplomatic/unnormalized text in one, and critical/normalized text in another, or both combined in one file — is a matter of case-by-case appreciation and something that needs to be experimented with more than we have so far had the time to do. For the time being, we only have experience with keeping the diplomatic edition of a *codex unicus* separate from that of the critical edition based on it.

We assume that the purpose of a critical edition is to transcend the state of a given text as transmitted to us in available manuscripts and to give readers access to an earlier state of the text, from which the extant transmission derives.[2] To evoke the useful distinction made by G. Thomas Tanselle in his beautiful essay *A Rationale for Textual Criticism*,[3] a critical edition is thus a hypothesis concerning the text of a work in the abstract, while a diplomatic edition aims to present the text of a concrete document.

The DHARMA project is never concerned with texts accessible in autograph manuscripts, and always with texts that have been transmitted over centuries of copying from one manuscript to the other, even though only a single witness may survive. The manuscripts we have are always characterized by some degree of difference from the text as it was intended to be read at the time of composition. The task of the critical editor is to present as much evidence as is necessary for establishing a scholarly hypothesis as to the text of the work being edited — and to reduce the volume of the evidence presented as much as possible. It is necessary to sift between evidence that is pertinent and evidence that is only a distraction. This Guide aims to enable you to present the edition of the text and the evidence on which your

---

[1] We will not dwell here on the reasons why a scholar might prefer to edit a text digitally. On this subject, see Roberto Alessi & Marcello Vitali-Rosati (eds.), *Les éditions critiques numériques*, Montréal: Presses de l'Université de Montréal, https://pum.umontreal.ca/catalogue/les_editions_critiques_numeriques, 2023. In specific relationship to Sanskrit philology and related fields, we may refer to Birgit Kellner, "On Editing Sanskrit Texts Digitally – Tools, Methods and Implications", in *Sanskrit Manuscripts in China III: Proceedings of a Panel at the 2016 Beijing International Seminar on Tibetan Studies, August 1 to 4*, edited by Birgit Kellner, Xuezhu Li, and Jowita Kramer, 93–106, Beijing: China Tibetology Publishing House, 2020.
(https://www.oeaw.ac.at/ikga/publikationen/sachgebiete/tibetologie/sanskrit-manuscripts-in-china-iii)

[2] The reconstructive aims that underlie critical editions vary from case to case. For instance, the aim may be to reconstruct the text of the work as written by its author, the text of the work as it was known by a given commentator, the state of the text that the witnesses allow to reconstruct, etc.

[3] Tanselle, G. Thomas. 1989. *A Rationale of Textual Criticism*. University of Pennsylvania Press. https://www.jstor.org/stable/j.ctt3fhcwk.

choices of reading are based, while also furnishing suggestions that may help you to reduce the volume of the evidence by weeding out what is not essential.

Because the aim of the critical edition is to facilitate access to the abstract content of the text of the work in question, and because trivial orthographic variation, that potentially distracts from accessing the intellectual contents, is inherent in all manuscript traditions with which we are concerned, just as it must have been even in any hypothetical autograph manuscripts now lost to us, your critical edition is expected to make use of **loose transliteration** (or **transcription**) as intended in the *DHARMA Transliteration Guide*.

Digital editions of the DHARMA project will be represented in XML files and comply with the encoding standards of the Text Encoding Initiative (TEI).[4] The TEI guidelines offer a very broad range of options for how to encode the kinds of information that the scholar making a critical edition needs to represent. The present *DHARMA Encoding Guidelines for Critical Editions* documents the choices made by the DHARMA project from among these options.[5] It is in some cases a specialization of the *DHARMA Encoding Guide for Diplomatic Editions* written by Dániel Balogh & Arlo Griffiths and in some others a complement to it. For this reason, some parts of the present document will be referring to this first guide. Other guides created and curated for the specific needs of the DHARMA project will also be referred to. Some of the recurring ones have been given an abbreviated form that you can find in the following list with the link to the latest release.

EGC	DHARMA Encoding Guide for Critical Editions
EGD	DHARMA Encoding Guide for Diplomatic Editions
FNC	DHARMA Files Naming Conventions
PVG	DHARMA Prosody and Verse Forms Guide
STS	DHARMA Symbol Taxonomy Supplement
TG	DHARMA Transliteration Guide
ZG	DHARMA Zotero Guide

To conclude these preliminaries, we wish to explain that the DHARMA project treats readings from manuscripts and published editions on equal footing. This position, which may appear counterintuitive to some, is based on both theoretical and practical considerations. Theoretically, it seems problematic to make a distinction between the two types of sources because premodern scribes often worked in a manner analogous to modern-age editors, basing a new copy on the consultation of multiple available manuscripts. Practically, it is in many situations

---

[4] For a very easy introduction, we recommend Lou Burnard's *What Is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources*, OpenEdition, 2014, https://doi.org/10.4000/books.oep.426, 2014.

[5] In making our choices, we have made use of the following guides in ascending order of frequency:
      <https://digitallatin.github.io/guidelines/LDLT-Guidelines.html>
      <https://www.digitalmanuscripts.eu//wp-content/uploads/2017/09/04-Textual-variants-MB.pdf>
      <https://betamasaheft.eu/Guidelines>
      <https://sarit.indology.info/sarit-pm/docs/encoding-guidelines-simple.html>
      <https://sarit.indology.info/sarit-pm/docs/encoding-guidelines-full.html>
      <https://pta.bbaw.de/pta/project/encoding-guidelines#building-a-digital-edition-according-to-tei>

impossible to make a distinction, because many published editions give no information about the manuscripts consulted, and even if they do, the manuscripts used by previous editors will often not be available to us, making it impossible for us to distinguish between cases where an editor accurately read his manuscript(s) and those where he misread them. This problem is further aggravated by the fact that many editors have not even attempted to indicate in which manuscripts the readings they adopted were observed, and indeed whether they observed them in any manuscript at all, or whether they consciously resorted to alteration of the transmitted text. We have no choice but to take the editor's readings as what they are: readings reflecting the editor's idea about how the text was formulated. The concrete implication of this position is that both manuscripts and printed editions will be declared as witnesses (<witness>) in the metadata section (<teiHeader>) of each file, where relevant in interaction with bibliographic data stored in the project's Zotero Group Library.[6]

Acknowledgments

This Guide is the product of three years of often intensive collaboration between the authors, in interaction with a group of scholars each involved, to some degree, in digitally editing texts from premodern India or Indonesia, some of them on the basis of the guidelines formulated here. We would like to express our thanks to Aditia Gunawan, Giulia Buriola (who has written drafts of a number of paragraphs included here), Martina Dello Buono, Dominic Goodall, Csaba Kiss, Andrew Ollett, Marine Schoettel, Amandine Wattelier-Bricout (who has furnished contents for one section) and Zakariya Pamuji Aminullah.

1. Structure of a DHARMA Critical Edition

The contents of DHARMA critical editions will be encoded in several files: the main file contains the edition itself, while translations form a secondary set of files linked to the first one. The main sections of this guide concern the edition, but you can nonetheless find information on how to deal with metadata and translations, respectively under §2. <teiHeader> — Recording Metadata and §9. Translation.
    Templates for these files are provided in the project-documentation repository of the DHARMA project.[7] All of them have a <TEI> element as root with the TEI namespace declared with @xmlns="http://www.tei-c.org/ns/1.0" as well as the default declaration of language set to English with @xml:lang="eng". You first need to add an @xml:id for the edition file. An @xml:id provides a unique identifier for the element bearing the attribute, and it may contain only alphanumeric characters from the ASCII tables. Moreover, @xml:id are case-sensitive: uppercase and lowercase letters are distinguished. Then, add an attribute @corresp link to this identifier on other files and prefix it with "#", to establish the correspondence between the files

---

[6] For the latest release of the ZG, check the GitHub project-documentation github repository.
[7] Currently, two templates in their first version are available in the templates folder, under CriticalEditions-mss.

you want to link together. Finally, provide an attribute @type with the value "edition" or "translation" on this root.

Example 1.A	
`<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="svayambhu" xml:lang="eng" type="edition">`	`<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="eng" type="translation" corresp="#svayambhu">`

Most of this content and preceding elements like the XML declaration and the processing instructions will be given in the templates and you won't have to fill them in. However, please note that the first one should always be formed with the attributes @version="1.0" and @encoding="UTF-8". This part of the will remain stable, unlike the processing instructions containing the URI of the validation schema.[8]

     The following example is a shortened version of the general structure of any edition file. More information is provided all through the present guide.

Example 1.B

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="https://raw.githubusercontent.com/erc-dharma/project-documentation/master/schema/latest/DHARMA_CritE
dSchema.rng" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model
href="https://raw.githubusercontent.com/erc-dharma/project-documentation/master/schema/latest/DHARMA_CritE
dSchema.rng" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<?xml-model
href="https://raw.githubusercontent.com/erc-dharma/project-documentation/master/schema/latest/DHARMA_CritE
dSQF.sch" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="eng" xml:id="id" type="edition">
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title type="main">Encoding Template for Critical Editions</title>
      <title type="sub">Digital Critical Edition of...</title>
      <editor ref="part:jodo">
        <forename>John</forename>
        <surname>Doe</surname>
      </editor>
      <respStmt>
        <resp>creation of the file</resp>
        <persName ref="part:jodo">
          <forename>John</forename>
          <surname>Doe</surname>
        </persName>
```

---

[8] Critical editions are validated with one RelaxNg file containing embedded schematron rules and one schematron file with some Quick Fixes for the most common cases.

```xml
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <authority>DHARMA</authority>
        <pubPlace>Paris</pubPlace>
        <idno type="filename">DHARMA_templateCriticalEdition_v01</idno>
        <availability>
          <licence target="https://creativecommons.org/licenses/by/4.0/">
            <p>This work is licenced under the Creative Commons Attribution 4.0 Unported Licence. To view a copy of the licence, visit https://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.</p>
            <p>Copyright (c) 2019-2025 by John Doe.</p>
          </licence>
        </availability>
        <date from="2019" to="2025">2019-2025</date>
      </publicationStmt>
      <sourceDesc>
        <listWit>
          <witness xml:id="id">
            <!-- Elements about a witness -->
          </witness>
        </listWit>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <projectDesc>
        <p>This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 809994).</p>
      </projectDesc>
                                        <schemaRef        type="guide"        key="        EGCv01"
url="https://docs.google.com/document/d/15HFxHJTOzIU1UDyVrB2yQYJ5wI6JyEshEkYgg5qwj8M/edit#"/>
    </encodingDesc>
    <profileDesc>
      <langUsage>
        <!-- choose the language(s) of your text, remove the others -->
      </langUsage>
      <textClass>
        <keywords>
          <!-- In due course we will can use this element to declare keywords as <term>s, but at the moment please ignore <keywords>  -->
        </keywords>
      </textClass>
    </profileDesc>
    <revisionDesc>
      <change who="part:axja" when="2020-01-06" status="draft">Creation of the template</change>
      <!-- replace this by a line that has you as @who, that has the @date on which you finished the first basic encoding of the file, and that has the following description of what was finished on that date: "Initial encoding of the file." Then add a <change> entry for each significant change to the file. -->
    </revisionDesc>
```

```
</teiHeader>
    <text xml:space="preserve" xml:lang="lg-code"><!-- update the language code -->
                <body>
                <!-- Structure as necessary with <div>, <p>, <ab>, <lg> depending the nature of the text -->
                </body>
        </text>
</TEI>
```

2. <teiHeader> — Recording Metadata

As editors of digital editions, we need to produce machine-readable information of the manuscripts or witnesses, editions and bibliographic references we are working with. These details should be encoded within the <teiHeader>, which works as a container for all the metadata regarding the materials in use for a digital edition. Note that a TEI-XML file without metadata does not have any value outside of the project. However, you will not need to build the <teiHeader> yourself, since the DHARMA project provides its own ready-made template for critical editions. We will now describe in sequence all TEI elements foreseen within the template.

2.1 <fileDesc> — Describing the XML file

The first element <fileDesc> is mandatory. It contains a description about the XML file itself, such as its title and the persons responsible for its content. The first, mandatory, child element of <fileDesc> is the <titleStmt>; as will be explained below, the <titleStmt> provides information on the title and those responsible for its content.

2.1.1 <titleStmt> — Titles and people involved

The <titleStmt> contains the title given to your digital edition as well as one or more optional statements of responsibility such as encoders, editors, authors, compilers and so on.

2.1.1.1 <title> and <author> — Declaring the title of the work and the name of its author
The element <title> indicates the title for the web publication of the digital edition. Following the TEI guidelines, we recommend that you indicate the title of the work being edited in its original language, if possible with a translation of the same into English, and a subtitle that should at least describe the nature of your edition, but may also contain words that describe the nature of the original text being edited.[9]
        Use an element <title> with @type="main" to state the title of the original work in the original language — but feel free to use stem forms of Sanskrit words — and add @xml:lang

---

[9] This will allow a distinction between the electronic work and the source text in citations and in catalogs which contain descriptions of both types of material.

with the relevant ISO language code.[10] To this can be added a corresponding <title type="alt"> for its English translation, no @xml:lang being required for it because English is set as our default language on the <TEI> root element in our template for critical editions.[11] Finally, it is recommended to work with a <title type="sub"> to state a subtitle that may include further description of the text and of the edition, in plain English. If the name of the work's author is known, you can encode it with <author>, after all the <title> elements. We recommend you use a normalized form of the name when possible.[12]

As the following two examples show, you should minimally give one <title type="main" xml:lang="..."> and one <title type="sub">, while you may optionally use a <title type="alt" xml:lang="..."> to translate the title of the original text.

Example 2.1.1.1.A
<titleStmt>   <title type="main" xml:lang="kaw-Latn">Svayambhu</title>   <title type="sub">Digital Critical Edition and Translation of an Old Javanese Juridical Text</title> </titleStmt>

Example 2.1.1.1.B
<titleStmt>   <title type="main" xml:lang="san-Latn">Naimittikakriyānusandhāna</title>   <title type="alt">Exposition of Occasional Rites</title>   <title type="sub">Digital Critical Edition and Translation of a Śaiva Ritual Manual</title>   <author>Brahmaśambhu</author> </titleStmt>

Even though we recommend keeping the usage of <title> elements simple, it is possible to work with more than one <title> element to provide a title for a base text and one or more dependent texts. In this situation, which should only occur if you are in fact editing base and dependent text in integrated form, the attribute @type="main" is still expected, but an attribute @subtype="base-text" or @subtype="commentary" must be added to connect the respective titles to the corresponding segments of the edition. When there is more than one dependent text, you need to add an extra @n attribute, in order to link the title to the text concerned.[13]

Example 2.1.1.1.C
<titleStmt>

---

[10] See EGD §10.3 and Appendix D.

[11] See §8.9.1 for more on use of @xml:lang.

[12] It might be worth considering adding the author in the authority list of the project, as well as linking it to international reference frameworks. Consult the XML-TEI Data Manager about this.

[13] To deal with the different titles, we are reusing the solution established for SARIT, and available in their Guidelines §4.1.1.

```
        <title type="main" subtype="base-text" xml:lang="san-Latn">Aṣṭāṅgahṛdayasaṁhitā</title>
        <title type="main" subtype="commentary" xml:lang="san-Latn" n="1">Sarvāṅgasundarā</title>
        <title type="main" subtype="commentary" xml:lang="san-Latn" n="2" >Āyurvedarasāyana</title>
        <title type="sub">Digital Critical Edition of a Sanskrit Medical Text with two Commentaries</title>
</titleStmt>
```

When you have a case in which several titles are provided, corresponding to base text and dependent texts, the different <author> elements can be associated with their respective works using their attribute @role, recommended values being "base-author" and "commentator". Once again, when there is more than one dependent text, you need to add the extra @n attribute.[14]

Example 2.1.1.1.D

```
<titleStmt>
        <title type="main" subtype="base-text" xml:lang="san-Latn">Aṣṭāṅgahṛdayasaṁhitā</title>
        <title type="main" subtype="commentary" xml:lang="san-Latn" n="1">Sarvāṅgasundarā</title>
        <title type="main" subtype="commentary" xml:lang="san-Latn" n="2" >Āyurvedarasāyana</title>
                <title type="sub" subtype="editorial">Digital Critical Edition of a Sanskrit Medical Text with two
Commentaries</title>
        <author role="base-author">Vāgbhaṭa</author>
        <author role="commentator" n="1">Aruṇadatta</author>
        <author role="commentator" n="2">Hemādri</author>
     </titleStmt>
```

2.1.1.2 <editor> and <respStmt> — Identifying scholarly responsibilities

The <editor> element is allowed to record the editors of the critical edition. Any person identified as editor and member of the project should be identified using the attribute @ref. Its value is built with the prefix "part:" followed by the identifier of the person. All the identifiers are given in the reference list stored in the project-documentation repository.

If you need to add as editor any colleague who cannot be considered as project participant, and is thus not present within the reference list, assign this <editor> an @xml:id and a @ref using the person's VIAF permalink as value.[15] Use the first two letters of her/his forename followed by the first two letters of the surname in the case of western names, and the two first letters of each part of the name in the case of non-western names.

In all cases, specify the editor's names inside the <editor> element, using the elements <forename> and <surname> for western names, but only with <name> when such a division doesn't apply.

Example 2.1.1.2.A

---

[14] This example is taken from SARIT Guidelines §4.1.1.
[15] To find a person's VIAF permalink, go to http://viaf.org/.

```
<editor ref="part:anac">
                <forename>Andrea</forename>
                <surname>Acri</surname>
 </editor>
 <editor ref="part:adgu">
                <name>Aditia Gunawan</name>
 </editor>
 <editor xml:id="coedes" ref="http://viaf.org/viaf/66465311">
                <forename>George</forename>
                <surname>Cœdès</surname>
 </editor>
```

For any other responsibility, use the <respStmt>. This element must include an element <resp>, which can be used to describe a person's role. The name of the person is then given with the element <persName> with the attribute @ref. If the person has already been encoded above, then use @ref in the same was as shown above under <editor> for project members, but for others use @ref="#ID" using the value of the @xml:id previously defined. The name of the person should be encoded with both elements <forename> and <surname>, or simply <name>, as necessary.

Example 2.1.1.2.B
```
<respStmt>
        <resp>EpiDoc encoding</resp>
        <persName ref="part:anac">
          <forename>Andrea</forename>
          <surname>Acri</surname>
        </persName>
        <persName ref="part:adgu">
          <name>Aditia Gunawan</name>
        </persName>
        <persName ref="#coedes">
          <forename>George</forename>
          <surname>Cœdès</surname>
        </persName>
</respStmt>
```

If several aspects of responsibility need to be distinguished, every <resp> must be wrapped in a separated element <respStmt>. In other words, repeat <respStmt> as many times as you need, with a single <resp> in each one.

Example 2.1.1.2.C
```
<respStmt>
```

```
        <resp>diplomatic edition of the codex unicus</resp>
          <persName ref="part:argr">
                  <forename>Arlo</forename>
                  <surname>Griffiths</surname>
          </persName>
          <persName ref="part:tilu">
                  <forename>Timothy</forename>
                  <surname>Lubin</surname>
          </persName>
</respStmt>
<respStmt>
           <resp>encoding of the edition</resp>
                  <persName ref="part:argr">
                          <forename>Arlo</forename>
                          <surname>Griffiths</surname>
                  </persName>
</respStmt>
```

2.1.2 <publicationStmt> — Metadata on the digital publication

The element <publicationStmt> contains metadata concerning the publication of the digital critical edition. It is a formal recommendation of the TEI that elements supplying information about publication place, address, identifier, availability and date be given following the name of the publisher, distributor or authority concerned, and preferably in that order.

In most cases, the element <authority> should at least identify the DHARMA project. Feel free to add elements regarding your specific situation, if your edition is associated with another project, institution or other, when necessary.

Please fill in the name of the city (or names of the cities) where you (and your co-editors) work. As you see in the example below, editors based in different locations may include the respective publication places in one <pubPlace>.

The element <idno> provides the name of the file itself for archiving purposes, since it helps keeping track of the record after the project. The template furnished the attribute @type with the mandatory value "filename". Follow the File Naming Conventions (FNC)[16] established for DHARMA, but do not keep the extension of the filename.

The <availability> is set by default to a Creative Commons licence identifying the author. It will only require your intervention in the second paragraph, where you will have to state the name(s) of the copyright holder(s).

Example 2.1.2.A
<publicationStmt>   <authority>DHARMA</authority>

---

[16] Available at the following link: https://bit.ly/2JytLbz.

```
        <pubPlace>Lyon (France) and Lexington, VA (USA)</pubPlace>
        <idno type="filename">DHARMA_CritEdSvayambhu</idno>
        <availability>
          <licence target="https://creativecommons.org/licenses/by/4.0/">
                <p>This work is licenced under the Creative Commons Attribution 4.0 Unported Licence. To view a
copy of the licence, visit https://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, 444
Castro Street, Suite 900, Mountain View, California, 94041, USA.</p>
            <p>Copyright © 2019-2025 by Arlo Griffiths and Tim Lubin.</p>
          </licence>
        </availability>
        <date from="2019" to="2025">2019-2025</date>
  </publicationStmt>
```

2.1.3 <sourceDesc> — Identifying and describing your sources

Inside the <sourceDesc>, you can record details about the original manuscripts and printed texts used as witnesses in establishing your critical edition. Start with a <listWit>, which is the container for all the witnesses that you will be referring to in the critical apparatus.

2.1.3.1 Declaring an individual witness

Each individual witness has to be declared with its own tag <witness>. The declaration requires, notably, the definition of a machine-readable siglum for each witness in the form of an @xml:id. As previously stated, the value of any @xml:id serves as a unique identifier for the element bearing the attribute; it can contain only alphanumeric characters from the ASCII tables and is case-sensitive.

Any specific information on the typographical rendering (e.g. use of superscript; see §8.8.1), symbols or non Latin characters can be given for the display using the <abbr> (abbreviation) element as shown in the example with the attribute @type="siglum". It should always be the first child of <witness>.

Example 2.1.3.1.A

```
<sourceDesc>
      <listWit>
            <witness xml:id="K2">
                    <abbr type="siglum">K<hi rend="superscript">2</hi></abbr>
                    <msDesc>
                            <msIdentifier>
                                    <settlement>Singaraja, Bali, Indonesia</settlement>
                                    <repository>Kirtya</repository>
                                    <idno>II A/3/774</idno>
                            </msIdentifier>
                    </msDesc>
            </witness>
            <witness xml:id="EdP">
```

```
                            <abbr type="siglum">Ed<hi rend="superscript">P</hi></abbr>
                            <bibl><ptr target="bib:Pollock1993_01"/></bibl>
                    </witness>
            </listWit>
    </sourceDesc>
```

The @xml:id's defined here will be used while encoding the critical apparatus, as explained in §5.4.

If you need to cite the readings of any published edition among your witnesses, note that you must furnish the bibliographic references within <witness> using the tags <bibl> and <ptr> as shown above. Should you need to limit the scope of the reference to certain specific parts or pages, etc, of the publication in question, use <citedRange> in the manner exhaustively presented in §8.4.1.

A single witness can be associated with more than one specific presentation of its text. This might be the case, for instance, when the witness is a work that presents what is intended to be the same edition in two scripts, such as the Raghu Vira edition presented in ex. 2.1.3.1.B with Balinese and Devanāgarī versions. The main <witness> is here given its @xml:id to record readings that are identical in both scripts. However, we have added in this main <witness> element a <listWit> in order to add two more <witness> elements representing the respective scripts with their respective @xml:id's and <abbr> elements to allow differentiation if differences need to be reported. To create the @xml:id, reproduce the main one and add the code to represent the script after a hyphen.

Example 2.1.3.1.B

```
<witness xml:id="EdRV">
        <abbr type="siglum">Ed<hi rend="superscript">RV</hi></abbr>
        <desc>Global siglum for the Balinese- and Devanāgarī-script editions</desc>
            <listBibl>
                    <bibl>
                                <ptr target="bib:RaghuVira1962_01"/>
                    </bibl>
            </listBibl>
            <listWit>
                    <witness xml:id="EdRV-bal">
                    <abbr type="siglum">Ed<hi rend="superscript">RV</hi><hi rend="subscript">bal</hi></abbr>
                    </witness>
                    <witness xml:id="EdRV-dev">
                    <abbr type="siglum">Ed<hi rend="superscript">RV</hi><hi rend="subscript">dev</hi></abbr>
                    </witness>
            </listWit>
    </witness>
```

2.1.3.2 Declaring a family of witnesses

Should you desire to regroup several witnesses because you identify them as belonging to the same branch of transmission, i.e. you consider them a family of sorts, open an element <listWit> inside which you will declare an element <witness> for every witness you wish to associate with this group of manuscripts. Since such a family is a concept rather than a physical object, the @xml:id to identify the group as a whole should be recorded at the <listWit> level.[17]

The TEI model doesn't foresee the possibility of providing an <abbr> element directly as a first child of <listWit>, as it does for <witness>, so you must use an element <head> to give a name to the family and add at the end, in an <abbr type="siglum">, the relevant encoding for its typographical rendering. If you want to explain your motivation behind the gathering of several witnesses, you can use the element <desc>.[18]

Example 2.1.3.2.A

```
<listWit xml:id="msj">
        <head>The Javanese Manuscripts  (<abbr type="siglum">J</abbr>)</head>
        <desc><!-- possible explanation --></desc>
        <witness xml:id="msja"><abbr type="siglum">J<hi rend="subscript">1</hi></abbr>
                <msDesc>
                        <msIdentifier>
                                <settlement>Jakarta</settlement>
                                <institution>Perpustakaan Nasional Republik Indonesia</institution>
                                <collection>Koleksi Naskah Nusantara</collection>
                                <idno>L631</idno>
                        </msIdentifier>
                </msDesc>
        </witness>
        <witness xml:id="msjb"><abbr type="siglum">J<hi rend="subscript">2</hi></abbr>
                <msDesc>
                        <msIdentifier>
                                <settlement>Jakarta</settlement>
                                <institution>Perpustakaan Nasional Republik Indonesia</institution>
                                <collection>Koleksi Naskah Nusantara</collection>
                                <idno>L298</idno>
                        </msIdentifier>
                </msDesc>
        </witness>
        <witness xml:id="msjc">
                <abbr type="siglum">J<hi rend="subscript">3</hi></abbr>
                <msDesc>
                        <msIdentifier>
                                <settlement>Jakarta</settlement>
                                <institution>Perpustakaan Nasional Republik Indonesia</institution>
```

---

[17] On this matter, DHARMA follows the model of the Digital Latin Library, §7.3.1.1.

[18] The use of <desc> is constrained and has the disadvantage of not being able to contain structuring elements such as <p> and ab>. See the possibilities in the *TEI Guidelines*: https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-desc.html.

```
                        <collection>Koleksi Naskah Nusantara</collection>
                        <idno>L241</idno>
                    </msIdentifier>
                </msDesc>
            </witness>
    </listWit>
```

This representation of the witnesses allows you to refer to both the branch of transmission and the witnesses that belong to it, as per convenience. With reference to the above example, referring in your apparatus to the identifier "msj" would be considered a shortcut to quoting the three witnesses at once, i.e. **J** would be equal to **J₁**, **J₂** and **J₃**. Declaring the family doesn't limit the possibility of quoting the witnesses J_1, J_2 and J_3 on their own.

Although there are no limits regarding the possibility of sub-categorizing the witnesses, we recommend you to limit yourself to the most meaningful ones. Please note that a witness can only be declared once. If you expect it to be associated to more than one family, one or several <listWit> will have to be used to internally structure those groups.

2.1.3.3 Declaring an archetype or a hyparchetype

An archetype represents the most recent common ancestor of known manuscripts while a hyparchetype is to be understood as the ancestor of a family of manuscripts. Though they are no longer surviving, you can still make it possible to encode the reading they are inferred to have had with the element <witness>. In theory such a transmissional situation could be represented using the family encoding described in §2.1.3.2. That solution lists the different witnesses both inferred and surviving without implying their relationships. However, if describing the relationships between branches of transmission is one of your preoccupations, you can embed a <listWit> inside a <witness>.[19]

For example, for the critical edition of the *Skandapurāṇa*, the editors use three manuscripts forming the recension S, meaning that they descend from to the same archetype: S1, S2 and S3/4. Yokochi had demonstrated that S1 stems from another hyparchetype than the common ancestor of S2 and S3/4.[20] The latter is in fact a single witness that has become separated over two collections. (See §2.1.3.4 on this kind of scenario.) The manuscripts can be analyzed into two families that stem from two hyparchetypes, to which Yokochi has assigned the sigla α and β. The latter has two branches, β1 and β2. The diagram below shows how Yokochi visualized the relationship between the witnesses.

---

[19] On this matter, DHARMA follows the model of the Digital Latin Library, §7.3.1.2. The example below is taken from that source.

[20] Yuko Yokochi (2013), *The Skandapurāṇa: the Vindhyavāsinī Cycle*, Leiden: Brill, figure 2, p. 51 and p. 58.
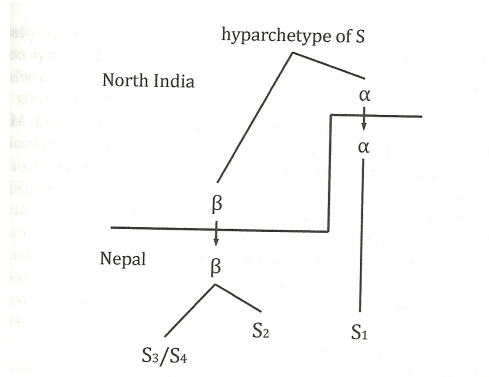
Figure 2   Revised Stemma Codicum of the S recension

If presented in the form of a list of witnesses, the result could be as follows, with the corresponding TEI code shown in the following example:

Archetype S: ancestor of the hyparchetypes α and β
- hyparchetype α
  - S1 National Archives, Kathmandu, MS2-229. (AD 810/811)
- hyparchetype β: ancestor of β1 β2
  - hyparchetype β1
    - S2 National Archives, Kathmandu MS 1-831 (Undated)
  - hyparchetype β2
    - S3 Bodleian Library, Oxford, MS Sanskrit.a.14 (R). (Undated)
    - S4 National Archives, Kathmandu, MS4-2260. (Undated)

Example 2.1.3.3.A

```
<listWit>
  <witness xml:id="S">
        <!-- archetype manuscript S -->
        <abbr type="siglum">S</abbr>
        <desc>Archetype S: ancestor of the hyparchetypes α and β</desc>
        <listWit>
          <witness xml:id="α">
                <abbr type="siglum">α</abbr>
                <desc>Hyparchetype α</desc>
            <listWit>
              <witness xml:id="S1">
                <abbr type="siglum">S<hi rend="subscript">1</hi></abbr>
                <msDesc>
                <msIdentifier>
                        <settlement>Kathmandu</settlement>
                        <repository>National Archives</repository>
                        <idno>MS2-229</idno>
```

```xml
            </msIdentifier>
          </msDesc>
        </witness>
      </listWit>
    </witness>
  </listWit>
<listWit>
  <witness xml:id="β">
        <abbr type="siglum">β</abbr>
        <desc>Hyparchetype β: ancestor of hyparchetypes β1 β2</desc>
      <listWit>
        <witness xml:id="β1">
        <abbr type="siglum">β1</abbr>
        <desc>Hyparchetype β1</desc>
          <listWit>
            <witness xml:id="S2">
            <abbr type="siglum">S<hi rend="subscript">2</hi></abbr>
            <msDesc>
            <msIdentifier>
                <settlement>Kathmandu</settlement>
                <repository>National Archives</repository>
                <idno>MS1-831</idno>
            </msIdentifier>
            </msDesc>
            </witness>
          </listWit>
      </witness>
    </listWit>
  <listWit>
    <witness xml:id="β2">
      <abbr type="siglum">β2</abbr>
      <desc>Hyparchetype β2</desc>
     <listWit>
       <witness xml:id="S3-S4">
       <abbr type="siglum">S<hi rend="subscript">3</hi>/S<hi rend="subscript">4</hi></abbr>
       <msDesc>
         <msIdentifier>
       <!-- instead of <msIdentifier>, you can also use <altIdentifier><idno n="..."/></altIdentifier> -->
           <msName>Constituent parts of S3/S4</msName>
         </msIdentifier>
       <msFrag xml:id="S3">
         <head><abbr type="siglum">S<hi rend="subscript">3</hi></abbr></head>
         <msIdentifier>
           <settlement>Oxford</settlement>
           <repository>Bodleian Library</repository>
           <idno>MS Sanskrit.a.14 (R)</idno>
         </msIdentifier>
       <!-- msContents and other elements possible -->
       </msFrag>
```

```xml
                    <msFrag xml:id="S4">
                      <head><abbr type="siglum">S<hi rend="subscript">4</hi></abbr></head>
                      <msIdentifier>
                        <settlement>Kathmandu</settlement>
                        <repository>National Archives</repository>
                        <idno>MS4-2260</idno>
                      </msIdentifier>
                    <!-- msContents and other elements possible -->
                    </msFrag>
                    </msDesc>
                    </witness>
                  </listWit>
                </witness>
              </listWit>
            </witness>
        </listWit>
      </witness>
<!-- End of hyparchetype S-->
</listWit>
```

2.1.3.4 Describing witnesses

Within each <witness> element that concerns a manuscript, you can supply a description of it, using the element <msDesc>. The description should furnish the basic information needed to identify the witness, such as its location, both geographical and institutional, and its identifier (shelfmark or identification pattern used by the holder), and should also furnish metadata on the script(s) deployed in the witness.

The <msDesc> is itself composed of four elements:

1. <msIdentifer>
2. <msContents>
3. <physDesc>
4. <history>

All four will be furnished in the encoding template and may not be removed, though it is not mandatory to fill in any contents for <msContents> and <history>. By contrast, <physDesc> must always be filled in, and <msIdentifier> must contain at least an <idno>, if the manuscript is kept in an institution, while it may contain other elements as well, such as <settlement> and <repository>. The encoding of <settlement> becomes mandatory if you are describing a manuscript kept in a private collection for which no identifier can be specified with <idno>.

In case of multi-text manuscripts it may be interesting to record which texts have been copied along with the edited text (and perhaps the foliation marking the beginning and end of this text). The element <msContents> offers the possibility to describe the content of any manuscript or part of it, and provides the <summary> element that is also used in the DHARMA project for encoding inscriptions. Note that if you associate a <summary> element with a

specific witness, this means that you need to provide a summary of the contents of your witness and not of your edition.

If your manuscript contains several texts or textual units, not all of which are covered by your edition, and you wish to make explicit which texts have been copied along with the text(s) you are editing, use the element <msItem> for each individual work within it. You can record basic information such as the author, the title, the location in the manuscript with folio references, the language of the text and others, respectively with the elements <author>, <title>, <locus> and <textLang>. Inside <msItem>, you may add the element <filiation> to provide information on the relationship between the manuscript and other surviving manuscripts of the same text, either specifically or in a general way. You can use <summary> to formulate any remark that pertains to the contents of the manuscript as a whole.[21]

Example 2.1.3.4.A
<msContents>   <summary>This is a multi-text manuscript. The constituent parts show distinct hands and three separate foliation ranges in Balinese script. Continuous pagination has been applied by pencil in a Roman hand. Both Balinese and Roman numberings show irregularities.</summary>   <msItem n="1">   <locus>fol. 1r-29v, p. 1-57</locus>   <title>Vratiśāsana</title>   </msItem>   <msItem n="2">   <locus>fol. 1v-11r, p. 61-80</locus>   <title>Ṛṣiśāsana</title>   </msItem>   <msItem n="3">   <locus>fol. 11r-21v, p. 80-106</locus>   <title>Śaivaśāsana</title>   </msItem>   <msItem n="4">   <locus>fol. 1v-5r, p. 109-115</locus>   <title>Śīlakrama</title>   </msItem> </msContents>

We furnish below (example 2.1.3.4.G) an alternative method for encoding composite manuscripts with <msPart> that allows a focused codicological description of the constituent parts.

Finally, you may also provide a <colophon> element to record any colophon. It should be structured with one or more <quote> elements on which you must provide an attribute @type with one of the following values: "diplomatic", "normalized" or "translation". Note that the @type="translation" can only be used if another type of <quote>, either "diplomatic" or "normalized", is also present. The element <quote> can contain tags typically used in a

---

[21] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-msContents.html

diplomatic edition, such as <unclear>, <add>, <del>, <gap/>, <supplied> etc. The element <note> should only be used at the end of the <quote>, as the last child. Finally, you should specify an attribute @xml:lang with the corresponding language code (see §Appendix B for the list of codes used in the project). Use <colophon> only for scribal statements not shared between your different witnesses; for text-final colophons that are a common part of the transmission, use the solution given in §3.3.2.2.

Example 2.1.3.4.B

```
<msContents>
    <msItem>
        <colophon>
         <quote xml:lang="kaw-Latn" type="normalized">iti tatva bima svarga, lulukatan parisamapta</quote>
        </colophon>
    </msItem>
</msContents>
```

Example 2.1.3.4.C

```
<msContents>
   <msItem>
     <colophon>
        <quote xml:lang="osn-Latn" type="diplomatic">Itiḥ kahuvusan·niṁ putaka bima Ḷpas· harani :ka,
sam·map·ta sam·pun· sinurata <gap reason="lost" unit="character" quantity="1"/>  Iṅ vulan· kasa, saṅ anurat·
panadaAn· sakapat· pun·, qəñcu nu ṅaheraṁ bukit· cikuray· samapun</quote>
     </colophon>
   </msItem>
</msContents>
```

The element <physDesc> is used to record metadata about the physical aspects of each of your witnesses, notably the script(s) that they use, encoded with the child element <handDesc>. It is possible to include in <physDesc> a simple free-text paragraph in a <p>; it is also possible to include such a <p> in more specialized child elements that can be part of <physDesc>. The simplest case is shown in the following examples:

Example 2.1.3.4.D

```
<physDesc>
    <p>Generic descriptive prose on the support on which this copy of the text is written.</p>
    <handDesc>...</handDesc>
</physDesc>
```

Although as a general rule it is not recommended in TEI to combine unstructured prose descriptions with the usage of more specialized elements, it may not always be possible to avoid combining specialized and generic elements that may be combined in a single <physDesc>.[22] Note that, for the document to be valid, all generic elements (i.e. <p>) must precede the first specialized element in the description, as in the following slightly more elaborate example:

Example 2.1.3.4.E

```
<physDesc>
   <p>Generic descriptive prose.</p>
   <objectDesc>
        <p>Generic descriptive prose on the support on which this copy of the text is written.</p>
   </objectDesc>
   <handDesc>...</handDesc>
</physDesc>
```

Among the numerous elements that can be nested inside your <physDesc>, the only one that is mandatory is <handDesc> (see §2.1.3.1), which can be used to describe the different hands used in the given manuscript but must be used to state the script that is used, even if there is only one hand. However, if you wish to provide details related to the artefact itself, use an element <objectDesc>. It is not mandatory that you structure it, so you can simply use a <p> element to encode relevant information such as dimensions, folio and so on. But if you need a finer structure, try to follow this model:

Example 2.1.3.4.F

```
<physDesc>
    <objectDesc form="concertina">
        <supportDesc>
          <support>Tree-bark</support>
            <extent>27 folios
                <dimensions scope="all" type="folio" unit="cm">
                        <height>27</height>
                        <width>24</width>
                </dimensions>
            </extent>
        </supportDesc>
        <layoutDesc>
            <layout>The texts run from folio A1 to A27, and from B1 to B26, B27 is left blank. The beginning of a
```

---

[22] There might be some circumstances in which avoiding the usage of such a mixed approach is not possible. This could happen, for instance, (1) if the description you are writing already exists in a prose form, where some of the specialized topics are treated together in paragraphs of prose, but others are treated distinctly; (2) if some parts of the description cannot be mapped to a pre-existing specialized element, although all parts of the description are clearly distinguished.

```
text is preceded by the <foreign>bindu na godang</foreign> <g type="B1"/> whereas the division into smaller
sections of the texts is stressed by the <foreign>bindu na metmet</foreign> <g type="B2"/>. From B1 to B9 the
internal divisions of the text are stressed by another type of <foreign>bindu na metmet</foreign> in which its upper
right-side has been elongated shaping a single black line <g type="BL"/>. The black line precedes and runs
parallel to the first line of the text, until the right-side margin. On these same folios, each small section of text is
accompanied by a drawing, representing different anthropomorphic figures.</layout>
                    </layoutDesc>
            </objectDesc>
            <handDesc>
                <handNote>
                            description of letters, possibly including
                            <height>letter-heights</height>
                </handNote>
            </handDesc>
                <bindingDesc>
                    <binding>
                            <p>Two wooden covers</p>
                    </binding>
                </bindingDesc>
    </physDesc>
```

In the case of a manuscript composed of parts that were originally separate and that have been
bound together at a later stage, it is possible, after describing the artefact as a whole in its
present condition in <physDesc> to furnish an element <msPart> for the constituents, making it
possible to formulate focused descriptions of the physical characteristic (dimensions, hands,
etc.) of each part. The following example is an alternative to the approach illustrated in
2.1.3.4.A:

Example 2.1.3.4.G

```
<msDesc>
    <msIdentifier>
            <settlement>Leiden, the Netherlands</settlement>
            <institution>Leiden University</institution>
            <repository>Leiden University Libraries</repository>
            <collection>Special Collections</collection>
            <idno>Or. 3632 (2)</idno>
        </msIdentifier>
        <msContents>
                <summary>This is a multi-text manuscript. The constituent parts show distinct hands and three
separate foliation ranges in Balinese script. Continuous pagination has been applied by pencil in a Roman hand.
Both Balinese and Roman numberings show irregularities.</summary>
        </msContents>
        <physDesc>
            <p>Palmleaf manuscript measuring 5 × 44 x 4  cm including the cover boards</p>
        </physDesc>
<msPart>
```

```
        <msIdentifier>
        <settlement>Leiden, the Netherlands</settlement>
                [...]
                <idno>Or. 3632 (2)</idno>
        </msIdentifier>
        <msContents>
                <msItem>
                        <locus>fol. 1r-29v, pp. 1-57</locus>
                        <title>Vratiśāsana</title>
                </msItem>
        </msContents>
        <physDesc>
                <p>...</p>
                <handDesc>
                        <p/>
                        <!--<handNote xml:id="A_H2" scriptRef="11111"/>-->
                </handDesc>
        </physDesc>
</msPart>
        </msDesc>
```

Conversely, if your text is transmitted in a fragmented manuscript, you can declare this with <msFrag>. This element works like <msPart>, but is used to emphasize the prior context of the manuscript, typically as a description of a virtual reconstruction of a manuscript or other objects whose fragments have been cataloged separately. This approach, more concerned by the physical aspect and history of the manuscript, is not mandatory. The choice of encoding these aspects or not is left to the encoder's discretion.

Example 2.1.3.4.H

```
<witness xml:id="S3-S4">
 <abbr type="siglum">S<hi rend="subscript">3</hi>/S<hi rend="subscript">4</hi></abbr>
    <msDesc>
      <msIdentifier><!-- instead of <msIdentifier>, you can also use <altIdentifier><idno n="..."/></altIdentifier> -->
       <msName>Constituent parts of S3/S4</msName>
      </msIdentifier>
    <msFrag xml:id="S3">
     <head><abbr type="siglum">S<hi rend="subscript">3</hi></abbr></head>
     <msIdentifier>
        <settlement>Oxford</settlement>
        <repository>Bodleian Library</repository>
        <idno>MS Sanskrit.a.14 (R)</idno>
     </msIdentifier>
     <!-- msContents and other elements possible -->
    </msFrag>
    <msFrag xml:id="S4">
     <head><abbr type="siglum">S<hi rend="subscript">4</hi></abbr></head>
```

```
        <msIdentifier>
          <settlement>Kathmandu</settlement>
          <repository>National Archives</repository>
          <idno>MS4-2260</idno>
        </msIdentifier>
        <!-- msContents and other elements possible -->
      </msFrag>
    </msDesc>
  </witness>
```

Finally, the element <history> can be used to supply a brief history of the manuscript as a physical artifact. If, within <history>, you wish to tag the date and place of creation of your manuscript, you can do so by incorporating the elements <origDate> and <origPlace> inside the free-text that makes up the <p>. Remember that all the information you store in these specialized elements must, in turn, be contained in <p>.

Here is a fully-fledged example:

Example 2.1.3.4.I[23]

```
<sourceDesc>
      <listWit>
        <witness xml:id="A1">
          <abbr type="siglum">A<hi rend="superscript">1</hi></abbr>
          <msDesc>
            <msIdentifier>
              <settlement>Leiden</settlement>
              <institution>Leiden University</institution>
              <repository>Leiden University Libraries</repository>
              <collection>Special Collections</collection>
              <idno>Or. 23732</idno>
            </msIdentifier>
            <msContents>
              <msItem>
                <author>Śāntarakṣita</author>
                <title>Vipañcitārthā</title>
                <colophon xml:lang="san-Latn">samvata Ācū2 śrāvaṇa-kṛṣṇa-Ekādaśyāṁ likhitaṁ
                  mayā| rājādhirāja-parameśvara-parama-bhaṭṭārakaḥ śrīmad-ānandade
                  vapādīya-vijaya-rājye śubha-dine</colophon>
              </msItem>
              <summary><!-- summary of the witness's contents --></summary>
            </msContents>
            <physDesc>
              <objectDesc>
```

```
            <p>Palm-leaf manuscript. 89 leaves in Siddhamātṛkā script.</p>
          </objectDesc>
          <handDesc>
              <handNote xml:id= "A1_H1" scriptRef="class:38774 maturity:83213">
              <abbr type="hand">A<hi rend="superscript">1</hi>H<hi rend="subscript">1</hi></abbr>
                    <p>main hand of the manuscript, black ink</p>
              </handNote>
          </handDesc>
        </physDesc>
        <history>
              <p>The manuscript was apparently written in <origDate when="1152">1152 CE</origDate> and
found, in June 1934, by Rāhula Sāṅkṛtyāyana in the monastery of Kun-de-ling (Lhasa).</p>
        </history>
      </msDesc>
    </witness>
    <witness xml:id="A2">...</witness>
    <witness xml:id="B">...</witness>
  </listWit>
</sourceDesc>
```

You might need to record metadata in even more detail than the above; in this case, contact the project's XML-TEI Data Manager.

2.1.3.5 <handDesc> — Encoding the hand(s) with its/their script(s)

As briefly introduced in §<sourceDesc>, the script(s) used in a manuscript must be encoded in <handDesc>. Below, we will offer a more detailed description of how to do so. The same element <handDesc> will be used if multiple hands are identifiable within the manuscript. If you need to refer to individual hands in your critical apparatus, then you should encode each hand as a child of <handDesc>. If you want to record and describe more than one hand, you must add the attribute @hands in <handDesc>; its value, which must be numerical, will indicate the total amount of encoded hands. Note that we consider each different script found within a single witness to represent a distinct hand.

Example 2.1.3.4.A

```
<witness xml:id="A1">
        <abbr>A<hi rend="superscript">1</hi></abbr>
        <msDesc>
            <msIdentifier>
              <settlement>Leiden</settlement>
            </msIdentifier>
            [….]
            <physDesc>
              <handDesc hands="3">
                    [...]
                </handDesc>
```

```
            </physDesc>
</witness>
```

Within <handDesc> you will record one <handNote> per hand: each of them will bear a unique @xml:id, an abbreviation encoded in <abbr> and a free-text description of the hand-type itself in a <p>. The main hand of the manuscript comes first:

<table>
<tr><td>Example 2.1.3.4.B</td></tr>
</table>

```
<witness xml:id="A1">
        <abbr>A<hi rend="superscript">1</hi></abbr>
        <msDesc>
                <msIdentifier>
                        <settlement>Leiden</settlement>
                </msIdentifier>
                [....]
        <physDesc>
            <handDesc hands="3">
                <handNote xml:id= "A1_H1" scriptRef="class:57470 maturity:83215">
                <abbr type="hand">A<hi rend="superscript">1</hi>H<hi rend="subscript">1</hi></abbr>
                        <p>main hand of the manuscript, black ink, Sundanese script</p>
                </handNote>
                <handNote xml:id="A1_H2"  scriptRef="class:57470 maturity:83215">
                <abbr type="hand">A<hi rend="superscript">1</hi>H<hi rend="subscript">2</hi></abbr>
                        <p>generally in red ink, placed on the margins of the manuscript, Sundanese script</p>
                </handNote>
                <handNote xml:id="A1_H3"  script="Latn">
                <abbr type="hand">A<hi rend="superscript">1</hi>H<hi rend="subscript">3</hi></abbr>
            <p>generally in blue ink, Roman script; interlinear transcriptions of selected words</p>
                </handNote>
            </handDesc>
        </physDesc>
    </msDesc>
</witness>
```

The @xml:id representing the hands shall be constructed as follows:

> hand 1 (the main hand) = siglum-of-the-manuscript_H1
> hand 2 = siglum-of-the-manuscript_H2, etc.

You will thus be able to refer to these hands by adding an attribute @hand to the relevant element within your edition (§5.5.13). The value of @hand will correspond to the value of the @xml:id you have chosen to represent that specific hand with. Note that the main hand is always understood to be active unless stated otherwise. This means that, in practice, you will rarely, if at all, be needing the xml:id for hand 1.

Unless you are dealing with a Latin-script hand, as in A1_H3 in the example above, the script should be provided with the attribute @scriptRef on each <handNote>. The content shall refer to DHARMA controlled vocabularies established for script classification as well as maturity, using a prefix to identify one of the two lists, and should be followed by the code associated to the concept.[24] The list established for script classification shall be prefixed with "class:", while a script maturity will be introduced with "maturity:". Note that both values are required to identify a script and should be separated by a blank space. As shown above, for exceptional cases like a Latin-script hand, you should use @script instead of @scriptRef, and use the script suffixes defined by ISO 15924,[25] i.e., in most cases -Latn.

2.1.4 Shortcut for citing sigla and hands

If you need to cite a witness in free-text sections, as in a <p> or in a <note>, where the @wit cannot be used (§5.4), you can use a <ptr/> tag as a shortcut to name the witness. See the following examples:

<table>
<tr><td>Example 2.1.4.A</td></tr>
</table>

```
<app>
   <lem type="emn">sakesi</lem>
   <rdg wit="#A2">sakeṁsi</rdg>
   <rdg wit="#A1">sakiṁsi</rdg>
   <note>A second hand has crossed out the <foreign>anusvāra</foreign> in <ptr target="#A2"/>.</note>
</app>


<note>While <ptr target="#A1_H1"/> appears accidentally to have left the ligature <foreign>ktya</foreign>
incomplete, <ptr target="#A1_H2"/> has furnished the missing <foreign>y</foreign> element.</note>
```

The empty elements <ptr target="#A2"/> and <ptr target="#A1_H2"/> work as pointers to the @xml:id of the manuscript A^2 and the hand A^1H_2 as defined in the <teiHeader>.

2.2 <encodingDesc> — Documenting your work

The <encodingDesc> follows the <fileDesc> and can contain several sub-elements. This element serves (1) to describe the project context in which your edition is being produced, (2) to keep a record of your editorial policies, and (3) to clarify the relationship between any external files and your digital edition, especially if there may be uncertainty or ambiguity related to the representation of the sources. The DHARMA template for critical editions will contain the following basic <encodingDesc>:

<table>
<tr><td>Example 2.2.A</td></tr>
</table>

---

[24] See https://opentheso.huma-num.fr/opentheso/?idt=th347 to access DHARMA controlled vocabularies.
[25] https://unicode.org/iso15924/iso15924-codes.html

```
<encodingDesc>
 <projectDesc>
            <p>The project DHARMA has received funding from the European Research Council (ERC) under the
European Union's Horizon 2020 research and innovation programme (grant agreement no 809994).</p>
        </projectDesc>
        <schemaRef type="guide" key="EGCv01" url="supplied_in_template"/>
<!-- UPDATE the link once the release is made official -->
        <listPrefixDef>
<prefixDef ident="bib" matchPattern="([a-zA-Z0-9\-\_]+)" replacementPattern="supplied_in_template">
 <p>Public URIs with the prefix bib to point to a Zotero Group Library named ERC-DHARMA whose data are open
to the public.</p>
</prefixDef>
<prefixDef ident="part" matchPattern="([a-z]+)" replacementPattern="supplied_in_template">
<p>Internal    URIs    using    the    part    prefix    to    point    to    person    elements    in    the
<ref>DHARMA_IdListMembers_V01.xml</ref> file.</p>
<prefixDef ident="txt" matchPattern="([a-zA-Z0-9]+)" replacementPattern="supplied_in_template">
<p>Internal URIs using the txt prefix to point to text-titles  in the <ref>DHARMA_IdListTexts_V01.xml</ref> file.</p>
</prefixDef>
 </listPrefixDef>
  </encodingDesc>
```

The section of <listPrefixDef> is mandatory, since it provides the correct paths to each external file containing the bibliographic references (as managed in the Zotero Group Library), the participants, and the textual sources for parallels, as will be further explained in §7.1 and sub-sections. As you see, each prefix definition contains (i) an attribute @ident, which provides the term that will function as the prefix for an abbreviated pointing scheme; (ii) an attribute @matchPattern that, supplies "a regular expression against which the values of other attributes can be matched"; (ii) @replacementPattern, which provides "the skeleton of a relative or absolute URI containing references to groups in the @matchPattern which, once subpattern substitution has been performed, complete the URI". The <p> is added for clarity, and simply explains in free-text what the encoding pattern does.

2.2.1 <projectDesc> — Describing the project

The <encodingDesc> should minimally have the sub-element <projectDesc>. This is used to explain, in one or two <p> elements of free-text, the context of the creation of the file. The first paragraph must be about DHARMA. A second paragraph can be used in any edition made in the framework of a personal research project such as a PhD.

Example 2.2.1.A

```
<encodingDesc>
        <projectDesc>
                <p>The project DHARMA has received funding from the European Research Council (ERC)
under the European Union's Horizon 2020 research and innovation programme (grant agreement no 809994).</p>
                <p>It has been carried out as part of my research towards a PhD thesis defended on the DDth of
```

```
MM, YYYY²⁶ at the École Pratique des Hautes Études, Paris.</p>
        </projectDesc>
</encodingDesc>
```

2.2.1.1 <editorialDecl> — Recording specificities of your own work

The element <editorialDecl>, a child-element of <encodingDesc>, can be used for formulating methodological choices underlying your edition. It is especially useful if you need to give more detail than, or need to make exceptions from, what is stated in the present guide, its Appendices and the DHARMA project's [Transliteration Guide](#).²⁷ We envisage the possibility that you, as editor, use one or more of the following elements to take notes during the editing process, and postpone till the final phase of your editing work which information, if any, are to be maintained in the digital publication.

The element <editoralDecl> may contain free-text descriptions wrapped in any number of <p>. Alternatively, the TEI offers a number of sub-elements that can be used to further structure the contents of <editorialDecl>, of which you may want to use: (1) <correction>, (2) <normalization>, (3) <punctuation> or (4) <interpretation>. The free-text contents of these sub-elements must likewise be wrapped in <p>.

1. The sub-element <correction> can be used to give specifications regarding the correction principles within your edition. It allows you to explain the logic applied for omission, truncation, correction, alternate readings and so forth, and to state whether such cases are handled silently or with mark-up.

2. The sub-element <normalization> can be used to explain the extent of normalization and regularization applied on the text as critically edited. You can use this element to document the specificities of your text and the way you have dealt with them as editor, regarding such topics as non-standard (dialectal) forms or non-standard spellings.

3. The sub-element <punctuation> can be used to explain the degree of correspondence between punctuation of your constituted text and the punctuation found in the witnesses — whether punctuation is entirely or largely editorial, whether you generally attempt to follow punctuation of the witnesses or whether issues of punctuation can trigger creation of an apparatus entry.

4. The sub-element <interpretation> can be used to describe any additions you are making to the text edition that relate to issues of analysis or interpretation.²⁸ This may be the place to describe what editorial hyphenation you implement.

Example 2.2.1.1.A

---

²⁶ The TEI guidelines allow a more in-depth encoding of such information, for example through the addition of a <date> with @when="YYYY-MM-DD" in case one wants to encode the date of publication of a bibliographic reference within a <p> in <projectDesc>. However, we do not advise our editors to do so.
²⁷ The relevant document is available on [HAL-SHS](#).
²⁸ Note that for this element a by-default paragraph about asterisk will be added in the template and should be reproduced in all the editions, Others could be added at a later stage, especially to record any practices that may not have been explicitly explained in one of the guides of the DHARMA project.

```
<encodingDesc>
  <projectDesc>
      <p>This project has received funding from the European Research Council (ERC) under the European
Union's Horizon 2020 research and innovation programme (grant agreement no 809994).</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p/>
    </correction>
    <normalization>
      <p>Original spelling and typography is retained, except that long s and ligatured
        forms are not encoded.</p>
    </normalization>
    <interpretation>
        <p>The asterisk * is used to flag lemmata or forms of lemmata not recorded in the <title>Old
Javanese-English Dictionary</title>.</p>
    </interpretation>
  </editorialDecl>
</encodingDesc>
```

2.2.1.2 <samplingDecl> — Explaining the sources chosen for your edition

Inside the <encodingDesc>, the <editorialDecl> is followed by the element <samplingDecl>. It is used to record free-text information about inclusion or omission of portions of the text, manuscripts, or witnesses. It is here that you may explain, for example, the reasons behind favoring one witness above others in your choice of readings or, to give another example, the fact that you wholly or partly omit from your collation the readings of a given witness because it is too damaged to be collated for all or a portion of the text.

Example 2.2.1.2.A
| ```
<samplingDecl>
   <p>Explanation</p>
</samplingDesc>
``` |

2.2.1.3 <schemaRef> — Declaring the schema

The element <schemaRef> can point to any external customization file of the TEI. We will use it to record the URL of the latest version of this encoding guide. More than one <schemaRef> can be provided, and we will do so, eventually, to link to the relevant ODD.[29]

The element works with the attributes @type, @key and @url. The first attribute characterizes the content (e.g. guide, schema or ODD). The second provides an identifier

---

[29] An ODD ("One Document Does it All") is a TEI XML file that states the set of elements used in a project, as well as the prose and reference documentation linked to it; the ODD documentation is, then, used to generate a relevant schema. A schema, in its turn, validates the "grammar" of the XML files according to constraints written for a project.

concerning the guide itself, normally the acronym established within the project (e.g. EGC), and the number of its version expressed according to rules stated in the [File Naming Conventions](#),[30] e.g. v01 for the first version.

Example 2.2.1.3.A
<schemaRef type="guide" key="EGCv01" url="*supplied_in_template*"/>

2.2.1.4 <variantEncoding> — Declaring the method used for building your apparatus

To comply with the TEI, it is obligatory to declare the methodology chosen to encode the apparatus, and notably the variant readings, especially since DHARMA works methodologies to accommodate all the practices of the project.

Depending on the method you have chosen to encode your apparatus, you will need to update the attribute @method in the <variantEncoding/> element.[31] To sum it up straightforwardly, if you are encoding your apparatus with the inline method (which is recommended for prose texts), you will select @method="parallel-segmentation", but if you opt for an apparatus coming at the end of the sections of your edition (which is really only an option in verse texts), you must select @method="location-referenced". For more on this topic, see §[5](#) as well as §[6.3](#).

Example 2.2.1.4.A
Inline apparatus as described in §5 <variantEncoding method="parallel-segmentation" location="internal"/> **Apparatus after text as described in §6.3** <variantEncoding method="location-referenced" location="internal"/>

2.2.2 <profileDesc> — Providing descriptive elements

The element <profileDesc> serves to provide metadata related to non-bibliographic aspects of the text. In DHARMA, it will mostly be used to record language usage and classificatory keywords.

The element <langUsage> is the mandatory container for <language>. This element serves to record the languages used in the text. In the attribute @ident, we shall indicate the ISO 693-3 **language codes** as listed in §[Appendix B](#).[32] However, since the element <langUsage> cannot be used to indicate the *actual* distribution of languages within the text you are editing, your digital edition will have to declare @xml:lang="eng" as base language for your

---

[30] Available at the following link: https://bit.ly/2JytLbz.
[31] You will find both methods provided in the template, and you will need to select the method that will be adopted in your edition.
[32] See also EGD §10.3 and Appendix D.

file (§1) and then mark containers for portions of text distinguishing the respective languages by the attribute @xml:lang with the values from the same list of language codes (see §8.10.1). This kind of encoding must be adopted, for instance, in the case of a Sanskrit *mūla* with gloss or commentary in another language.

The element <textClass> shall be used to contain the element <keywords>. Each one shall be encoded inside an element <term> and selected from the project's closed list of terms that will eventually be drawn up.

Example 2.2.2.A

```
<profileDesc>
        <langUsage>
        <!-- choose the language(s) of your text, remove the others -->
                <language ident="san-Latn">Sanskrit, romanized</language>
                <language ident="kaw-Latn">Old Javanese, romanized</language>
                <language ident="ori-Latn">Oriya, romanized</language>
                <language ident="osn-Latn">Old Sundanese, romanized</language>
                <language ident="tam-Latn">Tamil, romanized</language>
                <language ident="kan-Latn">Kannada, romanized</language>
                <language ident="tel-Latn">Telugu, romanized</language>
        </langUsage>
        <textClass>
                <keywords>
                <!-- In due course we will can use this element to declare keywords as <term>, but at
the moment please ignore <keywords>  -->
                        <term/>
                        <!-- Ignore this element for the time being. -->
                </keywords>
        </textClass>
</profileDesc>
```

2.2.3 <revisionDesc> — Keeping track of file History

The history of the file, from filling in the template to the final review, is to be recorded in the last high-level element of the <teiHeader>, namely <revisionDesc>. Recording changes can be useful for the management and control of the files or to have a quick overview of the latest changes made when you return to work on a file after some time has passed. Each important change should be encoded in an element <change> — the latest of which should always be added at the very top of the list — with the mandatory attribute @who, the value of which shall be the personal identifier[33] of the person(s) making the change, normally yours, with the prefix "part:" (as an abbreviated reference to the file listing participants of the project). It is possible to indicate more than one "part:" in one @who, distinguishing them with a blank space, as in the

---

[33] All identifiers are available in DHARMA's Members list.

example below. The attribute @when is also expected, and should have as value the date of the change in ISO format, that is YYYY-MM-DD.

You can use @status to help keep track of the major steps of your work through the values: draft, candidate, approved, published and withdrawn. This attribute can be used on both <change> and <revisionDesc>, but we recommend using it on <change> to begin with. We foresee @status being used with <revisionDesc> for project steps involving all the members, rather than just to document your own work. As long as you have not finalized your edition, the expected value is "draft".

Example 2.2.3.A
<revisionDesc>   <change when="2020-05-29" who="part:axja" status="draft">changing the structure: moving the @source content inside a note @type="parallels"</change>   <change when="2020-04-23" who="part:argr part:tilu" status="draft">started encoding of the edition</change> </revisionDesc>

At subsequent steps, the value of @status can be changed to one of the following values, depending on discussion with other team members.

- "candidate"
- "approved"
- "published"
- "withdrawn"

3. Text

3.1 <text> – Structuring the text

The edition proper is contained in the <text> element, whose opening tag must bear the attributes @xml:space with the value "preserve" and an @xml:lang. The value of the latter should be the main language of your text (see §Appendix B, for the relevant language ISO codes). This implies that, if not otherwise stated, textual data contained in your <text> will be in the language thus declared. Therefore, whenever there is a linguistic divergence from the declared main language (e.g. from Old Javanese to Sanskrit, if Old Javanese is the main language in your text), this should be stated by adding an @xml:lang to the relevant element.

The <text> contains the mandatory element <body>, within which you will encode your editions, translations and so on.[34] The basic layers of encoding within the <body> are provided by <div>, which will structure the various sections of your text.

Since project members are expected to work with separate XML files to respectively encode their edition and its translation, we have decided against employing a main <div> with @type="edition", @type="translation" and so forth, as is done in the case of EpiDoc editions for inscriptions. The following, then, represents the minimal containing structure for your critical edition:

Example 3.1.A
<text xml:space="preserve" xml:lang="san-Latn">   <body>   EDITED TEXT   EDITED TEXT   </body>   </text>

In the following paragraphs, we provide the lower-level textual units that you can use in your edition.

3.2 <div> – Sections of text

As previously stated, within your edition, textual divisions shall be wrapped in <div>. Such elements can be nested into one another, in order to represent the structure of your text. The level of the structure can be specified using the attribute @type, bearing values such as "chapter" and "canto", making it easier for you to navigate inside the XML file. Bear in mind that the attribute @type is expected each time a change of level occurs. However, also bear in mind that a small number of special divisions, viz. invocations and colophons, do not bear @type on <div> but on a child <ab> (see §3.3.2).

The value "chapter" is meant as the default value, while "canto" (Skt. *sarga*) is to be understood as a chapter of a verse text, if changes in meter mark its chapter-level divisions. In the case of texts structured by a base text (Skt. *mūla*) in verse or in prose (whether transmitted as such or implied) and a dependent text that glosses/paraphrases the former we shall use a dedicated <div> with @type="dyad" to bind the base text and the dependent text together. See below for the type "dyad" (§4.1). In case of a text mainly structured in <div type="dyad">, any opening or closing <div> that consists only in prose can be given the value "liminal".

---

[34]The TEI foresees the possibility of working with <front> and <back> to encode respectively the front matter and the backmatter of published books. For this reason, the element <body> is a part of the mandatory structure of the file itself and must be retained, however redundant it might be for our natively digital editions.

When verse text shows several changes of meters in the same chapter, you can use @type="metrical" to group together in one <div> all the stanzas following the same prosodic pattern. This <div type="metrical"> can only be used as a child of another <div>. See §3.2.3.

Should you need any other types of divisions lower than chapter/canto but higher than the individual paragraph or stanza, please let us know.

To offer freedom to encoders in determining the numbering pattern best suited to the text they are editing, the attribute @n, recording sequences, has not been made mandatory for any of the structural levels of your edition. It seems difficult to give general rules, that's why we recommend that you consult the project's XML-TEI Data Manager to decide what numbering scheme works best for your text. See §3.8, to access a description of the transformation with XSLT foreseen so far.

As for units below the <div> level, the mark-up will depend upon the characteristics of your text. In all cases where one or more prose sentences are involved, the lowest textual unit must be the <p> (paragraph), as explained in §3.3.1. Any stanzas contained in your text must be wrapped in the element <lg> (line group), as explained in §3.4.1; in this case, the lower textual unit within <lg> will be <l> (line), as explained in §3.4.3. Finally, in the case of chunks of texts that do not satisfy the minimum definition of a sentence and can for that reason not be contained by a <p>, you may use the element <ab>, as further explained in §3.3.2. In summary, here is a list of the recommended values for @type with <div>:

- chapter
- canto
- dyad
- group
- interpolation
- liminal
- metrical
- section

3.2.1 <head> – Section titles

Although some printed text editions in our fields of scholarship do state the title of the work at the top of the first page of the edition, we do not believe this ever reproduces anything similar observed in the same position in the witnesses and it certainly is redundant besides the declaration of the work's title in <titleStmt>. Therefore, we can limit ourselves here to headings for textual components.

We assume that your edition will treat text-internal headings of the type *atha śabdānuśāsanam* as regular sentences, perhaps forming a <p> on their own. Consequently, the DHARMA project's critical editions will use the element <head> only to represent editorial headers formulated in English. The element should always contain the @type="editorial" and @xml:lang="eng".

Example 3.2.1.A

```
<div n="4" type="chapter">
        <head type="editorial" xml:lang="eng">Privileges: ownership of controlled types of property</head>
            <p>...<p>
</div>
```

It is not required for a <div> (or any other text component) to have a <head>. However, if a <head> element is necessary within a container, it should be encoded right after the opening part of the relevant element, as the first child of <div>.

3.2.2 Declaring prosody for a canto

You may use the attribute @met on a <div> element to represent metrical information concerning all its children elements <lg>. The list of names allowed is given in the prosodic patterns authority file, on which, see the PVG. The canto header displayed will be like this:

Example 3.2.2.A

```
<div type="canto"  n="8" rend="met" met="śārdūlavikrīḍita" xml:id="candrakirana_04.01">
 <lg n="1">
     [...]
   </lg>
</div>
```

Display –

       Canto 8

       Śārdūlavikrīḍita: – – – ˘ ˘ – ˘ – ˘ ˘ ˘ – – – ˘ – – ˘ ˘̆

If you cannot give a name to the metrical pattern observed, enter the prosodic code as value in the @met, following the XML conventions presented at the top of the prosodic patterns file. Note that even in such cases, the pattern must be recorded in the mapping file, in the part for the relevant number of syllables per *pāda*, and the heading displayed will be like this:

Example 3.2.2.B

```
<div type="canto"  n="11" rend="met" met="--+-+---+------+-+-=" xml:id="rt_12">
 <lg n="1">
     [...]
   </lg>
</div>
```

Display –

       Canto 11

> Name unknown (*kṛti* class): ˘ ˘ — ˘ — ˘ ˘ ˘ — ˘ ˘ ˘ ˘ ˘ ˘ — ˘ — ˘ ˘

To make any of these types of display effective, you need to use the attribute @rend="met" on the containing <div>.[35]

Since the mapping list records the prosodic pattern corresponding to any name of a meter, the prosodic pattern itself does not need to appear in the encoding of a meter of known name. However, if you want to record the pattern for your own convenience, you can use a <note> element as a first child of the <div> with an attribute @type="prosody", or as a second child of the element <div> if a <head> element is also used.

Example 3.2.2.C

```
<div type="canto" n="1" met="jagaddhita" rend="met" xml:id="kakavin_11">
   <!-- If you want to add a head, do so here -->
   <note type="prosody">+++--+-+---+------+-+-+</note>
   <lg n="1">
     [...]
   </lg>
</div>
```

3.2.3 Identifying prosodic groups

When several groups of stanzas of the same meter are children of the same text division <div>, you may wrap those groups into a lower-level division also made with <div>, adding a @type="metrical". The encoding of @met works exactly as for <div type="canto"> (see §3.2.2). We foresee that this scenario is most likely to occur when the parent division is <div type="chapter">, but there is no constraint on the type of <div> that may contain <div type="metrical">.

Example 3.2.3.A

```
<div n="3" type="chapter" xml:id="candrakirana_04">
   <head>Long syllabes</head>
   <div type="metrical" met="śārdūlavikrīḍita">
     <lg xml:id="candrakirana_04.01" n="1"/>
     <lg xml:id="candrakirana_04.02" n="2"/>
     <lg xml:id="candrakirana_04.03" n="3"/>
     <lg xml:id="candrakirana_04.04" n="4"/>
```

---

[35] Although we will presumably always wish the metrical patterns for any <div> to be displayed, this requirement is made to ensure consistency with the case of @met on <lg>, where we need this attribute @rend. For the use of @met on <lg>, see §3.4.2.

```
    </div>
    <div type="metrical" met="drutavilambita">
        <lg xml:id="candrakirana_04.05" n="5"/>
        <lg xml:id="candrakirana_04.06" n="6"/>
    </div>
    <div type="metrical" met="pādānuṣṭubh">
        <lg xml:id="candrakirana_04.07" n="7"/>
        <lg xml:id="candrakirana_04.08" n="8"/>
        <lg xml:id="candrakirana_04.09" n="9"/>
        <lg xml:id="candrakirana_04.10" n="10"/>
    </div>
</div>
```

3.3 Prose Containers

3.3.1 <p> – Paragraphs

The basic container element for any prose text is the paragraph <p>. Coherent groups of sentences together form the typical paragraph, but you can use this element even if it contains no more than one complete sentence. End the current <p> and begin a new one when a new paragraph begins, in the same manner as you would do for modern text. As many elements <p> as needed can be encoded inside a <div>. A <p> can contain other elements, but it cannot contain other <p>: nesting is prohibited in this case. The element <p> will be provided with an XSLT-generated @xml:id; see §7.2.1.1. If you want to have full control of the way numbering appears in display (e.g., to circumvent numbering of paragraphs resuming at 1 at every new <div>) then manually encode the value to be displayed using the attribute @n on <p> (and preferably use full-stop as separator between counters, no need to use leading zeros); see §3.8.

Example 3.3.1.A

```
<div type="dyad" n="11" xml:id="svayambhu_02.11">
        <quote xml:id="svayambhu_02.11.01" type="base-text" xml:lang="san-Latn"><supplied reason="implied">
 <lg xml:id="svayambhu_02.11.01.01" met="anuṣṭubh">
        <l n="a">yasmin deśe niṣīdanti</l> <l n="b">viprā vedavidas trayaḥ |</l>
        <l n="c">rājñaś ca prakṛto vidvān</l> <l n="d">brāhmaṇas tāṁ sabhāṁ viduḥ ||</l>
 </lg>
 </supplied></quote>
 <p xml:id="svayambhu_02.11.02" n="11.1">saṅ brāhmaṇa tigaṅ siki, paḍa vruh maṅaji, hana brāhmaṇa sasiki
prajñān vidagdha, vinidhi muṅgva riṅ deśa saṅ prabhu, ikaṅ deśa papupulanira, yeka deśa sabhā ṅaranya de saṅ
paṇḍita.</p>
 </div>
```

3.3.2 <ab> – Anonymous blocks

As per TEI guidelines, the element <ab> (anonymous block) can be used to contain arbitrary component-level units of a text, acting as an anonymous container for phrase or inter-level elements analogous to a paragraph, but without the semantic baggage of labeling a unit as <p>. As with <p>, so too each <ab> will be furnished with an XSLT-generated @xml:id; see §7.3.1. If you want to have full control of the way numbering appears in display, then manually encode the value to be displayed using the attribute @n on <ab> (and preferably use full-stop as separator between counters, no need to use leading zeros); see §3.8.

In the DHARMA project's critical editions, we foresee the use of the <ab> element to encode invocations and colophons, and also for encoding speaker indications in texts cast as a dialogue (§3.5.3), as shown in the next two subsections. Finally, we foresee its use in the probably rare instances of text following a connective particle or phrase (e.g. *iti*, *api ca*) used to introduce or end stanzas and not functioning as an integral part of the unit adjacent to the stanza.[36]

3.3.2.1 Invocations and colophons

To represent an auspicious word, such as siddham, or an auspicious phrase, such as namo gaṇeśāya, if such a textual element is weakly connected to the adjacent text, we recommend that you use an appropriate prose container <ab> inside a <div> without any @type while assigning @type="invocation". In this case, assign no number to the <div>.

Example 3.3.2.1.A
 <ab type="invocation" xml:lang="san-Latn">namo gaṇeśāya|</ab>

However, in the case of invocations to specific divisions of the text, the invocation must be included at the top of the relevant <div> but after any furnished <head>. Depending on the structure of your text edition — namely whether the highest-level <div>s (e.g. <div type="chapter") themselves also contains child <div>s (e.g. <div type="dyad") — it may then be necessary to wrap the <ab> in its own <div> in order for the invocation to be brought into structural parallelism with the other child <div>s of the parent <div>.

Another type of text segment that is extraneous to the overall textual structure, but not classifiable as <head>, is the colophon, a term by which we mean here a short statement declaring the end of a text or of a text component. In these cases too, you have to use <ab>, and include these as final child within the <div> to which they belong, or within a separate <div> without @type or @n just before the end of the <body> if they pertain to the text as a whole. You must mark the <ab> element with @type="colophon". In this case too the need may arise to wrap the <ab> in a <div>. See example 3.3.2.1.C.

---

[36] See EGD §2.2.2 for how to encode this last scenario.

<table>
<tr><td align="center">Example 3.3.2.1.B</td></tr>
<tr><td>

```xml
<div>
        <ab type="colophon" xml:lang="san-Latn">iti strīsatyarakṣaṇa</ab>
</div>
```

</td></tr>
</table>

<table>
<tr><td align="center">Example 3.3.2.1.C</td></tr>
<tr><td>

```xml
<div type="chapter" n="2" xml:id="svayambhu_03">
    <head type="editorial">Protection of Faithful Women</head>
    <div type="dyad" n="32" xml:id="svayambhu_04.01">
        [...]
    </div>
        [...]
    <div xml:id="svayambhu_03.04">
    <ab xml:id="svayambhu_03.04.01" type="colophon" xml:lang="san-Latn">iti strīsatyasaṅrakṣaṇa.</ab>
</div>
</div><!-- end of chapter 2-->
```

</td></tr>
</table>

Note that the declaration of @xml:lang is not necessary if the language of the invocation and that of the text proper are identical.

3.3.2.2 Speaker indications

When dealing with a text that is not dramatic in nature (see §3.5.1) but is nevertheless structured in the form of a dialogue and the transmission itself dedicates structural elements to indicating the change of speaker, such elements — whether they can be analyzed as a full sentence or not — are to be wrapped in <ab> with @type="speaker".

<table>
<tr><td align="center">Example 3.3.2.2.A</td></tr>
<tr><td>

```xml
<ab type="speaker">devy uvāca</ab>
```

</td></tr>
</table>

It is possible to mark-up the speaker's name with <persName> and @key as per §8.6.2.1.

3.4 Verse Containers

Just as the main components of any prose in your edition need to be encoded, so do any verse components. The following example shows full encoding of a single Sanskrit stanza.

<table>
<tr><td align="center">Example 3.4.A</td></tr>
</table>

```
<lg met="anuṣṭubh" n="8.123">
        <l n="a">vyavahārān didṛkṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
 </lg>
```

In the following sections, you will find brief explanations on how to deal with different scenarios involving the encoding of text composed in verse form.[37]

3.4.1 <lg> — stanzas

One <lg>, or line group, corresponds to a single stanza as a whole. It can be nested either within a <div> or within a <p>. The <lg> can contain one or more <l> elements and must have the attribute @met to identify the meter of the stanza by its conventional name, if @met has not been declared on the parent <div>. Unlike what is prescribed for epigraphic editions in the EGD, in critical editions the numbering of the <lg> with the attribute @n is allowed but not mandatory. Notably, you should not apply @n to any stanzas that are only implied in the text you are editing but that are supplied in your edition. Another case where you will probably want to omit @n is when a whole text or a whole <div> consists in a single <lg>, which is the case for metrical composition structured essentially in lines and not in stanzas (e.g, when @met="free"), for in such cases arabic line numbering will appear in display. If you encode @n on any <lg> that are part of your text, then you must do so following a consistent pattern, the value of @n on <lg> always being displayed as such (see §3.8). An XSLT-generated attribute @xml:id will be provided also for <lg> (see §7.2.1.1).

The structural framework of a stanza must always be encoded, even when it is damaged or lost. Do not omit missing lines but identify them as lacunae.[38]

Example 3.4.1.A

```
<lg xml:id="candrakirana_14.01.03" n="3">
<l n="a">kita  makarūpa chanda,</l>
<l n="b">agave titi mvaṅ śloka,</l>
<l n="c">ya ta varṇanən kavətun,</l>
<l n="d">ya pinakasuluhiṅ rāt.</l></lg>

<lg xml:id="candrakirana_14.01.04" n="4">
<l n="1">lvirniṅ chanda ta ṅaranya,</l>
<l n="2">kvehiṅ vilaṅniṅ akṣara,</l>
<l n="3">rikanaṅ pāda sayati,</l>
<l n="4">pāda ṅaranya sakaṇḍa,</l>
<l n="5">pa<pb n="13v" edRef="#msba"/>taṅ pāda ya saśloka.</l></lg>
```

---

[37] Please note that the present guide follows the vocabulary established in EGD §2.3.1. All the elements discussed here have been introduced in greater detail in the EGD.

[38] As per the guidance in EGD §2.3.8.

> Display –
>
> kita makarūpa chanda,
> agave titi mvaṅ śloka,
> ya ta varṇanən kavətun,
> ya pinakasuluhiṅ rāt. 3
>
> [line 1] lvirniṅ chanda ta ṅaranya,
> [line 2] kvehiṅ vilaṅniṅ akṣara,
> [line 3] rikanaṅ pāda sayati,
> [line 4] pāda ṅaranya sakaṇḍa,
> [line 5] pa[B1:13v]taṅ pāda ya saśloka. 4

3.4.2 @met — metrical patterns

The metrical pattern, or "template", of a stanza will be encoded using the conventional name of the meter as values of the attribute @met in the element <lg>. The list of names allowed is given in the prosodic patterns authority file (see PVG). If a given <div> is entirely (or almost entirely) composed of stanzas of a given type, you may state the @met (or the dominant @met) on the <div>. The metrical pattern for each child <lg> is then inherited from the parent <div>, and doesn't need to be repeated — while any <lg> that has a divergent meter should receive its specific @met, and the divergent value can be displayed. However, as in the case of @met encoded on <div> (§3.2.2), so too the value of @met recorded in an <lg> element will only be displayed if you add a @rend="met".[39]

 If you cannot assign a name to the meter in question, you can still use @met to record the prosodic pattern using XML notation. Follow the conventions presented in the prosodic patterns authority file (see PVG). Prosodic XML notation is also used for the attribute @real on the <l> element, to which we now turn.

3.4.3 <l> — metrical lines

The line element <l> contains an individual metrical line (*pāda*). The attribute @n must be used to number the lines. For the numbering of Sanskrit, Prakrit or Old Javanese quantitative verse, you should use lower case Roman letters (a, b, c, d). In syllabo-quantitative verse, assign a letter to each <line>, but in quantitative verse, treat the hemistich as equivalent to a line, i.e. stanzas of the *āryā* family shall be marked up as consisting of two lines (ab, cd). In other traditions of versification (e.g. Old Sundanese or Tamil), especially when verse lines cannot meaningfully be grouped into stanzas, use a single Arabic number per line and wrap all lines in the relevant given text division together into a single <lg>. Metrically irregular lines are to receive the attribute @real.[40]

---

[39] This approach was chosen to be able to deal with texts mixing prose and verse where there is a strong preponderance of *anuṣṭubh* stanzas, for which the metrical pattern does not need to be displayed, besides a smaller number of stanzas using other patterns that do need to be displayed.
[40] EGD §2.3.5.

3.4.4 @real – deviant metrical patterns

The meter of a stanza or of an individual verse line shall not be encoded so long as it conforms to the pattern encoded for the group of stanzas or for the single stanza to which they belong. But when a line, or a group of lines, deviates from the expected pattern, and you believe the metrical irregularity was intended by the author(s) of the work you are editing, it is possible to record the phenomenon using the attribute @real on the stanza or the verse line concerned. The value of the attribute should be the actually observed prosodic pattern following the notation provided in the prosodic patterns authority file (see PVG).

Example 3.4.4.A

<lg met="pādānuṣṭubh">
   <l n="a">...</l>
   <l n="b">...</l>
   <l real="========" n="c">liṅ saṅ paṇḍita vruheṅ mārga</l>
</lg>

Display –
^273. liṅ […] marga:**]** Unmetrical line. The observed pattern is not *pādānuṣṭubh* but ˘ ˘ ˘ ˘ ˘ ˘ ˘ ˘.

If the above auto-generated observation, based on the values of @met and @real, does not yield an adequate result in the situation you are dealing with you, you can formulate a tailor-made observation on the metrical situation by encoding a <note> at the end of the <l>, which will annul the auto-generated observation.

Example 3.4.4.B

<lg met="pādānuṣṭubh">
   <l n="a">...</l>
   <l n="b">...</l>
   <l real="+++-+-++-" n="c">liṅ saṅ paṇḍita vruheṅ mārga<note>The line has nine syllables and is hence not compliant with the <foreign>pādānuṣṭubh</foreign> pattern.</note> </l>
</lg>

Display –
^273. liṅ […] marga:**]** The line has nine syllables and is hence not compliant with the *pādānuṣṭubh* pattern.

3.4.5 @enjamb – enjambment

In the context of the DHARMA project, enjambment is defined as the occurrence of a metrical line break within a word (usually between members of a compound; rarely within a morpheme). To mark it up, use the attribute @enjamb with the value "yes" when a break occurs at the end of

the line that does not coincide with the end of the word. This attribute should be added on the <l> element containing the initial part of the broken word, not to the one containing the final part like you would do with the attribute @break.

When the break falls between words that are members of a compound and without vowel fusion *sandhi*, put any editorial hyphens you might be using for compound analysis at the beginning of the second <l> involved. When a vowel fusion *sandhi* occurs at the transition from one <l> to the next, put the break after the fused vowel without adding a hyphen.

However, when a vowel fusion sandhi happens at the end of a stanza with text outside of it (typically the quotative *iti*), do not use @enjamb. Let us take as example the last quarter of a common stanza, which might appear as ... *pitṛbhiḥ saha majjatīti*|| in your witnesses, the *iti* being extraneous to the stanza yet written before the punctuation sign. In order to facilitate verse-indexation, the sandhi should be resolved and the metrically extraneous kept outside the stanza — meaning that it needs to be edited in an immediately following container.

Example 3.4.5.A

<lg>
    [...]
    <l n="d">pitṛbhiḥ saha majjati||</l>
</lg>
<ab>iti</ab>
	<lg>
	    [...]
	    <l n="d">pitṛbhiḥ saha majjati</l>
	</lg>
	<ab>iti ||</ab>

This must be done irrespective of whether any punctuation sign is included in your edition at the end of the stanza, and as shown in the two variants of the above example, the placement of the punctuation inside <l> or inside <ab> depends on your preference. If such phenomena are frequent in your text, we recommend that you state the fact that you are intervening as editor in the <teiHeader>, and apply one of the encoding solutions shown above without further comment, see §2.2.1.1. If such phenomena are exceptional, and if it is deemed necessary, you can provide a <note> element to explain any individual editorial intervention in human-readable terms:

Example 3.4.4.B
<lg>   [...]   <l n="d">pitṛbhiḥ saha majjati||</l> </lg>   <ab>iti<note>All witnesses read <foreign>majjatīti ||</foreign>.</note></ab>

3.5 Spoken Text

3.5.1 Dramatic elements

Plays are usually organized in structural and formal divisions in which various characters speak, taking turns. Their structure may be analyzed as hierarchic in more than one way: a scene may be divided into speeches and, at the same time, into verse lines. When you wish to emphasize the speech-based performance, you shall use the dramatic elements. However, if your text can't be qualified as dramatic, use the sections §3.5.2 and §3.5.3 to encode direct speech and dialogues.

Each individual speech of a play should be encoded with the element <sp>. It requires the element <speaker> as first child, which furnishes a label introducing the name of the speaker, while the spoken text itself should be tagged with the elements <p>, <ab> or <lg> and <l>, depending on whether it is in prose or in verse.

Any content that can be classified as stage direction should be encoded within the element <stage>. When the stage direction is related to a specific speech, <stage> should be contained in the element <sp> as a child or a descendent depending on the context.

If you ever need to encode drama, follow the few main principles formulated in this section and contact the project's XML-TEI Data Manager.

Example 3.5.1.A

```
<sp>
        <speaker>rāmaḥ</speaker>
        <stage>tataḥ praviṣaty upaviṣṭo rāma sītā ca</stage>
        <p>devi vaidehi samāśvasihi samāśvasihi te hi guravo na śaknuvanty asmān vimoktum</p>
        <lg met="anuṣṭubh">
          <l>kin tv anuṣṭhāna | nityatvaṃ</l>
          <l>svātantryam apakarṣati</l>
          <l>saṅkaṭā hy āhitāgnīnāṃ</l>
          <l>pratyavāyair gṛhasthatā</l>
        </lg>
</sp>
```

3.5.2 Direct speech

The above elements are specific to drama. If you are not sure your text could, or should, be classified as dramatic, you can use more neutral elements. Quotations of direct speech can be encoded with the element <said> which can be used for real people as well as fictional ones. The speaker can then be identified with the attribute @who if you find her/his identification ambiguous. Just write the name as you expect it to be displayed in the attribute @who.

3.5.3 Dialogue

If you are encoding a dialogue, which isn't part of a dramatic text, you might want to identify the changes of one interlocutor to another. We foresee two situations, one where the text itself dedicates a structural element to indicating the change of speaker, and another where the text does not but you wish to aid the reader by marking changes of speaker.

For the first scenario, see §3.3.2.2. In the second scenario, mark the place where an interlocutor begins to speak by inserting a <label> wrapped around the name of the interlocutor in normalized form. It will be treated as an editorial element and not as an actual part of the text.

Example 3.5.3.A
`<p><label type="speaker">Rāma</label>devi vaidehi samāśvasihi samāśvasihi te hi guravo na śaknuvanty asmān vimoktum</p>`

Note that the element <label> should always be a child of <div>, <ab>, <p> or <lg>. As in §3.3.2.2, it is possible to mark-up the speaker's name with <persName> and @key as per §8.6.2.1.

3.6 Mixed prose and verse

As we have just shown from a dramatic context in example §3.5.1.A, if your text mixes prose and verse, but you are not dealing with any scenario of the type discussed in §4, then every block of prose will be encoded in a <p> while every stanza will be encoded in <lg>.

Example 3.6.A

`<lg met="anuṣṭubh" xml:lang="san-Latn">`
`    <l n="a">vyavahārān didṛkṣus tu</l>`
`    <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>`
`    <l n="c">mantrajñair mantribhiś caiva</l>`
`    <l n="d">vinītaḥ praviśet sabhām ||</l>`
`</lg>`
`<lg met="puṣpitāgrā" xml:lang="san-Latn">`
`    <l n="a">daramukulitanetrapāli nīvī</l>`
`    <l n="b">nityamitabāhu kṛtoruyugmabandham</l>`
`    <l n="c">karakalitakucasthalaṃ navoḍhā</l>`
`    <l n="d">svapiti samīpam upetya kasya yūnaḥ ||</l>`
`</lg>`
`<p>samānalajjāmadanā madhyā
	eṣāivātipraśrayād ativiśrabdhanavoḍha	asyāś ceṣṭā sāgasi preyasi dhairye vakr' oktir adhairya paruṣavāk	yathā</p>`

3.7 Quotation made by author of the text being edited

If any of the stanzas is quoted from an external source, then wrap the stanza or group of stanzas in a <quote> element.

Example 3.7.A
| ```xml
<quote xml:lang="san-Latn">
      <lg met="anuṣṭubh">
        <l n="a">vyavahārān didṛkṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
      </lg>
</quote>
``` |

Note that this use of <quote> is not limited to whole stanzas; in fact, it is not limited to verse-context at all. Any word, string of words or block of words (<p>, <l>, <lg>) can be wrapped in <quote> to render a quotation from another work made by the author of your text. See §6.2 on how to encode any known source or parallels.

If you need the quoted segment to be displayed as an indented block quote, then apply the attribute @rend with the value "block" to the <quote> element. This will in principle only be necessary for long quotations of prose text within a prose <p>; in any scenario that involves quotations of stanzas or verse lines, the presence of <lg> will lead to verse-formatting in display. Quotations of shorter strings of prose without @rend="block" will be displayed within quotation marks.

3.8 Numbering the text units of the edition

The numbering of each block-level container must be handled by the encoder. It means that the @n attribute isn't mandatory in any element <div>, <p>, <ab> nor <lg>, but that if you want a specific value to appear in display, you can bring this about in complete freedom. That being said, we recommend that you number the relevant structural elements of your edition and if you do that you try to follow the following recommendations as consistently as possible.

Do not number any invocations or colophons.

If you want to fill in the numbering pattern as counters, then limit yourself to chains consisting of Arabic numbers for each counter, and with full-stop as separator in case more than one counter is relevant, e.g. @n="1.8.12". If you want to number only the current level without including counters for the other levels, then use plain Arabic numbers, e.g. @n="1".

Whatever system you choose to implement, we advise you to always check with the XML-TEI Manager of the project to get it approved. Bear in mind that the greater the diversity of

systems found in the files, the more it may become difficult to obtain a satisfactory display in all editions. Among the major downsides of giving the editor this liberty is not only the difference that will inevitably arise between various editions, but also the risk of your own unintentional inconsistency in using of @n.

For file-internal reference, an automated @xml:id will be provided to identify each level of the edition in order to overcome any problems of inconsistency or evolution of the encoding of @n. See §7.2.1.1.

4. Base text and dependent text

It may happen that a single text consists, in whole or in part, of a base text and a dependent text (commentary, gloss, etc.). You might want to encode only one of the two but, if you expect to encode both, you will need to make the relationship between the constituent texts explicit. Two solutions can be envisaged for doing so:

1. Encoding the texts together, as a single whole
2. Encoding the texts in separate files, and linking them to each other

How to choose which option is the most suitable? When base and dependent texts are at the same hierarchical level, or embedded into each other, it makes sense to encode them together. But when the dependent text is given in another hierarchical level, it would be easier to encode them separately. Choosing one solution does not, *a priori*, exclude the other: solution 2 could, for instance, implement solution 1. Let us say that you have already encoded the base text of your witness, embedded within its dependent text (solution 1), and you want to add a new layer of dependent text: encoding it in the same <div> might be rather complicated, but you could opt for encoding the additional dependent text in a separate file that refers back to the first one (solution 2). For the time being, this guide mostly provides details on solution 1.

4.1 Base text embedded in dependent text

When the base text is embedded in the dependent text, everything in your markup is assumed to form part of the latter, unless otherwise specified. Therefore, in order to mark a segment of text as part of the base text, you must use the element <quote> together with the attribute @type="base-text". You must also add an @xml:lang attribute when the base text and the commentary are not written in the same language (see §Appendix B for the relevant ISO codes).

Example 4.1.A

```
<div>
   <quote type="base-text" xml:lang="san-Latn">
      <lg met="anuṣṭubh">
```

```
        <l n="a">vyavahārān didr̥kṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
      </lg>
    </quote>
    <p>dependent text</p>
</div>
```

In many cases, the structure of the dependent text will map precisely onto that of the base text; this means that a chunk of base text can be encoded with the corresponding section of dependent text within a single structural container, namely a <div>.

Higher level divisions of text wrapped in <div> have been discussed in §3.2 with the appropriate @type attribute to be used for such sections. The same element <div> is to be used, but with different @type attributes, for pairing corresponding blocks of base text and dependent text. Although other values can be proposed if necessary, at this stage the only @type of <div> for such a lower hierarchical level occurring among the DHARMA project's digitally edited texts is that with the value "dyad". A dyad refers to a textual unit presenting a stanza together with dependent prose, the pair of which is itself part of a higher-level <div>. When necessary, you can number your lower-level <div> with the attribute @n. Note that both attributes @type and @n are optional at this level of XML structure.

The content of the <div>, in its substructure, should include one or more block-level containers; on the other hand, no text is allowed to stand outside of such containers within a <div>.

Example 4.1.B

```
<div>
  <div type="dyad" n="1">
   <quote type="base-text" xml:lang="san-Latn">
      <lg met="anuṣṭubh">
        <l n="a">vyavahārān didr̥kṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
      </lg>
   </quote>
    <p>kunaṅ hulaha saṅ prabhu, uniṅānira taṅ vyavahāranikaṅ rāt, arovaṅa ta sira brāhmaṇa vihikan maṅaji, lavan
mantri riṅ vruh mavivekā, sulakṣaṇaa ta sira tumamaa riṅ sabhā.</p>
  </div>
</div>
```

4.2 Several sections of base text for one section of dependent text

When dependent text is related to several sections of the base text at once, an enclosing division is required to wrap them together. This means that the element <quote type="base-text"> will contain all of the base text sections that are comprised in the division. Each should be encoded as an independent block according to its nature, by using an element <p> if it is a paragraph, an <lg> and <l> elements if it is a stanza, and so forth. All the chunks of base text, once encoded, will be siblings in terms of XML-TEI hierarchy.

Remember that any segment(s) of base text that your dependent text contains are considered, from the point of view of your edition, to be quoted from the former in the latter. Numbering furnished in any edition of the base text can be recorded in <listApp type="parallels">, see §6.2.

Example 4.2.A

```
<div type="dyad" n="5">
   <quote type="base-text" xml:lang="san-Latn">
       <lg>
          <l n="a">teṣām ādyam ṛṇādānaṁ nikṣepo 'svāmivikrayaḥ |</l>
          <l n="b">saṁbhūya ca samutthānaṁ dattasyānapakarma ca ||</l>
       </lg>
       <lg>
          <l n="a">vetanasyaiva cādānaṁ saṁvidaś ca vyatikramaḥ |</l>
          <l n="b">krayavikrayānuśayo vivādaḥ svāmipālayoḥ ||</l>
       </lg>
       <lg>
          <l n="a">sīmāvivādadharmaś ca pāruṣye daṇḍavācike |</l>
          <l n="b">steyaṁ ca sāhasaṁ caiva strīsaṁgrahaṇam eva ca ||</l>
       </lg>
    </quote>
  <p>dependent text that has reference to all three above stanzas at once</p>
</div>
```

4.3 Base text absent in dependent text

You may confront in your digital edition the scenario that the base text is not transmitted as such in the manuscripts of the dependent text, and was presumably never included at any stage of the dependent text's development. There are basic subtypes of this scenario, namely (1) that where a base text is absent as such but nevertheless clearly implied by the dependent text, and (2) that where the author of the dependent text shows no awareness of a block of base text that is present in one or more editions of the latter.

In the first subtype, should you feel that making the implicit text explicitly visible is useful for the readers' appreciation of the text that you are editing, then you should wrap the content of <quote> in a <supplied> tag, applying to the latter the attribute @reason with the value "implied".

Example 4.3.A
| ```xml
<quote type="base-text" xml:lang="san-Latn">
    <supplied reason="implied">
      <lg met="anuṣṭubh">
        <l n="a">vyavahārān didṛkṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
      </lg>
    </supplied>
</quote>
``` |

In the second subtype, when the dependent text lacks any trace of its author's awareness of a segment of base text, but you still wish to include this segment of base text in your digital edition, you must wrap the content of <quote> in a <supplied> tag, applying @reason with the value "omitted".

Example 4.3.B
| ```xml
<quote type="base-text" xml:lang="san-Latn">
    <supplied reason="omitted">
      <lg met="anuṣṭubh">
        <l n="a">vyavahārān didṛkṣus tu</l>
        <l n="b">brāhmaṇaiḥ saha pārthivaḥ</l>
        <l n="c">mantrajñair mantribhiś caiva</l>
        <l n="d">vinītaḥ praviśet sabhām ||</l>
      </lg>
    </supplied>
</quote>
``` |

We will now show how the coupling of <quote> with <supplied> is to be encoded in relationship to the dependent text, while also introducing how to encode a textual reference to the base text (a topic introduced more fully in §6.2).

4.4 Encoding base text ignored in the dependent text

If any base text is ignored in the dependent text yet you wish to include it in your edition, you face the problem that there is no natural block of dependent text to match it with inside a <div>. The inclusion of a <quote> containing <supplied reason="omitted"> into the edition can be

handled in two different ways. The first is to include more than one segment of base text in a single <div>. This is illustrated in the next example, which also introduces the use of <listApp type="parallels">.

```
                                    Example 4.4.A

<div type="dyad" n="199" xml:id="svayambhu_09.15">
   <quote xml:id="svayambhu_09.15.01" type="base-text" xml:lang="san-Latn">
      <supplied reason="omitted">
        <lg xml:id="svayambhu_09.15.01.01" met="anuṣṭubh">
          <l n="a">nānyad anyena saṁsṛṣṭaṁ</l> <l n="b">rūpaṁ vikrayam arhati |</l>
          <l n="c">na sāvadyaṁ na ca nyūnaṁ</l> <l n="d">na dūre na tirohitam ||</l>
        </lg>
        <listApp type="parallels">
          <app>
            <note sameAs="txt:MDhŚ08.203">Our text gives no paraphrase for the first stanza grouped in this
section.</note>
          </app>
        </listApp>
      </supplied>
   </quote>
   <quote xml:id="svayambhu_09.15.02" type="base-text" xml:lang="san-Latn">
     <supplied reason="implied">
       <lg xml:id="svayambhu_09.15.02.01" met="anuṣṭubh">
         <l n="a">anyāṁ ced darśayitvānyā</l> <l n="b">voḍhuḥ kanyā pradīyate |</l>
         <l n="c">ubhe ta ekaśulkena</l> <l n="d">vahed ity abravīn manuḥ ||</l>
       </lg>
        <listApp type="parallels">
          <app>
            <note sameAs="txt:MDhŚ08.204"/>
          </app>
        </listApp>
     </supplied>
   </quote>
     <p xml:id="svayambhu_09.15.03">ikaṅ vvaṅ amintonakən kanyā, donyan səḍaṅ tinonton, ..., maṅkana liṅ
bhaṭāra manu ṅūnin pavarah i bhagavān bhṛgu.</p>
   <quote xml:id="svayambhu_09.15.04" type="base-text" xml:lang="san-Latn">
      <supplied reason="implied">
        <lg xml:id="svayambhu_09.15.04.01" met="anuṣṭubh">
            <l n="a">nonmattāyā na kuṣṭhinyā</l> <l n="b">na ca yā spṛṣṭamaithunā |</l>
            <l n="c">pūrvaṁ doṣān abhikhyāpya</l><l n="d">pradātā daṇḍam arhati ||</l>
        </lg>
        <listApp type="parallels">
          <app>
            <note sameAs="txt:MDhŚ08.205">Our text gives no paraphrase for the third stanza grouped in this
section.</note>
          </app>
        </listApp>
      </supplied>
```

```
    </quote>
</div>
```

In the scenario encoded here, a single <p> of dependent text is grouped together in a <div> with two stanzas known from other transmissions of the base-text. The main purpose of the editor here is simply to include the vulgate base text integrally, despite the fact that the author of the dependent text ignores certain stanzas.

The second approach is to reserve a separate <div> for every segment of base text, regardless of whether any corresponding segment is found in the base text that is being edited. This scenario might be adopted if the editor believes that the dependent text intentionally does not comment on a stanza.

Example 4.4.B

```
<div type="dyad" n="199" xml:id="svayambhu_09.15">
        <quote xml:id="svayambhu_09.15.01" type="base-text" xml:lang="san-Latn">
                <supplied reason="omitted">
<lg xml:id="svayambhu_09.15.01.01" met="anuṣṭubh">
            <l n="a">nānyad anyena saṁsṛṣṭaṁ</l> <l n="b">rūpaṁ vikrayam arhati |</l>
            <l n="c">na sāvadyaṁ na ca nyūnaṁ</l> <l n="d">na dūre na tirohitam ||</l>
</lg>
<listApp type="parallels">
<app>
<note sameAs="txt:MDhŚ08.203">Our text gives no paraphrase for the first stanza grouped in this section.</note>
</app>
</listApp>
                </supplied>
        </quote>
</div>
<div type="dyad" n="200" xml:id="svayambhu_09.16">
        <quote xml:id="svayambhu_09.16.01" type="base-text" xml:lang="san-Latn">
                <supplied reason="implied">
<lg xml:id="svayambhu_09.16.01.01" met="anuṣṭubh">
        <l n="a">anyāṁ ced darśayitvānyā</l> <l n="b">voḍhuḥ kanyā pradīyate |</l>
        <l n="c">ubhe ta ekaśulkena</l> <l n="d">vahed ity abravīn manuḥ ||</l>
</lg>
<listApp type="parallels">
        <app>
                <note sameAs="txt:MDhŚ08.204"/>
        </app>
</listApp>
                </supplied>
        </quote>
        <p xml:id="svayambhu_09.16.02">ikaṅ vvaṅ amintonakən kanyā, donyan səḍaṅ tinonton, ..., maṅkana liṅ
bhaṭāra manu ṅūnin pavarah i bhagavān bhṛgu.</p>
</div>
<div type="dyad" n="201" xml:id="svayambhu_09.17">
```

```
<quote xml:id="svayambhu_09.17.01" type="base-text" xml:lang="san-Latn">
                <supplied reason="implied">
<lg xml:id="svayambhu_09.17.01.01" met="anuṣṭubh">
        <l n="a">nonmattāyā na kuṣṭhinyā</l> <l n="b">na ca yā spṛṣṭamaithunā |</l>
         <l n="c">pūrvaṁ doṣān abhikhyāpya</l><l n="d">pradātā daṇḍam arhati ||</l>
</lg>
<listApp type="parallels">
        <app>
        <note sameAs="txt:MDhŚ08.205">Our text gives no paraphrase for the third stanza grouped in this
section.</note>
        </app>
</listApp>
        </supplied>
        </quote>
</div>
```

Solution (2) tends to emphasize the structure of the vulgate of the base text and regards the fact that some of its elements are ignored in a dependent text as an accessory phenomenon, while solution (1) is more pertinent when a given base text is the focus of an edition.

4.5 Dependent text without base text

Some parts of the dependent text may not correspond to any particular section of the base text, as happens with introductory material or commentarial digressions. The same solutions that we have presented in the previous paragraph (§4.4) can be used to encode such scenarios.

If you decide that the subject is connected in some way to the preceding enclosing <div>, as in the case of digressions, then you can include the dependent text within that <div>, after the base text and the corresponding commentary with its own structural tags.

Example 4.5.A

```
<div type="dyad" n="31">
    <quote type="base-text" xml:lang="san-Latn">
        <lg n="31">
          <l n="a">verse line</l>
          <l n="b">verse line</l>
        </lg>
        <listApp type="parallels">
          <app/>
        </listApp>
    </quote>
    <p>dependent text related to the preceding <lg> of base text.</p>
    <p>further dependent text that doesn't correspond to the preceding <lg> of base text.</p>
  </div>
```

If, by contrast, you reckon that the text is too unconnected from the previous enclosing <div>, you can open a separate <div> to enclose this segment of dependent text on its own.

Example 4.5.B

```
<div type="dyad" n="31">
    <quote type="base-text" xml:lang="san-Latn">
       <lg n="31">
          <l n="a">verse line</l>
          <l n="b">verse line</l>
       </lg>
       <listApp type="parallels">
          <app/>
       </listApp>
    </quote>
    <p>dependent text related to the preceding <lg> of base text.</p>
</div>
<div n="32">
    <p>further dependent text that doesn't correspond to the preceding <lg> of base text.</p>
</div>
```

Note that introductory materials are always expected to form their own <div>. The best encoding solution is left to you and depends on your text's intrinsic structure. In case of doubt, we recommend the second solution, as it is less ambiguous.

4.6 Dependent text implies base text that is unknown

We distinguish from the scenario intended in the previous paragraph (§4.5) the case of implied but unknown base text. We mean a scenario where it is clear that your dependent text reflects a segment of another text but at least the precise wording of the segment of base text is unknown to you, in the current state of research — or perhaps it is even unknown in which work the reflected segment was contained. In this case, use an element <ab> for the unknown segment base text right before the relevant block of dependent text. By the nature of the case, the <ab> will contain no text as such, but an element <gap/> to represent it. This <gap/> will be wrapped in <supplied reason="implied"> and can be accompanied by the element <listApp> for which you must follow the rules given in §6.2.5 for unknown parallels. The <note> element will contain only textual content, but bear no attributes, as shown in the example.

Example 4.6.A

```
<div type="dyad" n="16" xml:id="svayambhu_02.16">
   <p xml:id="svayambhu_02.16.01">atəhər pramāṇastava.</p>
   <ab xml:id="svayambhu_02.16.02">
    <supplied reason="implied">
      <gap/>
```

```
        <listApp type="parallels">
          <app>
            <note>No Sanskrit stanza has been identified as corresponding to the following prose.</note>
          </app>
        </listApp>
      </supplied>
    </ab>
<p xml:id="svayambhu_02.16.03">apan tan hana ləviha sakeṅ dharma, sira pinakapramāṇaniṅ loka, sira ta
pinakasvarga saṅ magave hayu, sira pinakakaləpasan saṅ paṇḍita, tan maṅkana ikaṅ vvaṅ adharma, tan hana
gamāntyanya maṅkanāta saṅka pisana tuvi.</p>
</div>
```

4.7 Base text reconstructible from dependent text

It might happen that the text you are editing does not transmit a base text as such, but rather contains parts of it embedded within the dependent text. In this case, there might be enough constituent elements of the base text within the dependent text to allow a reconstruction of the base text itself. If such a reconstruction is among your research goals, your encoding should follow the model introduced here.

Example 4.7.A

```
<div>
      <div type="section">
        <quote type="base-text">
          <p xml:id="ID_02.27c" prev="#ID_02.27b" next="#ID_02.27d">some base text in prose.</p>
        </quote>
        <p>Some dependent text.</p>
      </div>
      <div type="section">
        <quote type="base-text">
          <p xml:id="ID_02.27d" prev="#ID_02.27c">some base text in prose</p>
        </quote>
        <p>Some dependent text.</p>
      </div>
</div>
```

The above scenario presents a main <div> (to which the attributes @type and @n can be added as needed), then divided into subsections, with a <div type="section"> for each pair consisting of base text and commentary. The internal organization of the <div type="section"> follows the model established in §4.1. The value "section" is just a proposition; in the scenario presented in the next section, we imagine that using "group" will make more sense for the encoder. Anyhow, feel free to contact the XML-TEI Data Manager to suggest new values.

    If you opt for this encoding strategy in order to reconstruct a base text that is fragmented between sections, it must be rebuilt by linking each part together. Specific models depending on whether you are dealing with dependent prose or dependent stanza are introduced below.

4.7.1 Prose interrupted by dependent prose

In case a block of text in prose is transmitted intertwined with (and interrupted by) the prose of a dependent text, it may be desirable to implement a system that will allow for this text to be reconstituted separately.

Even a fragmented paragraph should be encoded as such, with the element <p>, rather than using the element <ab>, as long as the reconstructed version can be assimilated at least to one complete sentence. Once each segment has received its XSLT-generated @xml:id, as explained in §7.3.1, you can interlink the segments with the attributes @prev and @next, which clarify the textual segments that precede and follow. Their value should always start with "#" followed by the @xml:id of the segment of prose they point to. This system allows all the parts to be chained together and can, if necessary, be used to reconstitute the base text without interspersal of dependent text.

<table>
<tr><td>Example 4.7.1.A</td></tr>
</table>

```
<div type="group" xml:id="ID_01.01">
        <p xml:id="ID_01.01.01" next="#ID_01.02.01" part="I">some text in prose</p>
        <p xml:id="ID_01.01.02">some dependent text here</p>
</div>
 <div type="group" xml:id="ID_01.02">
        <p xml:id="ID_01.02.01" prev="#ID_01.01.01" next="#ID_01.03.01" part="M">some text in prose</p>
        <p xml:id="ID_01.02.02">some dependent text here</p>
 </div>
 <div type="group" xml:id="ID_01.03">
        <p xml:id="ID_01.03.01" prev="#ID_01.02.01" part="F">some text in prose</p>
        <p xml:id="ID_01.03.02">some dependent text here</p>
 </div>
<div xml:id="ID_01.04">
        <p xml:id="ID_01.04.01">another paragraph in prose</p>
</div>
```

The example shows how three <p> elements corresponding to segments of the base text are connected to each other through their @prev and @next attributes; the first segment, with @xml:id="ID_01.01.01", is linked to the second, with @xml:id="ID_01.02.01", through @next; the second, with @xml:id="ID_01.02", is linked to the previous and the following bits with @prev and @next.

We do not foresee the order of segments of the base prose text being shuffled, as might happen if the base text is in verse form. This means that the fragmentation pattern should not be ambiguous — the order of the segments as preserved in the dependent text will match the content order of the base text as intended by the author of the latter — and so they can be marked up with the attribute @part, that renders explicit how all the parts interact. The permitted values for this attribute are "I", "M" and "F" in uppercase letter (for initial, medial and final), an

unlimited number of @part="M" being permitted but only one case each of @part="I" and @part="F" for one <p>.

In case you are dealing with segments of a base text in prose that are presented in the dependent text in an order which is not that of the base text as it was intended, then follow the encoding strategy with <join/> explained in §4.7.2 and contact the XML-TEI Data Manager.

4.7.2 Stanzas interrupted by dependent prose

In case a stanza of a base text is transmitted intertwined with (and interrupted by) the prose of a dependent text you should encode its constituent parts each with the element <lg>, and use the attributes @next and @prev as well as @part to chain them together as explained in §4.7.1. The usage of @prev, @next and @part is exactly as explained above, with the sole complication that an <lg> also has constituent verse lines that will interact in a more or less intricate fashion with the segmentation brought about by the dependent text.

Let us take as example a stanza with @n="3" contained in a <div n="33"> that is segmented into four parts by the dependent text requiring four subsections <div type="section">:

Example 4.7.2.A

```xml
<div type="section">
        <lg n="3" xml:id="ID_33.07" next="#ID_33.09" part="I">
                <l n="a">...</l>
        </lg><!-- end of the 1st part of the lg in question: an entire verse line-->
        <p xml:id="ID_33.08">some dependent text here</p>
</div>
<div type="section">
        <lg xml:id=ID_"33.09" prev="#ID_33.07" next="#ID_33.11" part="M">
                <l n="b">...</l>
                <l n="c">...</l>
        </lg><!-- end of the 2nd part of the lg in question: two entire verse lines-->
        <p xml:id="ID_33.10">some dependent text here</p>
</div>
<div type="section">
        <lg xml:id="ID_33.11" prev="#ID_33.09" next="#ID_33.13" part="M">
                <l n="d" part="I">...</l>
        </lg><!-- end of the 3rd part of the lg in question: the first part of a verse line-->
        <p xml:id="ID_33.12">some dependent text here</p>
</div>
<div type="section">
        <lg xml:id="ID_33.13" prev="#ID_33.11" part="F">
                <l n="d" part="F">...</l>
        </lg><!-- end of the 4th part of the lg in question: the last part of a verse line-->
        <p xml:id="ID_33.14">some dependent text here</p>
</div>
```

Note that @part is also used in the example for the <l> that is interrupted by dependent prose. This encoding allows us to reconstitute stanza number 3 during processing.

However, reconstituting the stanza directly in your file is possible by making use of the aggregating element <join/>. For instance, if you need to recreate the stanza because you want to cite it integrally in a comment or a note, you could either copy-paste its parts, or — more sophisticatedly — use <join/> to gather all or some of the parts of the <lg> element.

Example 4.7.2.B
<note>This is also seen in stanza 3: <join result="lg" scope="root" target="#ID_33.07 #ID_33.09 #ID_33.11 #ID_33.13"/>.</note>

The above will generate the equivalent of the following code:

Example 4.7.2.C
<note>This is also seen in stanza 3: <lg n="3"><l n="a">...</l><l n="b">...</l><l n="c">...</l><l n="d">...</l></lg>.</note>

Another scenario in which <join/> may be needed is if the order in which the dependent text presents the segments is not the one intended by the author of the stanza:

Example 4.7.2.D
| ```
<div type="chapter" n="33">
        <div type="dyad" n="15">
            <div type="section">
                <lg n="3" xml:id="ID_33.15.07" next="#ID_33.15.09" part="I">
                    <l n="a">..</l>
                    <l n="b">..</l>
                    <l n="c">..</l>
                </lg><!-- end of the 1st part of the lg in question: three entire verse line-->
                <p xml:id="ID_33.15.08">some dependent text here</p>
            </div>
            <div type="section">
                <lg xml:id="ID_33.15.09" prev="#ID_23.15.07" next="#ID_33.15.11" part="M">
                    <l n="d" part="M">...</l>
                </lg><!-- end of the 2nd part of the lg in question: an M part of verse d -->
                <p xml:id="ID_33.15.10">some dependent text here</p>
            </div>
            <div type="section">
                <lg xml:id="ID_33.15.11" prev="#ID_33.15.09" next="#ID_33.15.13" part="M">
                    <l n="d" part="M">...</l>
                </lg><!-- end of the 3rd part of the lg in question: an M part of verse d -->
                <p xml:id="ID_33.15.12">some dependent text here</p>
            </div>
``` |

```xml
                    <div type="section">
                        <lg xml:id="ID_33.15.13" prev="#ID_33.15.11" next="#ID_33.15.15" part="M">
                            <l n="d" part="I">...</l>
                        </lg><!-- end of the 4th part of the lg in question: the I part of verse d -->
                        <p xml:id="ID_33.15.14">some dependent text here</p>
                    </div>
                    <div type="section">
                        <lg xml:id="ID_33.15.15" prev="#ID_33.15.13" part="F">
                            <l n="d" part="F">...</l>
                        </lg><!-- end of the 5th part of the lg in question: the F part of verse d -->
                        <p xml:id="ID_33.15.16">some dependent text here</p>
                    </div>
                    <join result="lg" scope="root" target="#ID_33.15.07 #ID_33.15.13 #ID_33.15.09 #ID_33.15.11
#ID_33.15.15"/>
                </div>
    </div>
```

We don't actually foresee this element being used much, and we include it here mainly to let you know it exists. If you do need to use it, contact the project's XML-TEI Data Manager.

4.8 *Lemma*s of base text in dependent text

When the dependent text refers to individual terms or snippets from the base text, called *lemma* in the classical philological tradition,[41] resulting in an alternation of base and dependent text, without the whole base text being reconstructible on the basis of the latter, these *lemma*s shall be wrapped in the element <term>. It is possible to add an attribute @xml:lang to the tag, and we recommend doing so.

Example 4.8.A
`<term xml:lang="san-Latn">nikṣepa</term>`

It is possible, but not mandatory, to wrap the corresponding paraphrase or gloss in <gloss>. The attributes @xml:id and @target must be used to match a <gloss> with its <term>. The attribution of identifiers to terms will be handled by an XSLT process, as with other numbering (see §7.2.1.2), although for <term> the numbering pattern will be simple incremental numbering from the beginning of the file to its end. Once the @xml:id have been assigned, you can encode the values of @target, mandatorily with prefix "#".

Example 4.8.B
`<term xml:id="term004" xml:lang="san-Latn">nikṣepa</term> some text`

---

[41] Note that this is not exactly the same as a dictionary lemma, nor is it the same as the lemma of a critical apparatus, on which see §5.

```
<gloss target="#term004">gloss</gloss>
```

5. Critical apparatus

Many experienced editors of South and Southeast Asian texts are used to seeing printed editions with multiple "layers of apparatus", one or more of which are used to represent parallel passages.[42] It is important at this juncture to understand that such "layers of apparatus" are a convention of display. Digital editions, unlike printed ones, are not limited by the spatial features of a page-based layout and do not need to use anything like "layers of apparatus" to represent the relationship between the text as constituted by the editor and his observations on particular segments of that text. What is important, is for the types of information to be conveyed to be clearly defined and for the encoding to follow the respective categories of information. Anyhow, in the TEI context, the term *apparatus* has a more limited scope than you may be familiar with, as it concerns only the encoding of variant readings or notes on readings — to the exclusion of parallels, etc. It is to the apparatus in this narrow sense that we turn now.

Printed critical editions generally present the critical apparatus at the bottom of the page or in an appendix, using reference by page/line number or by elements of intrinsic structure of the text to make clear to which segment of text any variant readings (or comments) are to be matched. The TEI offers three main methods for encoding apparatus:

1. The **location-referenced method** assembles all contents of the apparatus separately from the edition (in a different part of the same file or even in a different file), and if the correspondence of any apparatus entry to a segment of text is to be made machine actionable, this must be achieved with reference to elements of intrinsic or extrinsic structure of the edition.[43]

2. The **double-end-point-attached method**[44] likewise assembles all contents of the apparatus separately from the edition (in a different part of the same file or even in a different file), but embeds anchors within the edition itself, to which the apparatus entries are linked.

3. The **parallel-segmentation method**,[45] by contrast, embeds the contents of each entry within the edition itself, thereby avoiding the need of any linking system.

In the framework of the DHARMA project, you might use either the first method (meaning that the entries are linked to a block of text without being embedded inside it) or the third method

---

[42] You will have noticed that the previous chapter has shown some of the relevant encoding strategies concerning parallel texts, and this subject will be dealt with more fully in §6.2.

[43] This method is explained in EGD §9.1.2 and §9.1.3.

[44] See TEI Guidelines, §12.2.2.

[45] See TEI Guidelines, §12.2.3. This method is the most advantageous from the computational perspective and seems to be favored in the TEI world for multi-witness editions in terms both of the number of specific tools that presuppose it and of the projects that have been applying it.

(meaning that the entries are embedded directly into the encoded text itself)[46] as well as a combination of the two. In what follows, we cover only method 3; however, in §6.3, we introduce method 1, which is easier to encode but only useful if you are editing a verse text.

5.1 Structural overview

Let us imagine a text edition based on three witnesses. Below, we illustrate the different ways in which an apparatus entry might be recorded, according to the different readings of the witnesses.

1. Two witnesses support the adopted reading, the third witness has a different reading

Example 5.1.A

preceding text
\<app>
   \<lem wit="#K #L">lemma\</lem>
   \<rdg wit="#M">variant reading\</rdg>
   \<note>optional observation\</note>
\</app>
following text

Display – lemma**] K L**, variant reading **M** • optional observation

2. None of the witnesses supports the adopted reading

Example 5.1.B

preceding text
\<app>
   \<lem type="emn">lemma\</lem>
   \<rdg wit="#K">variant reading 1\</rdg>
   \<rdg wit="#L #M">variant reading 2\</rdg>
   \<note>optional observation\</note>
\</app>
following text

Display – lemma**]** *em.*, variant reading 1 **K**, variant reading 2 **L M** • optional observation

3. One witness supports the adopted lemma, one reads it differently, one omits it

---

[46] Creating a digital edition from a printed one implies the conversion of its external apparatus to an internal one. Some exceptions may be necessary, as in the case of a printed edition with several layers of apparatus. If you find yourself in this situation, please consult the project's XML-TEI Data Manager before proceeding.

Example 5.1.C

preceding text
`<app>`
   `<lem wit="#K">`lemma`</lem>`
   `<rdg wit="#L">`variant reading 1`</rdg>`
   `<rdg wit="#M"><gap reason="omitted"/></rdg>`
   `<note>`optional observation`</note>`
`</app>`
following text

Display – lemma**] K**, variant reading 1 **L**, *om.* **M** • optional observation

4. All witnesses support the adopted reading, on which an observation is made

Example 5.1.D

preceding text
   `<app>`
      `<lem wit="#K #L #M">`lemma`</lem>`
      `<note>`optional observation`</note>`
   `</app>`
following text

Display – lemma**] K L M** • optional observation

If for some reason you do not wish the witnesses to be displayed in the apparatus, you can leave them unmentioned in the <lem> but in that case you need to use the attribute @type="note" (see §<u>5.3.2</u>):

Example 5.1.E

preceding text
   `<app>`
      `<lem type="note">`lemma`</lem>`
      `<note>`optional observation`</note>`
   `</app>`
following text

Display – Lemma**]** • optional observation

5. Two witnesses support the adopted reading, but one of them has a partial variant reading, while the third omits the whole segment

Example 5.1.F

preceding text
\<app>
      \<lem wit="#K #L">padāny
        \<app>
          \<lem wit="#K">aṣṭādaśaitāni</lem>
          \<rdg wit="#L">aṣṭadaśetani</rdg>
        </app>
      </lem>
      \<rdg wit="#M">\<gap reason="omitted"/></rdg>
\</app>
following text

Display — ^1. padāny aṣṭādaśaitāni] **K L**, *om*. **M**. ^2. aṣṭādaśaitāni] **K**, aṣṭadaśetani **L**, *om*. **M** (larger gap).

These five examples have introduced the elements <app>, <lem>, <rdg>, <gap> and <note> as well as the attributes @reason, @type, @wit — together, these are the most common building blocks you will use to create your apparatus entries. Some specific features require other building blocks that will be introduced below.

The element <app> will contain the whole apparatus entry, the first and mandatory child being <lem>, which is used to encode the reading of your choice, i.e., the lemma in philological terminology.[47] Generally, one or more variant readings will follow and be encoded inside an element <rdg> although, as shown in examples 5.1.D and 5.1.E, it is possible to create <app> entries consisting only of <lem> and <note> without any <rdg>. The lemma and any corresponding variant readings or observations are to be explicitly encoded together, i.e. wrapped within <app>, so that their interconnectedness is computationally explicit.

As shown in the example 5.1.F, apparatus entries may be nested, to record different phenomena applying to different parts of the same larger lemma.

None of your <app> elements will be numbered in the XML code: depending on what kind of display is chosen, and whether there is any need, this can be done automatically once the encoding of the file is finalized. On cross-referencing to a specific <app>, see §5.2.4 for its usage and §7.3.1 for its explanation.

---

[47] Note that the term 'lemma' in this context has a different meaning from the one intended in §4.8.

5.2 Which variant readings to record and which to ignore

5.2.1 Recording variant readings positively

In the DHARMA project, we opt for the use of a **positive apparatus**. This choice, reflected in the examples above, implies that, in any apparatus entry, all the witnesses should explicitly be listed and clearly identified, allowing the reader to be sure in every instance which witness reads what. Even though the process of encoding a positive apparatus can be more time consuming than encoding negatively, it has some advantages, among which the fact that it will enable implementing computer-aided verification of the consistency of your encoding.

This methodological requirement implies that each <app> should contain at least one, and normally (i.e. except in the case of scribal correction — see §5.5.13) only one instance, to be recorded with @wit, whether in <lem> (§5.3), in <rdg> (§5.5) or in <witDetail> (§5.10.2), of each witness listed in your <listWit> (§2.1.3). Be aware that any embedded <app> (see ex. 5.1F) will only inherit the witnesses declared on the embedding <app>.

5.2.2 Ignoring insignificant variant readings

An important aspect of your work as critical editor is deciding, case-by-case, whether a given variant reading for a given string of characters requires the creation of an <app> entry in your edition, or can simply be ignored. For instance, when all of your witnesses unanimously spell a word in a way that you wish to replace by a spelling that you consider "standard", or when you have a variety of substandard spellings none of which is exactly the same as the one you wish to adopt as "standard", you may deem none of the actual readings to merit being recorded as variants. In Appendix A, we furnish an overview of types of variants that we recommend ignoring if your are editing a Sanskrit and/or vernacular text from Java or Bali. The lists we furnish could be reused as the basis for your own further customized lists to be included in the <editorialDecl> of your edition (§2.2.1.1). However, bear in mind that if any variant reading needs to be recorded, then the principle of a positive apparatus requires that you represent all witnesses in your apparatus. This rule may lead you to explicitly record variant readings that you would have considered insignificant under other circumstances. However, it is also possible to abide in such a situation by the normalization rules declared in your <editorialDecl>: some readings can be considered to support the <lem> even if they are not strictly identical to it; consequently, the given witness would be encoded in <lem> and not in <rdg>. It is up to your discretion how you proceed in such situations. Perhaps the choice can be made on a case-by-case basis.

5.2.3 Hiding insignificant variants readings without deleting them

We offer you a way to experiment with the granularity of your editorial policy by applying the attribute @rend="hide" that may be used on any element <app>. Its effect is to only display the content of the <lem> without generating an apparatus entry. We foresee that this may be useful

if you initially create <app> entries to represent types of variant readings that you subsequently deem not to pass the threshold of significance for being included in your edition's apparatus. In this way, you will be able to reduce the number of apparatus entries so that only the more significant appear in the display, without needing to remove from the file itself the collation of the various witnesses once it has been made.

Example 5.2.3.A

sira dumadyakən vṛddhiniṅ janma kabeh,
    `<app rend="hide">`
      `<lem type="norm">`prāṇaprāṇī`</lem>`
      `<rdg wit="#msjb">`praṇapraṇī`</rdg>`
      `<rdg wit="#msba">`praṇapraṇi`</rdg>`
      `<rdg wit="#msbb">`pranaprani`</rdg>`
    `</app>`
 sarva hurip,

Display — text sira dumadyakən vṛddhiniṅ janma kabeh, prāṇaprāṇī sarva hurip, apparatus Not shown

5.2.4 Strategies to encode apparatus entries across block boundaries

Textual variants can occur at different levels of the text: although you will most frequently be encoding variants at the phrase level, they may just as well occur at a higher structural level, pertaining to verse lines or even whole stanzas, paragraphs or sections (<div>). If such a phenomenon occurs within your text, you need to exercise extra care while encoding the relevant apparatus entries, especially in case of a nested entry.

When your apparatus entry straddles a block-level container, such as <div>, <p>, <ab>, <lg> or even <l>, you might need to break it into two or more segments depending on its length. When this happens, you will have to provide an @xml:id consisting of the letters "app" followed by an incremental number formatted as a 5-digit number and add after an underscore a lower-case letter to identify the subparts. Once each part is clearly identified, you can link all the subparts together with the attribute @prev and @next, following the same mechanism that is used when a block level container is broken down to fit the base-text. See §7.2.1.2.

Example 5.2.4.A
`<div>`   `<p>`[...]

```
    <app xml:id="app00001_a" next="#app00001_b">...</app>
  </p>
</div>
<div>
  <p>
    <app xml:id="app00001_b" prev="#app00001_a">...</app>
    [...]</p>
</div>
```

5.3 The lemma

Within the element <app>, the adopted reading (or emendation/conjecture) with regard to which you are going to record one or more variant reading(s) and/or formulate a note is to be wrapped in the element <lem>.[48]

5.3.1 Extent of the lemma

There are no strict rules for the extent of your lemmas. Lemmas should remain sufficiently concise for the difference(s) between them and variant readings to stand out in your apparatus. There is no upper limit on the extent of a lemma. Indeed it may well occur that your lemma must extend over two or more adjoining words. Regarding the lower limit, your lemma should preferably consist of minimally a single whole word, but truncation of words is also admissible (§5.3.1.2), should the need arise.

In case of variation that involves multiple words, it is recommended to delimit your lemma in such a way as to include the first and last elements (normally, words) that are common to the lemma and all variant readings. See ex. 5.3.1.1.C and ex. 5.5.5.3.A for instance.

Metrical considerations will sometimes come into play: in a Sanskrit work in *anuṣṭubh* for which there are a large number of witnesses and lots of variants, it can make sense to take by default as the lemma the text of a complete eight-syllable quarter-verse (*pāda*), rather than producing five or six little *lemmata* for each half-line.

5.3.1.1 Ellipsis in the lemma's display

If you prefer any long lemmas to be shortened when displayed in the critical apparatus, you will not be able to indicate the shortened form inside <lem> itself, since we need a record of it in its full length to be displayed in the text edition. Instead, after recording the lemma in full with the element <lem>, insert an element <note> with a specific attribute @type="altLem" to express the truncated form of the lemma.[49] Let's imagine this lemma:

---

[48] Note that DHARMA has chosen not to allow declaring the empty element <lem/> in its encoding model for omissions and interpolation. See *TEI Encoding Guidelines*, §12.4.

[49] We present here the solution offered by Marjorie Burghart in her contribution "Textual Variants" to *Digital Editing of Medieval Manuscripts : A Textbook*, edited by Marjorie Burghart, Elena Pierazzo and James Cummings (2014–2017), p. 20.

Example 5.3.1.1.A
`<app><lem>`saṅ brāhmaṇa saṅ vruh riṅ aṣṭādaśavyavahāra,`</lem></app>`

If you desire in the apparatus the specific display saṅ brāhmaṇa ... -vyavahāra, then you must encode as follows.

Example 5.3.1.1.B
| ```
<app>
   <lem>saṅ brāhmaṇa saṅ vruh riṅ aṣṭādaśavyavahāra,</lem>
   <note type="altLem">saṅ brāhmaṇa ... -vyavahāra</note>
   <rdg wit="#A">...</rdg>
   <rdg wit="#B">...</rdg>
</app>
``` |

Note that you may simply type three full stops (...): the XSLT transformation will replace them by the proper ellipsis symbol (… U+2026).

As shown in the following, much more complex example, any `<note type="altLem">` should be placed immediately after the `<lem>` to which it applies. By contrast, any "normal" `<note>` should be the last child of `<app>` (see §8.2.1).

Example 5.3.1.1.C
| ```
 ... śūdra,
<app>
  <lem wit="#A #C">macaṅkramāta riṅ patapan, amava sabhūṣaṇanya,
                 sakavava ri patapan atah, pramāṇa sabhūṣaṇanya, kavāva riṅ patapan,
                 tan kilalanən de saṅ ratu, tan kavāva riṅ rāma, tan ucapən deniṅ śūdra,
                 ṅuniveh
    <app>
      <lem type="emn">sabanva</lem>
      <rdg wit="#A #C">saṁ banva</rdg>
   </app> uṅgvānira,
    <app>
      <lem type="conj">salinaraṅan</lem>
       <rdg wit="#A #C">valilaraṅan</rdg>
    </app>salinaraṅan
   </lem>
</app> saṅ prabhu</lem>
   <note type="altLem">macaṅkramāta ... saṅ prabhu</note>
   <rdg wit="#B"><gap reason="omitted"/></rdg>
         <note>The reading of <ptr target="#B"/>, jumping directly from <foreign>śūdra</foreign> to
<foreign>salviranya</foreign>, seems equally satisfactory.</note>
</app>
``` |

salviranya …

text
 ... śūdra, macaṅkramāta riṅ patapan, amava sabhūṣaṇanya, sakavava ri patapan ataḥ, pramāṇa sabhūṣaṇanya, kavāva riṅ patapan, tan kilalanən de saṅ ratu, tan kavāva riṅ rāma, tan ucapən deniṅ śūdra, ṅuniveh sabanva uṅgvānira, salinaraṅan saṅ prabhu, salviranya, ...

apparatus
^287. macaṅkramāta … saṅ prabhu] **A C**, *om*. **B** • The reading of **B**, jumping directly from *śūdra* to *salviranya*, seems equally satisfactory.
^288. sabanva] *em*., saṁ banva **A C**, *om*. **B** (larger gap).
^289. salinaraṅan] *conj*., valilaraṅan· **A C**, *om*. **B** (larger gap).

5.3.1.2 Truncation

It may be impossible or undesirable for the lemma boundary to be made to match with the most proximate word boundaries. In such cases, we offer the following ways to bring about display of a hyphen or of a raised circle to mark the point(s) of truncation, whereby hyphen shall mark natural truncation points (i.e. points where you would use hyphen if you would hyphenate your text) while raised circle shall mark any points that are not natural points for hyphenation. You must then use @rend inside <lem>, with one of the following six values:

- hyphenleft, hyphenright, hyphenaround
- circleleft, circleright, circlearound

Example 5.3.1.2.A
anya<app><lem type="emn" rend="hyphenleft" wit="#M">tobhāve</lem><rdg wit="#K #L">-tobhuve</rdg></app>
Display —
^1. -tobhāve] **M**, -tobhuve **K L**.

The needed hyphen before lemma *tobhāve* will be inserted by the transformation script, but in order to keep full control over display of <rdg>, no automatic insertion of any kind will apply to its contents. This is why you need to type the hyphen in <rdg>, as in *-tobhuve* in the example.

If your text edition systematically hyphenates compounds, and you wish to truncate at a break in a compound, then you have at your disposal the following two equivalent ways of bringing about proper display of the required hyphen in edition and apparatus:

Example 5.3.1.2.B

> *asvāmi*-*<app><lem type="emn" rend="hyphenleft">vikraya</lem><rdg wit="#L #K">*-*vikriya</rdg></app>*
> *<!-- if using @rend="hyphenleft", you need to type a hyphen outside of <lem> while typing one inside of <rdg> -->*
>
> *asvāmi<app><lem type="emn">*-*vikraya</lem><rdg wit="#L #K">*-*vikriya</rdg></app>*
> *<!-- if not using @rend="hyphenleft", you simply type the hyphens at the start of both <lem> and <rdg> -->*

Comparison of the two suggests that recourse to @rend="hyphenleft" will be less useful in the case of a fully hyphenated text edition.

The encoding of truncation at unnatural points is strictly analogous, except that here the issue of interaction with editorial hyphenation will never arise.

Example 5.3.1.2.C
any<app><lem rend="circlearound" wit="#M">atobh</lem><rdg wit="#K #L">°atheb°</rdg></app>āve
Display —
^1. °atobh°**] M**, °atheb° **K L**.

Note that none of the uses of @rend in the preceding examples can be applied to a <note> with @type="altLem". Should you require hyphens or raised circles in the display of the alternative lemma, you should explicitly type them in the contents of the <note>, as you need to do in <rdg>.

It may not, finally, be superfluous to point out explicitly that the extent of a variant reading should always correspond exactly to the extent of the lemma. In other words, you should never truncate an <rdg> if you do not truncate the corresponding <lem>, or truncate the <rdg> otherwise than you truncate the <lem>.

5.3.2 @type values usable with <lem>

If the adopted reading is not directly supported by any of the witnesses, then you must apply to the <lem> an attribute @type. The labels to express basic philological statements as to the status of the lemma are "norm", "emn" and "conj".

- "norm" — **Normalization** means any alteration to the text as transmitted in the witnesses in order to make it compliant with what is perceived as standard, without any of the actual readings being deemed to constitute a significant variant. Note that this @type will only be needed if for some reason you deem it useful to report any readings that do not constitute significant variants, because otherwise you would not be creating any <app> for the given segment of text at all (§5.2.2).
- "emn" — **Emendation** refers to a reading that is proposed by the editor to stand instead of the transmitted readings if these are deemed unacceptable, based on evidence that is deemed to point quite certainly to this specific choice of reading.

- "conj" — **Conjecture** means a reading that is proposed by the editor to stand instead of the transmitted readings if these are deemed unacceptable, or to fill in transmissional lacunae, without the editor disposing of strong supporting evidence for the specific alteration as opposed to other imaginable choices of reading.

Admittedly, the distinction between conjecture and emendation is not always sharp and can be rather subjective. Some editors indeed do not make the same three-way distinction, but we recommend it for critical editions produced by the DHARMA project. Here are some examples of how the @types can be used:

Example 5.3.2.A

`<app>`
        `<lem type="emn">`pariśrānto 'pi`</lem>`
        `<rdg wit="#B #C">`pariśrānte 'pi`</rdg>`
        `<rdg wit="#D"><gap reason="omitted"/></rdg>`
`</app>`

Example 5.3.2.B

`<app>`
        `<lem type="conj">`pariśrānto 'pi`</lem>`
        `<rdg wit="#B #C">`pariṣkṛto 'pi`</rdg>`
        `<rdg wit="#D"><gap reason="lost"/></rdg>`
`</app>`

We provide three further @type values, whose *raison d'être* is not only connected with philological appreciation of the status of the lemma but also with logical requirements of the TEI encoding model. These values are:

- "reformulated_elsewhere" — described in §5.5.6, this @type implies that you are adopting a lemma that is not attested as such in any witness but that is not deemed an emendation either, for instance a lemma that mixes the word order from one witness with the specific reading of one or more words from another witness.
- "omitted_elsewhere" — described in §5.5.9, this @type makes it possible on the one hand to oppose those witnesses which omit the whole segment to those which transmit it in some form while on the other hand allowing you to avoid giving the impression that there is no variation of reading for the lemma among the latter.
- "lost_elsewhere" — described in §5.5.11, this @type is intended to represent that one witness is affected by a larger lacuna, while the other witnesses keep transmitting variants.
- "note" — illustrated in §5.1, no 4 and §5.10.1, this type serves to indicate that the <app> entry in question contains only a <note> besides the <lem>, especially that it does not

contain any <rdg>. Use of this @type may or may not be accompanied by use of @wit on the <lem>.

From the point of view of the model, these values of @type allow you not to declare any @wit on <lem> and so to circumvent the problem of inheritance of values of @wit.[50]

5.4 Recording witnesses

If the adopted reading is supported by a witness (whether because the adopted reading is found as such in one or more of the witnesses or can be obtained by applying your normalization rules), then apply the attribute @wit inside <lem> using as its value the identifier of the relevant witness or witnesses as declared in the <sourceDesc> (see §2.1.3), preceded by the "#" that is needed to make the link to the list of sigla defined in your <listWit>. However, be careful not to refer to a family of witnesses and one of its members in the same attribute @wit. Any alternative readings corresponding to the adopted <lem> are to be reported by wrapping each one in its own element <rdg> and recording the relevant witness or witnesses in the same way as is done in <lem>.

Example 5.4.A
| ```
<app>
        <lem wit="#A #C">aṣṭādaśaitāni</lem>
        <rdg wit="#B">aṣṭadaśetani</rdg>
 </app>
``` |

In the next section, we will discuss in detail how to record variant readings.

5.5 Variant readings

In this section, you will find discussion of various specific features of our use of the element <rdg> in such cases as unclear or illegible readings, scribal additions and deletions, omissions due to scribal intervention or error, and lacunae due to physical deficiency of a witness.[51] Other phenomena such as interpolations or transpositions will be found in other sections of this guide, §5.6, and §5.8, as will instruction on how to supply untransmitted text segments, §5.7.

5.5.1 Quote witnesses as diplomatically as possible

When quoting a reading from a witness for which you have a TEI-encoded diplomatic edition, using what is called "strict transliteration" in the project's Transliteration Guide, you should copy verbatim from that diplomatic edition, paying attention to the following issues:

---

[50] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPSU
[51] Note that DHARMA has chosen not to allow declaring the empty element <rdg/> in its encoding model for omissions and interpolation. See *TEI Encoding Guidelines*, §12.4.

- you may remove any milestone elements (<lb/> and <pb/>)[52] if their presence has no incidence on the nature of the variant reading, but keep them if their presence seems relevant; you may copy them as they are in the file for your diplomatic edition along with any attributes (@n and @break) they bear.
- be aware of start-tags and end-tags:

    ○ add the start-tag for retained markup commencing before and ending inside your lemma;
    ○ add the end-tag for retained markup commencing inside your lemma and ending after it;
    ○ add start and end-tags for a lemma snipped from within a longer stretch of phrase-level markup.

If you do not have a separate TEI-encoded diplomatic edition, then represent the reading of the witness in question in strict transliteration, adapting the transliteration scheme of any published edition to DHARMA norms and citing any manuscript readings reported in a published edition as diplomatically as you can or with as much normalization as you deem necessary/useful, depending on the case.

5.5.2 Highlighting segments affected by variation

If more than one variant affects a lemma, it is possible to wrap the affected segments in the element <seg> to enable rapid identification of the segments of the <lem> affected by variation of reading encoded in the corresponding elements <rdg>. Apply the attribute @type with value "highlight" to such elements <seg>.

Example 5.5.2.A
<app>   <lem wit="#A #C">aṣṭādaś<seg type="highlight">ai</seg>tāni</lem>   <rdg wit="#B">aṣṭādaś<seg type="highlight">e</seg>tāni</rdg> </app>
Display — ^1. aṣṭādaś**ai**tāni] **A C**, aṣṭādaś**e**tāni **B**

5.5.3 Encoding causes of variance of reading

It is possible to implement a classification of the conscious interventions as well as the unconscious errors of scribes through the use of the attribute @cause on a rejected <rdg>. It may, for instance, be useful to mark readings which are identical in meaning to the adopted

---

[52] For more about milestone elements, we refer to §8.1 in this Guide.

lemma but express it through a different lexeme or with different (yet equivalent) morphological means. The values we propose for @cause are the following:[53]

- eye-skip (see §5.5.9)
- line_omission (see §5.5.9)
- haplography
- dittography
- lexical: the reading differs from the lemma by offering an entirely different lexeme
- orthographical: the reading differs from the lemma only in spelling (i.e., not in any grammatically significant way)
- morphological: the reading differs from the lemma in its grammatical form
- transposition: the reading changes the order in which text segments (letters/syllables, words, forms, phrases, sentences, lines) occur in the lemma
- syntactic: the reading differs from the lemma in its grammatical construction or arrangement
- subtractive: omission of some unit of text through oversight, erasure, or some other cause — use eye-skip or haplography if one of those two seems to be the specific cause
- additive: an interpolation, i.e., some text judged by the editor to have entered the tradition through the mistaken or deliberate copying of auxiliary material (e.g., glosses, marginalia, etc.) — use dittography if that seems to be the specific cause

Here is an example of a lexical variant, where the *prabu* in two witnesses is a synonym of *ratu* in the witness that supports the lemma. The grammatical constructions are also slightly different (with suffix -*niṅ* before *ratu* in the lemma against free-standing morpheme *saṁ* before *prabu* in the rejected reading), but the editor has here chosen not to encode that type of variation nor even to remark upon it in a <note>, though that would have been possible.

Example 5.5.3.A
dravya<app>   <lem rend="hyphenleft" wit="#L">niṅ ratu</lem>   <rdg wit="#K #M" cause="lexical">saṁ prabu</rdg> </app> tinuvavakən,

In the following example, the only significant difference between the rejected reading and the lemma is that the same words are arranged differently, which the editor has decided to underline with the attribute @cause="transposition".

---

[53] We adopt most of the values and definitions from the Digital Latin Library Guidelines, §12.15 Tagging Readings for Analysis. Note, however, that the Digital Latin Library offers these values to be used with the attribute @ana, while we have favored @cause. On this point, we follow the TEI Guidelines, where the following values are proposed: homeoteleuton, homeoarchy, paleographicConfusion, haplography, dittography, falseEmendation. See §att.textCritical.

<table>
<tr><td>Example 5.5.3.B</td></tr>
</table>

```
<app>
        <lem wit="#K #M">likhita, sākṣi, bhukti</lem>
        <rdg wit="#L" cause="transposition">sakṣi, likittha, bhūkti</rdg>
</app>
```

See also §5.8 for this and other scenarios of transposition of text segments. Contact the project's XML-TEI Data Manager if the proposed values of @cause do not meet your requirements.

5.5.4. Grouping variant readings

The TEI model offers the possibility of grouping variant readings based on their classification with the element <rdgGrp>. We have not made any concrete recommendations for its use in this guide. Nonetheless, encoding groups of variant readings with <rdgGrp> may be useful in analyzing the proximity or divergence between the respective witnesses. If you see any use for it in your edition, please contact the XML-TEI Data Manager. Note that it is possible to use it only on a few apparatus entries deemed particularly revealing.

The <rdgGrp> element should always be used inside an <app> entry after the <lem>. All variant readings need to enter one of the groups thus established (no <rdg> element can be declared as a sibling of <lem>, all must be descendants of an <rdgGrp>). Add an attribute @cause to the <rdgGrp> to make explicit that a given group of variants belong to the same class of variants (see §5.5.3).

<table>
<tr><td>Example 5.5.3.A</td></tr>
</table>

```
<app>
      <lem wit="#El #Ra2">though</lem>
      <rdgGrp cause="orthographic">
                <rdg wit="#La">thogh</rdg>
                <rdg wit="#Hg">thouh</rdg>
      </rdgGrp>
      <rdgGrp cause="lexical">
                <rdg wit="#Lb">although</rdg>
                <rdg wit="#Hh">whereas</rdg>
      </rdgGrp>
</app>
```

5.5.5 Encoding readings *ante* and *post correctionem*

Even after a scribal intervention, it is often possible to determine how a witness read before a scribe intervened to correct it. This is conventionally called the *ante correctionem* (*ac*) reading,

while the result of correction is conventionally labeled *post correctionem* (*pc*). There are major advantages to encoding the before/after reading each in its own <rdg>, namely:

- that either one may also be attested in another witness and that it will be consistent with our general approach in recording variant readings to mark agreement between two or more witnesses on a single reading by associating all those witnesses to a single <rdg> through @wit.
- that the *scribal* correction may be the result of the intervention of a hand distinct from the hand that was responsible for the initial reading, a fact that could not be represented if you made use of <choice><sic>...</sic><corr>...</corr></choice> as is done for *editorial* interventions in the project's epigraphic editions.

Before we move on to showing how you can apply the labels *ac* and *pc* to the before and after readings, we must emphasize that we do not see any objection to ignoring *ac* readings if they are not found as such in any other witness and if you encode the *pc* reading in <lem> — in other words, if you believe the scribe did not see the *ac* reading in any manuscript from which he was copying and agree with the correction that he made. Another situation in which you may feel free to ignore the *ac* reading (or leave unmarked that a given <rdg> is *pc*) is when it has become irrecoverable in the course of the scribal process of correction.

However, there will often be situations where you wish to encode both *ac* and *pc* readings, most especially if it is the former that supports the text you accept in <lem>. Let's say you are aiming at a result that, in print, would resemble the following:

$$\text{vijayaḥ P}_{pc}\text{ Q }]\text{ vajayaḥ P}_{ac}\text{ vajayo R}$$

You should then make use of the element **<witDetail>** (see §5.10.2) which can be linked to a reading as well as to one or more witnesses for that reading, or even to a specific hand responsible for a given reading. The link is inferred from the element's position, immediately after the element <lem> or <rdg> bearing the attribute @wit for the witness in which a correction has occurred.[54]

Example 5.5.5.A
| ```
<app>
    <lem wit="#P #Q">vijayaḥ</lem>
        <witDetail wit="#P" type="pc"/>
    <rdg wit="#P">vajayaḥ</rdg>
        <witDetail wit="#P" type="ac"/>
``` |

---

[54] A draft version of this guide also allowed the use of @varSeq, i.e. varSeq="1" for *ante correctionem* and varSeq="2" for *post correctionem* readings. We have removed this possibility since it didn't seem to bring any versatility to the encoding that is not also furnished by use of <witDetail/>. If you need its reintroduction for a case you can't fully express with <witDetail/>, contact the project's XML-TEI Data manager.

```
    <rdg wit="#R">vajayo</rdg>
 </app>
```

Display –

^1. vijayaḥ**] P_{pc} Q**, vajayaḥ **P_{ac}**, vajayo **R**

Example 5.5.5.B

```
<app>
        <lem wit="#P #Q">vijayaḥ</lem>
                <witDetail wit="#P" type="pc"/>
        <rdg wit="#P">v<subst><del>a</del><add place="above">i</add></subst>jayaḥ</rdg>
                <witDetail wit="#P" type="ac"/>
        <rdg wit="#R">vajayo</rdg>
 </app>
```

Display –

^1. vijayaḥ**] P_{pc} Q**, v⟨⟨[[a]]i⟩⟩jayaḥ **P_{ac}**, vajayo **R**

The following examples concern a scenario where the correction was applied by a second hand, and in the last example this fact is the object of an editorial comment:

Example 5.5.5.C

```
<app>
    <lem wit="#P #Q" hand="#P_H2">vijayaḥ</lem>
        <witDetail wit="#P" type="pc"/>
        <rdg wit="#P" hand="#P_H1">vajayaḥ</rdg>
        <witDetail wit="#P" type="ac"/>
    <rdg wit="#R">vajayo</rdg>
 </app>
```

Display –

^1. vijayaḥ**] P_{pc-sm} Q**, vajayaḥ **P_{ac-pm}**, vajayo **R**

Example 5.5.5.D

```
 <app>
        <lem wit="#P #Q" hand="#P_H2">vijayaḥ</lem>
        <rdg wit="#P" hand="#P_H1">vajayaḥ</rdg>
        <witDetail wit="#P" type="ac"/>
        <rdg wit="#R">vajayo</rdg>
        <note>In <ptr target="#P"/>, the second hand has added the i-vowel marker.</note>
```

<table>
<tr><td></app></td></tr>
<tr><td>Display –

^1. vijayaḥ] **P**_{pc-sm} **Q**, vajayaḥ **P**_{ac-pm}, vajayo **R** • In **P**, the second hand has added the i-vowel marker.</td></tr>
</table>

As you can see in the examples, it is possible, within the <rdg> elements, to make use of the full gamut of elements described in the chapter on the "Representation of primary sources" of the TEI Guidelines. In the examples above, we have used <gap>, <del> and <add>. For more on these elements, see §5.5.7 through §5.5.13.

5.5.6 Encoding a rejected branch of transmission

Should you face a scenario where different (groups of) witnesses offer different formulations (e.g., ordering of words) for a given segment, you will be forced to accept one and reject one or more other formulations, or propose a reading that is not attested as such in any of the witnesses. In the latter case, you may be forced to emend in a manner not supported by any witness, but the reading of the segment retained in your edition may also be a compromise between varying branches of transmission. If your reading is indeed based to a significant extent on one branch of transmission, while another branch shows a reformulation stretching over multiple words, you can indicate this by assigning the attribute @type="reformulated_elsewhere" to the <lem> in your edition. At the moment, we foresee two specific scenarios in which this attribute must be used.

5.5.6.1 Adopting the formulation from one branch but a reading from another
Should you wish to adopt the overall formulation of one branch of transmission while still needing a reading from a rejected branch for a specific word that is part of the segment reformulated in that branch, then you may encode the situation along the lines of the following example:

<table>
<tr><td align="center">Example 5.5.6.1.A</td></tr>
<tr><td>

```
<app>
  <lem type="reformulated_elsewhere">ḍaṇḍa saṅ prabhu,
  amaṅgih <app><lem wit="#L">duryaśa</lem><rdg wit="#K #M">ḍūryyaṣan</rdg></app> sira,
   <app><lem wit="#K">amaṅgih</lem><rdg wit="#L">hamaṅguḥ</rdg><rdg wit="#M">amaṅguḥ</rdg></app>
pāpa</lem>
 <witDetail wit="#K #M" type="retained"/>
 <note type="altLem" rend="hyphenright">ḍaṇḍa saṅ prabhu ... amaṅgih pāpa</note>
 <rdg wit="#L">ḍəṇḍa, hamaṅguḥ dūryyaṣa: saṁ prabhū, makaḍḍi hamaṅguḥ papa:-</rdg>
  <note>The <ptr target="#L"/> reading is smoother with regard to the placement of <foreign>saṅ prabhu</foreign>;
we have nevertheless decided to adopt the word order and punctuation of <ptr target="#K #M"/>. The somewhat
awkward placement of <foreign>saṅ prabhu</foreign> could be smoothened by moving the punctuation mark to
stand before instead of after <foreign>saṅ prabhu</foreign>.</note>
 </app>
```

</td></tr>
</table>

In the <app> for the passage which is reformulated in some witness(es), the element <lem> that bears the attribute @type="reformulated_elsewhere" doesn't take any attribute @wit. Rather, the witness(es) which furnish(es) the basic structure of your lemma should be declared in a <witDetail type="retained"/> following the main <lem> and before any <note type="altLem">, should you need one (see §5.3.1.1). The integral reading(s) of the witness(es) showing reformulation are represented in <rdg>. You can then cite all witnesses in any <app> that needs to be included within the <lem> marked with @type="reformulated_elsewhere". The reason for declaring the witness(es) whose formulation is retained in <witDetail/> rather than in <lem> is connected with the logic regarding inheritance of the attributes: to allow embedding an <app> in a <lem> all the witnesses must have been indicated in the parent <app>, which isn't the case here. That's why we need to declare the witnesses in a different manner than usual.

5.5.6.2 Adopting the formulation from one branch but a reading from another across boundaries

Should you need to declare a reformulation on a larger segment of text crossing block-level containers, use the element <span type="reformulationStart"/> and <span type="reformulationEnd"/>. Unlike the twin elements <span type="omissionStart"/> and <span type="omissionEnd"/>, we don't expect them inside an <app> entry. The <span/> should be used as an anchor to declare the beginning and the end of the reformulation. In order to match the start element with the end element corresponding to it, any <span type="reformulationStart"/> should be given an @xml:id formed on the abbreviated term "reform" followed by a three-digit incremental number. The <span type="reformulationEnd"/> should be linked to it with the attribute @corresp with the value of the @xml:id preceded by "#". These attributes are mandatory in order to match the reformulation's boundaries correctly, because it may happen that two or more start/end pairs come to overlap. An <app> entry should be opened right after the <span type="reformulationEnd"/>. It should contain an element <lem> associated with the witness whose reformalation has been kept in an attribute @wit and identified clearly with the attribute @type="retained". The rejected variants can be encoded as usual with the element <rdg> (see §5.5) and the <app> may also contain any necessary <note> (see §5.10.1).

<div style="border:1px solid;">

Example 5.5.6.2.A

```
 <div>
  [...]
 <span type="reformulationStart" xml:id="reform001"/>
  ya kaṅkən guru
   <app>
      <lem wit="#B">kautamaan</lem>
      <rdg wit="#A #EdASD">Utama</rdg>
   </app>, ya kaṅkən guru
  [...]
</div>
<div>
   <app>
      <lem wit="#B">kaprəmanaan</lem>
      <rdg wit="#A">primana</rdg>
      <rdg wit="#EdASD">prəmana</rdg></app>, ya kaṅkən guru kaupadesaan
   <span type="reformulationEnd" corresp="#reform001"/>
    <app>
       <lem type="retained" wit="#A"/>
       <rdg wit="#B" >ya guru kamulyin· ṅarana guru kaprəmanaAn·, ya kaṁkən· guru kaUtamaAn·</rdg>
       <note>It seems that <ptr target="#B"/> favors forms with circumfix <foreign>ka-...-an</foreign> and arranges
the elements in a different sequence, while omitting an element corresponding to <foreign>kaupadesaan</foreign>
in <ptr target="#A"/>.</note>
    </app>
  [...]
</div>
```

Display —

text
  ya kaṅkən guru kautamaan, ya kaṅkən guru
[...]
 kaprəmanaan, ya kaṅkən guru kaupadesaan

apparatus
^1. ya kaṅkən guru … (§**2**) ya kaṅkən guru kaupadesaan**]** Thus formulated in **A**, ya guru kamulyin· ṅarana guru
kaprəmanaAn·, ya kaṁkən· guru kaUtamaAn· **B** • It seems that **B** favors forms with circumfix *ka-...-an* and
arranges the elements in a different sequence, while omitting an element corresponding to *kaupadesaan* in **A**.
^2. kamulyan**] B**, mulya **A Ed**[ASD]
^3. kaprəmanaan**] B**, primana **A**, prəmana **Ed**[ASD]

</div>

5.5.6.3 Encoding the intended reading of a rejected branch of transmission

In case another branch of transmission is clearly discernible for a given passage, with its own
reading that makes sense and must have been intended by whoever modified what was the
previously received text, but the witnesses that attest to such a reformulated reading have small
differences, it might be interesting to offer also the critical edition of the alternative reading that
would have come in place of the <lem> if the underlying reading of those witnesses had been

adopted. In such a scenario, you may record the underlying reading that you have observed but decided not to adopt in your edited text, as a <rdg> with @type="paradosis" and @wit for all the witnesses that reflect this reading.[55]

<table>
<tr><td colspan="1" align="center">Example 5.5.6.3.A</td></tr>
</table>

   \<app\>
  \<lem wit="#L"\>śeṣaniṅ ḍaṇḍa, śeṣaniṅ *tukon, ika tan tumusa riṅ anak putuniṅ ahutaṅ\</lem\>
     \<note type="altLem"\>śeṣaniṅ tukon ... anak putuniṅ ahutaṅ\</note\>
    \<rdg wit="#K"\>śeṣaniṁ ḍaṇḍa, śeṣa vlan·, Ika ta kabeḥ, yan matikaṁ mahutaṁ, tan yogya tagihən hanaknya\</rdg\>
     \<rdg wit="#M"\>sesaniṁ, sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya\</rdg\>
      \<rdg type="paradosis" wit="#K #M"\>śeṣaniṅ ḍaṇḍa, śeṣa vlyan, ika ta kabeh, yan matīkaṅ mahutaṅ, tan yogya tagihən hanaknya\</rdg\>
       \<note\>The two alternative ways of reading the last sentence, represented respectively by \<ptr target="#L"/\> and \<ptr target="#K #M"/\>, seem equally acceptable. It is remarkable that where \<ptr target="#L"/\> reads \<foreign\>śeṣaniṅ tukon\</foreign\>, \<ptr target="#K #M"/\> seem to reflect \<foreign\>śeṣa vlyan\</foreign\> in their hyparchetype, while they agree with \<ptr target="#L"/\> on reading \<foreign\>śeṣaniṅ ḍaṇḍa\</foreign\>. There is external textual evidence in Perpusnas L882 (\<foreign\>mvah hutaṅ tan kaliliranā deniṅ putra, hutaṅ tan paputra, hutaṅ totohan, śeṣadaṇḍa, hutaṅ tukon, hutaṅ sajāṅ\</foreign\>) that might support conjecturing \<foreign\>śeṣa ḍaṇḍa\</foreign\> in our text. There is also evidence that \<foreign\>tukon\</foreign\> and \<foreign\>vəlyan\</foreign\> were felt to be equivalent in this genre of literature (UBL Or 5037 \<foreign\>ana vvaṅ istrī linamar saptapayu sinrahan ikaṅ tukon mati taṅ istrīkaṅ sinrahan tukon vəlyan ika si baṅavan ika tan vaṅsulakna ikaṅ tukon\</foreign\>).\</note\>
   \</app\>

Display –

^1040. śeṣaniṅ tukon … anak putuniṅ ahutaṅ**]** Thus formulated in **L**, śeṣaniṁ ḍaṇḍa, śeṣa vlan·, Ika ta kabeḥ, yan matikaṁ mahutaṁ, tan yogya tagihən hanaknya **K**, sesaniṁ, sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya **M**, śeṣaniṅ ḍaṇḍa, śeṣa vlyan, ika ta kabeh, yan matīkaṅ mahutaṅ, tan yogya tagihən hanaknya **K M** • *Paradosis* of **K M**: śeṣaniṅ ḍaṇḍa, śeṣa vlyan, ika ta kabeh, yan matīkaṅ mahutaṅ, tan yogya tagihən hanaknya • The two alternative ways of reading the last sentence, represented respectively by **L** and **K M**, seem equally acceptable. It is remarkable that where **L** reads *śeṣaniṅ tukon*, **K M** seem to reflect *śeṣa vlyan* in their hyparchetype, while they agree with L on reading śeṣaniṅ ḍaṇḍa. There is external textual evidence in Perpusnas L882 *(mvah hutaṅ tan kaliliranā deniṅ putra, hutaṅ tan paputra, hutaṅ totohan, śeṣadaṇḍa, hutaṅ tukon, hutaṅ sajāṅ)* that might support conjecturing *śeṣa ḍaṇḍa* in our text. There is also evidence that *tukon* and *vəlyan* were felt to be equivalent in this genre of literature (UBL Or 5037 *ana vvaṅ istrī linamar saptapayu sinrahan ikaṅ tukon mati taṅ istrīkaṅ sinrahan tukon vəlyan ika si baṅavan ika tan vaṅsulakna ikaṅ tukon*).

5.5.7 Encoding (segments of) a reading as unclear

Many manuscripts contain text that is difficult to read. It might be a single grapheme, a word, a sentence or even longer passages. You may use the element <unclear> to represent unclearness (and degrees of unclearness) of reading.[56] Note that the only child allowed within

---

[55] We are slightly adapting the term "paradosis" here from the way it is used in M. L. West, *Textual Criticism and Editorial Technique* (1973), pp. 53–59.
[56] EGD §5.3.1.

<unclear> is <g/>; in case the unclear segment straddles some other TEI element, the unclear text around it must be wrapped separately in <unclear>.

To encode a segment that would be illegible out of context (e.g. because of physical damage or sloppiness of the scribe), but as to whose reading you are quite confident in the context, use the element <unclear> without any attributes.

Example 5.5.7.A
<rdg wit="#L642">ya ta mas <unclear>r̥kr̥t</unclear> masǝkar vañva</rdg>

You can use the attribute @reason with the value "eccentric_ductus" if your confidence regarding the reading is affected by any unusual or incomplete aspect of a glyph.

Example 5.5.7.B
<app>   <lem wit="#L642">ceva masva ceva mibuḥ</lem>   <rdg wit="#L633">cava<unclear reason="eccentric_ductus">ma</unclear>so sevavibu</rdg> </app>

Any tentative reading can be indicated using the attribute @cert, the only admitted value for which is "low".[57] You can encode segments whose reading is ambiguous with <choice> and <unclear>.[58] If the given portion is really impossible to read, you can use the element <gap/> element with @reason="illegible" as described in the following section.

5.5.8 Encoding (segments of) a reading as illegible

When the reading corresponding to a given lemma in a witness is wholly or partially illegible (e.g. because of physical damage or extreme sloppiness of the scribe), then encode the illegible segment in <gap> with the value "illegible" for @reason inside the <rdg> for the relevant witness. In this case, you will typically be able to know or estimate the number of illegible characters, and need to specify @quantity and @unit within <gap reason="illegible"/>. Here, @quantity must express the number of lost characters in numerical form. The default value for the attribute @unit is set on "character" (which must be understood to mean a whole *akṣara in our* context), but it can also be used with the value "component" depending on the case.[59] If you estimate the number of illegible characters, add an attribute @precision with the value "low" to underline the uncertainty. It is also possible to use @extent="unknown" with the attribute @unit="character".

---

[57] In other words, we exclude the values "medium" and "high" foreseen by the TEI, especially because a hypothetical @cert="high" would mean that you do not need to encode <unclear> at all. See EGD §5.3.2.
[58] See EGD §5.3.3.
[59] See EGD §5.4.5.

<table>
<tr><td colspan="1">Example 5.5.8.A</td></tr>
</table>

```
<app>
        <lem wit="#K">tinurunakən sakeṅ svarganya</lem>
        <rdg wit="#L">tinurunakən <gap reason="illegible" quantity="4" unit="character"/>nya</rdg>
</app>
```

Display —

^1. tinurunakən sakeṅ svarganya] **K**, tinurunakən [4x]nya **L**.

<table>
<tr><td colspan="1">Example 5.5.8.B</td></tr>
</table>

```
<app>
        <lem wit="#K">tinurunakən sakeṅ svarganya</lem>
        <rdg wit="#L">tinurunakən <gap reason="illegible" extent="unknown" unit="character"/>nya</rdg>
    </app>
```

Display —

^1. tinurunakən sakeṅ svarganya] **K**, tinurunakən [...]nya **L**.

5.5.9 Text of lemma absent from a witness due to scribal omission

When you wish to indicate that a given lemma has no corresponding reading in a given witness, because the scribe has omitted the string of characters in question, then you can use <gap> with the value "omitted" on @reason inside the <rdg> for the relevant witness. Let us discuss the following provisional example:

<table>
<tr><td colspan="1">Example 5.5.9.A</td></tr>
</table>

```
<app>
        <lem wit="#K">tinurunakən sakeṅ svarganya deniṅ devata kabeh</lem>
        <note type="altLem">tinurunakən ... kabeh</note>
        <rdg wit="#L" cause="eye-skip"><gap reason="omitted"/></rdg>
        <note>This sentence seems to have been omitted in <ptr target="#L"/> due to eye-skip from
<foreign>kavaḥ</foreign> to <foreign>kabeḥ</foreign>.</note>
</app>
```

Display —

^1. tinurunakən ... kabeh] **K**, *om*. **L** (eye-skip). This sentence seems to have been omitted in **L** due to eye-skip from *kavaḥ* to *kabeḥ*.

In the case of text omitted due to eye-skip, as in this example, the mandatory elements are <app>, <lem>, <rdg> and <gap> with @cause="eye-skip", while the two elements <note> are optional. As you see, there is a redundancy in the example between the use of

@cause="eye-skip" (cf. §5.5.3) and the contents of the final <note>, which you may want to avoid by avoiding the explicit term "eye-skip" in the way you formulate the note that explains for human readers what process of eye-skip you imagine. So let us propose the following slight modification:

Example 5.5.9.B
<app>   <lem wit="#K">tinurunakən sakeṅ svarganya deniṅ devata kabeh</lem>   <note type="altLem">tinurunakən ... kabeh</note>   <rdg wit="#L" cause="eye-skip"><gap reason="omitted"/></rdg>   <note>The scribe of <ptr target="#L"/> seems to have skipped from <foreign>kavaḥ</foreign> to <foreign>kabeḥ</foreign>.</note> </app>
Display —
^1. tinurunakən ... kabeh] **K**, *om*. **L** (eye-skip). The scribe of **L** seems to have skipped from *kavaḥ* to *kabeḥ*.

In case the process of eye-skip seems self-evident, there will of course be no need to use any explanatory <note>.

On the other hand, you may wish to make the process even more transparent by including in the lemma both segments which have triggered the *saut du même au même*, in which case, however, you may not use <gap> in your <rdg>.

Example 5.5.9.C
ikaṅ avyavahāra kalih, tan ana kumavruhi <app>   <lem wit="#K">vicāranya, tātan vruh saṅ prāgvivāka ri hutaṅnya, konən</lem>   <note type="altLem">vicāranya, tātan .... hutaṅnya, konən</note>   <rdg wit="#L" cause="eye-skip">vicāranya konən</rdg> </app> asatyaha
Display —
^1. vicāranya, tātan ... hutaṅnya, konən] **K**, *vicāranya konən* **L** (eye-skip).

A model similar to <rdg wit="#L" cause="eye-skip"><gap reason="omitted"/></rdg> in ex. 5.5.8.A applies to the omission of an entire line of verse by one or more witnesses. In this case it is possible to add the attribute @cause="line_omission" to the <rdg> element in order to display such a reading with an indication like "(line omission)" instead of "(eye-skip)".

Example 5.5.9.D
<lg n="23">

```
    <l n="a">nāsti satyāt paro dharmo</l>
        <app>
           <lem wit="#A"><l n="b">nānṛtāt pātakaṁ param |</l></lem>
           <rdg wit="#B" cause="line_omission"><gap reason="omitted"/></rdg>
        </app>
    <l n="c">sthitiś ca loke dharmaś ca</l>
    <l n="d">tasmāt satyaṁ viśiṣyate ||</l>
</lg>
```

Display —

^1. nānṛtāt pātakaṁ param |] **A**, *om*. **B** (line omission).

In case any variant reading occurs among the witnesses that do transmit the segment affected by a line omission in another witness, you will need to make use of the @type with value "omitted_elsewhere" on <lem> (see §5.3.2), as in the next example:

Example 5.5.9.E

```
<l n="c">
  <app>
   <lem type="omitted_elsewhere">kvehnya savarga
    <app>
     <lem wit="#msjb">muṅguh ikanaṅ</lem>
     <rdg wit="#msbb">muṅgu rikanaṅ</rdg>
    </app>
   <app>
     <lem wit="#msjb">vacan aniga-</lem>
     <rdg wit="#msbb">jati maniga-</rdg>
   </app>niga</lem>
<witDetail type="retained" wit="#msjb #msbb"/>
  <note type="altLem">kvehnya ... aniga-niga</note>
  <rdg wit="#msba" cause="line_omission"><gap reason="omitted"/></rdg></app>,
</l>
```

Display —

^1. kvehnya ... aniga-niga**]** transmitted in **J$_2$ B$_2$**, *om*. **B$_1$** (line omission).
^2. muṅguh ikanaṅ**] J$_2$**, muṅgu rikanaṅ **B$_2$**, *om*. **B$_1$** (larger gap)
^3. vacan aniga-**] J$_2$**, jati maniga- **B$_2$**, *om*. **B$_1$** (larger gap)

Note that this last encoding strategy is a customization for the case of a line omission but otherwise exactly analogous to ex. 5.5.11.1.B. Remember also that the list of permitted values of @cause is limited. See §5.5.3.

5.5.10 Text of lemma absent from a witness due to physical deficiency

When you wish to indicate that a given lemma has no corresponding reading in a given witness, because the manuscript has physically lost the string of characters in question, then use <gap> with the value "lost" on @reason inside the <rdg> for the relevant witness. Example:

Example 5.5.10.A
<app>   <lem wit="#K">tinurunakən sakeṅ svarganya deniṅ devata kabeh</lem>   <note type="altLem">tinurunakən ... kabeh</note>   <rdg wit="#L"><gap reason="lost"/></rdg> </app>
Display — ^1. tinurunakən ... kabeh] **K**, *lac*. **L**.

If necessary, you can also use the element <gap/> to express the partial loss of the lemma in a given witness. In this case, you will typically be able to know or estimate the number of lost characters, and need to specify @quantity and @unit within <gap reason="lost"/>. Here, @quantity must express the number of lost characters in numerical form. You can also add the attribute @precision="low", if you need to underline the fact that the quantity of lost characters can only be estimated. For details, see §5.5.8.

Example 5.5.10.B
<app>   <lem wit="#K">tinurunakən sakeṅ svarganya</lem>   <rdg wit="#L">tinurunakən <gap reason="lost" quantity="4" unit="character"/>nya</rdg> </app>
Display — ^1. tinurunakən sakeṅ svarganya] **K**, tinurunakən [4+]nya **L**.

Example 5.5.10.C
<app>   <lem wit="#K">tinurunakən sakeṅ svarganya</lem>   <rdg wit="#L">tinurunakən <gap reason="lost" extent="unknown" unit="character"/>nya</rdg> </app>
Display — ^1. tinurunakən sakeṅ svarganya] **K**, tinurunakən [...]nya **L**.

Note that we distinguish this kind of case where the physical deficiency corresponds to a part of the lemma from the scenario of larger physical deficiency within which an expected text segment falls. This second case is explained in §5.5.10.

5.5.10.1 Text absent from a witness due to physical deficiency but with known meter
If the number of syllables is known thanks to its prosodic pattern for any text segment that is irretrievably lost or illegible because of a physical deficiency, the following encoding must be used: wrap the element <gap/> within a <seg> with the attribute @met. The value of the attribute should match the prosodic conventions in the external authority file (see PVG). For the case of a segment of text of known metrical structure not in the reported reading of a witness but in the text edition itself (for instance if the text is being edited on the basis of a single witness that is lacunose), see §5.7.2.

5.5.11 Absence of a witness for a lemma due to larger physical lacuna

5.5.11.1 Textual loss within a block-level container
It may also happen that a lemma finds no corresponding reading in a given witness for the reason that the segment in question falls within a larger physical lacuna in that witness. If the entire lemma falls within a single textual block, you may nest the relevant <app> in a larger <app>, as follows:

Example 5.5.11.1.A
<app>   <lem>   <app>   <lem type="norm" wit="#EdASD">məṅpəṅ</lem>   <rdg wit="#A">məpəṁ</rdg>   </app>, məlah,   <app>   <lem type="norm" wit="#EdASD">mañcir</lem>   </app>, midvakən,</lem>   <rdg wit="#B"><gap reason="lost"/></rdg>   </app>
Display — text   məṅpəṅ, məlah, mañcir, midvakən, apparatus ^938. məṅpəṅ, məlah, mañcir, midvakən,**]** *lac*. **B** ^939. məṅpəṅ**]** *norm*. **Ed**ASD, məpəṁ **A**, *lac*. **B** (larger gap) ^940. mañcir**]** *norm*. **Ed**ASD, *lac*. **B** (larger gap)

You should not encode in the embedded <app> an <rdg> for the witness affected by the larger lacuna. The indication "*lac*. **L** (larger gap)" in the example is generated by the fact that <lem> falls within the <app> to which <rdg wit="#L"><gap reason="lost"/></rdg> applies. In other words, this scenario is an exception to the general rule that all witnesses should explicitly be listed and clearly identified in a given <app> (see §5.2.1). A slightly more complicated but also more rigorous manner of encoding this scenario is as follows:

Example 5.5.11.1.B
<app>   <lem type="lost_elsewhere">   <app>   <lem wit="#EdASD">məṅpəṅ</lem>   <rdg wit="#A">məpəṁ</rdg>   </app>, məlah,   <app>   <lem wit="#EdASD">mañcir</lem>   </app>, midvakən,</lem>   <witDetail type="retained" wit="#A #EdASD"/>   <rdg wit="#B"><gap reason="lost"/></rdg>   </app>
Display — text   məṅpəṅ, məlah, mañcir, midvakən, apparatus ^938. məṅpəṅ, məlah, mañcir, midvakən,**]** preserved in **A Ed**ASD, *lac*. **B** ^939. məṅpəṅ**] Ed**ASD, məpəṁ **A**, *lac*. **B** (larger gap) ^940. mañcir**] Ed**ASD, *lac*. **B** (larger gap)

This second solution makes it possible to generate "preserved in A Ed[ASD]" in the display. For @type="lost_elsewhere", see §5.3.2.

5.5.11.2 Textual loss across boundaries between block-level containers
When a lacuna straddles boundaries between block-level containers, you will need to use an alternative approach, namely to mark the beginning and end of the physical lacuna with <lacunaStart/> and <lacunaEnd/> inside <rdg> elements corresponding to <lem> elements for the words affected by the beginning and end of the gap in the witness:

Example 5.5.11.2.A
<p n="3">...kalaṅkaṅ saṅ hyaṅ   <app><lem wit="#A #EdASD">dasasila</lem><rdg wit="#B">dasa<lacunaStart/></rdg></app>, maya-maya ...

```
        <app><lem wit="#A">di</lem><rdg wit="#EdASD"><gap reason="omitted"/></rdg></app> bumi lamba, di
bumi
        <app><lem type="emn">parək</lem><rdg wit="#A">paR̥<unclear>ka</unclear></rdg><rdg
wit="#EdASD">tan parək</rdg></app>.</p>

<p n="4">
        <app><lem wit="#A #EdASD">ini pakən</lem><rdg wit="#B"><lacunaEnd/> pak·kən</rdg></app> uraṅ...
</p>
```

Display —

text

1.3 i…kalaṅkaṅ saṅ hyaṅ dasa-sila, maya-maya saṅ hyaṅ dasa-marga, kaprətyaksaan na dasa-indriya, pakən
ṅrətakən di bumi lamba, di bumi parək.
1.4 ini pakən uraṅ...

apparatus
^24. dasa-sila] **A Ed**[ASD], dasa[... **B**
^25. dasa-sila … (**§1.4**) ini pakən] A gap due to loss intervenes in **B**
^26. di] **A**, *om*. **Ed**[ASD], *lac*. **B** (larger gap)
^27. parək] *em*., paR̥ka **A**, tan parək **Ed**[ASD], *lac*. **B** (larger gap)
^28. ini pakən] **A Ed**[ASD], ...] pak·kən **B**

Note that the <rdg> containing <lacunaStart/> should always include at least the last preserved word or compound member before <lacunaStart/>; inversely, the <rdg> containing <lacunaEnd> should always include at least the first preserved word or compound member after <lacunaEnd/>. In this approach, too, you may not encode any <rdg> for the given witness in *loci* falling between <lacunaStart/> and <lacunaEnd/>. The two indications "*lac*. **B** (larger gap)"' in the example are generated by the fact that the two <lem>s fall between the two elements and no <rdg> for #B has been explicitly encoded.

Note also that any <lacunaStart/> must be closed at some point, before the end of your edition, with a <lacunaEnd/>. When a witness breaks off a few paragraphs before the end of the text, the <lacunaEnd/> must be added in an <app> entry at the last word. If an apparatus entry is required at that point for other reasons, then proceed as usual. Otherwise add an attribute @rend="hide" to this <app>. This is the only scenario in which the <lacunaEnd/> does not need to be encoded with a word or compound member.

Example 5.5.11.2.B
`<app rend="hide"><lem wit="#EdASD #B">palipurna.</lem><rdg wit="#A"><lacunaEnd/></rdg></app>`

5.5.12 Absence of a witness for a lemma due to larger scribal omission

The following scenarios are strictly analogous to those treated in the previous section, and differ from them only in that the reason for textual loss in the given witness is not physical damage but

scribal omission, meaning that the <gap> is in this case to be encoded with @reason="omitted" instead of @reason="lost".

5.5.12.1 Textual loss within a block-level container

As above, we offer two encoding strategies, again with a slight preference for the second:

Example 5.5.12.1.A

```
<app><lem>saṅ vṛddhah yan tan vruh
    <app>
        <lem wit="#L">iṅ</lem>
        <rdg wit="#K">riṅ</rdg>
    </app> dharma, tan vṛddhaḥ ṅaranira,</lem>
    <note type="altLem">saṅ vṛddhah ... tan vṛddhaḥ ṅaranira,</note>
    <rdg wit="#M" cause="eye-skip"><gap reason="omitted"/></rdg>
</app>
```

Display –

Text

saṅ vṛddhah yan tan vruh iṅ dharma, tan vṛddhaḥ ṅaranira,

Apparatus
^1. saṅ vṛddhah ... tan vṛddhaḥ ṅaranira,] *om.* **M** (eye-skip).
^2. iṅ] **L**, riṅ **K**, *om.* **M** (larger gap).

Example 5.5.12.1.B

```
<app><lem type="omitted_elsewhere">saṅ vṛddhah yan tan vruh
    <app>
        <lem wit="#L">iṅ</lem>
        <rdg wit="#K">riṅ</rdg>
    </app> dharma, tan vṛddhaḥ ṅaranira,</lem>
    <witDetail type="retained" wit="#L #K"/>
    <note type="altLem">saṅ vṛddhah ... tan vṛddhaḥ ṅaranira,</note>
    <rdg wit="#M" cause="eye-skip"><gap reason="omitted"/></rdg>
</app>
```

Display –

Text
saṅ vṛddhah yan tan vruh iṅ dharma, tan vṛddhaḥ ṅaranira,

Apparatus
^1. saṅ vṛddhah ... tan vṛddhaḥ ṅaranira,] transmitted in **L K,** *om.* **M** (eye-skip).
^2. iṅ] **L**, riṅ **K**, *om.* **M** (larger gap).

5.5.12.2 Textual loss across boundaries between block-level containers

Since the TEI clearly assumes that <lacunaStart/> and <lacunaEnd/> shall be used only in cases where physical damage is the reason for textual loss, but does not provide any strictly analogous pair of elements for the situation where scribal omission is the reason, we have decided to introduce the tailor-made element-attribute combinations <span type="omissionStart"/> and <span type="omissionEnd"/>.[60]

<table>
<tr><td align="center">Example 5.5.12.2.A</td></tr>
</table>

```
<p n="1">...kalaṅkaṅ saṅ hyaṅ
<app>
        <lem wit="#A #EdASD">dasasila</lem><rdg wit="#B">dasa<span type="omissionStart"/></rdg>
</app>, maya-maya ...
<app>
        <lem wit="#A">di</lem><rdg wit="#EdASD"><gap reason="omitted"/></rdg>
</app> bumi lamba, di bumi
<app>
        <lem  type="emn">parək</lem><rdg  wit="#A">paR̥<unclear>ka</unclear></rdg><rdg  wit="#EdASD">tan
        parək</rdg>
</app>.</p>


<p n="2">
<app>
        <lem wit="#A #EdASD">ini pakən</lem><rdg wit="#B"><span type="omissionEnd"/> pak·kən</rdg>
</app> uraṅ... </p>
```

Display —

text

1.1 ... kalaṅkaṅ saṅ hyaṅ dasasila, maya-maya saṅ hyaṅ dasa-marga, kaprətyaksaan na dasa-indriya, pakən ṅrətakən di bumi lamba, di bumi parək.
1.2 ini pakən uraṅ...

apparatus

^1. dasasila] **A Ed**[ASD], dasa[... **B**
^2. dasasila … (**§1.2**) ini pakən] A gap due to omission intervenes in **B**
^3. di] **A**, *om.* **Ed**[ASD], *om.* **B** (larger gap)
^4. parək] *em.*, paR̥ka **A**, tan parək **Ed**[ASD], *om.* **B** (larger gap)
^5. ini pakən] **A Ed**[ASD], ...] pak·kən **B**

---

[60]See TEI §12.1.5 and the comment on the potential for confusion inherent in the term "lacuna" made in Library of Digital Latin Texts §12.9 (Gaps in Witnesses). We have favored an approach with <span> rather than <milestone/> to avoid the need of adding the mandatory attribute @unit with the value "absent" as suggested by the TEI Guidelines §att.milestone for cases which mark the beginning of a piece of text not present in the edition.

Again (see §5.5.11), there will be no need to encode any <rdg> for the given witness in *loci* falling between the two elements <span>. The two indications "*om*. **B**" in the example are generated by the fact that the apparatus entries fall between the opening and closing <span/>, and no <rdg> for #B has been explicitly encoded.

It is possible to combine the tailor-made element-attribute <span type="omissionStart"/> for a larger textual omission with the values available for the attribute @cause (see §5.5.3). You must then add the attribute @cause with the chosen value on the parent element <rdg> in which you declare the starting point of the omission, as you would do for a smaller omission. You don't need to add the @cause on the <rdg> containing its end. The value of @cause will be reflected in the apparatus entry that makes the connection between beginning and end of the omission.

Example 5.5.12.2.B

<div type="dyad" n="118" xml:id="svayambhu_07.10">
[...]
<app><lem wit="#L #M">ṅaranya, limaṅ</lem><rdg wit="#K" cause="eye-skip">ṅaranya, <span type="omissionStart"/></rdg></app>
        kṛṣṇalam,<app><lem type="emn">samāṣa</lem><rdg wit="#L #M">sama:s·</rdg></app> ṅaranya,
[...]
</div>
<div type="dyad" n="119" xml:id="svayambhu_07.11">
[..]
pirak <app><lem wit="#L #M">ṅaranya, ikaṅ</lem><rdg wit="#K"><span type="omissionEnd"/> ikaṅ</rdg></app>
sapuluh tahil
[...]
</div>

Display —

^808. ṅaranya, limaṅ**] L M**, ṅaranya, [... **K** (eye-skip)
^809. ṅaranya, limaṅ … (**119.1**) ṅaranya, ikaṅ**] A gap due to omission intervenes in **K** caused by eye-skip
^810. samāṣa**] *em*., sama:s·**L M**, *om*. **K** (larger gap).
[...]
^815. ṅaranya, ikaṅ**] L M**, ...] ikaṅ **K**

For the time being, we only allow the use of @cause="eye-skip" for interaction with <span type="omissionStart"/> and <span type="omissionEnd"/>. Contact the XML-TEI Data Manager should you require another value of @cause.

5.5.13 Reporting scribal additions and deletions

5.5.13.1 <add> – Additions

Several possibilities are offered to indicate that text has been added to a witness. It can be done of course with a simple prose description, even though we rather recommend using semantic markup, namely the element <add>. Do not encode with <add> any segments supplied by a previous editor or to supply any elements yourself. This element should be used only for additions made in the text by the scribe or someone else who handled the manuscript prior to

an edition. Such additions are often easily identified by their placement above or below the line, or in the margin. Making use of the attribute @place is mandatory, and the following values are available:[61]

- "inline": addition made within the original text (*in textu*)
- "below": addition below the line (*subscr.*)
- "above": addition above the line (*suprascr.*)
- "top": addition in the top margin (*in mg. sup. = in margine superiore*)
- "bottom": addition in the bottom margin (*in mg. inf. = in margine inferiore*)
- "left": addition in the left margin (*in mg. = in margine*)
- "right": addition in the right margin (*in mg. = in margine*)
- "overstrike": addition made in the space where a previous string of text has been erased (*in ras. = in rasura*). This value can only be used with the element <add> when it goes hand in hand with the element <del> (see §5.5.13.2)
- "unspecified": no information on the place is currently available

If you want to specify the scribal hand making the addition, it is possible to do so using the attribute @hand. You need to have declared this hand in the <handDesc> part of the <teiHeader>, see §2.1.3.1. Then, provide its @xml:id as value for @hand. Here are two examples:

Example 5.5.12.1.A
<app>   <lem wit="#K">deśa</lem>   <rdg wit="#L">deśa <add place="below">ṅaranya</add>,</rdg> </app>
Display — ^1. deśa] **K**, deśa ⟨⟨*subscr.* ṅaranya⟩⟩, **L**.

Example 5.5.13.1.B
<app>   <lem wit="#K">deśa</lem>   <rdg wit="#L">deśa <add place="below" hand="#L_H2">ṅaranya</add>,</rdg> </app>
Display — ^1. deśa] **K**, deśa ⟨⟨H$_2$ *subscr.* ṅaranya⟩⟩, **L**.

---

[61] Some values are the same as in EGD §4.4.2. Note that all values will be displayed except if the element <add> is used within <subst>.

When the addition is marked in the witness with any non-alphanumeric mark (such as the cross-shaped called *kākapada* in Sanskrit), you may use @rend with the value "mark" to encode this. This encoding method shall apply regardless of where such a scribal mark appears (at the locus of insertion, next to the inserted text, or at both places) and of what shape the mark takes.[62] The particular shape or shapes used by a given hand in a given witness can be described in <handDesc> as well.

5.5.13.2 <del> – Deletions

Scribal deletions are to be wrapped in <del> analogously to the use of <add> for additions. In this section, deletions are taken into account only by themselves but, as the section below explains, you can also encode <del> together with <add>, when deletions and additions co-occur. To qualify how deletions have been carried out, you can use the attribute @rend with one of the following values:[63]

- "strikeout": text struck through, erased or scraped off, without it having become impossible to detect what was deleted (*ante ras.*)
- "dots": dots added next to the characters to be deleted (*exp. = expunction*)
- "ui": for the combined application (typical in Javanese and Balinese manuscripts) of vowel markers *u* and *i* under and above characters to be deleted (*ui*)

The use of the attribute @rend is optional. For the time being, no differentiation is foreseen in the display.

Example 5.5.13.2.A
<app>   <lem wit="#L">maṅaku dr̥vya, juga ya,</lem>   <rdg wit="#K">maṅaku dr̥vya, <del rend="ui">Ikamaṅkana:,</del>juga: ya:,</rdg>   <note>The canceled string was initially written due to eye-skip from the preceding paragraph.</note> </app>
Display —
^159 . maṅaku dr̥vya, juga ya,**] L**, maṅaku dr̥vya, [[*ui* Ikamaṅkana:,]] juga: ya:, **K** • The canceled string was initially written due to eye-skip from the preceding paragraph.

Note that any text marked as deleted must be legible. If the passage is in fact illegible, it should be encoded with the empty element <gap/> as per §5.5.9. However, if you want to cite a reading that presents a deleted *akṣara* which has become illegible in the deletion process, you can follow the following encoding model:

Example 5.5.13.2.B

---

[62] See EGD §4.4.2.
[63] See EGD §4.4.1.

```
<app>
  <lem type="emn">aṅalāntaranana</lem>
  <rdg wit="#L">haṅala<del><gap reason="illegible" unit="character" quantity="1"/></del>ntarana:</rdg>
  <rdg wit="#K">aṅalantarana:</rdg>
  <rdg wit="#M">aṅalantara</rdg>
</app>
```

Display —

^23 . aṅalāntaranana] em., haṅala[[1x]]ntarana: **L**, aṅalantarana: **K** , aṅalantara **M**.

5.5.13.3 Combining additions and deletions

A deletion occurring in combination with an addition may be understood as a corrective substitution process. You may deem it preferable to encode *ante* and *post correctionem* readings as per §5.5.5 in order to bring out the correspondence of the uncorrected or corrected state of the text with the readings of other witnesses in <lem> or <rdg>. However, you may also choose to encode such cases using <add> in combination with <del> and the wrapper <subst>. Here are two examples, the first involving deletion of a vowel marker yielding inherent vowel *a* as substitution:

Example 5.5.13.3.A

```
<app>
  <lem type="emn">samagrig asambevara</lem>
  <rdg wit="#A">samagrig<subst><del>ə</del><add place="below">a</add></subst>sambehara</rdg>
  <rdg wit="#B #C">samagrigə sambehara</rdg>
</app>
```

Display —

^78 . samagrig asambevara] em., samagrig⟨⟨[[ə]]a⟩⟩sambehara **A**, samagrigə sambehara **BC**

Example 5.5.13.3.B

```
<app>
  <lem wit="#A #B">parasparopasarpaṇa</lem>
  <rdg wit="#C">parasparopasa<subst><del>na</del><add place="inline">pa:</add></subst>na</rdg>
</app>
```

Display —

^270 . parasparopasarpaṇa] **AB**, parasparopasa⟨⟨[[na]]pa:⟩⟩na **C**

However, not all cases of co-occurrence of addition and deletion are interpretable in a straightforward manner. When you are not sure how a deletion interacts with an addition, it is recommended to use the elements <del> and <add> without nesting them in <subst>. When

you do use <subst>, the values of @place on <add> and of @rend on <del> will not be displayed but must still be encoded.

5.6 Reporting extraneous text segments (interpolations)

When a (group of) witness(es) has a segment of text that is absent in one or more other (groups of) witnesses and you consider the state of the text in the latter to be original, then the more expansive (group or groups of) manuscript(s) bear testimony to a process often called interpolation. This is different, in theory, from an addition for which a single scribe in a manuscript is responsible. But when you are editing a text on the basis of a small number of witnesses, it may not be possible to discern groups of manuscripts and it may hence not be possible to distinguish interpolations from scribal additions. We offer the following ways of encoding such extraneous text segments.

5.6.1 Extraneous text block with @type="interpolation"

Any text block (div, p, ab, lg) that is transmitted in all your witnesses or in a whole group of witnesses, or even only in a single witness, but which you consider to have been interpolated during a stage of the text's transmission posterior to the one you are trying to reconstruct, may be included in your edition in the suitable block-level container to which you will apply @type="interpolation". Do not furnish any @n for such interpolated elements. The display will assign to such elements the number of the previous element of the same level, followed by an asterisk.

Example 5.6.1.A
<div type="dyad" n="54">[...]</div>   <div type="interpolation">   <app><lem wit="#L"><p>ya tapva ahutaṅ, pinet kinuṅkuṅ ta ya de pradhana, ... </p>   <p>kunəṅ pvekaṅ tan tinuron, ḍaṇḍa kaṅ sinambat vrat iṅ apadu.</p></lem>   <note type="altLem">ya tapva ahutaṅ, ..., ḍaṇḍa kaṅ sinambat vrat iṅ apadu.</note>   <note>These paragraphs are absent in <ptr target="#K"/> and seem extraneous to our text.</note>   </app>   </div>

5.6.2 Extraneous text in <rdg> (optionally with <seg type="interpolation">)

A segment, that may or may not correspond to a whole semantic block, found in one or more of your witnesses, which you do not wish to display in your edited text,[64] can be recorded in the apparatus with reference to a word or words common to the accepted and to the rejected

---

[64] Note that this is different from the case of a scribal edition taking place "before our eyes" in a witness, and which would be encoded using the <add> element (§5.5.13.1). In the situation we are dealing with here, the segment is present in one or more witnesses, but you as editor decide that it was absent in the stage of the text's history that you are trying to reconstruct.

readings that serve to anchor the extraneous segment in question.[65] Such segments may be as short as a single word or stretch over multiple phrases. It is possible, but not mandatory, to flag the extraneous segment with <seg @type="interpolation">, which may be useful for querying your XML data on interpolations, though we do not foresee any special display of such segments.

<table>
<tr><td colspan="2" align="center">Example 5.6.2.A</td></tr>
<tr>
<td>

```
<app>
  <lem wit="#L">tan</lem>
  <rdg wit="#K">riṁ tan</rdg>
</app>
```

</td>
<td>

```
<app>
  <lem wit="#L">tan</lem>
  <rdg wit="#K"><seg type="interpolation">riṁ</seg>
tan</rdg>
</app>
```

</td>
</tr>
</table>

<table>
<tr><td align="center">Example 5.6.2.B</td></tr>
<tr><td>

```
<app><lem wit="#K #M">kunaṅ ikaṅ
      <app><lem wit="#L #K">pirak paṇa sapuluh</lem><rdg wit="#M">sapuluh pana pirak</rdg></app>
      <abbr>ma</abbr>,</lem>
      <note type="altLem">kunaṅ ikaṅ ... sapuluh <abbr>ma</abbr></note>
      <rdg wit="#L"><seg type="interpolation">kunaṁ Ikaṁ pirak ṣapuluḥ pana:, ma:, || ○ || <lg
      xml:lang="san-Latn" met="anuṣṭubh"><l>R̥ṇadeve vratijñante,</l> <l>pañcama siṁtam ahartti,</l>
      <l>hapanave dadviguṇiṁ,</l> <l>tan manom anuṣasaniṁ,</l></lg> ka, riṅ mavutaṁ yan masaṅketa:, yapva
      ta mityeṁ vuhus riṁ purvvaka:, daṇḍanika: de bhūpatya, pañca satapa<del><gap reason="illegible"
      unit="character" quantity="1"/></del>ṇa, kunaṁṅ apuvara haṅasi<unclear>ḥ</unclear> dvigunotama daṇḍa
      || ○ || <lg xml:lang="san-Latn" met="anuṣṭubh"><l>niptiṁ vak nityaṁ* krodaṣṭaṁ*,</l> <l>valat śobr̥ṁ*
      namaṣṭaḍḍi,</l> <l>niṣṭe maḍyamaṁ* hutamaṁ*,</l> <l>purvvakaṁ* mituva daṇḍaṁ*,</l></lg> ka, hana ta
      ya: vaṁṅ apihutaṁ, tinagiḥ pihutaṁnya, tan paveḥ vetan drəvenya, hasr̥ṁ saṁṅ apihutaṁ, maṅalap
      sadr̥venya, maṅalap histri, sunu, bhūmi, nūn pasu sakalviriṁ patik vənaṁ yata hinalap·, vetniṁ kaṁniṣṭa
      vinaṅūn mityeṁṅ ūjar,, maka:don hinira:-hira:, haṅiṅindəti, haṅlindihi, tan druḥ kaṁṅ adr̥ve vaṁṅ apihutaṁ,
      lumka:s kaṅśa:seṁ tan hambava cihna Utər· daḺm, hikya ṅaran valat sahaṣa havalat śobra, hiṅar:vakən
      vinalik rantas vitya:, mvaṁ tinibakna daṇḍa mahira<del>ntaṁ</del>n kaṁ sa:hasobraḍḍi, kunəṁ
      pinaraṅgvakna vutaṁ, lavan daṇḍanekaṁ sahaṣa, vnaṅ pasaṅana vrat niṣṭa maḍyotama, ye niṣṭa vit hutaṁ
      mvaḥ paṅamet· sinahaṣa:, niṣṭa: daṇḍa,, 5000, yen madya paṅamete, daṇḍa, 10000, hutama paṅamete,
      daṇḍa, 20000, sapaṅamete haṅsula:kna riṁ kaṁṅ avutaṅ, vaṣana, daṇḍa riṁ saṁ bhūpatya || 0 || kunaṁ
      Ikaṁ pratekaniṁ harta:, sapaṇna:, 20, limaṁ paṇa:, 2, ku, tkeṁ hartha,</seg> kunaṁ ikaṁ pirak pana
      sapuluḥ, ma:,</rdg>
      <note>Manuscript <ptr target="#L"/> inserts a substantial interpolation between two iterations of the lemma.
```

The interpolation notably contains the full text of MDhŚ 8.139 and an unidentified Sanskrit stanza. Although having stanza 8.139, with paraphrase, in this general part of the text is potentially attractive, the fact that the stanza is quoted in full is suspect, and the precise locus where it is inserted interrupts the paraphrase of

</td></tr>
</table>

> 8.137. On these grounds, we favor the hypothesis that the segment of text found only in this witness is extraneous to our text.</note>
>
> </app>

5.6.3 <surplus> – identifying small segments of extraneous text

Any smaller segment that you consider, as editor, to be inauthentic or superfluous but nevertheless need to retain in the edition, for some reason (for instance if all your witnesses have it), should be encoded as <surplus>.

Example 5.6.3.A
*lələbakna<surplus>,</surplus> mvaṅ
Display –   *lələbakna{,} mvaṅ

Example 5.6.3.B
maṅkana, <surplus><term xml:lang="san-Latn"><app><lem type="norm">dānapratibhūḥ</lem><rdg wit="#L">da:napratibhuḥ</rdg><rdg wit="#K #M">dana:pratibuḥ</rdg></app></term> ṅaranya,</surplus> yan mati ikaṅ maṅaku,
Display –   maṅkana, {**dānapratibhūḥ** ṅaranya,} yan mati ikaṅ maṅaku,

5.6.4 <secl> — misplaced text segments

The TEI Guidelines offer a further element <secl> (meaning "secluded") for "the case of an interpolation which the editor regards as genuine (i.e. written by the author in question), but out of its original place"[66] which you can't determine. In such a case, <secl> allows taking its contents out of the text flow and marking it as not belonging where it is found in the witnesses. Should this element seem convenient for transmissional phenomena that you need to encode, then please contact the XML-TEI Data Manager.

---

[66] See TEI Guidelines, §11.3.1.7.

5.7 Supplying untransmitted text or indicating your inability to supply lost text

The <supplied> element can be used to mark any segment that you wish to include in your edited text but that is not found in any of the manuscripts. Besides the specific scenario of a dependent text (like a commentary) that implies but does not cite its base text fully or at all (see §4.3), we envisage the use of <supplied> when:

1. a segment is supposed to have been present at the stage of the text's development that your edition is trying to reconstruct but assumed to have been lost or expunged in the subsequent course of transmission
2. you wish to indicate the purely editorial nature of untransmitted punctuation that you may wish to insert to help the reader, without implying that it was ever present in the transmission

It is possible to use <supplied> directly in your edition in cases where the editorial intervention does not require the creation of an apparatus entry; <supplied> can also be used inside a <lem> (and even inside an <rdg>, if a printed edition indicates it has supplied a given segment), which, in turn, is to be wrapped in <app>. In any scenario, the element must be used with the attribute @reason having one of the following values:

- "lost": to be used when text is conjecturally restored to fill a physical lacuna which affects the given locus in all witnesses
- "omitted": to be used when it is assumed that a given string as been voluntarily expunged or involuntarily omitted in the course of transmission
- "implied": to be used when a given string of text or punctuation sign, not found in any of your witnesses, is not assumed ever to have been present in the transmission of your text but somehow implied and needs to be displayed in your publication
- "undefined": to be used for instance when a printed edition indicates that the editor has supplied a segment without indicating the reason
- "unknown": to be used when you cannot determine the reason behind loss or omission

Furthermore, it is possible to indicate on which evidential basis you supply the segment by adding the attribute @evidence, with one of the following three values:

1. "internal"
2. "external"
3. "conjecture"

The first value means that there is an analogous passage elsewhere in your text that supports the hypothesis that a segment needs to be supplied in the current context; the second means

that there is such evidence but that it is external to your text. If you use @evidence="external", then you are expected to encode the parallel passage(s) in question as per §6.2.

Example 5.7.A

... sadaṅayan<supplied reason="implied">,</supplied> sajamus ...

Example 5.7.B

... vvaṅ atuha
<app>
  <lem><supplied reason="omitted" evidence="internal">dahat, rare</supplied></lem>
  <rdg wit="#A #B #C"><gap reason="omitted"/></rdg>
   <note>The two words supplied here are assumed to have been omitted in transmission due to eye-skip. Their restoration is supported by two other occurrences of the pair <foreign>atuha/rare</foreign> in our text, in §12 and §91.</note>
</app> dahat, ...

Example 5.7.C

... yan pahutaṅ,
<app>
  <lem><supplied reason="omitted" evidence="external">mvaṅ managih riṅ tan pahutaṅ</supplied></lem>
  <rdg wit="#A #B #C"><gap reason="omitted"/></rdg>
</app> adharma ṅaranya, ...

5.7.1 Reporting a locus of suspected textual loss

A particular scenario arises when you have reason to believe, against unanimous evidence of the manuscripts, that a certain amount of text has been omitted or lost in transmission at a given locus, but are unable to determine what precise segment of text is to be supplied to fill the gap. This is basically the opposite of an interpolation. Such a situation may be encoded as follows:

Example 5.7.1.A

salvirniṅ <supplied reason="omitted" evidence="conjecture"><gap/></supplied> hyaṅ ṛṣiśāsana

---

Or, if you need to render the reasons behind your assumption of textual loss explicit in a note:

Example 5.7.1.B

tan paṅilaṅakən sira iri ya,
 maṅkana liṅ <app><lem><supplied reason="omitted" evidence="conjecture"><gap/>.</supplied></lem><note>All

---

mss. have \<foreign\>atəhər\</foreign\> (spelled \<foreign\>at:hər\</foreign\>) directly following \<foreign\>liṅ\</foreign\>. This seems grammatically impossible, as \<foreign\>maṅkana liṅ\</foreign\> is always used in a possessive construction immediately followed by a designation of the speaker. We can only guess that words like \<foreign\>bhaṭāra manu\</foreign\> have been omitted.\</note\>\</app\>
\</p\>
 \</div\>\<div type="dyad" n="16"\>
\<p xml:id="svayambhu_02.16.01"\>atəhər ...\</p\>
 \</div\>

You may use "lost" or "unknown" as values of @reason (in many cases, it will be impossible to know the reason for a segment's absence). Use of @evidence is not mandatory and if you do find it useful, other values may be relevant than "conjecture". See the list of suggested values above.

5.7.2 Marking up lacunae with known meter

If you are editing a text on the basis of a *codex unicus* which is incompletely preserved, some (parts of) verse may be lost. You might wish to mark this lost text that can't be restored, but for which the prosodic pattern is known. To do so, wrap a \<gap/\> element with \<seg\>. Add the attribute @met to it and record the prosodic pattern of the lacuna (see PVG for the conventions to be used in @met). In syllabic verse, the number of lost characters can be calculated accurately. In other cases, such as moraic verse, you might need to use the attribute @precision="low".[67]

Example 5.7.2.A
\<seg met="++-+---+-"\>\<gap reason="lost" quantity="7" unit="character"/\>\</seg\>
Display – [– – ˘ – ˘ ˘ ˘ – ˘]

5.8 Encoding transposition of text

The term transposition is used here to designate phenomena of inversion of the order of words, sentences or paragraphs *vis-à-vis* the order of your edited text. It may be possible to distinguish purely physical processes of displacement of folios, leading to a witness no longer presenting the text in the proper order, from cases where witnesses show varying sequences because some person involved in transmitting the text made a change in the text's sequence, intentionally or not. (Our encoding model does not take into account the issue of intentionality.) Complex transpositions are difficult to encode with the parallel segmentation method, if they involve variants occurring in different locations. For simpler cases, such as inversion between

---

[67] EGD §5.4.4.

two segments of text following each other in close succession (§5.8.1), physical displacement of folios in an antegraph (§5.8.2) or changes in the order of lines within a stanza (§5.8.3), you can make use of the methods for encoding apparatus described so far, with the specifications we offer below. We address a slightly more complicated scenario in §5.8.4. But for cases that are even more intricate, it may be most pragmatic to decide in favor of a prose description in a <note> rather than an apparatus entry with complex TEI encoding.[68]

5.8.1 Transposition of text segments within a single block

In simple cases, i.e., as long as the transposed segments all fall within a single block of text, the usual pattern for creating an apparatus entry should be followed. The reading accepted by the editor as the lemma goes into the element <lem>, while any variant reading is recorded in <rdg> with @cause="transposition" as per §5.5.3 (example 5.5.3.B). It is possible to highlight transposed segments with <seg type="highlight"> as per §5.5.2.

Example 5.8.1.A
<app>   <lem wit="#K">ika ta saṅ prabhu, tinut <seg type="highlight">sira</seg> deniṅ rat kabeh</lem>   <rdg wit="#L" cause="transposition">Ika: ta <seg type="highlight">sira</seg> saṁ prabhū tinut deniṁ rat kabeḥ</rdg>   <rdg wit="#M">ika ta saṁ prabu <seg type="highlight">maṅkana</seg>, tinut <seg type="highlight">sira</seg> de<seg type="highlight">nikaṁ</seg> rat kabeḥ</rdg> </app>
Display —
^933. ika ta saṅ prabhu, tinut **sira** deniṅ rat kabeh] **K**, Ika: ta **sira** saṁ prabhū tinut deniṁ rat kabeḥ **L** (transposition), ika ta saṁ prabu **maṅkana**, tinut sira de**nikaṁ** rat kabeḥ **M**

5.8.2 Transposition through physical displacement

If you have determined that one or more folios must have been displaced in an antegraph of a given witness, so that the text jumps in the middle of one folio to the middle of another in a *prima facie* random manner, then you can record such textual displacement with the element <locus> using the attribute @type with the value "displacement". To document the place in the text where the reader has to flip from one folio and line to another, you must create an <app> to accommodate an <rdg> for the witness in question, and should use the attributes @from and @to for filling in the page/folio and line references. You are likely to come up with corresponding <app> entries, for where the text jumps and then jumps back, and small errors are liable to occur around the *loci* of displacement, as in the following example:

---

[68] We have taken inspiration from https://digitallatin.github.io/guidelines/LDLT-Guidelines.html (§12.12.4) on certain points in determining our encoding model.

<table>
<tr><td colspan="2" align="center">Example 5.8.2.A</td></tr>
<tr>
<td>

```
<app>
  <lem wit="#L #K">piraknya</lem>
  <rdg wit="#M">si<locus type="displacement"
  from="21r12" to="12r2"/>raknya</rdg>
</app>
```

</td>
<td>

```
<app>
  <lem wit="#K">alapənira</lem>
  <rdg wit="#L"><gap reason="omitted"/></rdg>
  <rdg wit="#M">a<locus type="displacement"
  from="12r2" to="21r2"/>pənira</rdg>
</app>
```

</td>
</tr>
<tr>
<td>Display –</td>
<td>Display –</td>
</tr>
<tr>
<td>^850. piraknya] **L K**, si[displacement from 21r12 to 12r2]raknya **M**</td>
<td>^911. alapənira] **K**, *om*. **L**, a[displacement from 12r12 to 21r2]pənira **M**</td>
</tr>
</table>

5.8.3 Transposition of the lines of a stanza

Let us now consider a case where two hemistichs appear in reverse order. In the following example, witness A shows the accepted order of the four verse lines, while witness B puts the second hemistich before the first. Except for this change in order, the text of all verse lines is strictly identical in both witnesses. For this reason, you do not need to reproduce the same text and can use the attribute @corresp on the <l> elements in <rdg>. Note that, in this case, the <lem> element must bear the attribute @type="transposed_elsewhere" and shall always contain the text in full.

It is only possible to use @corresp inside the <rdg>, not *vice versa*. The value of the attribute @corresp should refer to the @xml:id of the element to which it corresponds, prefixed by "#". An XSLT transformation process that can be launched and relaunched by the project's XML-TEI Data Manager will generate @xml:ids for all <lg> and <l> elements in our XML files.[69] Note that in the following example, the <lg> element must be added inside the <app> rather than outside, since the TEI model requires that an <lg> element must at least contain one direct child <l>, <lg> or <gap>.[70]

<table>
<tr><td align="center">Example 5.8.3.A</td></tr>
<tr><td>

```
<app>
    <lem wit="#A" type="transposed_elsewhere">
      <lg met="anuṣṭubh">
          <l n="a" xml:id="siksaguru_33a">dharmakṣetre kurukṣetre</l>
          <l n="b" xml:id="siksaguru_33b">dharmajñair bahubhir vṛte|</l>
          <l n="c" xml:id="siksaguru_33c">uñchavṛttir dvijaḥ kaś cit</l>
          <l n="d" xml:id="siksaguru_33d">kāpotir abhavat purā||</l>
      </lg>
    </lem>
    <rdg wit="#B">
      <lg met="anuṣṭubh">
```

</td></tr>
</table>

---

[69] This process can be launched or relaunched upon request.
[70] See TEI Guidelines, https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lg.html.

```
                    <l corresp="#siksaguru_33c"/>
                    <l corresp="#siksaguru_33d"/>
                    <l corresp="#siksaguru_33a"/>
                    <l corresp="#siksaguru_33b"/>
                </lg>
              </rdg>
          </app>
```

Display —
Apparatus
^1. **B** presents the lines in order cdab.

If the variation between the witnesses does not only concern the order of the verse lines but also the text of one or more lines, then encode the transposition as above, using the attribute @corresp, while encoding any variant readings of specific text segments in the usual manner, meaning that you will embed one or more <app> elements in the <app> representing the transposition. In view of this embedding, it is necessary to encode the witness(es) supporting the order of text as retained in your edition following the pattern presented in (on <witDetail type="retained"/>).

Example 5.8.3.B

```
<app>
    <lem type="transposed_elsewhere">
       <lg met="anuṣṭubh">
          <l n="a" xml:id="siksaguru_33a">dharmakṣetre
             <app>
                <lem rend="hyphenright" wit="#B">kuru</lem>
                <rdg wit="#A">giri-</rdg>
             </app>kṣetre</l>
          <l n="b" xml:id="siksaguru_33b">dharmajñair bahubhir vṛte|</l>
          <l n="c" xml:id="siksaguru_33c">uñchavṛttir dvijaḥ kaś cit</l>
          <l n="d" xml:id="siksaguru_33d">kāpotir abhavat purā||</l>
       </lg>
       </lem>
       <witDetail wit="#A"  type="retained"/>
        <rdg wit="#B">
       <lg met="anuṣṭubh">
          <l corresp="#siksaguru_33c"/>
          <l corresp="#siksaguru_33d"/>
          <l corresp="#siksaguru_33a"/>
          <l corresp="#siksaguru_33b"/>
       </lg>
       </rdg>
    </app>
```

Display —

Apparatus
^1. **B** presents the lines in order cdab.
^2. kuru-**]** **B**, giri- **A**.

5.8.4 Discontinuous transposition

Now we will consider discontinuous transposition, that is a given witness placing a text segment in a different location of the text that does not overlap with the lemma, and may even be situated in a different text block. It is possible to limit yourself to explaining the differences of position solely in a <note> (see §6.1) at the end of the <p> or <l> at which point one or more witnesses start to diverge from the accepted text. However, it is also possible to encode this kind of scenario, but it means that a single <app> element will not be sufficient to indicate the various positions in the text where the segment is found in the respective witnesses. If you do decide to encode the scenario rigorously, it means that you will have an apparatus entry that only contains a <lem> linked to another <app> that only contains an <rdg>.

The reading you are adopting as a lemma should, as usual, be encoded in a <lem> element. Whether it comes before or after the *locus* at which the transposed <rdg> is recorded in your edition, the <lem> should bear the attribute @type with the value "transposed_elsewhere" also in cases of discontinuous transposition. The <lem> affected by the transposition should be furnished with an attribute @xml:id whose pattern will be "trsp###", with an incremental three-digit number. For now, we expect you to attribute the numbers manually and don't foresee any auto-generation of those identifiers with an XSLT script, but should you think you would benefit from it, please contact the XML-TEI Data Manager of the project. This identifier should be represented with an @corresp preceded by a "#" in all the variants readings rejected by the encoder, as in the following example.

Example 5.8.4.A

```
<app xml:id="app00755">
    <lem wit="#A"  type="transposed_elsewhere" xml:id="trsp001">
        <lg met="anuṣṭubh" n="12">
            <l n="a">dharmakṣetre kurukṣetre</l>
            <l n="b">dharmajñair bahubhir vṛte|</l>
            <l n="c">uñchavṛttir dvijaḥ kaś cit</l>
            <l n="d">kāpotir abhavat purā ||</l>
        </lg>
    </lem>
    <note type="altLem">dharmakṣetre… purā||</note>
     <note>The MS A provides the stanza as the 12th, while in the MS B it is the 23rd. We have decided to keep the stanza's position as in A.</note>
</app>
[...]
```

```
<app xml:id="app00856">
    <rdg wit="#B" corresp="#trsp001" cause="transposition"/>
    <note>The stanza is number 12 in our edition.</note>
</app>
```

Display —

^755. dharmakṣetre… purā||**]** **A** (transposition) • The MS A provides the stanza as the 12th, while in the MS B it is the 23rd. We have decided to stanza's position as in A.

^856. dharmakṣetre… purā||**]** **B** (transposition, see st. 12) • The stanza is number 12 in our edition.

If you need to encode any variant readings for the segment affected by transposition, then you will face the problem of inheritance of attribute values and need to make use of <witDetail> to encode the @wit for the witness whose position you retain (compare the encoding strategy described in §5.5.6.1). In the following example, the editor keeps the position of the stanza following one witness and yet makes use of a variant reading from a witness that places the segment elsewhere.

Example 5.8.4.B

```
<app xml:id="app00755">
  <lem  type="transposed_elsewhere" xml:id="trsp001">
      <lg met="anuṣṭubh" n="12">
          <l n="a">dharmakṣetre <app><lem rend="hyphenright" wit="#B">kuru</lem>
  <rdg wit="#A">giri-</rdg></app>kṣetre</l>
          <l n="b">dharmajñair bahubhir vṛte|</l>
          <l n="c">uñchavṛttir dvijaḥ kaś cit</l>
          <l n="d">kāpotir abhavat purā||</l>
      </lg>
  </lem>
  <witDetail wit="#A" type="retained"/>
  <note type="altLem">dharmakṣetre… purā||</note>
   <note>The MS A provides the stanza as the 12th, while in the MS B it is the 23rd. We have decided to keep the
stanza's position as in A.</note>
</app>
[...]
<app xml:id="app00856">
      <rdg wit="#B"  corresp="#trsp001"  cause="transposition"/>
      <note>The stanza is number 12 in our edition.</note>
</app>
```

Display —

^755. dharmakṣetre… purā||**]** **A** (transposition) • The MS A provides the stanza as the 12th, while in the MS B it is the 23rd. We have decided to stanza's position as in A.

^756. kuru-**]** **B**, giri- **A**.

^856. dharmakṣetre… purā||**]** **B** (transposition, see st. 12) • The stanza is number 12 in our edition.

Now let's imagine the editor wants to retain a transposed segment neither in the position that it has in A nor in the one it has in B.

<table>
<tr><td colspan="1" align="center">Example 5.8.4.C</td></tr>
</table>

```
<app xml:id="app00755">
    <rdg wit="#A"  cause= "transposition" corresp="#trsp001"/>
</app>
[...]
<app xml:id="app00856">
<rdg wit="#B"  cause="transposition" corresp="#trsp001"/>
</app>
[...]
<app xml:id="app00955">
  <lem type="emn transposed_elsewhere" xml:id="trsp001">
     <lg met="anuṣṭubh" n="12">
         <l n="a">dharmakṣetre kurukṣetre</l>
         <l n="b">dharmajñair bahubhir vṛte|</l>
         <l n="c">uñchavṛttir dvijaḥ kaś cit</l>
         <l n="d">kāpotir abhavat purā||</l>
     </lg>
  </lem>
  <note type="altLem">dharmakṣetre… purā||</note>
   <note>We have decided to emend the transposed stanza to avoid the variants proposed by the witnesses A and
B because ….</note>
</app>
```

Display —

^755. dharmakṣetre… purā||**]** **A** (transposition, see st. 12)
^856. dharmakṣetre… purā||**]** **B** (transposition, see st. 1212)
^955. dharmakṣetre… purā||**]** *em.* (transposition) • We have decided to emend the transposed stanza to avoid the variants proposed by witnesses **A** and **B** because …

5.9 Marking an irreparably corrupt segment as crux

Any passage that is transmitted in unintelligible or otherwise apparently corrupt form, and which you are not able to repair by conjectural emendation, is to be marked as a so-called "crux" by wrapping it in <sic>. Make sure that readings of all witnesses are recorded in an <app> wrapped around the <sic>, using <lem> and <rdg> in the usual way, unless the crux passage is found as such without variant readings in all your witnesses. Should the witnesses not be in precise agreement, yet none evidently be more acceptable than the rest, try to represent in <sic> what seems to be the corrupt reading from which they all derive (no need for any @type on <lem> in this case), or adopt the reading of one of the witnesses (with or without applying any cosmetic normalization). Any text segment ABC wrapped in <sic> will be displayed in obeli: †ABC†.

Example 5.9.A
kaṇṭhāravam, tan paṅhvab, <app><lem wit="#L642"><sic>akadanəm</sic></lem><rdg wit="#L633">Akadaranəm·</rdg></app> tan paṅulet ri harəpən saṅ guru, <sic>sudəm</sic> phalarahasyam tan
Display — text: kaṇṭhāravam, tan paṅhvab, †akadanəm†(234) tan paṅulet ri harəpən saṅ guru, †sudəm† phalarahasyam tan apparatus: ^234 . †akadanəm†] **L642**, Akadaranəm· **L633**

5.10 Annotations in apparatus

5.10.1 <note> – Apparatus Notes

Using <note> to insert free-text observations into your edition requires the element either being enclosed in an <app> with variant readings or it being placed as the final child of an encoding block (<p>, <ab> or <lg>). Such a <note> will normally comment on some philological problem raised by a lemma, by one or more variant reading(s), or by the block of text as a whole. For more on the use of <note>, see §6.1 and §8.2. It is understood that the contents of <note> are formulated in English by default, and for this reason there is no need to use @xml:lang="eng" (by way of exception to the general principles stated in §8.10.1).

If you want to assign a <note> to a sequence of words defined as a <lem>, without assigning any @wit to the <lem>, and without encoding any <rdg> (e.g., because the remark applies to a reading shared by all witnesses), you need use the attribute @type="note" to be conformant with the encoding model (see §5.3.2 and §5.1, no. 4).

Example 6.1.C
<app><lem type="note"><app><lem wit="#L #M">vetanasya</lem><rdg wit="#K">veka:nasya</rdg></app> na ca <app><lem wit="#K #M">dānam</lem><rdg wit="#L">daniṁ</rdg></app></lem><note>All mss. indicate that this was the received Sanskrit text, which makes perfect sense, but is nevertheless problematic because it is unmetrical.</note></app>

5.10.2 <witDetail> – Witnesses Notes

The <witDetail> is defined as a "specialized note" by the *TEI Guidelines*. Unlike <note>, it can provide specific information related to a particular aspect of the witness you would like to evoke for the witness or witnesses supporting a given reading, e.g. a note explaining that the variant readings might involve two hands. It may only be used to comment on the <lem> or the <rdg> immediately preceding it and must bear the attribute @wit identifying the witness or witnesses with which the observation is concerned. The DHARMA schema prohibits use of <witDetail> other than as an empty element, though this is not a TEI constraint. We have made this choice because we so far use <witDetail> only as a solution to encode *ante* and *post correctionem* readings (see §5.5.5 or @type "ac" and "pc"), as well as specific information regarding reformulation on a larger segment of text (see §5.5.6.2 for an explanation for @type "retained" and "rejected"). This being the case, we are not inclined to allow use of this element in other contexts. However, if you feel like you might benefit from using it, contact the XML-TEI Data Manager.

6. Editorial material applied to entire blocks of text

6.1 Text critical notes

If you need to apply a note to a whole paragraph, an invocation/colophon, a stanza or a verse line of edited text, out of the context of associating parallel passages to the edited text (a topic to which we turn in the immediately following paragraphs), you can use a <note> as final child of the respective block-level container (<p>, <ab>, <lg> or <l>), to be placed after any final punctuation (and after the final </l> in the case of <lg>). The note will be displayed as part of the apparatus with a lemma generated automatically on the basis of the first and last words of the block without your needing to apply <app> and <lem>.

Example 6.1.A
kunaṅ ikaṅ strī ... śūdra <app><lem wit="#L">sākṣinya</lem><rdg wit="#K">sakṣyanya</rdg></app>, yan <app><lem wit="#L">caṇḍāla</lem><witDetail wit="#L" type="pc"/><rdg wit="#L">cacaṇḍala</rdg><witDetail wit="#L" type="ac"/><rdg wit="#K">caṇḍala mavyavahara</rdg></app>, caṇḍāla <app><lem wit="#L">sākṣinya</lem><rdg wit="#K">sakṣyanya</rdg></app>. <note>The readings of both witnesses cannot be reconciled with each other without doing violence to one or the other. It is noticeable that the reading of <ptr target="#K"/> is much more coherent than that of <ptr target="#L"/>. By the principle of <foreign>lectio difficilior potior</foreign>, <ptr target="#L"/>'s readings should probably be preferred: they are less consistent, somewhat elliptical, and compressed, whereas <ptr target="#K"/>'s readings look like an expansion and systematization.</note>

Example 6.1.B

<lg xml:id="rt_49.02" n="2">
<l n="a"><lb n="6"/>sampuh rakva ghinositan sahananiṅ bala marəka kabeh manaṅkila,</l>
<l n="b"><lb n="7"/>mvaṅ raṅkaṅ hana bhakti marpat ahaləp rinacana sinugandha lepana,<note>After this line, <ptr target="#JK"/> has: <foreign>(*malya-baṅ pinako-pa sobba mideriṅ taretepan ivaṅ*)</foreign>, the brackets clearly being meant to cancel the words in question, which are identical to 48.3c. The scribe realized he had skipped ahead.</note></l>
<l n="c"><lb n="8"/>śobha listiṅ i madhyaniṅ vataṅan énjah arəja tinulis bhinūṣaṇan,</l>
<l n="d"><lb n="9"/>mvaṅ luṅsir kanaka dravāraṅ i hiriṅnya kumuliliṅ arambatan tətē.</l><note>Stanza numbered 294 in <ptr target="#JK"/>.</note></lg>

Display –

^380. mvaṅ […] lepana,**]** After this line, **JK** has: *(malya-baṅ pinako-pa sobba mideriṅ taretepan ivaṅ)*, the brackets clearly being meant to cancel the words in question, which are identical to 48.3c. The scribe realized he had skipped ahead.
^381. sampuh […] tətē.**]** Stanza numbered 294 in **JK**.

6.2 Parallel passages: quotations, testimonia, etc.

A specific case of the above concerns parallel passages. The term "parallel" designates any text that is not in the direct tradition of the edited text. The term is used here to cover such phenomena as quotations or adaptations from your text in a later one, possible or certified sources of your text figuring in earlier texts, transmission of parts of your text in anthologies, or similar textual material in another text that may point to that text as well as yours having reworked the same pre-existing material.

Testimonia, namely quotations of a text in other ancient sources, can be considered as witnesses. If a quotation of your text attests to variant readings, the relevant information can be provided as an annotation in the critical apparatus (§5), but our discussion here will focus on how to record parallel passages separately from your critical apparatus. In terms of how your edition would be displayed this might mean a separate layer at the bottom of a page of edited text, or an appendix in a book, not to mention the broad range of possibilities that will be at our disposal in the visual presentation of an online edition.

It should be noted that encoding parallel passages is optional: the editor must decide whether listing them is necessary or useful for her/his own edition or not. It is difficult to make general recommendations regarding the choice between reporting parallel passages in the manner described here and citing them in annotation to your critical apparatus, or (in a different file, see §9.1) to your translation. But if you do opt for recording parallel passages separately, we recommend limiting yourself to reporting material that constitutes direct evidence for the transmission/reception history of your text, and/or to textual material that has had direct impact on the way you have chosen to constitute your text (i.e. your rejection, acceptance or emendation of transmitted readings).

The method outlined below enables you to document with some nuance the relationship between the text you are editing (including any base text embedded in or implied by it) and any

parallel texts. We mean hereby that a parallel passage which is similar, but not identical, to the text of your edition will not be encoded in the same way as a strictly identical one. In this connection, note that some variants (e.g. ones that merely pertain to spelling, or that involve an obvious infelicity in the edition of reference that you would weed out if you were re-editing that text) may be ignored as insignificant for evaluating the question of identity or difference between mutual parallels.

6.2.1 Associating a known parallel to a whole block of text

Any type of parallel passage can be given within the element <listApp> with a dedicated attribute @type="parallels".[71] This element uses the children <app> and <note> already familiar from the chapter Critical Apparatus, but their use for recording parallels is slightly different, among other things because it does not involve <lem>. The <listApp type="parallels"> will be the first following sibling (even in the case of an externally encoded apparatus, see §6.3) of the block containing the text for which a parallel will be reported, namely <p> or <lg>. With <quote>, when the passage in question is itself quoted from another text, the <listApp type="parallels"> can follow the <p> or <lg>, rather than the <quote> itself. This last scenario is shown in the following example:

```
Example  6.2.1.A

<div type="dyad" n="32">
   <quote type="base-text" xml:lang="san-Latn">
      <lg n="31" met="anuṣṭubh">
         <l n="a">arjayej jñānam arthāṁś ca</l>
         <l n="b">vidvān amaravat sthitaḥ |</l>
         < ln="c">keśeṣv iva gṛhītaḥ san</l>
         <l n="d">mṛtyunā dharmam ācaret ‖31‖</l>
      </lg>
      <listApp type="parallels">
         <app>
            <note corresp="txt:NarSm_04.12"/>
            <note corresp="txt:ViSm_03.11"/>
         </app>
      </listApp>
   </quote>
   <p>mataṅnya deyanika saṅ meṅət, apagəh kadi tan kneṅ pāti, lviraniran paṅarjana jñāna, artha, kunaṅ yan paṅarjana dharma, kadi katona rumaṅgut mastakanira, ta pva ikaṅ mṛtyu denira, ahosanā palayvana juga sira.</p>
</div>
```

As shown in the example, the element <listApp type="parallels"> shall be structured internally with at least one <app> element, which will, in turn, contain one or more <note> elements; you shall use one <note> element per parallel. Note that, while the above example shows

---

[71] Note that we have decided not to follow TEI guidelines §16.1.2, although its model has been applied in the SARIT project.

@type="base-text" on <quote>, the same structure can be used — without that attribute — for quotations of preexisting textual material which do not fall in the category of base text interacting with dependent text.

Example 6.2.1.B

```
<quote xml:lang="san-Latn">
   <lg met="anuṣṭubh">
       <l>// 0 // <app><lem wit="#B">yatīnāṁ</lem><rdg wit="#A #C">ya tana:ṁ</rdg></app> putra yat dravyaṁ,
yat kiñ cit <app><lem type="norm">pustakādikam</lem><rdg wit="#A #C #B">pustakaḍikaṁ*</rdg></app>,</l>
     [...]
   </lg>
 <listApp type="parallels">
       <app>
          <note corresp="txt:..."/>
          <note corresp="txt:..."/>
       </app>
     </listApp>
</quote>
```

However, outside a quoted passage, the parallels are indeed at the same structural level than the <p> or <lg> elements.

Example 6.2.1.C

```
<p xml:id="siksaguru_11.01"><term xml:lang="san-Latn">kurvanti
       <app><lem type="emn">kṛtāyām</lem><rdg wit="#A">kata ya</rdg><rdg wit="#B">kata yan</rdg><rdg
wit="#C">tata yan</rdg></app> dharma</term>, gavayakəna sakagavayan saṅ hyaṅ dharma, lvirnya ta nihan,
<term xml:lang="san-Latn">dāna</term>, <term xml:lang="san-Latn">śīla</term>,
<term xml:lang="san-Latn"><app><lem type="emn">kṣānti</lem><rdg wit="#A">yanti</rdg><rdg wit="#B
#C">kyanti</rdg></app></term>, <term xml:lang="san-Latn">vīrya</term>, <term xml:lang="san-Latn"><app><lem
type="emn">dhyāna</lem><rdg wit="#A #B #C">byana</rdg></app></term>,
<term xml:lang="san-Latn">prajñā</term>,
<term xml:lang="san-Latn">karuṇā</term>, <term xml:lang="san-Latn">mudita</term>, <app><lem
type="norm"><term xml:lang="san-Latn">metrī</term></lem><rdg wit="#A #C">metri</rdg><rdg
wit="#B">mes·tri</rdg></app>, <term xml:lang="san-Latn">upekṣā</term>.
       </p>
<listApp type="parallels">
       <app>
         <note corresp="txt:KAS.10§17">KAS 10§17: <foreign>ikaṅ metrī karuṇā muditā upekṣā, ya tikā sinaṅguh
caturpāramitā ṅaranya. papupulni caturpāramitā mvaṅ ṣaṭpāramitā, lvirnya: dāna, śīla, kṣānti, vīrya, dhyāna,
prajñā, metrī, karuṇā, muditā, upekṣā. ya tikā sinaṅguh daśapāramitā ṅaranya, ya tikā matattva
pañcadevī</foreign>.</note>
       </app>
</listApp>
<p xml:id="siksaguru_11.02">kaliṅanya, [...]</p>
```

It is also possible to apply a typology of parallels, in which case you will differentiate <app> elements by their attributes @type. This will make it possible, for instance, to sort your parallel texts in order to enable multiple display scenarios. However, since we do not foresee this option being used much, the value of @type is left to the encoder's discretion. Please contact the XML-TEI Data Manager if you want to use this option.

Example 6.2.1.D

```xml
<listApp type="parallels">
          <app type="imitation">
                  <note corresp="txt:NarSm_04.12">
                          <lg>properly encode the parallel stanza here</lg>
                  </note>
          </app>
          <app type="quotedby">
                  <note corresp="txt:KKT_07.111">
                          <lg>properly encode the parallel stanza here</lg>
                  </note>
          </app>
          <app type="adaptation">
                  <note corresp="txt:MaBh_07.59">
                          <lg>properly encode the parallel stanza here</lg>
                  </note>
                  <note corresp="txt:Ram_04.12">
                          <lg>properly encode the parallel stanza here</lg>
                  </note>
          </app>
  </listApp>
```

6.2.2 <note> — Empty or with Contents

The content of the element <note> itself varies depending on the encoding context. On the one hand, it can be left empty if you are using one of the attributes to provide a link toward an external resource (§7.2.3) containing the relevant segment of parallel text. Several scenarios should be considered here:

1. you want to quote the parallel text by retrieving it automatically in display from an electronic edition of the text without actually duplicating it in your file — *for the time being, this is only a theoretical option that cannot be implemented yet*;
2. you want to quote the parallel text by duplicating it into your file;
3. you just want to display a textual reference without actually quoting the parallel text.

In the example that follows, there are two identical parallels for the passage in question and you, as editor, think it would be redundant to display their readings as they are identical to the relevant segment of the text you yourself are editing.[72]

Example 6.2.2.A

```
<listApp type="parallels">
        <app>
                <note copyOf="txt:MDhŚ_08.00"/>
                <note copyOf="txt:NarSm_04.12"/>
        </app>
</listApp>
```

Even if you do not need to quote the parallel text, you may wish to comment on it and in that case your <note> shall no longer be empty by contain your free-text discussion of the parallel:

Example 6.2.2.B

```
<div type="dyad" n="4">
   <quote type="base-text" xml:lang="san-Latn">
     <supplied reason="implied">
       <lg met="anuṣṭubh">
         <l n="ab">hiṁsāṁ yaḥ kurute kaś cid deyaṁ vā na prayacchati |</l>
         <l n="cd">sthāne te dve vivādasya bhinne 'ṣṭādaśadhā punaḥ ||</l>
       </lg>
     </supplied>
     <listApp type="parallels">
       <app>
         <note>This stanza is absent from the constituted text of <bibl><ptr target="bib:Olivelle2005_01"/></bibl>,
though it is referenced in his app. crit. for one manuscript. The stanza occurs in ed. <bibl><ptr
target="bib:Mandlik1886_01"/>Mandlik</bibl> 8.4 with reading <foreign>bhinno</foreign>, where our text implies
<foreign>bhinne</foreign>.</note>
       </app>
     </listApp>
   </quote>
   <p>hiṁsāṁ [...]</p>
</div>
```

On the other hand, the <note> element can be given contents by providing the parallel text when this cannot be retrieved automatically and/or by giving complementary indications through a further embedded <note>. A simple option is to encode the parallel passage as a string of plain text, wrapped only in <foreign> (§8.10.2).

---

[72] For the use made here of @copyOf, see §6.2.3.1.

<div style="border:1px solid">

Example 6.2.2.C

```
<listApp type="parallels">
        <app>
        <note>Cf. Vṛh 61: <foreign><term xml:lang="san-Latn">akrodha</term> ṅaranya tan bvat sərəṅən, <term
xml:lang="san-Latn">guruśuśrūṣā</term> ṅaranya bhakty aguru, <term xml:lang="san-Latn">śauca</term>
ṅaranya nitya majapa maradina śarīra, <term xml:lang="san-Latn">āhāralāghava</term> ṅaranya tan abvat iṅ
pinaṅan, <term xml:lang="san-Latn">apramāda</term> ṅaranya tan paləhpaləha, pəṅpəṅən ikaṅ hurip sādhana
niṅ magavaya yogasamādhi, hayva hinələmhələm, gavayakəna tekaṅ sādhana, sādhana ṅaranya ikaṅ
yogamārga, makalarapan daśaśīla, ikaṅ daśaśīla umaṅunakən ikaṅ yoga</foreign>.</note>
        <note>Cf. VS.04: <foreign><term xml:lang="san-Latn">akrodha</term> ṅaraniṅ tan kataman srəṅən.
<term xml:lang="san-Latn">guruśuśrūṣā</term> ṅaraniṅ lot umulahakən siddhaniṅ svakāryaniṅ guru, gurubhakti
kaliṅanya riṅ maṅkana, makanimitta hyuniran ruməṅvakən sarinahasyaniṅ varahvarah saṅ guru. <term
xml:lang="san-Latn">śauca</term> ṅaraniṅ nityaśah mācamana sūryasevana ṅarcana ri bhaṭāra. <term
xml:lang="san-Latn">āhāralāghava</term> ṅaraniṅ tan baraṅbaraṅ iṅ pinaṅan, <term
xml:lang="san-Latn">apramāda</term> ṅaraniṅ tan paləhpaləh, an aṅabhyāsa ri saṅ hyaṅ kabhujaṅgan. ika ta
kalima niyamabrata ṅa, liṅ bhaṭāra śiva</foreign>.</note>
        </app>
 </listApp>
```

</div>

But if it seems convenient (e.g., because you can copy-paste from a text edition already available in TEI encoding) or useful for some other purpose, you may encode any parallel verse text that you wish to include in your <listApp> using <lg> with as many <l> as needed (see §3.4). If the parallel is in prose, you must use as many elements <p> or <ab> as necessary (see §3.3). In addition, you can add an embedded <note> to highlight specific features of the quoted passage. As you see, the embedded <note> element might also contain bibliographic references, encoded as per §8.4.

<div style="border:1px solid">

Example 6.2.2.D

```
<listApp type="parallels">
        <app>
                <note copyOf="txt:MDhŚ_08.004"/>
                <note corresp="txt:MDhŚ_08.058">
                        <lg>properly encode the parallel stanza here, using as many <l> as needed</lg>
        <note>The paraphrase suggests that the received reading was <foreign>bandhyo</foreign>. But no such
variant is reported in <bibl><ptr target="bib:Olivelle2005_01"/></bibl>.</note><!-- end of the embedded note -->
                </note><!-- end of the embedding note -->
        </app>
</listApp>
```

</div>

As shown in this example, the scenarios of an empty and a non-empty <note> can be combined within the same <app>.

References can be provided using several solutions in the context of <note>, especially considering that our use of the "txt:" prefix is currently theoretical, as explained in §7.1.3. You

can use <ptr/> or <ref> elements as described in §7.1.3.3 for citing text or in §7.2 for internal links as shown in the following example.

Example 6.2.2.E
<listApp type="parallels">   <app><note corresp="txt:SMG_09.02">Cf. <ref target="DHARMA_CritEdSasanaMahaguru.xml#smg_09.02">SMG 9.2</ref>.</note>   </app> </listApp>

6.2.3 <note> — Attributes

To qualify the nature of the parallelism, a set of attributes is available for use on the <note> for the parallel in question. The values for such attributes should always be built with the prefix "txt:" as described in §7.1.3.[73]

1. @sameAs or @copyOf: both of these attributes are used to identify identical content. See §6.2.3.1 below for further details.
2. @corresp: points to elements that correspond to the current element in some way but are not necessarily identical. See §6.2.3.2 below for further details.

The potential uses of such notation include the possibility of making concordances between other texts and yours. In the following paragraphs, we will furnish more detail on when and how to use @sameAs, @copyOf and @corresp.

6.2.3.1 @sameAs and @copyOf — Declaring an identical parallel passage

Two attributes can be used to express the identity of content: @sameAs and @copyOf. The former allows you to represent not only the fact that a segment of another text is identical in reading to the current segment of your text but also to reproduce the identical segment for analytical purposes (e.g. if its reading has been an argument for your choice between or even against transmitted readings of your text), while the latter enables you to represent the simple fact that identical contents is found in a parallel text, without repeating the textual segment in question.

Let us start with an example of @copyOf:

Example 6.2.3.1.A
<listApp type="parallels">

---

[73] Although it is allowed by the TEI, we have decided not to use the attribute @source in this context, since it does not allow qualifying the nature of parallels.

```
            <app>
                    <note copyOf="txt:MDhŚ_08.004"/>
            </app>
</listApp>
```

You should use @sameAs if the parallel passage pointed to with this attribute is repeated in the current <note> (and, if necessary, commented upon in an embedded <note>):

```
<listApp type="parallels">
        <app>
                <note sameAs="txt:MDhŚ_08.004"><!-- Manu08-04 from GRETIL -->
                        <lg>
                                <l n="a">teṣām ādyam ṛṇādānaṁ</l>
                                <l n="b">nikṣepo 'svāmivikrayaḥ |</l>
                                <l n="c">saṁbhūya ca samutthānaṁ</l>
                                <l n="d">dattasyānapakarma ca ||</l>
                        </lg>
                </note>
                <note>if relevant, some comment here</note>
        </app>
    </listApp>
```

6.2.3.2 @corresp — Declaring a non-identical parallel text

Any correspondence other than that of identity between a textual segment of your edition and a segment of another text can be established through the attribute @corresp. This attribute implies nothing about the degree or the nature of the correspondence other than that the two segments match without being strictly identical. If they were, you would use @sameAs or @copyOf, as explained earlier in §6.2.4. As shown in several of the examples above, you can record any observations on the correspondence you have marked up by adding a <note>.

6.2.4 Associating a parallel to a segment of a stanza or paragraph

Besides the method of recording parallel passages which matches a parallel to an entire <p> or <lg> of text, discussed so far, it is also possible to match a parallel passage to a smaller segment. Two solutions are offered here, the choice between them depending on how complex it will be to encode the delimitation of the segment in question.

6.2.4.1 Declaring a parallel for a segment within a text block

First, you can identify the segment for which there is a parallel using <lem> in an <app> within <listApp>, still to be declared (cf. §6.2.1) as the first sibling of the containing <p> or <lg>.

```
<p xml:id="skk_01.02">ini saṅ hyaṅ dasa-krəta, kundaṅən uraṅ reya, asiṅ nu dek nañjərkən saṅ hyaṅ sasana
krəta, pakənən həbəl hirup⟨,⟩ həbəl ñovana, jadyan kuras, jadyan tahun, dəgdəg tañjər, jaya praṅ, ñovana na uraṅ
reya.</p>
<listApp type="parallels">
        <app>
            <lem>jadyan kuras, jadyan tahun</lem>
                <note corresp="txt:PNJ_129"><foreign>jadyan kuras jadyan tahun</foreign>.</note>
                <note  corresp="txt:TB_35v"><foreign>pada  ñovana,  dadian  kuras,  dadian  tahun,  dadian
kaneneh</foreign>.</note>
                <note corresp="txt:MSJ_21"><foreign>maṅka jadiyan kuras jadiyan tahun</foreign>.</note>
                <note corresp="txt:LK_137"><foreign>tətəṅər jadian tahun, saṅ atuaṅ dina taraṅ, tətəṅər jadian
kuras, putər kuruṅ dina iruṅ</foreign>.</note>
        </app>
</listApp>
```

The correspondence between the contents of <lem> and a segment of the text is not encoded in
any way and needs to be made by the reader of the edition. For this reason, the <lem> does not
need to be identical to the corresponding string in the body of your <p> or <lg>: you can
truncate it or modify it in any other way that seems useful (in the manner of <note
type="altLem"> in §5.3.1.1). As before, it is possible to assign a category to the parallelism with
an attribute. This method is recommended if the segment falls within a relatively short <lg> or
<p> and the reader will easily be able to match any lemma between the body of the text edition
and the display of parallels for that <lg> or <p>.

6.2.4.2 Declaring a segment of text using <anchor>

If you need to match your parallels with a longer paragraph (or stanza), or even across
structural boundaries, it may become difficult for you to match the contents of <lem> to the
relevant segment of edited text. In this case, we recommend that you use <anchor> elements at
the start and end of the segment and refer to these anchors in the <app> using the attributes
@from and @to, respectively for the beginning and the endpoint of the lemma.

Example 6.2.4.2.A

```
<p>kalaṅkaṅ  saṅ  hyaṅ  dasasila,  <anchor  xml:id="anchor001"/>maya-maya  saṅ  hyaṅ  dasa-marga,<anchor
xml:id="anchor2"/> kaprətyaksaan na dasa-indriya, pakən ṅrətakən di bumi lamba, di bumi parək.
 </p>
<listApp type="parallels">
            <app from="#anchor001" to="#anchor002"><note sameAs="txt:AmGu_34"/></app>
</listApp>
```

As before, it is possible to comment on the parallelism with a further <note>.

Example 6.2.4.2.B

```
<p>kalaṅkaṅ saṅ hyaṅ dasasila, <anchor xml:id="anchor001"/>maya-maya saṅ hyaṅ dasa-marga,<anchor
xml:id="anchor2"/> kaprətyaksaan na dasa-indriya, pakən ṅrətakən di bumi lamba, di bumi parək.</p>
<listApp type="parallels">
                <app from="#anchor001" to="#anchor002">
                        <note>The same words also appear in AmGu.34.</note>
                </app>
</listApp>
```

Please follow the numbering pattern for @xml:id shown above, raising the number of each new <anchor> by 1 as you advance in your file, formatting it as a 3-digit number. Remember that you may never declare two identical @xml:id in a single file, and that Oxygen will show your file to be invalid if you do, mistakenly, reproduce a value that is already taken.

6.2.5 Associating parallel texts that are implied but unknown

As discussed in §4.6, if you expect the existence of a parallel text (for instance because your author explicitly states that he is quoting another work) but no actual quotation from and/or reference to a source text can be given, this shall be marked in a plain-text note:

Example 6.2.5.A
``` <div type="dyad" n="...">` `   <p n="1">ACTUALLY TRANSMITTED TEXT.` `   <listApp type="parallels">` `   <app>` `   <note>your note stating that parallels are expected but unknown</note>` `   </app>` `   </listApp>` `   </p>` ` </div>` ```

Note that no attributes should be given for the element <note> in this case.

6.3 Encoding an external apparatus stanza-by-stanza

If your text is rigorously structured in verse form and you prefer not to adopt the parallel segmentation method for encoding your critical apparatus, you can follow a structural approach analogous to the one described in the section §6.2: each block of text will be immediately accompanied by its associated apparatus, wrapped in <listApp> with @type="apparatus". This <listApp> element should be the very first following sibling of the element (<lg>, <p> or <ab>) in which the segment of text is wrapped. Inside <listApp>, each apparatus entry should be encoded in a <app> element, on which you may provide a specification of the lemma's locus in

the edition using the attribute @loc.[74] Do so by using the values given in @n attributes for your verse lines as in example 133. Copy the text of each lemma into <lem> and use the attribute @wit in the manner described in §5.4. The lemma declared for the <lem> element should normally be a copy of the relevant string inside the block of text.[75] Variant readings should be encoded in an <rdg> element, as per §5.5, using the attribute @wit in the same way as in <lem>.

Example 6.3.A

```
<lg n="1" met="āryā">
    <l n="ab">śaśadharasadṛśasaroje alake devākṣayānake caraṇau</l>
    <l n="cd">sitarajabhāsvaravimalau śambhor nityaṁ [ ] pāntu janān
    </l>
  </lg>
    <listApp type="apparatus">
      <app loc="ab">
        <lem wit="#msB #msD #msE #msF">alake</lem>
        <witDetail wit="#msF" type="pc"/>
        <rdg wit="#msA">a<unclear>lake</unclear></rdg>
        <rdg wit="#msC">a<unclear>la</unclear>ke</rdg>
      </app>
      <app loc="cd">
        <lem wit="#msA #msB #msC #msE #msF">pāntu janān</lem>
        <witDetail wit="msF" type="pc"/>
        <rdg wit="#msD">pāntu janā<gap reason="lost" quantity="1" unit="character"/></rdg>
        <rdg wit="#msF">pāta janān</rdg>
        <witDetail wit="msF" type="ac"/>
      </app>
    </listApp>
```

Example 6.3.B

```
<ab type="speaker">anarthayajña uvāca</ab>
    <listApp type="apparatus">
      <app loc="uvaca">
        <lem wit="#msCa #msCb #msCc #msNa #msNb #msNc #Ed">anarthayajña uvāca</lem>
        <witDetail wit="msNa" type="pc"/>
          <rdg wit="#msNa"><gap reason="omitted"/></rdg>
          <witDetail wit="msNa" type="ac"/>
      </app>
    </listApp>
  <lg n="3.13" met="anuṣṭubh">
```

---

[74] The attribute @loc isn't mandatory, but it is recommended for verse lines <l> to earn some processing time.
[75] Though the same exceptions can be made here as those outlined in EGD §9.1.3.

```
    <l n="a">atisaṁśayakaṣṭaṁ te</l> <l n="b">pṛṣṭo 'haṁ dvijasattama</l>
    <l n="c">durvijñeyaṁ manuṣyais tu</l> <l n="d">devadānavapannagaiḥ</l>
  </lg>
<listApp type="apparatus">
      <app loc="a">
        <lem wit="#msCb #msNa #msNb #msNc">atisaṁśayakaṣṭaṁ te</lem>
        <rdg wit="#msCa">atiśaṁsa<unclear>ya</unclear>kaṣṭan te</rdg>
        <rdg wit="#msCc #Ed">atiśaṁsayakaṣṭam me</rdg>
      </app>
      <app loc="b">
        <lem wit="#msCa #msCb #msNa #msNb #msNc #Ed">dvijasattama</lem>
        <rdg wit="#msCc">ca dvijottamaḥ</rdg>
      </app>
      <app loc="c">
        <lem wit="#msCa #msCb #msNa #msNc">°jñeyaṁ</lem>
        <rdg wit="#msCc #msNb #Ed">°jñeya</rdg>
      </app>
      <app loc="c">
        <lem wit="#msCa #msNa #msNb #msNc #Ed">manuṣyais tu</lem>
        <rdg wit="#msCb">manuṣaiś ca</rdg>
        <rdg wit="#msCc">maṇukṣe<unclear>ptu</unclear></rdg>
      </app>
    </listApp>
```

7. Referencing and cross-referencing

7.1 Links based on prefixes

It will happen often that you need to encode some piece of information that is stored outside of your TEI-XML working file. In order to facilitate project-wide consistency in storing and citing this kind of information, we have decided to create three different reference lists, containing relevant data on:

1. DHARMA project members ("part:");
2. Bibliographic references ("bib:");
3. Text editions, both machine-readable and non machine-readable ("txt:").

The declarations "part:", "bib:" and "txt:" are shortcuts to the files that list information of the these types. In brief, they work as pointers to the relevant external reference lists. In the following paragraphs, we give more detail regarding their respective functions. The correct path to the external files for participants, bibliographic resources, and text editions are provided in our

encoding template, in the <encodingDesc> within the element <listPrefixDef> as previously explained (§2.2).

7.1.1 part: — Referring to other DHARMA team members

The prefix "part:" stands for "participants" and works as an abbreviated reference to the file that lists the participants of the project, i.e. "DHARMA_IdListMembers_v01", which you can find in the project documentation folder in GitHub.[76]

Within this reference file, each participant is encoded through the element <persName> bearing a specific @xml:id. Each @xml:id consists of four letters joined together, all lower-case: in simple scenarios, the two initial letters of a person's first name joined to the two initial letters of her/his surname, will form the identifier, but it is also possible to form identifiers from initials according to another pattern. For instance, "dogo" stands for Dominic Goodall and "zapa" for Zakariya Pamuji Aminullah, while "amwb" stands for Amandine Wattelier-Bricout.

You will need to point to the values of these @xml:id's for instance if your digital edition is the fruit of multiple hands and, within your <fileDesc>, you need to declare who the editors are (see §2.1.1.2). Another situation in which these identifiers need to be used is when, in the <revisionDesc>, you want to signal who made certain changes to the text (see §2.2.3). Please bear in mind that, in order to correctly point to the relevant information, the value of the attribute you use (e.g. @ref, @who, etc.) will have to be built with the prefix "part:" followed by the identifier of the person.

7.1.2 bib: — Referring to items in the DHARMA Zotero Group Library

The prefix "bib:" stands for "bibliographic reference" and works as an abbreviated link to an item in the DHARMA project's Zotero group library. These bibliographic data are not encoded in an xml list, by contrast with our use of the prefixes "part:" and "txt:". When you want to point to a specific bibliographic item from the common library, you must use its short title as declared in the Zotero item itself.[77]

7.1.3 txt: — Referring to editions of other texts

Reference to editions of other texts may need to be made more or less frequently in your own edition, for instance if you want to cite parallels to given segments of your digital edition or simply to justify some philological decision. When such a reference is made, you may be content with a simple abbreviation plus numerical sequence to indicate the text you are referring to, and the passage in it, but it may also be necessary to cite the words of the passage referred to, and in the latter case it might be convenient to be able to retrieve the passage to the cited automatically. In the context of a collective research project, it is desirable for the same text editions to be used whenever the given text needs to be cited, and for the titles of texts to be referred to by a uniform set of abbreviations and with a uniform reference pattern. Such uniform indications of passages in other editions are here called "canonical references" and are built

---

[76] Available in the project-documentation repository.
[77] For the project's use of Zotero's Short Title field, see the ZG §4.10.

with the prefix "txt:" standing for "text edition". After this prefix shall follow the canonical reference that itself consists of two parts. The first part (e.g. MaBh for *Mahābhārata*) is the main identifier of the record and must match a declared @xml:id in the project's authority file for text editions (see §7.1.3.1). The second part provides the textual locus or range with reference to the numerical position of structural elements of the edition.

7.1.3.1 Entries in the txt: authority file

In the authority file,[78] each text is encoded within a <bibl> element, bearing a specific <abbr type="siglum"> and, built on the siglum defined therein, an @xml:id for the <bibl> entry as a whole.

<div style="border:1px solid black">

Example 7.1.3.1.A

```xml
<bibl corresp="bib:Olivelle2005_01" xml:id="EdMaDhaSa2005">
                <abbr type="siglum">MāDhŚā</abbr>
                <title>Mānava-Dharmaśāstra</title>
                <ptr target="https://raw.githubusercontent.com/sarit/SARIT-corpus/master/manusmrti.xml"/>
 </bibl>
<bibl corresp="bib:Acri2017_01" xml:id="EdDhPa2017">
        <abbr type="siglum">DhPāt</abbr>
        <title>Dharma Pātañjala</title>
        <ptr
target="https://raw.githubusercontent.com/erc-dharma/tfd-nusantara-philology/master/editions/DHARMA_CritEdDh
armaPatanjala.xml"/>
        <msIdentifier>
                <settlement>Berlin</settlement>
                <repository>Staatsbibliothek</repository>
                <idno>MS Schoemann I-21</idno>
        </msIdentifier>
 </bibl>
```

</div>

As the example shows, within each <bibl>, you can encode the following types of information:

1. the existence of a digital representation of the text edition (although we show only examples of .xml format, digital representations in .txt or .html formats can be referred to in analogous fashion)
2. the relevant Zotero item for the published edition of reference
3. the manuscript on which the edition of reference is based (an option that should be used only if the edition of reference is based on a known *codex unicus*)

In defining the siglum, you are free to use diacritical marks. Although we recommend constructing sigla by using the first two letters of each element of the text's title, and a capital letter on the first letter of each segment, we recognize that it will be very hard across a large

---

[78] Available in the project-documentation repository.

scholarly community to arrive at a consensus on how to form sigla, and that the choices made in the example will look strange to many team members.[79]

The @xml:id will then be constructed using the same siglum, but stripped of any diacritical marks, with the letters "Ed" as prefix, and with the year in which the edition was published as suffix. If you need to encode more than one edition per text, you will need to create a separate <bibl> for each edition.

A <ptr/> element can be provided for the edition if a digital representation thereof is available, and it will be a requirement to furnish such a <ptr/> element if any automated text retrieval is envisaged (see §7.1.3.3).

7.1.3.2 Quoting entries from the txt: authority file

You will need to point to the values of the @xml:id of these bibliographic items when encoding in your xml file the reference to the given text edition, using either the element <note> bearing the attributes @sameAs, @copyOf, @corresp (see §6.2.3) or the element <ptr/> (§7.1.3.3). The values of these attributes should always be prefixed by "txt:", and end with the textual reference separated from the @xml:id by an underscore, as shown in the following example.

Example 7.1.3.2.A

```
<listApp type="parallels">
        <app>
                <note sameAs="txt:EdMaDhSa2005_08.069"/>
        </app>
</listApp>
<!-- this example shows a case where you are content with the structure of the digital representation and don't
need to specify anything more -->
<listApp type="parallels">
        <app>
                <note sameAs="txt:EdMaDhSa2005_08.069-08.070"/>
        </app>
</listApp>
<!-- this example shows a case where you are citing a parallel that is two consecutive stanzas long. Any range is
expressed with a hyphen separating the full expression of the starting and ending points -->
<listApp type="parallels">
  <app>
    <note sameAs="txt:EdMaDhSa2005_08.069.cd txt:EdMaDhSa2005_08.080.ab"/>
  </app>
</listApp>
<!-- this example shows a case where you are citing a parallel that is discontinuous, e.g. the second half of a
stanza 8.69 plus the first half of a stanza 8.80
-->
```

---

[79] For the moment, we offer no solution that would allow team members to display sigla other than those defined in <abbr> according to the pattern recommended here, or various team members each to define their own <abbr> for the same text title. But it can probably be done by removing the @xml:id from the containing <bibl> and defining a separate @xml:id in each respective <abbr>. The pattern for forming xml:id's will then need to be redefined.

```
<listApp type="parallels">
        <app>
<note sameAs="txt:EdMaDhSa2005_08.070">
<bibl><ptr target="bibl:Olivelle2005_01"/>
<citedRange unit="chapter">8</citedRange><citedRange unit="stanza">69</citedRange>
</bibl>
</note>
        </app>
</listApp>
<!-- this example shows a case where the stanza numbering of the digital resource diverges from that of the
printed edition -->
<listApp type="parallels">
        <app><note sameAs="txt:EdAmGa1981_03">
<bibl><ptr target="bibl:Atja+SalehDanasasmita1981_01"/>
<citedRange unit="page">1</citedRange><citedRange unit="line">6</citedRange>
</bibl>
</note>
</app>
</listApp>
<!-- this example shows a case where the digital resource is structured with xml elements while the corresponding
printed edition needs to be cited by page and line numbers -->
```

Any locus or sequence of text should be expressed in two parts separated by an underscore.[80] The first part consists in the @xml:id of the text edition, while the second part defines a locus or a sequence (expressed as starting point plus ending point separated by a hyphen). This second part needs to represent the structural pattern of the chosen text edition. For instance, if you want to reference stanzas 29 and 30, in chapter 12 of book 1 of the *Mahābhārata,* the reference would be MaBh_01.12.29-01.12.30. The sequence is to be given in its full length and may not be shortened. This is to ensure that ranges across structural boundaries are handled efficiently.

Thus far the theory of our system. In reality, as the previous example with various scenarios suggests, there is probably no limit to the potential complexity of references that you may need to make. For this reason, the system proposed above with location references after an underscore is only an imperfect prototype. While we recommend using the above system as much as possible, your files will also validate if you limit yourself to citing the edition's @xml:id and omit the underscore and numerical specification. However closely or less closely you follow the above encoding model, it will in all cases be necessary to spell out in human-readable manner what you wish to appear in display.

Example 7.1.3.2.B

---

[80] This separator is chosen in order to comply with some of the recommendations established by Canonical Text Services (CTS). This is a protocol to identify texts or fragments of text following the classical notions of "work" and "citation" used in the humanities with the aim to make them retrievable through a network service. It was built by Christopher Blackwell and Neel Smith as part of their Homer Multitext project.

```
<listApp type="parallels">
  <app>
        <note sameAs="txt:EdMaDhSa2004_08.069">MāDhŚā 8.69</note>
        <note corresp="txt:EdDhPa2017">DhPāt (p. 124) read the same stanza with variant
<foreign>guru</foreign> instead of <foreign>ācārya</foreign> in the last <foreign>pāda</foreign>.</note>
  </app>
   <app>
        <note corresp="txt:SHP_5b">Cf. SHP 5b: <foreign>ika sarva janma, ikaṅ janma tumuvuh, janma triyak,
janma mrəga pasu paksi, janma mavoṅ, janma siṅ voṅ, vastuniṅ voṅ, samaṅkana kvehiṅ janma pamətuanikeṅ
lanaṅ vadon</foreign>. Possibly <foreign>mavoṅ</foreign> should be emended to <foreign>voṅ</foreign> and
vastuniṅ to <foreign>vastu siṅ</foreign>.
        </note>
  </app>
  <app>
        <note corresp="txt:SMG_21.1">
Cf. <ref target="DHARMA_CritEdSasanaMahaguru.xml#sasanamahaguru_21.01">SMG 21.1</ref>: <foreign>ini
catur-rahayu ṅaranya, silokanya nihan, mas pirak komala hintən, kaliṅanya, mas ta ma ṅaranna, sabda tuhu təpət
bənər, byakta pañcaksara, pirak ta ma ṅaranna, ambək krəta, yogya rahayu, komala ta ma ṅaranna, gəiṅna caaṅ
padaṅ, heraṅ ləga loganda aṅən-aṅənna, hintən ma ṅaranna, caṅciṅ səri, səmu imut, səmu guyu, rame
ambək.</foreign>.</note>
      </app>
</listApp>
```

On the use of <ref> in the last scenario, see §<u>7.2.2.1</u>.

7.1.3.3 Shortcuts for citing texts
If you need to cite a text declared in the authority list in any free-text section, such as in a <p> or in a <note>, you can use an empty tag as a shortcut to mention its title or siglum. The logic is the same as in section §<u>2.1.4</u>. See the following examples:

Example 7.1.3.3.A
<note>While <ptr rend="title" target="txt:EdAmGa1981"/> appears …</note>
Display —
While *Amanat Galuṅguṅ* appears …

Example 7.1.3.3.B
<note>While <ptr rend="siglum" target="txt:EdAmGa1981"/> appears …</note>
Display —
While **AmGa** appears …

The empty element <ptr target="txt:EdAmGa1981"/> works as a pointer to the @xml:id of the text defined in the authority list from which we will retrieve the title. You can add an attribute @rend to select a specific rendering, whether only the siglum or the whole title of the work: the latter might be preferable for instance when the siglum is ambiguous or not universally recognized. Choose between the values "title" and "siglum" to fill in the attribute @rend. Note that the <ptr/> element calling an item from the txt authority file must have one of these values for @rend, if not nothing will be displayed at all.

7.1.3.4 Automatic retrieval?

Automatic retrieval of the text of another edition — a theoretical desideratum alluded to in the introduction to §7.1.3 — is basically only manageable if the other edition is rigorously structured, and preferably encoded in TEI-XML, a condition which is only rarely fulfilled in our field at this time. The practical consequences are that you will always need to insert passages to be cited into your file and that the encoding of canonical text references as described in the §7.1.3.2 helps to lend more semantic specificity to your TEI file but serves no computational process in our system for the time being.

7.2 Other links

Note that in this section, you will only find a solution for ad-hoc linking that should only appear in freeform text elements such <note>. Several linking mechanisms are available in the DHARMA encoding model. The main two possibility at your disposal are:

1. <ptr/> — an empty element that points to an external resource with the attribute @target
2. <ref> — a text-containing element whose contents will be hyperlinked to the resource declared in the attribute @target

When using the first solution, you must always be aware of how textual content will be generated automatically to anchor the link identified in the attribute, while in the second solution, you will be able to decide on your own which text is used as anchor for the hyperlink.

7.2.1 Making links inside your edition using identifiers

In various contexts, both a human text editor and a computer will need to be able to establish unambiguous connections between distinct blocks of text. To make this possible, our system requires unique identifiers for block elements (defined through @xml:id attributes). We have opted for a referencing system based on the mark-up of your text edition. The unique reference pattern is autogenerated by stringing together the sequential values corresponding to each level of the document hierarchy, each number being separated from the next by a full-stop.[81] Each

---

[81] This means that our system uses identifiers in the form of an untyped path. The other solution would have been to use identifiers with typed paths. It would have required the identifier of a block element to be

such identifying pattern is anchored by the value of the @xml:id declared on the <TEI> root of the edition, followed by an underscore.[82]

7.2.1.1 Identifying patterns for block-level containers

In order to identify the structural hierarchy of the xml, attributes @xml:id will be inserted into your code to allow cross-referencing to <div>, <p>, <ab>, <lg> or <quote> segments of your text edition.[83] This will be done through an XSLT transformation process that can be launched and relaunched by the project's XML-TEI Data Manager at the encoder's request.[84] The values assigned to the elements will reference the sequential numbers of the parent <div> and of the elements <ab>, <p>, <lg> and <quote> within that <div>. This means that in some cases, the <div> identifier can diverge from the editorial numbering pattern reflected in your @n attributes. It also means that if you are making any changes to the structure, all the identifiers will have to be updated. Such updating is not without consequences since it will break any link based on the previous identifiers. So be careful before making structural changes and do not forget that you will have to update any encoding that references internal identifiers.

In the following example, the first <div> element used to enclose an invocation has received an @xml:id numbered in relation to its position in the file's internal hierarchy, even though it is the second <div> that is marked by the attribute @n as the first chapter of the text.

Example 7.2.1.1.A

```
<div xml:id="rsisasana_01">
        <ab xml:id="rsisasana_01.01" type="invocation" xml:lang="san-Latn">[...]</ab>
</div>
<div n="1" type="chapter" xml:id="rsisasana_02">
         <lg xml:id="rsisasana_02.01" met="upajāti">[...]</lg>
         <p xml:id="rsisasana_02.02">kaliṅanya, [...]</p>
</div>
```

The @xml:ids can be used to interlink or reference segments of texts elsewhere in the file (see §7.2.2.3). They will become indispensable if you are dealing with prose or stanzas interrupted by base text, scenarios discussed in §4.7.1 and §4.7.2. Once each segment has its own @xml:id, you can interlink the segments with the attributes @prev and @next, used to identify the previous segment and the next. Their value should always start with "#" followed by the @xml:id of the segment of prose they point to. This system allows all the parts to be chained

---

built with the element name, a hyphen, and a number, for instance p-2, which seemed undesirably verbose to us. See TEI Guidelines, Chapter 16.1 on links and Chapter 16.2 on pointing mechanisms.

[82] See TEI Guidelines, Chapter 1.3, part 1.1.1 about Element Identifiers and Labels.

[83] Note that the identifiers generated by our system will take the file's element <text> as their hierarchical starting point. However, the element <text> itself is omitted from the identifying patterns. We have opted for this starting point rather than the <body> element to prevent identifiers being assigned to components inside any <front> and <back> elements that are present (see §10.1 and §10.2). If the need to refer to front/back matter appears at some point in the project, contact the project's XML-TEI Data Manager.

[84] The XSLT file to generate the numbering pattern can be found in the project-documentation github repository.

together and can, if necessary, be used to reconstitute the base text without interspersal of dependent text.

7.2.1.2 Identifying patterns for other containing elements

Some other elements can also be given an identifier to handle cross-referencing and linking elements across block boundaries. Among those elements, we foresee such a pattern could be useful for <term>, <lem> and <app>.[85] For those three elements, the identifying pattern should be the name of the element followed by a 3-digit number for <term> and a 5-digit number for <app> and <lem>, i.e. term001, lem00001 and app00001. See §5.2.4 and §4.8 for more about their use.

7.2.1.3 File internal links

Once the @xml:id attributes have been generated (as per §7.2.1.1) for every structural element of your edition (<div>, <lg>, <p> and <ab>), it will be possible to reference any of those elements by their unique identifier. To create a link toward another part of your edition from a free-text part, you can use the element <ref> with the attribute @target filled with the identifier of the element you want to mention with the hashtag symbol. For instance, if you want to link a note with the chapter 2, you will provide the identifier of this chapter in the @target attribute.

Example 7.2.1.3.A
<ref target="#svayambhu_03">See the same phenomenon in chapter 2</ref>

7.2.1.4 Shortcuts for citing a part of the edition

If you want to quote a part of your edition, you can use the <ptr/> element with the identifier of the said part, preceded by a hashtag, as value for @target. The display will then be the name or the attribute @type of the element and the @n attribute provided with it. For instance, if no attributes have been associated with the element you are referencing, a number based on the occurrence of the element in your edition will be automatically generated. If you require a custom display, we recommend you use the solution with <ref> explained just above (§7.2.1.3).

Example 7.2.1.4.A
<app>   <lem type="conj"><supplied reason="omitted" evidence="internal">daṇḍa riṅ trivarṇa,</supplied></lem>   <rdg wit="#L #K #M"><gap reason="omitted"/></rdg>   <note>Our restitution is based on the parallel in <ptr target="#svayambhu_07.01"/>.</note> </app>
Display —
^782. ⟨daṇḍa riṅ trivarṇa,⟩**]** *conj.*, *om*. **L K M** • Our restitution is based on the parallel in **dyad 109**.

7.2.2 Making links with resources outside the edition

7.2.2.1 Create a link between two DHARMA editions

The solution provided above to create a link between two elements in the same file can be used to create a link between two XML files created in the DHARMA framework. You should use the element <ref> around a human-readable reference like the title of the text you want to link to. To generate the hyperlink, you must also reference an attribute @target in which you should give the full name of the file of the text.

Example 7.2.2.1.A
<ref target="DHARMA_CritEdSvayambhu.xml">Svayambhu</ref>

If you want to provide a specific reference to a part of the text encoded in TEI in DHARMA framework, you will use the element <ref>. The value of the attribute @target shall be built using the filename, followed by a hashtag symbol and then by the @xml:id of the part you want to reference. In the following example, a link is made with the *Siksa Guru*, Chapter 7, first paragraph, encoded in XML-TEI by Aditia Gunawan. In the relevant XML file, the *locus* is identified as @xml:id="siksaguru_07.01".

Example 7.2.2.1.B
<note corresp="txt:SiGu_7.1"><ref target="DHARMA_CritEdSiksaGuru.xml#siksaguru_07.01">SiGu 7.1</ref>: <foreign>hana kapiṅpitunya sih saṅ paṇḍita, ri saṅ sevaka dharma, saṅ hyaṅ caduśakti saṅ paṇḍita, ṅa, apa kari mataṅ yan caduśakti, ṅa, pat śakti, ṅa, bisanya, ya ta pinakasañjata saṅ paṇḍita, ndya ya nihan, cakra jalakuna, cakra jalapakṣa, cakra sambharaṇa, cakra taraṅgabāhu</foreign>.</note>

7.2.2.2 Making a hyperlink

The element <ref> can also be used for regular hyperlinks with any external resources to the project. In this case, the @target attribute shall contain a well-formed URL, ideally a permalink. Be extra careful about filling any exterior link since we would like as much as possible to avoid such links being broken.

Example 7.2.2.2.A
<ref target="https://erc-dharma.github.io/">See DHARMA temporary static website</ref>

8. Encoding additional information in the edition

8.1 Milestone elements: pages, folios and lines

In the TEI context, a milestone is an empty element that marks the transition from one physical or textual unit to another.

Dedicated milestone elements to encode page beginnings <pb/> — the term "page" here also includes manuscript folios — and line beginnings <lb/> are not expected in most of the DHARMA project's digital critical editions, since we are generally dealing with critical editions based on more than a single witness. An exception can be made if you have a limited number of witnesses and you wish to encode the place where, in relation to your edited text, a given witness switches from one folio, or folio-side, to another (see §8.1.2).

An exception must also be made when you are encoding a critical edition that has already been published in printed format, or are using a printed edition among your witnesses; for such cases, see §8.1.1.

8.1.1 Lines and pages for printed editions

We do not recommend that you encode any manuscript line beginnings in a critical edition. However, if you are editing a text for which there is a **published edition** of reference, you can encode its page and line beginnings using the <pb/> and <lb/> elements, in combination with the attribute @source to identify the edition in question (see §7.1.2).

We suppose that generally it will make sense to encode page beginnings of the edition of reference, and if this seems useful, then it may also also be useful to encode its line beginnings. If you do encode any page (and line) beginnings, then you must make sure that every one of them is represented in your digital edition.

You should use the attribute @n, converting any non-Arabic printed numbers to their Arabic equivalents, to record the page numbers of the printed edition. For numbers indicating printed lines, it is sufficient to mark the line beginnings with <lb/> without applying @n — the machine can do the counting, provided that you do not skip any line beginnings. Page or line beginnings occuring in the middle of a word must bear the attribute @break with the value "no". [86]

8.1.2 Folios

In some cases, it may also be deemed useful to represent the folio beginnings of your manuscript witnesses, though we do not expect the need to encode manuscript line beginnings to arise in many critical editions. For folio beginnings, you must use the same element <pb/> as used for the pages of printed editions. But, contrary to what might be expected on the basis of how witnesses are to be identified in your critical apparatus (§5), in this case you cannot use the

---

[86] EGD §3.2.4, §3.5.5 as well as §8.1.2, §8.1.3, §9.1.6.

attribute @wit to identify them, because the TEI has not foreseen this option. As a workaround, you should use <pb edRef="#..."/>, and fill in the siglum of the witness as defined in <sourceDesc>, e.g. <pb edRef="#B"/>. Note that the attribute @edRef is a pointer, so that the identifier must be prefixed here with "#".

To indicate the number and side of **manuscript folios**, you should add the attribute @n whose value shall be the folio number and one-letter indication of its side (recto/verso): "1r", "1v", "2r", "2v" and so on. Manual encoding of @n is always necessary on elements <pb edRef="#..."/>. As in the case of printed editions, if your <pb/> occurs in the middle of the word, you should add the attribute @break with the value "no".

In cases where two witnesses move from one folio or folio-side to the next at exactly the same locus — a situation that is especially common at the beginning of texts — you must use two <pb/> elements (<pb edRef="#A"/><pb edRef="#B"/>) and you must specify the respective folio numbers and sides by using @n (<pb edRef="#A" n="3r"/><pb edRef="#B" n="4r"/>). Note that it is this need to use @n that stands in the way of encoding in the usual way, which would be <pb edRef="#A #B"/>.

The following is an example of the beginning of an edition encoded from a printed edition based on a codex unicus:

Example 8.1.2.A
<pb edRef="#EdA" n="194"/> <ab type="invocation"><pb edRef="#msA" n="0v"/><lb n="1"/>siddhir astu</ab>

8.2 Notes

Insertion of free-text notes into your edition is constrained. It is only possible to do so at the end of any <app> element and at the very end of one of only four block-level containers (after any final punctuation), namely after prose paragraphs (<p>), after invocations or colophons (<ab>), after whole stanzas (<lg>) or after verse-lines (<l>). See respectively §5.10.1 and §6.1. Of course it is also possible to insert notes into paragraphs of your translation (see §9.1.4.5) and your <teiHeader>. In all cases, you will use the element <note>. Generally, it will be necessary to make sure the contents of a <note> is a full sentence, beginning with a capital letter and ending with a punctuation mark (though not in a <note> used to record a parallel passage, §6.2). Where relevant, insert the <note> after any adjacent punctuation mark, not before it. The contents of notes may be structured into paragraphs by creating <p> elements within the <note>, when the note amounts to more than a few sentences. Keep in mind that in this case, all the contents of <note> should be contained inside a <p> element.

If you are not the author of the note's contents, you may use the attribute @resp or @source to declare someone else's authorship. To give credit to another member of the project, use @resp, while @source should be used only if the note is adopted verbatim from a publication. Alternatively, it is possible to declare the author of the original note using the regular

bibliographic citation system (see §8.4), or, depending on the context, simply by referring to the author's name.

We do not require the systematic use of @xml:lang with <note>. Its use would be expected to set the correct language on the content to avoid the inheritance from the one declared at the <text> level (see §8.10.1). However, to avoid repeating the same information in every <note> element, our system assumes that @xml:lang="eng" is implied by default. It means that, if you provide a note in another language, for whatever reason, the declaration of the language becomes mandatory with @xml:lang. Also, note that a similar omission of declaration of language is tolerated on elements such as <q> and <foreign> (see respectively §8.3.2 and §8.10.2), even if explicit declaration is recommended.

Finally, you can differentiate your <note>s using the @type attribute based on a subject, e.g. <note type="commentary">. The values are left to your own discretion based on the needs of your edition. Please contact the XML-TEI Data Manager before doing so.

8.3 Quotations

Generally, while encoding quotations, do not add any quotation marks, as they will be generated by the transformation process, provided that proper tags have been applied. However, if for any reason you do not encode the quotation with TEI tags, then type quotation marks explicitly.

When citing original-language text from a publication, you should generally follow DHARMA transliteration conventions. This also holds, for instance, when citing readings from published sources in your apparatus, when citing translations or when stating paraphrased/summarised opinions. But if they stand inside a <quote>, any transliterated words in the cited text shall be retained in their original form rather than being converted to the DHARMA system.

8.3.1 <cit> and <quote> – Cited Quotation

To encode quotations from an external source, with accompanying information about its origin, it is good practice (though not mandatory) to use the element <cit> as a container and a <quote> element inside the <cit>. The second child element of <cit> may be <bibl>, <ptr/> or <ref>, depending on whether the source cited is a bibliographic entry or digital resource.[87]

For the first situation, use the element <bibl>, if you are quoting this material from a published work (see §8.4.1 on how to do use <bibl> with <ptr/>).

Example 8.3.1.B
<cit>   <quote></quote>   <bibl><ptr/></bibl> </cit>

---

[87] See TEI Guidelines, chapter 3, part 3.3.

For the second situation, use the element <ref> around the reference in the manner described in §7.2.2.

Example 8.3.1.A

```
<cit>
    <quote></quote>
    <ref></ref>
</cit>
```

Note that it is also possible to use <quote> without using <cit>. Whether or not you use <cit>, the use of <quote> should be limited to actual (multiword) quotations from another scholar's work. In all other situations, we recommend using <q>.

8.3.2 <q> – Quoted Material

The element <q> is semantically less specific than <quote>. It can be used quite freely and the transformation routine will represent it as single quotation marks. We expect <q> to be applied for stating the meanings of words or phrases in commentary, for representing in an edition that a given segments is direct speech, and no doubt there are other scenarios too where <q> will be useful.

Example 8.3.2.A

<note>There is no verbal form in the entry s.v. <foreign>śanku</foreign> 1 (<q>a particular kind of weapon</q>). The passive irrealis form used here is spelled with an <foreign>h</foreign>, which serves to bridge the hiatus between base and suffix <foreign>-ən</foreign>.</note>

When using <q>, as when using <quote>, you should not type quotation marks. Nonetheless, if you need more control on the display, it is possible not to use any tag and directly to write the desired characters. This method should be reduced to special cases as much as possible.

8.4 Bibliographic References

8.4.1 Bibliographic citations with Zotero

Bibliographic citations may be used in any part of your XML file. Their main container is <bibl>. The empty element <ptr/> should mandatorily appear as the first element within <bibl> and contain the attribute @target, whose value shall be the Zotero Short Title of the cited publication, prefixed with the string "bib:" (see §7.1.2).

To give further specification regarding the part of the publication concerned, add the element <citedRange> as a second child of <bibl>. Its content should specify the relevant

structural units of the publication being cited. If you need to refer only to page numbers, there is no need to add any other content: <citedRange>103</citedRange>. If you do need to refer to any other textual units, add the attribute @unit and fill it in with one of the following values:

- "appendix"
- "book"
- "entry"
- "figure"
- "item"
- "line"
- "note"
- "page"
- "part"
- "plate"
- "table"
- "section"
- "volume"

The content of each element should normally consist mainly in numbers. Any range should be recorded using a hyphen to separate the starting point from the ending one, both points being expressed in full (e.g. <citedRange>123-124</citedRange>, not 123-4 or 123-24). To list non-adjacent elements, use a comma followed by a space, e.g. <citedRange>12, 24</citedRange>.

Example 8.4.1.A
<bibl><ptr target="bib:Naerssen1976_01"/><citedRange>301-302</citedRange></bibl>
Display —
van Naerssen 1976: 301–302.

If the name(s) of the author(s) of the work has/have already been mentioned in the sentence and need not appear again in the hyperlinked display of the reference, use the attribute @rend with the value "omitname" in the <bibl> element; if you expect to show "*ibid*.", use the value "ibid" for @rend.[88]

Example 8.4.1.B
<bibl rend="omitname"><ptr target="bib:Naerssen1976_01"/></bibl>
Display —

---

[88] All this section is following recommendations established by the EGD, see §10.4.5.

Example 8.4.1.C
<bibl rend="ibid"><ptr target="bib:Naerssen1976_01"/><citedRange>301-302</citedRange></bibl>
Display —   *ibid.*: 301–302

8.4.2 <title> – Encoding titles

Should you need to quote a title in any notes or commentary, tag it with the element <title>. It is basically up to you what you consider to be a title, whether designations of epigraphic or non-epigraphic primary sources, designations of secondary sources regardless of whether they are cited in full, in abbreviated form or with a known acronym, etc.[89] The default rendering of segments tagged with <title> will be italic.

Example 8.4.2.A
<note>This is the second of to two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMd1 KSS Dave. Cf. <title>Ślokāntara</title> 7 <foreign>nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ | triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||</foreign>. The latter stanza is also transmitted in <title>Tantri Kamandaka</title> §63.</note>
Display —   This is the second of two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMd1 KSS Dave. Cf. *Ślokāntara* 7 *nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ | triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||*. The latter stanza is also transmitted in *Tantri Kamandaka* §63.

When this is not desired, add the attribute @level with the value "a" (for "analytic"), in which case the content will be displayed with quotation marks rather than italics, or use the attribute @rend with the value "plain" to obtain a title without any kind of typographic distinction from its surrounding text.

Example 8.4.2.B
<note>This is the second of to two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMd1 KSS Dave. Cf. <title level="a">Ślokāntara</title> 7 <foreign>nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ | triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||</foreign>. The latter stanza is also transmitted in <title level="a">Tantri Kamandaka</title> §63.</note>

---

[89] Follow the rules given in the EGD, §10.4.2

<table>
<tr><td>

Display —

This is the second of two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMd1 KSS Dave. Cf. "Ślokāntara" 7 *nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ | triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||*. The latter stanza is also transmitted in "Tantri Kamandaka" §63.

</td></tr>
</table>

Example 8.4.2.C
`<note>`This is the second of two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMD1 TMd4 Tr1 MTr4 MTr6 [Jolly Gr] Mandlik [ṭa] KSS Dave. Cf. `<title` rend="plain"`>`Ślokāntara`</title>` 7 `<foreign>`nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ	triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||`</foreign>`. The latter stanza is also transmitted in `<title` rend="plain"`>`Tantri Kamandaka`</title>` §63.`</note>`
Display — This is the second of two stanzas interpolated after 8.82 recorded in Olivelle's edition, found in his witnesses GMd1 KSS Dave. Cf. Ślokāntara 7 *nāsti satyāt paro dharmo nānṛtāt pātakaṁ paraṁ	triloke ca hi dharma[ḥ] syāt tasmāt satyaṁ na lopayet ||*. The latter stanza is also transmitted in Tantri Kamandaka §63.

8.5 @resp – Attributing responsibility

It is possible to declare intellectual responsibility with the attribute @resp. It can be added to any XML element to encode the fact that a particular project participant is the author of that particular item. At the first stages of the project, since you will work mainly on your own, you will surely not need this attribute often. But when reviews will start, it will allow us to identify interventions made by various members of the project.

The attribute @resp should contain the personal identifier of the project member preceded by the prefix "part:" (see §7.2.1). To credit more than one participant, separate each identifier with a blank space.

8.6 Encoding diplomatic features

All features described in the project's Encoding Guide for Diplomatic Editions are also available for use in apparatus entries, particularly in <rdg>. Besides elements such as <gap>, <supplied>, <surplus>, <add>, <del>, <unclear> that we have already presented above,[90] we may mention here the possibility of using <space> and <g> in the manners prescribed in the EGD.

It is also possible to expressing uncertainty using the attribute @cert="low" with the elements <unclear>, <supplied>, <num> and <seg>, as well as @precision="low" and <certainty/> with the element <gap/>, as in the following example.[91]

---

[90] EGD §4, §5 and §6.
[91] EGD §2.3.4 and §5.4.6

<table>
<tr><td align="center">Example 8.6.A</td></tr>
<tr><td>

```
<gap reason="lost" quantity="30" unit="characters">
   <certainty match=".." locus="name"/>
</gap>
```
</td></tr>
</table>

8.7 Encoding indexables

All elements useful to distinguish certain segments from the surrounding text because they can be of interest in a database, to facilitate the creation of an index, or even to support search functionality, such as names, dates, numbers and abbreviations should be encoded in the same manner as has been described in the EGD.[92] Thus, the elements <persName>, <roleName> and <placeName> for places and people, as well as <abbr>, <num>, <measure> and <date>, are all available for use in a critical edition, inside or outside apparatus entries. Note that <lem> or <rdg> can be used within them as needed to record variant readings. For instance, see the following example with <abbr>. You can decide which one of those elements, if any, should be encoded depending on your text and research aim.

<table>
<tr><td align="center">Example 8.7.A</td></tr>
<tr><td>

```
<app>
   <lem type="emn">
      <abbr>mā</abbr>
   </lem>
   <rdg wit="#L #K #M">ma:s·</rdg>
</app>
```
</td></tr>
</table>

8.8 About formatting

8.8.1 <hi> – Typographical formatting

It is possible to state explicitly how characters should be formatted in the display without making any statement as to why they are formatted that way. We foresee this manner of encoding being used for defining the formatting of sigla for witnesses (see §2.1.3) and whenever you wish to cite as such the formatting of a passage quoted from another scholar's work. Use the element <hi> with the attribute @rend with one to several of the following values:

- "italic"
- "bold"
- "subscript"
- "superscript"

---

[92] EGD §7.

We also offer the possibility of using <hi> with @rend="check" during the preliminary stage of the editing process, as shown in §<u>8.11</u>.

8.8.2 <list> – Encoding lists

If needed, you can format text as a structured list using the elements <list> and <item>. Although our schema allows their use in most contexts, we recommend avoiding use of these elements in your text edition to avoid the code (and the display) becoming too complex. By contrast, in the context of translation, or in the metadata to your edition, you may feel free to make use of list elements.

Within an enclosing element, typically <p> or <ab>, create a list with the element <list> and inside it an element <item> to wrap each component of the list. Several styles are available by adding the attribute @rend. For instance, @rend="bulleted" creates a bulleted list and @rend="numbered" will create an auto-numbered one. If you require a list with headwords and their descriptions, you will need to add an element <label> around the first and an element <item> around the second, both working as an alternating sequence inside the main <list>.

Without any attribute or <label>, the default display of a list will be that each item is displayed in a new line with an indent.

8.9 Marking *metri causa* spellings or word forms

If your text contains spelling anomalies because the author needed to resort to such means to accommodate the meter, it is possible to mark anomalous spellings by using the diplomatic element <orig> within a <seg> and apply to the latter the attribute @type with the value "mc", abbreviated for *metri causa*.

Example 8.9.1.A
yak<seg type="mc"><orig>a</orig></seg>ragaṇa

Should you wish to record the normalized (but contextually unmetrical) spelling, you must add the element <reg> and wrap it together with <orig> in an element <choice>. This will notably allow you to harvest even anomalously spelled attestations of a given word when lemmatizing a text's vocabulary. [93]

Example 8.9.1.B
yak<seg type="mc"><choice><orig>a</orig><reg>ā</reg></choice></seg>ragaṇa

---

[93] Even though they are verbose, those solutions seemed the best compromise between possible data treatment, TEI encoding model and coherence with DHARMA practices so far.

8.10 About languages

8.10.1 @xml:lang – identifying languages

The language of each TEI file for a critical edition is set as English. This is done through an attribute @xml:lang="eng" furnished in the root element <TEI> in our templates. This default setting is to be understood as the language applied to all children elements. The language code "eng" is taken from the ISO standard 639-3.

Since the default language of the file is set as English, the general rule is that any segment whose contents is in another language should be explicitly marked with its own suitable @xml:lang.[94] This means that your critical edition should be given its own @xml:lang on the <text> element.[95] A text originally written in any Indic script and edited in Romanised transliteration should be suffixed with "-Latn", as per the ISO standard 15924. Without this suffix, it will be assumed that the edition is given in the native script associated with the language given for the region and time period.

Example 8.10.1.A
<text xml:space="preserve" xml:lang="osn-Latn">

Your edition being contained in <text>, whenever you yourself are speaking as editor, in English, the above logic would require you to redeclare English as the current language using attribute @xml:lang. However, we have made an exception to this rule in the case of editorial use of <note>.

Translations, being encoded in <text> elements that do not bear an @xml:lang (§9.1.2.1), automatically inherit the language value of the root element <TEI> and are hence by default understood to be in English. But if you provide a translation in French, for instance, or in another language than English, you will need to provide the language code on the <text> element.

8.10.2 Tagging text segments in another language

If the text you are editing takes recourse more or less exceptionally to other languages, which might be labeled 'foreign' from the point of view of the text's main language (as defined in the element <text>), it will be interesting to use the element <foreign> in order to facilitate specific modes of display and data treatment.

Any smaller text segment (a word, a sentence or an even bigger portion of text below the paragraph level) that is formulated in such a 'foreign' language, should be marked with <foreign> in combination with the attribute @xml:lang if more than two languages have been

---

[94] All of the language codes relevant for the project are shown in §Appendix B.
[95] The <div type="edition">, which may be familiar to some from EpiDoc, has not been retained in the present guide (see §3.1).

identified in <langUsage> in the <teiHeader> (see §2.2.2). If only two languages have been identified, then it is not mandatory to encode @xml:lang on <foreign>, though we still recommend doing so systematically, especially if you require language sensitive treatment of the textual data (e.g. for indexing), and use of @xml:lang on <foreign> is mandatory in any case when more than two languages have been identified in <langUsage>. In case you are facing a segment as large as a whole block-level container (<ab>, <p>, <l>, <lg>) or even a <div>, then it is expected that the @xml:lang for the relevant 'foreign' language is indicated on the container. In that case, it is still possible to use <foreign> though its sole purpose will then be to trigger a special mode of display such as italics.

Example 8.10.2.A

`<text xml:space="preserve" xml:lang="kaw-Latn">`
    [...]
`<p xml:id="siksaguru_13.03" xml:lang="osn-Latn"><foreign>`ma`<pb edRef="#B" n="38v"/>`na duka ku ulahta, ku [...] `<app><lem wit="#A #C">`duka`</lem><rdg wit="#B"></rdg></app>` ku ulahta, ku sabdata, ku ambəkta.`</foreign></p>`
    [...]
`</text>`

8.11 Validation and review process

To help you keep track of segments of your edition that need to be double checked in the review process, it is possible to add the attribute @rend with the value "check" to the relevant element. The resulting display of the content will be a yellow highlight.

Example 8.11.A	

`<app>`
  `<lem type="conj" rend="check">`aṅkən`</lem>`
  `<rdg wit="#L #K">`hankan`</rdg>`
`</app>`	

We recommend using the attribute @rend="check" as high as possible in the file's hierarchy in order to avoid needing to repeat the attribute several times on children elements. To identify a specific part of the text which does not match any already existing tags, you can add the

element <hi>. It is understood that the editor will gradually clear the XML file from any occurrences of @rend="check" as the finalization of the editions comes into sight.

Within verse lines where a given reading is not yet acceptable because non-compliant with the meter, it is also possible to use the more specific value "unmetrical", yielding a pink highlight, to mark readings that will need to be rechecked and perhaps modified further down the editorial process. If your final conclusion is that the metrical irregularity is authorial, then encode the line (or stanza) as per §3.4.4.

Example 8.11.B	

\<l n="c"\>\<lb n="4"/\>
  \<app\>\<lem rend="check"\>svīkān\</lem\>
  \<rdg wit="#L"\>stika:n\</rdg\>
  \<rdg wit="#JK"\>\<hi rend="unmetrical"\>stikan\</hi\>\</rdg\>
\</app\>rakṣa rəṅihniṅ adyah anahən lara pinuputan antyaniṅ gati,\</l\>	

Be extra careful when using both @rend="check" and @rend="unmetrical" together. The first value will always have the priority over the second, meaning that yellow highlighting will render pink highlighting invisible.

Finally, we remind you of the need to encode significant changes in the status of your work in the manner described in §2.2.3.

9. Encoding translations

Translations of critical editions as well as diplomatic editions are to be done in separate files. This decision has been taken to reduce the length of the files and to facilitate working with files for edition and translation in parallel. The edition file will be understood as the main source file, while translation files will be regarded as secondary and linked to the main file. Every translation file should have a structure exactly matching the structure chosen for the edition. If, for whatever reason, you decide to encode two distinct translations (for instance, one into French and one into English), they shall be encoded in separate files. Creating a secondary file to go with your edition will require you to follow some recommendations regarding metadata.[96]

---

[96] A template is available in the project-documentation gitHub repository.

9.1 <teiHeader> for translations

The metadata of your edition as a whole shall be declared in the <teiHeader> element of the main source file. See section §2 of this guide. The translation <teiHeader> shall not contain any repetition of those metadata already declared about the edition and the witnesses for instance, though some basic information is required for the sake of traceability. Indeed, one of the downsides of separating the edition and translation into distinct files is the risk of losing the link between the translation(s) and the edition. For this reason, metadata need to be provided to identify the file and allow restoring the relationship with the main source file.

The main parts of the <teiHeader> are respected, meaning that you will find a <fileDesc> gathering a <titleStmt>, a <publicationStmt> and a <sourceDesc>, followed by an <encodingDesc> as well as <revisionDesc>. This section will only give rules specific to the translation and suppose you know the basics of encoding a <teiHeader> as already explained above for the edition.

It is expected that you give a <title> with attributes @type="main" and @subtype="translation". If necessary, you can reuse the possibilities offered by the model of the edition, by adding another title with the attribute @type="subtitle" and @subtype="translation". We don't expect any copy and paste of the <editor> and <respStmt> elements declared in the main file, but rather that you declare the people involved specifically with the translation work and provide a description of their role with the element <resp>. However, if the <editor>s are the same, feel free to declare them identically.

You will also need to provide a full <publicationStmt>; follow the section §2.1.2 of this guide to fill it in. The only difference will concern the <idno type="filename">. Indeed, it has to follow the file naming conventions set for the translation file and not for the critical editions files. See FNC §2.2.4.

The main changes will be situated in the <sourceDesc> element. First, you will need to declare a <biblFull> element and copy and paste the <titleStmt> and <publicationStmt> sections of the edition file. If your translation reproduces one that has already been published in another format, add an element <biblStruct> to declare it. Declare sources for the translation even if you have reworked it, and without consideration for the length of elements that have been integrated into your translation. But if it is unpublished, you can declare all the authors in the <teiHeader> thanks to the <respStmt>. If you are dealing with a complicated case, you may provide a <notesStmt> section, with a <note> to explain more precisely the work of each person involved, in free-text.

Example 9.1.1.A

```
<sourceDesc>
    <biblFull> <!-- Copy and paste the titleStmt and publicationStmt of the edition file -->
      <titleStmt>
        <title type="main">Svayambhu</title>
          <title type="sub" subtype="editorial">Digital Critical Edition and Translation of an Old Javanese Juridical
```

```
Text</title>
        <editor ref="part:argr">
          <forename>Arlo</forename>
          <surname>Griffiths</surname>
        </editor>
        <editor ref="part:tilu">
          <forename>Timothy</forename>
          <surname>Lubin</surname>
        </editor>
        <respStmt>
          <resp>structuring of the TEI file</resp>
          <persName>Arlo Griffiths</persName>
        </respStmt>
      </titleStmt>
      <publicationStmt>
        <authority>DHARMA</authority>
        <pubPlace>Lyon (France) &amp; Lexington, VA (USA)</pubPlace>
        <idno type="filename">DHARMA_CritEdSvayambhu</idno>
        <availability>
          <licence target="https://creativecommons.org/licenses/by/4.0/">
            <p>This work is licensed under the Creative Commons Attribution 4.0 Unported Licence. To view a
copy of the licence, visit https://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, 444
Castro Street, Suite 900, Mountain View, California, 94041, USA.</p>
            <p>Copyright (c) 2019-2025 by Arlo Griffiths &amp; Timothy Lubin.</p>
          </licence>
        </availability>
        <date from="2019" to="2025">2019-2025</date>
      </publicationStmt>
    </biblFull>
    <!-- if your translation is a copy of already published one: fill a biblStruct -->
  </sourceDesc>
```

The <encodingDesc> will at least contain the <projectDesc> element with its first mandatory <p> about the DHARMA project. If you have provided another explanation about your own project, please copy and paste it into your translation file as well. Any other elements used for the main edition in <projectDesc> are available for the translation, but none are mandatory. Use them depending on your needs to describe specific choices regarding the translation itself. See the section §2.2 to identify all the elements at your disposal.

Finally, keeping track of the important steps in the development of your work is still recommended with the element <revisionDesc> and the use of the <change> elements. See the section §2.2.3.

9.2 Structuring the translation

9.2.1 <text> and <body> elements

To structure the translation, follow the main rules set out for the edition and explain under §3. Add the mandatory elements <text> (with an attribute @xml:space="preserve") and <body>.

If your translation is given in English, then you won't need to declare a new attribute @xml:lang on the <text> element, since it will be declared by default of the <TEI> root. However, if your translation is given in any other language, please declare the @xml:lang with the matching ISO standard 639-3 code. See §Appendix B.

9.2.2 <div> – Sections of translations

Then use the <div> elements to reproduce the main sections of your edition. We expect you to keep the <div> used for invocations and colophons, as well as those used for chapters and cantos, keeping both attributes @n and @type.

The editorial element <head type="editorial" xml:lang="eng"> can be kept in your translation. It is not required that you keep it, if you have added one in your edition. However, if you decide to keep it, make sure to encode it right after the opening part of the relevant element, as the first child of <div>; delete the @xml:lang element if you are in an English translation. If you want to keep the <head type="editorial">, even if it does not match the language of the translation itself, then keep the attribute @xml:lang="eng".

9.2.3 Structural organization of the translation

The main unit level for translated text will be the paragraph. All three of the elements <p>, <ab> and <lg> used in the edition will be represented in the translation with <p> elements.

For paragraphs translating an original in verse form, you can use the attribute @rend="stanza" to obtain a verse-type display. In that case, it is necessary to reproduce the attribute @n from the <lg> in the edition in its corresponding <p>.

9.3 Identifying correspondence with editions

9.3.1 Parallelism with the edited text

To match segments of translation with the corresponding segments of edited text, we will make use of the @xml:id attributes on the elements <div>, <p>, <ab>, <lg> and <quote> declared in the edition. In the corresponding <p> elements of the translation, apply the attribute @corresp and assign as value the @xml:id of the relevant block of edition preceded by a "#".

<table>
<tr><td colspan="2" align="center">Example 9.1.3.1.A</td></tr>
<tr>
<td>

```
<div type="chapter" n="1" xml:id="#siksaguru_01">
        <head type="editorial" xml:lang="eng">Pure
Circle of Three Bodily Domains</head>
        <p xml:id="siksaguru_01.01"><pb edRef="#A"
n="1v"/><pb edRef="#B" n="1v"/>ndah nihan
varahakna mami rumuhun ri saṅ sevaka dharma,
katəkan kita ruməṅəkən kavivekan saṅ hyaṅ dharma,
mvaṅ kabyavasthānikaṅ saṅ hyaṅ ajñāna, hana trikāya
maṇḍala pariśuddha*, ṅaranya, inujarakən saṅ paṇḍita
ri saṅ sevaka dharma, nyāpan tahan kveki ulah śabda
lavan ambək, yan kva hana ma<pb edRef="#B"
n="2r"/>līna ri hr̥dayanta*, ya teka prih
pəgəṅən<supplied reason="omitted">,</supplied>
rakṣa pahayu juga ya, apan hetuniṅ manəmvakən
sukhaduhkha halahayu.</p>
[...]
</div>
```

</td>
<td>

```
<div corresp="#siksaguru_01">
     <p n="1" corresp="#siksaguru_01.01">I will first
teach the servant of the <foreign>dharma</foreign> as
follows. You will succeed in listening to the discussion of
the holy <foreign>dharma</foreign> and the determined
state of the holy gnosis. There is what is called the pure
circle of three bodily domains,<note><foreign>trikāya
maṇḍala pariśuddha</foreign>: On the concept of the
pure circle of three bodily domains, see Chapter 10,
§Dharma and Three Bodily Domains.</note> uttered by
the teacher to the servant of the
<foreign>dharma</foreign>. That is possibly the action,
speech, and mind. If there is a feeling of loss in your
heart,<note>The emendation of
<foreign>pr̥dayanta</foreign> into
<foreign>hr̥dayanta</foreign> is not only easily
justifiable on palaeographic grounds (the similarity of
<foreign>aksara pa</foreign> and
<foreign>ha</foreign>), but also based on the
connection of <foreign>līna</foreign> 'feeling of lost'
with <foreign>hati</foreign> 'heart' in several OJ texts.
Cf. RY 24.41: <foreign>aḍā kalana līna riṅ
hati</foreign>.</note> it is them that should be
practiced and restrained. You should just guard them
and fix them, for they are the causes of finding joy and
sorrow, good and bad.</p>
</div>
```

</td>
</tr>
</table>

9.3.2 Declaring the authorship of on a translated segment of text

In case of a translation written by several authors and/or encoding parts or a whole previously published translation, you can use both the attributes @resp and @source to provide a sharper delimitation of the authorship than what is possible with declaring it in the <teiHeader>:

- use the attribute @source on structural elements, such as <div> or <p>, for which you want to declare the author and its publication. Fill it in with the prefix "bib:" and the Zotero Short Title of the publication.
- Use the attribute @resp on structural elements to credit a member of the DHARMA project. Add the prefix "part:" and her/his DHARMA unique identifier.

If several publications or project members are involved, it is possible to declare several references in those two attributes, as long as you add a space between each.

9.4 Markup available with a <p> of translation

If you have encoded an edition, you do not need to replicate all its encoded features in the translation. You may, however, want to mark up segments of the translation itself, for instance when the translation skips segments of the edition or, inversely, when the translation contains segments that find no correspondence in the edition. Note that miscellaneous encoding elements such as <foreign> or <hi> are available as well as <note> to insert remarks of the translator that are not part of the translation per se.[97]

9.4.1 Additions into the translated text

Restoration of segments in edited texts should be translated. When the restored text corresponds to a lost content, use the attribute @reason with the value "lost" in combination with the element <supplied>. But if the content was omitted by the scribe, use the value "omitted" for the attribute @reason.

Any addition made in the translation that does not find its equivalent in the edition should be identified with the element <supplied>. But we will distinguish between additions made in order to clarify the translation in the target language, those made to obtain a clear syntax and those used for disambiguation purposes. The first case shall be expressed with the attribute @reason with the value "subaudible". Those parts don't find any equivalent in the edited texts. Their use isn't mandatory and it is left to the encoders to use them or not depending on how literal the translation is.

The second case should be identified with the value "subaudible". It must be used for words that are not explicit in the edited texts, but that need to be made explicit to arrive at a proper sentence in the target language. Use it for cases where you need to signal accuracy of your translation.

The final case will be encoded with the value "explanation" for the attribute @reason. You should use it to tag any supplementary words provided for clarification which are motivated by the syntax of the target language. The <supplied reason="explanation"> can also be used in combination with the element <foreign> to provide the original words from the edited text inside your translation.

9.4.2 Lacunae

Lacunae that can't be restored should be encoded with the <gap/> element and the attribute @reason with value "lost" (cf. the rules concerning use of <gap/> in various subsections of §5.5). If your intent is to represent a segment of text that has been skipped in the translation because it is not intelligible, then use the element <gap/> with @reason="ellipsis". In the former case, we might expect either, for short segments of text, the <supplied reason="explanation">

---

[97] Most of this section will follow features already made available in DHARMA EGD from section §9.2.9 to §9.2.14.

with the child element <foreign> replicating the unintelligible text, or else, in case of larger segments of text, a <note> element to provide an explanation of what the missing part is.

The approach with <gap reason="ellipsis"/> can also be used to encode any extent of texts that you deem unnecessary to translate (e.g. in the case of highly repetitive texts). We recommend in this case to also provide an explanatory <note> element right after the ellipsis.

9.4.3 Expressing doubts and flagging problems

It is possible to express your doubts or uncertainty for any translated word or sentence by adding an attribute @cert="low" on the semantic elements. If no element is available around the segment of text you would like to identify as an attempt, use the element <seg cert="low">.

This doubt might arise from your own tentativeness or from the edited text being unclear, illegible, or restored, or by the language itself.

When text seems inappropriate as well as unexpected, tag the equivalent in the translation with the element <sic>.

9.4.4 Indicating bitextuality

It is possible to represent both meanings in a double entendre in your translation. To do so, wrap the secondary and less literal meaning between an element <seg> with the attribute @rend with the value "pun". The meaning you have chosen to identify as the most literal won't bear any markup.

9.4.5 Providing notes

You may add any necessary <note> in your translation as long as they are contained inside the <p> that structure the translation and as long as the basic rules already stated in the section §8.2 are followed. It is possible to distinguish notes based on their content to distribute them in different blocks or levels of footnotes. In this case, use an attribute @type which values are left to your consideration. Please contact the XML-TEI Data Manager, if you intend to use @type on <note> in your translation.

10. Encoding published books

If you are encoding an edition that has already been published in traditional book format, you might wish to encode all its sections like the preface, the introduction and so on. Two elements are available to reproduce what comes before and after the edition itself: <front> and <back>. They should be used as children of the <text> element. Logically, the element <front> should come before the <body>, while the <back> must come after it. Two possibilities are offered to you:

1. you adapt the edition to fully match the rules stated for natively digital editions, i.e. separating in different files the content and declaring the witnesses in the <teiHeader>.
2. you favor a sort of facsimile version of the published critical edition. In this scenario, follow the rules stated here to organize the content in a single file using the elements <front> and <back>. Whatever the published edition does, it is expected that you reproduce the apparatus as an internal one, but if you deem it impossible, you can always revert to an external kind of apparatus following rules provided under §6.3.

10.1 <front> – Front Matter

If you have chosen the second encoding strategy to deal with an already published critical edition, you will store all the content coming before the edition itself inside the <front> element. It can contain an introduction with a survey of witnesses and sources, and possibly other content as well. Each subsection of the front matter should be given its own <div> with a unique @xml:id identifying it.

In this case, which amounts to creating a kind of facsimile of the published critical edition, your <front> will probably contain a survey of the witnesses and sources, the bibliography, and other information that may be of interest. It means that no witness information shall be declared in the <teiHeader>. Instead, please follow rules to encode the witnesses as if they were declared in the <teiHeader> (see §2.1.3), especially the rules for making machine-actionable sigla. Contact the XML-TEI Data manager if you are in this situation. Every globally permitted element is available to encode the content of the front matter.

Example 10.1.A

```
<text>
  <front>
    <div type="section" xml:id="preface">
      <head>Preface</head>
      <p>General introductory remarks about the edition.</p>
    </div>
    <div type="section" xml:id="bibliography">
      <head>Bibliography</head>
    </div>
  </front>
  <body>
    <div></div>
  </body>
</text>
```

10.2 <back> – Back Matter

By contrast, the content of the back of an already published edition shouldn't be encoded. Indeed, the edition being encoded, all of the traditional content found at the end of a text edition (glossary, indexes, bibliography) can be generated automatically and adding it manually in the xml file will be redundant. Nevertheless, if it is necessary, the <back> element is still available as the last child of <text>. Before starting to encode <back>, we recommend that you contact the XML-TEI Data manager.

If a <back> element is added, use the same kind of structure as the one used for the front matter. A <div> shall enclose each subsection and needs to have its own @xml:id attribute.

<div style="border:1px solid #999">

Example 10.2.A

```
<text>
   <front>
      <div></div>
   </front>
   <body>
      <div></div>
   </body>
   <back>
      <div type="section" xml:id="bibliography">
         <head>Bibliography</head>
      </div>
   </back>
</text>
```

</div>

Appendix A: Editorial Conventions

1. How to represent punctuation?

For **Indian texts** entirely formulated in Sanskrit, follow the convention of ending half stanzas with a single *daṇḍa* | and whole stanzas with a double *daṇḍa* ||. In prose contexts, use single *daṇḍa*s as lowest level punctuation mark while reserving double *daṇḍa* to mark paragraph ends. A text edition shall not normally make use of any western punctuation marks such as colon or semicolon. Many editors declare that the punctuation of their editions are entirely editorial and make no attempt to record the actual punctuation found in their witnesses.

For **texts transmitted on Java and Bali**, by contrast, we recommend making exclusive use of western punctuation marks, especially comma and full-stop, and you may also use colon and semi-colon if they seem useful. In our (limited) experience, the punctuation of witnesses tends to be relatively uniform from one witness to another and tends to be a useful guide to

parsing texts. If you find this to be the case also in the text you are editing, then you may make it a principle to retain use comma when your witnesses use a lower-level punctuation sign and a full-stop when they use a higher-level sign. Should you adopt the principle that punctuation of your edition is normally compliant with the transmission, the you must use <supplied> to insert any untransmitted punctuation and can use <surplus> to suppress any transmitted punctuation. Or you can decide upon a policy that leads to fewer editorial brackets in your edition.

Whatever you do, declare explicitly what is your policy, using the <punctuation> element in <editorialDecl> (see §2.2.1.1).

2. Spelling normalization in editing Indonesian texts

As stated in the preliminaries, a critical edition is supposed to follow the loose transliteration scheme offered in the DHARMA Transliteration Guide. This means, i.a., that a critical edition should not contain any vowel killers, and that some distinctions made in strict transliteration need to be abandoned. Notably, strict *ṁ* and *ṅ* should merge into loose *ṅ* while strict *ḥ* and *h* should merge into loose *h*. If your text contains any Sanskrit, retain these distinctions only in the Sanskrit context. The distinction between lowercase and uppercase for independent vs. dependent vowels should also be abandoned in your critically edited text.

The following nornalization rules apply to the constituted Old Javanese and Old Sundanese text:

- freely insert *ə* in order to makes lexical items and affixes closer to the normalized spelling familiar from OJED
  - some examples within a base: *tmən → təmən*, *vnaṅ → vənaṅ*
  - a frequently occurring case involves the suffix *-akən/-akəna*, often spelled *-akn/-akna*, to be normalized as *-akən/-akəna*.
- freely adjust manuscript readings involving the signs *s*, *ś*, and *ṣ* to effect compliance with dictionary spellings
- freely remove spelled *h-* if the dictionary spelling of a word as an initial vowel
- freely simplify geminate spellings at word or morpheme boundaries to effect compliance with dictionary spellings
  - The *visarga* (*ḥ*) followed by the fricative (h) is merged into h, for example
    - maḥhala → mahala
  - The *anusvāra* (*ṁ*) followed by the velar nasal (*ṅ*) becomes a single *ṅ*
    - paṁṅaduA → *paṅadua*
  - The combination of *layar* followed by *r* is reduced to r
    - *kasampəṙran → kasampəran*
- interpret any *ulu ricem* in **Sanskrit** segments as plain *m*
- normalize vowel length of words borrowed from Sanskrit as well as indigenous words
- in **Old Sundanese** context, any vowel marker added to a ligature with *pasangan y*, intended to be pronounced before the *y* (*Cyi* and *Cye*) is to be rendered as intended (*CViya* or *CVeya*), for example *syi → siya*, *rye → reya*.

3. Variant readings deemed insignificant in Indonesian texts

Unless there is any other reason to report witness readings, feel free not consider any instances of the following list as insignificant variants and hence to ignore them altogether in your apparatus (§):

- the difference between <lem> and <rdg> would concern only the normalization implemented pursuant to the above rules for the constituted Old Javanese text
- the difference between <lem> and <rdg> would concern only presence/absence of comma punctuation
- the difference between <lem> and <rdg> would only concern vowel length
- the difference between <lem> and <rdg> would only concern difference between single and geminate consonant
- the difference between <lem> and <rdg> would only concern the fact that one or more witnesses show a *ṁ* on the wrong syllable without the misplacement of *ṁ* generating any possible meaning
- the difference between <lem> and <rdg> would only concern difference between *dm* and *nm*
- the difference between <lem> and <rdg> would only concern sibilants (*ś/ṣ/s*)
- the difference between <lem> and <rdg> would only concern *a* and *ə*
- the difference between <lem> and <rdg> would only concern aspirated/unaspirated consonants
- the difference between <lem> and <rdg> would only concern intrusion of *h* in *pasangan* for vowel initial suffixed (type *tinulishakən* for *tinulisakən*)
- the difference between <lem> and <rdg> would only concern dental/retroflex consonants
- the difference between <lem> and <rdg> would only concern *e/ai* or *o/au*
- the difference between <lem> and <rdg> would only concern *Cva/Co* or *Cya/Ce* (with C standing for any consonant)
- the difference between <lem> and <rdg> would only concern *Cṛ/Cra/Crə/Cru* (with C standing for any consonant)
- the difference between <lem> and <rdg> would only concern *a/aha/ā*
- the difference between <lem> and <rdg> would only concern the difference *v/uv/h/Ø*
  - include under this heading variants such as
        mavyavahāra/mavyava:ra
        mamuhaka/mamvaka
        sa:kṣyakna/sakṣihakna

As per §5.4, the reading of any @wit inside <lem> should in most circumstances be silently normalized on the same points. E.g., even if witness L actually shows manavādi, it may still be cited in support of a <lem>mānavādi</lem> because the difference between *ā* and *a* is among the cases listed above.

As per §5.5.1, strict transliteration is to be used when you do cite any readings from a witness in <rdg>.

Appendix B: Language Codes

Language	Code
Undetermined language	unknown
Arabic	ara
Balinese, old or modern	ban
Batak, old or modern	btk
Burmese, modern	mya
Burmese, old	obr
Cham, modern (of Phanrang)	cjm
Cham, modern (of Cambodia)	cja
Cham, old (also known as "Old Cam")	ocm
Dutch	ndl
English	eng
French	fra
German	deu
Indonesian	ind
Japanese	jpn
Javanese, modern	jav
Javanese, old	kaw
Kannada, old or modern	kan
Khmer, Middle	xhm
Khmer, modern	khm
Khmer, old	okz
Malay, modern (Bahasa Malaysia)	zlm
Malay, old	omy
Mon, old	omx
Oriya	ori
Pali	pli
Prakrit	pra
Pyu	pyx
Sanskrit	san
Sasak	sas
Sundanese, old	osn
Tagalog, old or modern	tgl
Tamil, old or modern	tam
Telugu, old or modern	tel
Vietnamese	vie