Encoding Guide for Critical Editions

Draft Version 0.1, 2020-2021

Giulia Buriola, Arlo Griffiths & Axelle Janiak

Version History

Author(s), Contributor(s)	Version	Changes	Date
Giulia Buriola, Arlo Griffiths, Axelle Janiak	0.1	Redaction of the first draft	2020-2021

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).

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 normalized/critical text in another, or both combined in one file — is a matter of case-by-case appreciation, and something the project will have to experiment with.

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 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

---

[1] We will not dwell here on the reasons why a scholar might prefer to edit a text digitally. On this subject, 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.

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 project's 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). 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 *Encoding Guidelines for Critical Editions* documents the choices made by the DHARMA project from among these options.[4] It is in some cases a specialisation 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 DHARMA 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 links to the latest release.

EGC	DHARMA Encoding Guide for Critical Editions
EGD	DHARMA Encoding Guide for Diplomatic Editions
FNC	DHARMA Files Naming Conventions
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 chooses to treat readings from manuscripts and published editions on equal footing. This choice, which may appear counterintuitive to some, is based on both theoretical and practical considerations. Theoretically, it seems problematic to make such a distinction 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 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

[4] In making our choices, we have made use of the following guides in ascending order of frequency:
<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://digitallatin.github.io/guidelines/LDLT-Guidelines.html>

metadata section (<teiHeader>) of each file, where relevant in interaction with bibliographic data stored in the project's Zotero Group Library.[5]

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, commentary and bibliography 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, translations, commentary and bibliography respectively under §<teiHeader> — Recording Metadata, §translation, §commentary and §bibliography.

Templates for all those files are provided in the project-documentation repository of the DHARMA project.[6] 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 you want to link together. Finally, provide an attribute @type with the value "edition", "translation", "commentary" or "bibliography" on this root.

Example 1	
<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.[7]

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 2

[5] For the latest release of the Zotero Guide, check the GitHub project-documentation repository: https://github.com/erc-dharma/project-documentation/tree/master/guides/zotero

[6] Currently, four templates in their first version are available in the templates folder, under CriticalEditions-mss.

[7] Currently, critical editions are validated with the tei all schema. Be aware that at least two lines will be given but more can be added depending on the technologies used.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="https://raw.githubusercontent.com/erc-dharma/project-documentation/master/schema/latest/DHARMA_CritEdSchema.rng" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model
href="https://raw.githubusercontent.com/erc-dharma/project-documentation/master/schema/latest/DHARMA_CritEdSchema.rng" 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="subtitle">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>
         </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/15HFxHJTOzIU1UDyVrB2yQYJ5wl6JyEshEkYgg5qwj8M/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>
```

<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.

<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.

<titleStmt> — Titles and people involved

<title> — Declaring the title

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

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.[8] See the examples below.

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 with the relevant ISO language code.[9] To this can be added a corresponding <title type="main"> 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. (See §@xml:lang – identifying languages for more on use of @xml:lang.) Finally, use another element <title @type="sub"> to state a subtitle for your edition that may include further description of the text and of the edition, in plain English.

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

Example 3

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

Example 4

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

[9] See EGD §10.3 and Appendix D.

```
<titleStmt>
        <title type="main" xml:lang="san-Latn">Naimittikakriyānusandhāna</title>
          <title type="main">Exposition of Occasional Rites</title>
          <title type="sub" subtype="editorial">Digital Critical Edition and Translation of a Śaiva Ritual
Manual</title>
 </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 text. 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 title to the proper segments of the edition. When there is more than one dependent, you need to add an extra @n attribute, in order to link the title to the text concerned.

Example 5

```
<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" xml:lang="eng">Digital Critical Edition of a Sanskrit Medical Text with two
Commentaries</title>
</titleStmt>
```

<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.[10] 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 6

[10] 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="geco" 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 7

```
<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="#geco">
          <forename>George</forename>
          <surname>Cœdès</surname>
        </persName>
</respStmt>
```

<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 Convention (FNC)[11] 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

```
<publicationStmt>
        <authority>DHARMA</authority>
        <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>
```

<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.

Each 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.

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

Any specific information on the typographical rendering (e.g. use of superscript; see §Typographical formatting), 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

```
<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 §Recording witnesses.

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 EGD §10.4.5.

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, <physDecs> must always be filled in, and <msIdentifier> must contain an <idno> while it may contain other elements as well, such as <settlement> or <repository>.

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.

Example

```
<msContents>
  <msItem n="1">
    <locus>fol. 1r-29v</locus>
    <title>Vratiśāsana</title>
  </msItem>
  <msItem n="2">
    <locus>fol. 1v-11r</locus>
    <title>Ṛṣiśāsana</title>
  </msItem>
  <msItem n="3">
    <locus>fol. 11r-21v</locus>
    <title>Śaivaśāsana</title>
  </msItem>
  <msItem n="4">
    <locus>fol. 1v-5r</locus>
    <title>Śīlakrama</title>
  </msItem>
</msContents>
```

Finally, you may also provide a <colophon> element to record any colophon. You should specify an attribute @xml:lang with the corresponding language code (see §Appendix C 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 §colophons.

Example

```
<msContents>
    <msItem>
        <colophon xml:lang="ozn-Latn">iti siksa guru, ña, lamun vruh di vvitniñ bənər tar kəna ku upaḍrava, lamun tar vruh di vvitniñ bənər, kəna ku na upadrava, nihan kavuvusan siksa guru, ña, kayatnakna saṅ sevaka dharma</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

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

It is not recommended to combine unstructured prose descriptions with the usage of more specialized elements. However, if you cannot avoid it,[12] specialized and generic elements may be combined in a single <physDesc>. Remember 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 example:

Example

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

[12] 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.

```
    </supportDesc>
  </objectDesc>
  <handDesc>...</handDesc>
</physDesc>
```

Among the numerous elements that can be nested inside your <physDesc>, the only one that is mandatory is <handDesc> (see §<handDesc>), 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.

Following the <physDesc>, the element <msPart> can be used to declare that your witness belongs to a manuscript composed of parts that were originally separated and that have been bound together at a later stage, as shown in the example below.

Example

```
<msDesc>
  [...]
  <physDesc>
    [...]
  </physDesc>
  <msPart>
    <msIdentifier>
      <idno>ms. 10066-77 ff. 140r-156v</idno>
    </msIdentifier>
    <msContents>
      <summary xml:lang="la">Physiologus</summary>
      <textLang mainLang="la">Latin</textLang>
    </msContents>
  </msPart>
  <msPart>
    <msIdentifier>
      <idno>ms. 10066-77 ff. 112r-139r</idno>
    </msIdentifier>
    <msContents>
      <summary xml:lang="la">Prudentius, Psychomachia</summary>
      <textLang mainLang="la">Latin</textLang>
    </msContents>
  </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 catalogued 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

```
<msDesc>
   <msIdentifier>
      <msName xml:lang="la">Codex Suprasliensis</msName>
   </msIdentifier>
   <msFrag>
      <msIdentifier>
         <settlement>Ljubljana</settlement>
         <repository>Narodna in univerzitetna knjiznica</repository>
         <idno>MS Kopitar 2</idno>
      </msIdentifier>
      <msContents>
         <summary>Contains ff. 10 to 42 only</summary>
      </msContents>
   </msFrag>
   <msFrag>
      <msIdentifier>
         <settlement>Warszawa</settlement>
         <repository>Biblioteka Narodowa</repository>
         <idno>BO 3.201</idno>
      </msIdentifier>
   </msFrag>
   <msFrag>
      <msIdentifier>
         <settlement>Sankt-Peterburg</settlement>
         <repository>Rossiiskaia natsional'naia biblioteka</repository>
         <idno>Q.p.I.72</idno>
      </msIdentifier>
   </msFrag>
</msDesc>
```

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

```
<sourceDesc>
      <listWit>
        <witness xml:id="A1">
          <abbr>A<hi rend="superscript">1</hi></abbr>
          <msDesc>
```

```
            <msIdentifier>
              <settlement>Leiden</settlement>
              <institution>Leiden University</institution>
              <repository>Leiden University Libraries</repository>
              <location>Special Collections</location>
              <idno>Or. 23732</idno>
            </msIdentifier>
            <msContents>
              <msItem>
                <author>Śāntarakṣita</author>
                <title>Vipañcitārthā</title>
                <colophon xml:lang="ozn-Latn"></colophon>
              </msItem>
              <summary>summary of the contents of the witness</summary>
            </msContents>
            <physDesc>
              <objectDesc>
                <p>Palm-leaf manuscript. 89 leaves in Siddhamāṭṛkā 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.

<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

```
<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:

Example

```
<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"  scriptRef="class:57470 maturity:83215">
                <abbr type="hand">A<hi rend="superscript">1</hi>H<hi rend="subscript">3</hi></abbr>
             <p>generally in blue ink, Sundanese script; this hand is more present starting in the second half the
manuscript</p>
                </handNote>
```

```
            </handDesc>
        </physDesc>
    </msDesc>
</witness>
```

The @xml:ids 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 (§Scribal additions and deletions in Critical Apparatus). 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.

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. 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.

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 (§Recording witnesses), you can use a <ptr/> tag as a shortcut to name the witness. See the following examples:

Example

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

<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>:

Example

```
<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="to_be_filled_in_later"/>
<!-- UPDATE the link once the release is made official -->
        <listPrefixDef>
<prefixDef ident="bib" matchPattern="([a-zA-Z0-9\-\_]+)" replacementPattern="to_be_filled_in_later"">
 <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=to_be_filled_in_later"">
<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=to_be_filled_in_later"">
<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 §Referring to digital sources external to your file 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.

<projectDesc> — Describing the project

The <encodingDesc> should minimally have the sub-element <projectDesc>. This is used to explain, in one or two 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
<encodingDesc> <projectDesc> 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).   It has been carried out as part of my research towards a PhD thesis defended on the DDth of MM, YYYY[13] at the École Pratique des Hautes Études, Paris. </projectDesc> </encodingDesc>

<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.[14] 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 . 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 .

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

[13] 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 in <projectDesc>. However, we do not advise our editors to do so.
[14] You can find the relevant document at the following link: https://hal.archives-ouvertes.fr/halshs-02272407.

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.[15]

Example

```
<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>
```

<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

[15] 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 explain in one of the guides of the DHARMA project.

```
<samplingDecl>
   <p>Explanation</p>
</samplingDesc>
```

<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.[16]

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 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,[17] e.g. v01 for the first version.

Example
<schemaRef type="guide" key="EGCv01" url="*to_be_filled_in_later*"/>

<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. For more on this topic, see the section §Apparatus. The DHARMA project uses only a single methodology for encoding apparatus in critical editions, so you can leave the following element in the template untouched.

Example
<variantEncoding method="parallel-segmentation" location="internal"/>

<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 C.[18] However, since the element <langUsage> cannot be used to indicate the *actual* distribution of languages within the text you

[16] 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 constraints written for a project.

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

[18] See also EGD §10.3 and Appendix D.

are editing, your digital edition will have to declare a base language (§Structuring the text) and then mark containers for portions of text using a different language by the attribute @xml:lang choosing the values from the same list of language codes. 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

```
<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="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>
```

<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[19] 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

[19] All identifiers are available at https://github.com/erc-dharma/project-documentation/blob/master/DHARMA_IdListMembers_v01.xml

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. During the initial phase of the project, the recommended value is “draft”.

Example

```
<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>
```

Edition Text

<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 C, 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.[20] The basic layers of encoding within the <body> are provided by , 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 with @type=“edition”, @type=“translation” and so forth, as is done in the case of EpiDoc editions for

[20]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.

inscriptions. The following, then, represents the minimal containing structure for your critical edition:

Example

```
<text xml:space="preserve" xml:lang="san">
        <body>
          <div>EDITED TEXT</div>
          <div>EDITED TEXT</div>
        </body>
 </text>
```

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

– Sections of text

As previously stated, within your edition, textual divisions shall be wrapped in . 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.

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 with @type="dyad" to bind the base text and the dependent text together. See below for the type "dyad" (§Dependent text matching its base exactly). **Should you need any other types of divisions lower than chapter/canto but higher than the individual paragraph or stanza, please let us know**.

As for units below the 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 §Paragraphs. Any stanzas contained in your text must be wrapped in the element <lg> (line group), as explained in §Encoding stanzas; in this case, the lower textual unit within <lg> will be <l> (line), as explained in §Encoding lines. 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 §Anonymous Blocks.

It is necessary to add an attribute @n on one or more structural levels of your edition, but it seems difficult to give general rules. We recommend that you consult the project's XML-TEI Data Manager and your academic supervisor to decide what numbering scheme works best for your text. Only Arabic numbers shall be used as values of @n, and be shown in display.

For file-internal reference, an automated @xml:id will be provided to identify each level of the edition (§Identifying Pattern for Containers).

For the special case of unnumbered to wrap around <p> or <ab> for invocations and colophon (§Invocation and §Colophons).

<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

```
<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 (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 .

Prose Containers

<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.[21] As many elements <p> as needed can be encoded inside a . A <p> can contain other elements, but it cannot contain other <p>s: nesting is prohibited in this case. The element <p> will be provided with an XSLT-generated @xml:id; see §Identifying Pattern for Containers. Do not encode any attribute @n on <p>.

[21] EGD §2.2.1.

<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 labelling a unit as <p>. As with <p>, so too each <ab> will be furnished with an XSLT-generated @xml:id; see §Identifying Pattern for Containers. Do not encode any attribute @n on <ab>.

In the DHARMA project's critical editions, we foresee the use of the <ab> element 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.[22] You may occasionally need to make use of <ab> in invocations and colophons, as shown in the next two subsections.

Invocations

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 without any @type while assigning @type="invocation". In this case, assign no number to the . However, in the case of invocations to specific divisions of the text, the invocation must be included at the top of the relevant but after any furnished <head>.

Example
 <ab type="invocation" xml:lang="san-Latn">namo gaṇeśaya|</ab>

Example
 <ab type="invocation" xml:lang="san-Latn">siddham|</ab>

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

Colophons

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>,

[22] EGD §2.2.2.

and include these as final child within the to which they belong, or within a separate 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".

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

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.

Example
<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. All the elements discussed here have been introduced in greater detail in the EDG, §2.3.[23]

<lg> — stanzas

One <lg>, or line group, corresponds to a single stanza as a whole. It can be nested either within a or within a <p>. The <lg> can contain one or more <l> elements and has the mandatory attribute @met to identify the meter of the stanza by its conventional name. 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.[24] 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. If you do encode @n on <lg>s that are part of your text, then you must do so consistently. As for , <p> and <ab>, an XSLT-generated attribute @xml:id will be provided also for <lg>, see §Identifying Pattern for Containers.

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, as per the guidance in EGD §2.3.8.

[23] Please note that the present guide follows the vocabulary established in EGD, §2.3.1.
[24] EGD, §2.3.3.

@met — metrical patterns

The metrical pattern, or “templates”, of stanzas will be encoded according to the conventional name of meters as values of the attribute @met in the element <lg>. The list of names allowed is given in the EGD, Tables 3 and 6 of Appendix B. Contact the EGD’s authors if you require any additions.

If you cannot put a name to the meter in your edition, you can still use @met to record the prosodic template using prosodic code. Follow the conventions presented in Table 2 of Appendix B in the EGD. This prosodic notation is also used for the attribute @real on the <l> element, to which we now turn.

<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 verses, you should use lower case Roman letters (a, b, c, d). In other traditions of versification (e.g. Old Sundanese or Tamil), use a single Arabic number per line. Metrically irregular lines are to receive the attribute @real (EGD §2.3.5).

@real – actual metrical patterns

The meter of individual verse lines shall not be encoded separately so long as they conform to the meter encoded for the stanza to which they belong. But when a line deviates from the expected pattern, it is possible to record the phenomenon using the attribute @real on the verse line concerned. The value of the attribute should be the prosodic pattern given following the notation provided in the Table 2 of Appendix B of the EGD.

@enjamb – enjambment

In the context of the DHARMA project, enjambment had been defined as “the occurrence of a line break within a word (usually between members of a compound; rarely within a morpheme)” (EGD §2.3.1). 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 (see §EGD Example 2.3.9.A and Example 2.3.9.C). 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 ... *piṭ̣rbhiḥ 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	
<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 shows above without further comment, see §<editorialDecl> — Recording specificities of your own work. 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
<lg> [...] <l n="d">pitṛbhiḥ saha majjati||</l> </lg> <ab>iti<note>All witnesses read <foreign>majjatīti ||</foreign>.</note></ab>

Spoken Text

Dramatic elements

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

Plays are usually organized in divisions in which various characters speak, taking turns. Each individual speech should be encoded with the element <sp>. It requires the element <speaker> as 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 could be assimilated with stage directions 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.

Example

```
<sp>
        <speaker>rāmaḥ</speaker>
        <stage>tataḥ praviśaty 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>
```

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.

Dialogue

If you are facing a text with dialogue, you might want to identify the interlocutors using the editorial element <label> with the attribute @type="speaker". This element provides an easy way to indicate each change of speaker as well as the name of the speaker in both verse and prose texts. Note that the element <label> should always be a child of , <ab>, <p> or <lg>. We foresee two cases in which it can be useful: the text underlines the turn to speak and a text in which they are absent, but you wish to add them for clarity's sake.

In the first scenario, any interlocutors explicitly mentioned should be encoded as <persName> associated with the attribute @key to provide the normalized form of the name, if necessary. Then, wrap the name with <label>.

Example

```
<ab type="speaker"><persName key="to-be-added">Devī</persName> devy uvāca</ab>
```

In the second case, add the name of the interlocutor written as you expect it to be displayed. It will be treated as an editorial element and not as an actual part of the text.

Example
<label type="speaker">Devī</label> devy uvāca

Mixed prose and verse

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

Example

```
<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>
```

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

```
<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 XYZ 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.

Base text and dependent text

It may happen that a single text consists 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. However, 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 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.

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 C for the relevant ISO codes).

Example

```
<div>
   <quote type="base-text" 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>
   <p>dependent text</p>
</div>
```

Base text implied in dependent text

Another scenario you may confront in your digital edition is when the base text is not transmitted as such in the manuscripts of the dependent text, and was presumably never present at any stage of the dependent text's development. If you deem the base text to be implicitly present and feel that making it explicitly visible is useful for the readers' appreciation of the text 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

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

Base text untransmitted by author of dependent text

It may also occur that a dependent text lacks any trace of its author's awareness of a segment of base text, but that you wish to show this segment nonetheless. In this case, wrap the content of <quote> in a <supplied> tag, applying the attribute reason with the value “omitted”.

Example

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

Dependent text matching its base text exactly

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 .

Higher level divisions of text wrapped in have been discussed in § — Sections of text with the appropriate @type attribute to be used for such sections. The same element 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 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 . When necessary, you can number your lower-level with the attribute @n. Note that both attributes @type and @n are optional at this level of XML structure.

The content of the , 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 .

Example

```
<div n="1">
```

```
  <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>
  <p>kunaṅ hulaha saṅ prabhu, uniṅānira taṅ vyavahāranikaṅ rāt, harovaṅa ta sira brāhmaṇa vihikan maṅaji,
lavan mantri <app><lem
    wit="#L">riṅ</lem><rdg wit="#K">,</rdg></app>vruh <app><lem wit="#L">mavivekā</lem><rdg
    wit="#K">mavikvekā</rdg></app>, sulakṣaṇaha ta sira tumamaha riṅ sabhā.</p>
</div>
```

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 entity 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.

The numbering of your should increase by one, regardless of the numbering inherent in any base text that it may contain. 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”>.

Example

```
<div type="dyad" n="5">
  <quote type="base-text" xml:lang="san-Latn">
    <supplied reason="implied">
      <lg>
        <l>teṣām ādyam ṛṇādānaṁ nikṣepo ‘svāmivikrayaḥ |</l>
        <l>saṁbhūya ca samutthānaṁ dattasyānapakarma ca ||</l>
      </lg>
      <lg>
        <l>vetanasyaiva cādānaṁ saṁvidaś ca vyatikramaḥ |</l>
        <l>krayavikrayānuśayo vivādaḥ svāmipālayoḥ ||</l>
      </lg>
      <lg>
        <l>sīmāvivādadharmaś ca pāruṣye daṇḍavācike |</l>
        <l>steyaṁ ca sāhasaṁ caiva strīsaṁgrahaṇam eva ca ||</l>
```

```
        </lg>
      </supplied>
    </quote>
  <p>Some dependent text</p>
</div>
```

Base text without dependent text

When the dependent text skips, or appears to skip, one or more sections of the base text, you must make a choice between one of the following two representations.

(1) Including more than one segment of base text in a single .

Example

```
<div>
  <quote type="base-text" xml:lang="san-Latn">
    <lg n="272" xml:id="svayambhu_07.13.02.01">
      <!-- Note that this first <lg> encodes the segment of base text without commentary -->
      <l>verse line</l>
      <l>verse line</l>
    </lg>
    <lg n="273" xml:id="svayambhu_07.13.02.02">
      <l>verse line</l>
      <l>verse line</l>
    </lg>
    <listApp type="parallels">
      <app/>
    </listApp>
  </quote>
  <p>Dependent text on <ptr target="#svayambhu_07.13.02.02"/> and, according to the editor, at least partly on
<ptr target="#svayambhu_07.13.02.01"/>.</p>
</div>
```

In the scenario encoded here, the stanza bearing the attribute @n with value “272” finds no explicit reference in the dependent text, since the latter openly comments only on <lg n=“273”>. However, the editor feels that the dependent text silently comments on <lg n=“272”> and wants to mark this up. The same structure could also be used if the dependent text openly comments on stanza 272 but does so only after stanza 273 has been quoted.

(2) Reserving a separate for every segment of base text, regardless of whether any dependent text is given. This scenario might be adopted if the editor believes that the dependent text intentionally does not comment on a stanza.

Example

```
<div xml:id="ID_01">
    <quote xml:id="ID_01.01" type="base-text" xml:lang="san-Latn">
      <!-- the first <lg> element would be the base text without commentary-->
      <lg xml:id="ID_01.01.01" n="272">
        <l>verse line</l>
        <l>verse line</l>
      </lg>
    </quote>
  </div>
  <div xml:id="ID_02">
    <quote xml:id="ID_02.01" type="base-text" xml:lang="san-Latn">
      <lg xml:id="ID_02.01.01" n="273">
        <l>verse line</l>
        <l>verse line</l>
      </lg>
      <listApp type="parallels">
        <app/>
      </listApp>
    </quote>
    <p xml:id="ID_02.02">Dependent text on <ptr target="#ID_02.01.01"/> that does not refers to <ptr
target="#ID_01.01.01"/>
    </p>
  </div>
```

Solution (2) tends to emphasize the absence of dependent text on a given stanza, while solution (1) tends to dissimulate it.

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 (§Base text without dependent text) can be used to encode such scenarios.

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

Example

```
<div n="31">
    <quote type="base-text" xml:lang="san-Latn">
      <lg n="31">
```

```
            <l>verse line</l>
            <l>verse line</l>
         </lg>
         <listApp type="parallels">
            <app/>
         </listApp>
      </quote>
      <p>Dependent text related to the preceding p of base text.</p>
      <p>Further dependent text that doesn't correspond to the preceding p of base text.</p>
   </div>
```

(2) If, however, you reckon that the text is too unconnected from the previous enclosing , you can open a separate to enclose this segment of dependent text on its own.

Example

```
<div 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 p of base text.</p>
</div>
<div n="32">
      <p>Further dependent text that doesn't correspond to the preceding p of base text.</p>
</div>
```

Note that introductory materials are always expected to form their own . 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.

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. However, 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

```
<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 (to which the attributes @type and @n can be added as needed), then divided into subsections, with a for each pair consisting of base text and commentary. The internal organization of the follows the model established in §Dependent text matching exactly the base text. 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.

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 §Identifying Pattern for Containers, 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.

Example

```
<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 §Stanzas interrupted by dependent prose and contact the XML-TEI Data Manager.

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 §Base prose interrupted by dependent prose. 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 that is segmented into four parts by the dependent text requiring four subsections :

Example

```
<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

```
<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

```
<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

```
<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>
            <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.

*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 (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 §Critical Apparatus), 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
<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 §Identifying pattern for containers), 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
<term xml:id="term004" xml:lang="san-Latn">nikṣepa</term> some text <gloss target="#term004">gloss</gloss>

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. You will have noticed that the previous chapter has covered the relevant encoding needs. It is important 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. This method is explained in EGD§9.1.2 and §9.1.3.
2. The **double-end-point-attached method**[25] 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**,[26] 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 bound to a block of text without being embedded inside it) or the third method (meaning that the entries are embedded directly into the encoded text itself)[27] as well as a combination of the two. In what follows, we cover only method 3; however, in §Encoding an external apparatus verse-by-verse, we introduce method 1, which is easier to encode but only useful if you are editing a verse text.

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

[25] See TEI Guidelines, here: https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPDE (accessed: 2020-12-17).
[26] See TEI Guidelines, here: https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPPS (accessed: 2020-12-17). This method seems to be favored as an encoding technique for multi-witness editions for both specific tools and projects that have been applying it.
[27] 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.

```
preceding text
<app>
   <lem wit="#K #L">lemma</lem>
   <rdg wit="#M">variant reading</rdg>
   <note>optional observation on adopted and/or rejected readings</note>
</app>
following text
```

2. None of the witnesses supports the adopted reading

Example

```
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 on adopted and/or rejected readings</note>
</app>
following text
```

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

Example

```
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 on adopted and/or rejected readings</note>
</app>
following text
```

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

Example

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

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

Example

```
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**, padāny aṣṭadaśetani **L**, *om*. **M**.

These five examples have introduced the elements <app>, <lem>, <rdg>, <gap> and <note> and the attributes @reason, @type and @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, the lemma in philological terminology.[28] Generally, one or more variant readings will follow and be encoded inside an element <rdg> although, as shown in the fourth example above, it is possible to create <app> entries consisting only of <lem> and <note> without any <rdg>. The lemma and any corresponding variant readings are to be explicitly encoded together, i.e. wrapped within <app>, so that their interconnectedness is made clear.

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

Do not number your <app> elements: 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.

Which variant readings to record and which to ignore

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

[28] Note that the term ‘lemma’ in this context has a different meaning from the one intended in §Lemmas of base text in dependent text.

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 automatic 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 §Readings involving scribal correction) only one instance, to be recorded with @wit, whether in <lem> or in <rdg>, of each witness listed in your <witList>. Be aware that, for any embedded <app>, the children <app> will only inherit the witnesses declared on the element anchoring this embedded <app>. Nonetheless, all the witnesses should be declared on the parent <app>.

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 the Appendices, we furnish overviews of types of variants that we recommend ignoring in the respective cases of editing a South Asian Sanskrit text and of editing a Sanskrit and/or vernacular text from Java or Bali. These lists could be reused as the basis for your own further customized lists to be included in the <editorialDecl> of your edition (§Recording specificites of your own work). 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.

Strategies to encode variants

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 ( ). If such a phenomenon occurs within your text, you need to exercise extra care while encoding the relevant apparatus entry, especially for cases of nested apparatus entries.[29] By the same logic,

[29] For instance, it would be an error if an apparatus entry put a <p> element inside another <p> or an <l> element inside another one. In the same way, be careful to avoid declaring different structural levels as variants in the same apparatus entry; for instance a element can't be declared in a <lem> and be matched with a <p> element in the corresponding <rdg>. The same could be said about the witnesses. You can't provide information about one witness in a nested apparatus entry if it isn't identified in the @wit

you must be careful about recording phenomena with different requirements than the default behavior.[30]

Strategies to encode apparatus entries across block boundaries

When your apparatus entry straddles a block-level container, such as , <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 §Identifying pattern for containers.

Example

```
<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>
```

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>.

attribute of the enclosing <lem>. If you do so, your variant is controversial regarding the inheritage pattern between parents and children elements, on top of being an unnecessary repetition of the <rdg> of the enclosing <app>.

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

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 will often 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, 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 example XXX.

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 mark by default as the lemma the text of each complete eight-syllable quarter-verse (*pāda*), rather than producing five or six little *lemmata* for each half-line.

Ellipsis in the lemma's display

If you need 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, record the lemma in full but, immediately after the element <lem>, insert an element <note> with a specific attribute @type="altLem" to express the truncated form of the lemma.[31] Let's imagine this lemma:

Example
<app><lem>saṅ brāhmaṇa saṅ vruh riṅ aṣṭādaśavyavahāra,</lem></app>

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

Example
<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>

[31] 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. (https://www.digitalmanuscripts.eu//wp-content/uploads/2017/09/04-Textual-variants-MB.pdf, accessed: 2021-01-15).

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 §<note> – Generic Notes).

Example

```
 ... śūdra,
<app>
  <lem type="absent_elsewhere" 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> 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 ...
```

Display —

^1. macaṅkramāta ... saṅ prabhu**]** only in **A C**.

On the use of @type="absent_elsewhere" on <lem>, see §Absence of a witness for a lemma due to larger physical lacuna. On the use of @type on <lem>, see §Other @type values usable with <lem>.

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:

- hyphenfront, hyphenback, hyphenaround
- circlefront, circleback, circlearound

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

Comparison of the two suggests that recourse to @rend="hyphenfront" 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
any<app><lem type="emn" 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>.

Normalization, emendation and conjecture of the transmitted readings

If the adopted reading is not directly supported by any of the witnesses, then you must apply to the <lem> an attribute @type. The permitted values are “norm”, “conj” and “emn”, respectively for normalization, emendation and conjecture.

- “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 however 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 (§Ignoring insignificant variant readings).
- “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 example of how the @types can be used:

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

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

```
</app>
```

Other @type values usable with <lem>

The element <lem> can take the attribute @type with the following values:

- “absent_elsewhere”
- “refomulated_elsewhere”

The first value allows to generate an apparatus entry pointing out that the given lemma is only found in one witness (see the section about §Absence of a witness for a lemma due to larger physical lacuna) and the second, described in §Adopting word order from one witness but a reading from another, allows to underline that you are adopting a lemma that mixes the word order from one witness with the specific reading of one or more words from another witness.

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 §Identifying and describing your sources), preceded by the “#” that is needed to make the link to the list of sigla defined in your <witList>. 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

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

In the next chapter, we will discuss the recording of variant readings in detail.

Variant readings

In this chapter, 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. Other phenomena such as interpolations or transpositions will be found in other sections of this guide, $Reporting extraneous text, and §Encoding transpositions of text, as will instruction on how to supply untransmitted text segments, §Supplying untransmitted text.

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:

- you may remove any milestone elements (<lb/> and <pb/>)[32] 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 manuscript readings as diplomatically as you can or with as much normalization as you deem necessary/useful.

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
<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**

[32] To know more about milestone elements, check the section on §Milestone elements in this Guide.

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 lemma but express it through a different lexeme or with different (yet equivalent) morphological means. The values we propose for @cause are the following:[33]

- eye-skip (see §Text of lemma absent from a witness due to scribal omission)
- line_omission (see §Text of lemma absent from a witness due to scribal omission)
- haplography
- dittography
- lexical
- morphological
- ordinal
- syntactic
- subtractive
- additive

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
dravya<app><lem rend="hyphenfront" wit="#L">niṅ ratu</lem><rdg wit="#K #M" cause="lexical">saṁ prabu</rdg></app> tinuvavakən,

Contact the project's XML-TEI Data Manager if the proposed values of @cause do not meet your requirements.

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,

[33] See Digital Latin Library Guidelines, 12.15, Tagging Readings for Analysis, https://digitallatin.github.io/guidelines/LDLT-Guidelines.html, but note that the Digital Latin Library offers these values to be used with the attribute @ana, while in the case of DHARMA, 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 https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.textCritical.html.

while the result of correction is conventionally labelled *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 ignore 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>. If you are aiming at a result that, in print, would resemble the following:

vijayaḥ P_{pc} Q **]** vajayaḥ P_{ac} vajayo R

We recommend you opt for one of the following solutions.

1) Use of the element **<witDetail>.** This device serves for what the TEI Guidelines call a "specialized note", 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.

Example

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

Example

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

The following examples concern a scenario where the correction was applied by a second hand:

Example

```
<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>
```

Example

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

The display could be something like this, with labels sm and pm meaning *secunda manus* and *prima manus*:

vijayaḥ **P**$_{\text{pc-sm}}$ **Q]** vajayaḥ **P**$_{\text{ac-pm}}$ vajayo **R**

2) Use of the attribute **@varSeq.** Within <app> you are allowed to treat each state of the text as a distinct reading; therefore, to indicate the various stages of the representation of the text in a witness, such as *ante* and *post correctionem* readings, the <rdg> can be given the attribute @varSeq, attributing a sequence to the various changes in your witness. The following two examples would yield the same displays as the encodings just described with <witDetail>.

Example

```
<app>
        <lem xml:id="lem001" wit="#Q">vijayaḥ</lem>
        <rdg wit="#P" varSeq="1"><del>va</del>jayaḥ</rdg>
        <rdg wit="#P" varSeq="2" sameAs="#lem001"><add>vi</add>jayaḥ</rdg>
        <rdg wit="#R">vajayo</rdg>
      </app>
```

Example

```
      <app>
        <lem xml:id="lem001" wit="#Q">vijayaḥ</lem>
        <rdg wit="#P" varSeq="1" hand="#P_H1"><del>va</del>jayaḥ</rdg>
        <rdg wit="#P" varSeq="2" sameAs="#lem001" hand="#P_H2"><add>vi</add>jayaḥ</rdg>
        <rdg wit="#R">vajayo</rdg>
      </app>
```

The following example concerns a case where the part of the *ac* reading that has undergone scribal corrections is no longer legible.

Example

```
<app>
    <lem xml:id="lem001" wit="#Q">vijayaḥ</lem>
    <rdg wit="#P" varSeq="1"><gap reason="illegible" quantity="1" unit="character"/>jayaḥ</rdg>
    <rdg wit="#P" varSeq="2" sameAs="#lem001">vijayaḥ</rdg>
    <rdg wit="#R">vajayo</rdg>
</app>
```

Let us now clarify the purpose of several other elements in this encoding: the @xml:id in <lem> and the attribute @sameAs in one of the <rdg> elements are used to make the connection between witness Q and witness P clear: they both give the lemma, i.e. *vijayaḥ*. However, while Q has the lemma's spelling as such, P only has it as the result of correction by the scribe from *vajayaḥ* to *vijayaḥ*. You shall encode such processes of correction in your markup by encoding two <rdg>s, which will represent, respectively, the labels P_{ac} and P_{pc} from the display example above. Both <rdg>s should bear the attribute @varSeq: the one for *ac* shall have as its value the number "1", while the one for *pc* shall have the value "2". The latter should bear the attribute @sameAs, whose value matches the @xml:id in <lem>, indicating that this corrected reading in P also supports the adopted reading. As in all other cases where @xml:id is used, a unique label must be used for every instance. Please follow the model of lem001 shown here, and

make sure you assign an incrementally higher number for every new @xml:id that you create for a <lem>.[34]

Finally, within the <rdg> elements, you can use the full gamut of elements described in the chapter on the "Representation of primary sources" of the TEI Guidelines,[35] for which we refer also to EGD §4.4 and §5. In the examples above, we have used <gap>, <del> and <add>.

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.

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:

Example

```
<app>
  <lem wit="#L" type="reformulated_elsewhere">kunaṅ sakaton
    <app>
      <lem wit="#K">sakarəṅə denta</lem>
      <rdg wit="#L">sakaR̥ṅəntenta</rdg>
    </app> sākṣi, yen tuhu, śīghra
  </lem>
    <note type="altLem">kunaṅ sakaton ... śīghra</note>
    <rdg wit="#K">kunaṁ yan satya: kitta:, sakaton sakaR̥ṅə: denta:, tuhuta:, śigr̥</rdg>
    <note>The two witnesses offer two alternative ways of expressing essentially the same meaning.</note>
</app> ujaraknanta, ...
```

Display —

text

[34] You can use the regular expression [l][e][m]\d\d\d to search all occurrences of @xml:id="lem..." in your xml file and then assign a number one higher than the previous highest number.
[35] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html

kunaṅ sakaton sakarəṅə denta sākṣi, yen tuhu, śīghra ujaraknanta, ...

apparatus
^1. kunaṅ sakaton ... śīghra] Thus formulated in **L**, kunaṁ yan satya: kitta:, sakaton sakaṟṇə: denta:, tuhuta:, śigṛ **K**. • The two witnesses offer two alternative ways of expressing essentially the same meaning.
^2. sakarəṅə denta] **K**, sakaṟṇəntenta **L**.

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 reading of those witnesses had been adopted. In such a scenario, you may record the reformulation 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 attest to this reformulation.[36]

Example

<app><lem wit="#L" type="reformulated_elsewhere">śeṣaniṅ ḍaṇḍa, śeṣaniṅ *tukon, ika tan tumusa riṅ anak putuniṅ ahutaṅ</lem><note type="altLem">śeṣaniṅ tukon ... anak putuniṅ ahutaṅ</note><rdg type="#K">śeṣaniṁ ḍaṇḍa, śeṣa vlan·, Ika ta kabeḥ, yan matikaṁ mahutaṁ, tan yogya tagihən hanaknya</rdg><rdg type="#M">sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya</rdg><rdg type="#M">sesaniṁ, sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya</rdg><rdg type="paradosis" wit="#K #M">śeṣaniṅ daṇḍa, 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 –

^813 . śeṣaniṅ tukon … anak putuniṅ ahutaṅ] Thus formulated in **L**, śeṣaniṁ ḍaṇḍa, śeṣa vlan·, Ika ta kabeḥ, yan matikaṁ mahutaṁ, tan yogya tagihən hanaknya , sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya , sesaniṁ, sesa vəlyan, ika ta kabeḥ yan mati kaṁ mahutaṁ, tan yogya tagihən anaknya , śeṣaniṅ daṇḍa, śeṣa vlyan, ika ta kabeh, yan matīkaṅ mahutaṅ, tan yogya tagihən hanaknya **K M** • *Paradosis* of **K M**: śeṣaniṅ daṇḍa, 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*

[36] 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.

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*).

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, following the rules stated in EGD §5.3.1. Note that the only child allowed within <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
<rdg wit="#L642">ya ta mas <unclear>r̥kr̥t</unclear> masəkar 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
<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”. (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. You can encode segments whose reading is ambiguous with <choice> and <unclear> as explained in EGD §5.3.3. 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.

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” on @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”, but it can also be used with the value “component” depending on the context.[37] 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".

Example

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

Display —

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

Example

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

Display —

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

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:

Example

```
<app>
        <lem wit="#K">tinurunakən sakeń svarganya deniń devata kabeh</lem>
        <note type="altLem">tinurunakən ... kabeh</note>
        <rdg wit="#L" cause="eye-skip"><gap reason="omitted"/></rdg>
```

[37] See EGD, §5.4.5

```
        <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" 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 eyeskip you imagine. So let us propose the following slight modification:

Example

```
<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 eyeskip 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

```
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 example XYZ 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
<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).

Note that the list of permitted values of @cause is limited. See §Encoding causes of variance of reading.

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
<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 §Encoding (segments of) a reading as illegible.

Example
<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
<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**.

Text absent from a witness due to physical deficiency but with known metre

If any text lost or illegible because of a physical deficiency that cannot be restored, is known thanks to its prosodic pattern, the following solution applies: you can wrap the element <gap/> within a <seg> with the attribute @met. The value of the attribute should match the prosody conventions chosen by DHARMA as per Table 2 of the Appendix B of the EGD. To know more about this, see EGD, §5.4.4.

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

Textual loss within a block-level container

It may also happen that a lemma finds no corresponding reading in a given witness, if 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
<app> <lem wit="#K" type="absent_elsewhere">ya vacananta, inastuti- <app> <lem type="emn" rend="hyphenfront">kīrti</lem> <rdg wit="#L"><gap reason="lost"/></rdg>

```
                        <rdg wit="#K">-kitta:</rdg>
                    </app>
                    den bhaṭāra brahmā.</lem>
        <note type="altLem">ya vacananta, ... brahmā.</note>
        <rdg wit="#L"><gap reason="lost"/></rdg>
</app>
```

Display —

text
ya vacananta, inastuti-kīrti den bhaṭāra brahmā.

apparatus
^1. ya vacananta, ... brahmā.] only in **K**.
^2. -kīrti] *em.*, *lac.* **L**, -kitta: **K**.

You will observe that the use of @type="absent_elsewhere" serves to generate an apparatus entry pointing out that the given lemma is found “only in” the witness that supports it. (See also our use of @type="reformulated_elsewhere" in §Adopting word order from one witness against others.)

Although, in order to obtain the desired display, it would not be necessary to encode an <rdg> for the witness affected by the lacuna in the embedded <app>, we have decided to make it mandatory because doing so avoids the need to make an exception to the general rule that all witnesses should explicitly be listed and clearly identified in a given <app> (see §Recording variant readings positively).

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

```
<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"></gap></rdg></app> bumi
lamba, di bumi
        <app><lem type="emn">parək</lem><rdg wit="#A">paṟ̥<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 ... kalaṅkaṅ saṅ hyaṅ dasasila(25), maya-maya saṅ hyaṅ dasa-marga, kaprətyaksaan na dasa-indriya, pakən ṅrətakən di(26) bumi lamba, di bumi parək(27).
1.4 ini pakən(28) uraṅ...

apparatus
^25. dasasila] **A Ed**^{ASD}, dasa[... **B**
^26. di] **A**, *om.* **Ed**^{ASD}, *lac.* **B**
^27. parək] *em.*, paṚka **A**, tan parək **Ed**^{ASD}, *lac.* **B**
^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, there will be no need to encode any <rdg> for the given witness in *loci* falling between <lacunaStart/> and <lacunaEnd/>. The two indications "*lac.* **B**" 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.

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".

Textual loss within a block-level container

Example

```
<app>
        <lem wit="#K" type="absent_elsewhere">ya vacananta, inastuti-
                <app>
                        <lem type="emn" rend="hyphenfront">kīrti</lem>
                        <rdg wit="#K">-kitta:</rdg>
                </app>
                        den bhaṭāra brahmā.</lem>
        <note type="altLem">ya vacananta, ... brahmā.</note>
        <rdg wit="#L"><gap reason="omitted"/></rdg>
</app>
```

Display —

text
ya vacananta, inastuti-kīrti den bhaṭāra brahmā.

apparatus
^1. ya vacananta, ... brahmā.] only in **K**.

^2. -kīrti] *em.*, *om.* **L**, -kitta: **K**.

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 and .[38]

Example

```
<p n="3">...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"></gap></rdg></app> bumi lamba, di bumi
<app><lem type="emn">parək</lem><rdg wit="#A">paṟ<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"><span type="omissionEnd"/>
pak·kən</rdg></app> uraṅ... </p>
```

Display —

text

1.3 ... kalaṅkaṅ saṅ hyaṅ (24-)dasasila(25), maya-maya saṅ hyaṅ dasa-marga, kaprətyaksaan na dasa-indriya, pakən ṅrətakən di(26) bumi lamba, di bumi parək(27).
1.4 ini pakən(28)(-24) uraṅ...

apparatus
^24. dasasila ... ini pakən] an omission intervenes in **B**
^25. dasasila] **A Ed**^{ASD}, dasa[... **B**
^26. di] **A**, *om.* **Ed**^{ASD}, *om.* **B**
^27. parək] *em.*, paṟka **A**, tan parək **Ed**^{ASD}, *om.* **B**
^28. ini pakən] **A Ed**^{ASD}, ...] pak·kən **B**

Again (see §Absence of a witness for a lemma due to larger physical lacuna), there will be no need to encode any <rdg> for the given witness in *loci* falling between the two elements . The two indications "*om.* **B**" in the example are generated by the fact that the apparatus

[38]See https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPMI 12 Critical Apparatus and the comment on the potential for confusion inherent in the term "lacuna" made in https://digitallatin.github.io/guidelines/LDLT-Guidelines.html 12.9 (Gaps in Witnesses). We have favored an approach with rather than <milestone/> to avoid the need of adding the mandatory attribute @unit with the value "absent" as suggested by the TEI Guidelines for cases which mark the beginning of a piece of text not present in the edition. https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.milestoneUnit.html.

entries fall between the opening and closing , and no <rdg> for #B has been explicitly encoded.

It is possible to combine the tailor-made element-attribute for a larger textual omission with the values available for the attribute @cause (see §Encoding causes of variance of reading). 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
 [...] <app><lem wit="#L #M">ṅaranya, limaṅ</lem><rdg wit="#K" cause="eye-skip">ṅaranya, </rdg></app> kr̥ṣṇalam,<app><lem type="emn">samāṣa</lem><rdg wit="#L #M">sama:s·</rdg></app> ṅaranya, [...]   [..] pirak <app><lem wit="#L #M">ṅaranya, ikaṅ</lem><rdg wit="#K"> ikaṅ</rdg></app> sapuluh tahil [...]
Display — ^203. ṅaranya, limaṅ ... ṅaranya, ikaṅ] an omission due to eye-skip intervenes in **K**

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

Reporting scribal additions and deletions

<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 who handles the manuscript prior to your 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:[39]

- “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 §<del> – Deletions)
- “unspecified”: no information 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 <handDesc> – Encoding different hands. Then, provide its @xml:id as value for @hand. Here are two examples:

Example
<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
<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**.

[39] Some values are the same as the EGD, §4.4.2.

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.[40] The particular shape or shapes used by a given hand in a given witness can be described in <handDesc> as well.

<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:[41]

- "strikeout": text struck through, slashed, erased or scraped off, but still offering the possibility to detect what was deleted (*ante ras*.)
- "dots": dots added next to the characters to be deleted (*exp*. = *expunction*)
- "ui": for the combined application of vowel markers u and i under and above characters to be deleted

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

Example
<app> <lem wit="#L">maṅaku dṛvya, juga ya,</lem> <rdg wit="#K">maṅaku dṛvya, <del rend="ui">lkamaṅkana:,</del>juga: ya:,</rdg> <note>The cancelled string was initially written due to eye-skip from the preceding paragraph.</note> </app>
Display — ^159 . maṅaku dṛvya, juga ya,] **L**, maṅaku dṛvya, ⟦lkamaṅkana:,⟧ juga: ya:, **K** • The cancelled 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 §Text of lemma absent from a witness due to physical deficiency. 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

[40] See EGD, §4.4.2.
[41] 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**.

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 such a process as per §Encoding readings ante and post correctionem 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, we give you the option 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 inherently as substitution:

Example

```
<app>
  <lem type="emn">samagrig asambevara</lem>
  <rdg wit="#A">samagrig<subst><del>ə</del><add>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

```
<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>.

Reporting extraneous text segments (interpolations)

When a group of witnesses 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 of manuscripts bear witness 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.

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, 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

```
<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>
```

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,[42] can be recorded in the

[42] 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 (§<add> – Additions). In the situation we are

apparatus with reference to a word or words common to the accepted and to the rejected readings that serve to anchor the extraneous segment in question.[43] 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.

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

Example
<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 lkaṁ pirak ṣapuluḥ pana:, ma:, || ○ || <lg xml:lang="san-Latn" met="anuṣṭubh"><l>Ṛṇadeve 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:, ḍaṇḍanika: de bhūpatya, pañca satapa~~<gap reason="illegible" unit="character" quantity="1"/>~~ṇa, kunaṁṅ apuvara haṅasi<unclear>ḥ</unclear> dvigunotama ḍaṇḍa || ○ || <lg xml:lang="san-Latn" met="anuṣṭubh"><l>nipṭiṁ vak nityaṁ* krodaṣṭaṁ*,</l> <l>valat śobṛṁ* namasṭadḍi,</l> <l>niṣṭe maḍyamaṁ* hutamaṁ*,</l> <l>purvvakaṁ* mituva ḍaṇḍaṁ*,</l></lg> ka, hana ta ya: vaṁṅ apihutaṁ, tinagiḥ pihutaṁnya, tan paveḥ vetan drəvenya, hasṛṁ saṁṅ apihutaṁ, maṅalap sadṛvenya, 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 adṛve vaṁṅ apihutaṁ, lumka:s kaṅśa:seṁ tan hambava cihna Utər· daḷm, hikya ñaran valat sahaṣa havalat śobra, hiṅar:vakən vinalik rantas vitya:, mvaṁ tinibakna ḍaṇḍa mahira~~ntaṁ~~n kaṁ sa:hasobradḍi, kunəṁ pinaraṅgvakna vutaṁ, lavan daṇḍanekaṁ sahaṣa, vnaṅ pasaṅana vrat niṣṭa maḍyotama, ye niṣṭa vit hutaṁ mvaḥ paṅamet· sinahaṣa:, niṣṭa: ḍaṇḍa,, 5000, yen madya paṅamete, ḍaṇḍa, 10000, hutama paṅamete, ḍaṇḍa, 20000, sapaṅamete haṅsula:kna riṁ kaṁṅ avutaṅ, vaṣana, ḍaṇḍa riṁ saṁ bhūpatya || 0 || kunaṁ lkaṁ 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 8.137. On these

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.

[43] The usual method to encode such a phenomenon is to provide an empty <lem> element for this additional text. Nonetheless, since no empty element has been used in the DHARMA encoding model, we have decided in favor of keeping the usual printed strategy. Anway, the missing anchoring lemma would have raised a particular display issue, if members of the project expect to display the main edition with notes in a way that reproduces the printed apparatus.

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

<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>.

<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”[44] 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.

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 §Base text implied in dependent text), 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:

[44] See TEI Guidelines, §11.3.1.7

- “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”

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. “conjectural”

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 external to your text. If you use @evidence="external", then you are expected to encode the parallel passage(s) in question as per §Parallel Passages.

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

Example
... 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
... 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, ...

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

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

Example
salvirniṅ <app><lem><supplied evidence="conjectural"><gap/></supplied></lem> <note>A rather long passage seems to be missing here. The preceding objects appear in epigraphy in lists of ornaments the beneficiary of grants are allowed to display, so one would expect a term for a general category of objects to follow after <foreign>salvirniṅ</foreign>.</note> </app> hyaṅ ṛṣiśāsana

Use of @evidence is not mandatory and if you do find it useful, other values may be relevant than “conjectural”. See the list of suggested values above.

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 (intentionally) made a change in the text’s sequence. Complex transpositions are difficult to encode with the parallel segmentation method, if they involve variants occurring in different locations. For simpler cases, such as a physical displacement of folios in an antegraph or an inversion between two segments of text following each other in close succession, you can make use of the methods for encoding apparatus described so far, with the specifications we offer below; for more complex cases, it may be most pragmatic to decide in favor of a descriptive prose method using <note> rather than an apparatus entry with complex TEI encoding.

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 an at first sight 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:

Example	
<app><lem wit="#L #K">piraknya</lem><rdg wit="#M">si<locus type="displacement" from="21r12" to="12r2"/>raknya</rdg></app>	<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>
Display – ^850. piraknya] **L K**, si[displacement from 21r12 to 12r2]raknya **M**	Display – ^911. alapənira] **K**, *om.* **L**, a[displacement from 12r12 to 21r2]pənira **M**

Transposition within a sentence

In simple cases, 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>. You should identify the transposed segment with the attribute @type associated with the value “transposition”.

Example
<app> <lem wit="#K">ika ta saṅ prabhu, tinut <seg type="highlight">sira</seg> deniṅ rat kabeh</lem> <rdg wit="#L" type="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**, ika ta saṁ

prabu **maṅkana**, tinut sira de**nikaṁ** rat kabeḥ **M**

Transposition within 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 shall always contain the text in full and that 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:id's for all <lg> and <l> elements in our XML files. (This process can be launched upon your request, if necessary.)

Example

```
<lg met="anuṣṭubh">
      <app>
         <lem wit="#A">
            <l n="a" xml:id="33a">dharmakṣetre kurukṣetre</l>
            <l n="b" xml:id="33b">dharmajñair bahubhir vr̥te|</l>
            <l n="c" xml:id="33c">uñchavr̥ttir dvijaḥ kaś cit</l>
            <l n="d" xml:id="33d">kāpotir abhavat purā||</l>
         </lem>
         <rdg wit="#B" type="transposition">
            <l corresp="#33c"/>
            <l corresp="#33d"/>
            <l corresp="#33a"/>
            <l corresp="#33b"/>
         </rdg>
      </app>
 </lg>
```

Display —

Apparatus
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.

Example

```
<lg met="anuṣṭubh">
      <app>
         <lem wit="#A #B">
            <l n="a" xml:id="33a">dharmakṣetre
               <app>
               <lem rend="hyphenback" wit="#A">kuru</lem>
               <rdg wit="#B">giri-</rdg>
               </app>kṣetre</l>
            <l n="b" xml:id="33b">dharmajñair bahubhir vṛte|</l>
            <l n="c" xml:id="33c">uñchavṛttir dvijaḥ kaś cit</l>
            <l n="d" xml:id="33d">kāpotir abhavat purā||</l>
         </lem>
          <rdg wit="#B" type="transposition">
            <l corresp="#33c"/>
            <l corresp="#33d"/>
            <l corresp="#33a"/>
            <l corresp="#33b"/>
         </rdg>
      </app>
 </lg>
```

Display —

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

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. Such a scenario means that a single <app> element will not be sufficient to delimit the scope of the affected text. All the variants should be encoded where the witnesses place them, meaning that you will have some apparatus entries that only contain a <lem> linked to apparatus entries that only contain an <rdg>. The reading you are adopting as a lemma should, as usual, be encoded in a <lem> element. All the <app> entries concerning a case of discontinuous transposition should bear the attribute @type with the value “transposition” and the fact that the one alternative excludes the other is to be represented by matching attributes @xml:id and @exclude, as in the following example.

Example

```
<app  type="transposition" xml:id="app00755">
   <lem wit="#A">
      <seg type="transposition-static" xml:id="trsp001">
         <lg met="anuṣṭubh" n="12">
            <l n="a" xml:id="33a">dharmakṣetre kurukṣetre</l>
            <l n="b" xml:id="33b">dharmajñair bahubhir vṛte|</l>
            <l n="c" xml:id="33c">uñchavṛttir dvijaḥ kaś cit</l>
            <l n="d" xml:id="33d">kāpotir abhavat purā||</l>
         </lg>
      </seg>
   </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 variant A as the lemma for the following reasons ...</note>
</app>
[..]
<app type="transposed_elsewhere" xml:id="app00856">
      <lem wit="#A"/>
      <rdg wit="#B">
         <ptr target="#trsp001"/>
      </rdg>
      <note>Note if necessary</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 keep the variant A as the lemma for the following reasons ..

^856. dharmakṣetre… purā||**]** **B** (transposed segment) • Note if necessary.

Example

Let's imagine the editor doesn't want to keep the position of the transposition matching the reading in A nor B.

```
<app type="transposed_elsewhere" xml:id="app00755">
      <lem type=”emn"/>
      <rdg wit="#A" type= "transposition-orig">
         <ptr target="#trsp001"/>
      </rdg>
</app>
[...]
<app type="transposed_elsewhere" xml:id="app00856">
      <lem type=”emn"/>
      <rdg wit="#B" type= "transposition-orig">
         <ptr target="#trsp001"/>
      </rdg>
```

```
</app>
[...]
<app  type="transposition" xml:id="app00955">
   <lem type="emn">
      <seg xml:id="trsp001">
         <lg met="anuṣṭubh" n="XX">
            <l n="a" xml:id="XX">dharmakṣetre kurukṣetre</l>
            <l n="b" xml:id="XX">dharmajñair bahubhir vṛte|</l>
            <l n="c" xml:id="XX">uñchavṛttir dvijaḥ kaś cit</l>
            <l n="d" xml:id="XX">kāpotir abhavat purā||</l>
         </lg>
      </seg>
   </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** (transposed segment)
^856. dharmakṣetre… purā||**]** **B** (transposed segment)
^955. dharmakṣetre… purā||**]** *emn.* (transposition) • We have decided to emend the transposed stanza to avoid the variants proposed by witnesses **A** and **B** because …

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> without that element in turn being wrapped in <choice> will be displayed in obeli: †ABC†.

Example

```
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**

Applying notes to entire blocks of text

If you need to apply a note to a whole paragraph or stanza 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 block-level container (or <lg>) and 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
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 the 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>

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 (§Critical Apparatus), 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) 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.

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".[45] 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 last element before the closing tag of the block containing the text for which a parallel will be reported, namely <p> or <lg> — or <quote> if the passage in question is itself quoted from another text. This last scenario is shown in the following example:

Example

```
<div type="dyad" n="32">
   <quote type="base-text" xml:lang="san-Latn">
      <lg n="31" met="anuṣṭubh">
         <l>arjayej jñānam arthāṁś ca</l>
         <l>vidvān amaravat sthitaḥ |</l>
         <l>keśeṣv iva gṛhītaḥ san</l>
         <l>mṛtyunā dharmam ācaret ॥31॥</l>
      </lg>
      <listApp type="parallels">
```

[45] Note that we have decided not to follow the TEI consortium recommendation, although its model was applied in the SARIT project: “16.1.2 Using Pointers and Links”, TEI Guidelines P5, version 4.0 from 2020-02-13, https://www.tei-c.org/release/doc/tei-p5-doc/fr/html/SA.html#SAPTEG, accessed 2020-12-17.

```
        <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 mṛ̥tyu 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 @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

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

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

```
<listApp type="parallels">
                <app type="imitation">
                        <note corresp="txt:NarSm_04.12">
                                <lg>properly encode the parallel stanza here</lg>
                        </note>
                </app>
                <app type="adaptation">
                        <note corresp="txt:MBh_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>
```

<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 (§txt: — Referring to machine readable versions of other texts) 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 (for proper use of @copyOf, see §Declaring an identical parallel passage).

Example

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

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 an embedded <note>.

As shown in the following example, these scenarios can be combined within the same <app>. Let us imagine that there are two parallels for the passage in question: one of them is identical and you do not wish to show its reading; the other is only a partial correspondence, so you wish to cite it, and you wish to add your philological observations on the differences. To do so, you should proceed as follows:

Example

```
<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>
```

As indicated in the example, you must encode any parallel verse text that you wish to include in your <listApp> using <lg> with as many <l> as needed (and §Verse containers). If the parallel is in prose, you must use as many elements <p> or <ab> as necessary (see §Prose containers). 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 §Bibliographic References.

Here is one more example where <note> is used only for free-text discussion of the parallel:

Example

```
<div type="dyad" n="4">
   <quote type="base-text" xml:lang="san-Latn">
      <supplied reason="implied">
         <lg met="anuṣṭubh">
            <l>hiṁsāṁ yaḥ kurute kaś cid deyaṁ vā na prayacchati |</l>
            <l>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>
```

Observe that it is not necessary in such a case to embed <note> in another <note>.

<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 §Referring to External Digital Resources.[46]

1. @sameAs or @copyOf: both of these attributes are used to identify identical content. However, 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.
2. @corresp: points to elements that correspond to the current element in some way but are not necessarily identical.

In the following paragraphs, we will furnish more detail on when and how to use @sameAs, @copyOf and @corresp.

@sameAs and @copyOf — Declaring an identical parallel passage

Two attributes can be used to identify the identity of content: @sameAs and @copyOf. Although they are both meant to express that your text’s segment and its parallel are identical in content, the first indicates that the text is repeated and presented in your digital edition, while @copyOf does not have this implication. Let us start with an example of @copyOf:

Example

```
<listApp type="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 <quote> (and, if necessary, commented upon in a further <note>):

Example

[46] 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; for the same reason, we will not use <ptr/> with @target inside <item>.

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

@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 §Declaring an identical parallel passage. As shown in several of the examples above, you can record any observations on the correspondence you have marked up by adding a <note>.

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 the complexity of encoding the delimitation of the segment in question.

Declaring a segment of text with <lem> but without encoded anchor in the text edition

First, you can identify the segment for which there is a parallel using <lem> in an <app> within <listApp>, still to be declared as the last child of the containing <p> or <lg>.

Example

```
<listApp type="parallels">
        <app>
            <lem>jadyan kuras, jadyan tahun</lem>
                <note corresp="txt:PNJ.129">jadyan kuras jadyan tahun.</note>
                <note corresp="txt:TB.35v">pada ñovana, dadian kuras, dadian tahun, dadian kaneneh.</note>
                <note corresp="txt:MSJ.21"> mańka jadiyan kuras jadiyan tahun.</note>
                <note corresp="txt:LK.137">tətəńər jadian tahun, sań atuań dina tarań, tətəńər jadian kuras,
```

```
putər kuruń dina iruń.</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 §Ellipsis in the lemma's display). 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>.

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

Example with @sameAs

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

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

Example

Example with <note>

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

Please follow the numbering pattern for @xml:ids 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:ids 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.

Associating parallel texts that are implied but unknown

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

Example

```
<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>
```

Encoding an external apparatus stanza-by-stanza

If you favor a location-reference approach to encoding your critical apparatus against a parallel segmentation method, you must follow the structural approach described in the section §Parallel passages: each block of text will be immediately accompanied by its associated apparatus. Open a new <listApp> element as the last child 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. Copy the text of each lemma into <lem> and use the attribute @wit in the manner described in §Recording witnesses. To be sure the apparatus entry and the edited text can be matched together, the lemma declared for the <lem> element should normally be a copy of the relevant string inside the block of text, though the same exceptions can be here made as those outlined in EGD §9.1.3. Variant readings should be encoded in an <rdg> element, as per §Variant readings, using the attribute @wit in the same way as in <lem>.

Example

```
<lg n="1.1" met="āryā">
  <l n="ab">śaśadharasadṛśasaroje alake devākṣayānake caraṇau</l>
```

```
  <l n="cd">sitarajabhāsvaravimalau śambhor nityaṁ [ ] pāntu janān</l>
  <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×</rdg>
      <rdg wit="#msF">pāta janān</rdg>
        <witDetail wit="msF" type="ac"/>
    </app>
  </listApp>
</lg>
```

Example

```
<ab type="speaker">anarthayajña uvāca
  <listApp>
    <app loc="uvaca">
       <lem wit="#mssCaCbCc #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>
</ab>
<lg n="1.13" met="anuṣṭubh">
  <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>
  <listApp>
    <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>
</lg>
```

Referring to digital resources

General Structure for Canonical References[47]

The references to textual sources, such as an electronic version of a text you want to refer to, consist of two parts. The first part (e.g. MBh for *Mahābhārata*) consists of the main identifier of the record, while the second part provides the path and documents parameters to navigate inside the text (e.g. 01.12.29 for book 1, canto 12, stanza 29). The two parts are separated by an underscore (e.g. MBh_01.12.29). You can provide the first part of the reference on its own as well as a reference to the text with a specific passage. Note that this second parameter is a specific implementation, and needs to be mapped on the electronic version of a specific edition.

Any range or sequence of text should be expressed by providing the starting point and the ending point separated by a hyphen, after the main identifier of the text and the separating colon. For instance, if you want to reference stanza 29 and 30, in the canto 12 of the book 1 in *Mahābhārata,* the reference would be MBh_01.12.29-01.12.30. The sequence is to be given in its full length and can't be shortened to allow ranges working across structural boundaries to work efficiently.

The first part shall match a declared @xml:id in the primary textual sources file, the MBh for *Mahābhārata* in our example. The second part, introduced by the underscore (the choice of which was made in order to comply with some of the recommendations established by the Canonical Text Services protocol)[48] provides the reference to a specific passage or range within the selected text and needs to represent the formatting of the chosen digital edition as well as the formatted pattern. Levels of citations should then match the encoded digital edition. Each

[47] This section is theoretical and has not been tested. For now, if we recommend you try to follow as much as possible the described citation pattern. The implementation and the obvious changes that it might require will be handled in a later stage of the project.

[48] Also called CTS, Canonical Text Services is a protocol to identify texts or fragments of text following the classical notions of "work" and "citation" used by 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. See http://cite-architecture.org/.

level must be separated by a dot and must match the structuring elements such as , <p>, <lg> or, in case of a diplomatic edition, boundary-marking elements such as <pb/> and <lb/>. The elements can be nested and, for each of them, you will need to declare a new citation level. Empty elements, here, only matter by their position in the hierarchy. The value of a level can be expressed by attributes such as @n, @xml:id, @corresp or sometimes the concatenation of @type or @subtype with @n.

To be valid, your reference must follow a hierarchical structure: this means that you cannot refer to a line or a paragraph without referring to the entire structure of the referred passage. This recommendation is made to avoid any possible ambiguity in the identification of a passage.

Repetitive links based on prefixes

It will happen often that you need to encode some pieces of information that are 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. Primary textual sources, both machine-readable and non machine-readable, that appear you need to cite for some reason in your edition or translation, for instance as parallels to given segments of your digital edition or simply to justify some philological decision ("txt:").

The prefix declarations "part:", "bib:" and "txt:" are shortcuts to the files that list the information introduced above. In brief, they work as pointers to identifiers defined in 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 primary sources (texts) are provided in our encoding template, in the <encodingDesc> within the element <listPrefixDef> as previously explained (§<encodingDesc>).

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.[49]

Within this reference file, each participant is encoded through the element <person> bearing a specific @xml:id. Each @xml:id consists of four letters joined together, all lower-case: the two initial letters of a person's first name joined to the two initial letters of her/his surname. For instance, "argr" stands for Arlo Griffiths, and "dogo" for Dominic Goodall.

[49] https://github.com/erc-dharma/project-documentation/blob/master/DHARMA_IdListMembers_v01.xml

You will need to point to the values of these @xml:id when, for instance, your digital edition is the fruit of multiple hands and, within your <fileDesc>, you need to declare who the editors are (see §Identifying scholarly responsibilities); or when, within the <revisionDesc>, you want to signal who made certain changes to the text (see §Keeping track of File History). Please bear in mind that, in order to correctly point to the relevant information, the value of the attribute you will use (e.g. @ref, @who, etc.) will have to be built with the prefix "part:" followed by the identifier of the person.

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.[50]

txt: — Referring to machine readable versions of other texts

The prefix "txt:" stands for "text" and shall be used as an abbreviated link to the project's authority file listing all the editions of reference for primary sources, along with any machine-readable representation thereof that may be available. You will need to use this prefix for instance when encoding the existence of a parallel to a passage in your edition. Within the authority file,[51] and following the logical structure explained for "part:", 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.

Example

```
<bibl corresp="bib:Atja+SalehDanasasmita1981_01" xml:id="EdAmGa1981">
            <abbr type="siglum">AmGa</abbr>
            <title>Amanat Galuṅguṅ</title>
            <ptr
target="https://github.com/arlogriffiths/jawakuno/blob/master/texts/xml/AmanatGalunggung_inprocess.xml"/>
              <msIdentifier>
               <settlement>Jakarta</settlement>
               <repository>Perpusnas</repository>
                 <idno>L 632a</idno>
              </msIdentifier>
          </bibl>
```

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

[50] For the project's use of Zotero's Short Title field, see the Zotero Guide (§4.10).
[51] Available in the project documentation folder in GitHub: https://github.com/erc-dharma/project-documentation/blob/master/DHARMA_IdListTexts_v01.xml.

1. the existence of a digital representation of the text edition, whether it is in .txt, .html or in .xml format
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)

Note that 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 element — in defining the siglum, you are free to use diacritical marks. 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.

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 §Attributes) or the element <ptr/> (§Shortcut for citing other texts). 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

```
<listApp type="parallels">
        <app>
                <note sameAs="txt:EdMaDhSa2004_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:EdMaDhSa2004_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:EdMaDhSa2004_08.070"><bibl><ptr target="bibl:Ollivelle2004"/><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 -->
```

The empty <ptr/> element provided for each work can contain more than one URL, separated with a blank space. If a white space character is part of a URL, make sure to escape it by transforming it to "%20". If several values have been recorded for a work, we will be facing the difficulty of selecting one value from another. Our system will process only the first one. It means you need to put the most important URL as the first value of the attribute @target. If you are dealing with a case where one work needs to be cited with true differentiation between all its editions, contact the TEI-XML data manager of the project.

Shortcut for citing other texts

If you need to cite a text declared in the authority list in free-text sections, as in a <p> or in a <note>, you can use an empty tag as a shortcut to name it. The logic is the same as in section §Shortcut for citing sigla and hands; see the following examples:

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

Example
<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 use the attribute @rend to select the rendering you expect, 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.

Internal identifiers

To produce identifiers for block elements, several possibilities are available in TEI. We favored a referencing system based on the mark-up of the text to match the general set to produce canonical references. This reference system is intended for long-term use, so it must be seen as an unchanging point in the text, once the first version of your edition has been published.

One of the available methods to provide automatic unique values for @xml:id attributes is to build for each element a domain-style address with one component for each level of the document hierarchy separated by full-stop. DHARMA favored a untyped path form of identifier.[52] Each component will be formed by a number, indicating the position of the XML node inside the document hierarchy. Since to make the resulting identifier conformant, it requires an anchoring alphabetic letter, we decide to align the identifying pattern with the canonical reference system by adding the @xml:id of the edition declared on the <TEI> root of the file, followed by an underscore.

Finally, those identifiers generated will take the element <text> as a starting point.[53] However, the element <text> itself is omitted from the identifiers pattern.

About the identifying pattern for 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 , <p>, <ab>, <lg> or <quote> segments of your text edition. 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.[54] Their values will reference the sequential numbers of the parent and of the elements <ab>, <p>, <lg> and <quote> within that . This means that in some cases, the identifier can diverge from the editorial numbering pattern reflected in your @n attributes. In the following example, the first 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 that is marked by the attribute @n as the first chapter of the text.

Example

```
<div n="0" type="chapter" xml:id="rsisasana_01">
        <ab xml:id="rsisasana_01.01" type="invocation" xml:lang="san-Latn">[...]</ab>
```

[52] The other solution would have been a typed path form of identifier. It would have required the identifier of a block element to be built with the element name, a hyphen, and a number, for instance p-2, which can become verbose.

[53] We favored such a starting point rather than the <body> element to avoid preventing identifiers to be assigned to components inside the front and back matters, if such a need appears at some point in the project.

[54] The XSLT file to apply the numbering pattern is available on github in the project-documentation repository.

```
</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:id is provided so you can interlink or reference segments of texts elsewhere in the file. They will become indispensable if you are dealing with prose or stanzas interrupted by base text, scenarios discussed in §Prose interrupted by dependent prose and §Stanzas interrupted by dependent prose. 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 together and can, if necessary, be used to reconstitute the base text without interspersal of dependent text.

About the identifying pattern for elements

Some elements can also be given an identifier to handle cross-reference and link elements across block boundaries. Among those elements, we foresee such a pattern could be useful for <term> and <app>, but feel free to contact the XML-TEI Data Manager if you need another element to be numbered.

For those two elements, the identifier pattern should be the name of the element followed by a 3-digits number for <term> and a 5-digits number for <app>, i.e. term001 and app00001. See §Strategies to encode apparatus entries across block boundaries and §Lemmas of base text in dependent text to know more about their use.

Punctual links

Note that in this section, you will only find a solution for ad-hoc linking. Any repetitive links should be declared in the manner described in §Referring to digital resources external to your file and should only appear in freeform text elements such <note>.

Before starting this section, you should be aware that several linking mechanisms are available in a DHARMA conformant encoding. The main two possibility at your disposal are:

1. <ptr/> is a self-closing pointer with the attribute @target
2. <ref> contains text to be hyperlinked with the resource declared in the attribute @target.

When using the first solution, you must always be sure that 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 anchoring it.

Links to external resources or files

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 name of the text you want to link with your current work. To generate the hypertext link, you must also reference an attribute @target in which you should give the full name of the file of the text. Be sure to follow the FNC guide to name your file.

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

If you want to provide a specific reference to a part of the text, 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.

Example
<ref target="DHARMA_CritEdSvayambhu.xml#svayambhu-03">See Svayambhu, chapter 2</ref>

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, or better a permalink. Be extra careful about filling any exterior link since we would like as much as possible to avoid breaking of this link.

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

Hyperlinking to texts with URNs

The element <ref> should also be used for punctual hyperlinks with any external texts. However when available, we recommend you provide a link to a Universal Resource Name (URN). CTS is the most common schema for providings URNs for digital editions in the Humanities. Classical Latin and Medieval Latin texts are extensively covered by the CTS URN, for others languages and period references have a less exhaustive coverage. Depending on the scholarly fields, a URN isn't always available and you might need to revert to using a simple permalink, url link or even a specific project identifying pattern. In case of doubt, contact the XML-TEi Data Manager of the project.

File internal links

The technical team has implemented a system that will process your XML file and automatically assign an @xml:id to every structural element of your edition (, <lg>, <p> and <ab>), so that, subsequently, it will be possible to reference any of those elements by their unique identifier preceded by a hashtag symbol “#” inside the attribute @target.

Create a file internal reference

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
<ref target="#svayambhu-03">See the same phenomenon on the chapter 2</ref>

Shortcut 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 attribute @target filled with the identifier of the said part. The display will then be the name of the element and the @n attribute provided with it. If no attributes have been associated with the element you are referencing, an automatic number based on the occurrence of the element in your edition will be automatically generated. Then, if you want another display, we recommend you use the solution with <ref> explained above §Create a file internal reference.

Example
<note>Dependent text on <ptr target=”#svayambhu_07.13.02.02"/> and, according to the editor, at least partly on <ptr target=”#svayambhu_07.13.02.01"/></note>
Display — Dependent text on stanza 273 and, according to the editor, at least partly on stanza 272

Encoding Additional information in the Edition

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 §“Folios”).

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; in such cases, see §“Lines and pages for printed editions”.

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 §bib: -- Referring to items in the DHARMA Zotero Group Library).

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 any printed line numbers, it is sufficient to mark the line beginnings with <lb/> without applying @n. Page or line beginnings occuring in the middle of a word must bear the attribute @break with the value “no”, as described in more detail in the EGD.[55]

Folios

In some cases, it may also be deemed useful to represent the folio beginnings of your manuscript witnesses; we do not expect there will often be the need to encode manuscript line beginnings in a critical edition. 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 (§Apparatus criticus), in this case you cannot use the attribute @wit to identify them, because the TEI does not allow it. 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), as in “1r”, “1v”, “2r”, “2v” and so on. Manual encoding of @n is always necessary on elements <pb

[55] EGD, §3.2.4, §3.5.5 as well as §8.1.2, §8.1.3, §9.1.6.

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 place in the text — a situation that is especially common at the beginning of texts — you must use two <pb/> elements, like this: <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
<pb edRef="#EdA" n="194"/> <ab type="invocation"><pb edRef="#A" n="0v"/><lb n="1"/>siddhir astu</ab>

Notes

<note> – Generic 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 any block-level container (after any final punctuation). To do so, add the element <note>. Make sure the contents of a <note> is a full sentence, beginning with a capital letter and ending with a punctuation mark. Place it after any adjacent punctuation mark, not before it. The contents of notes may be structured into paragraphs by creating <p> elements within <note> when the note becomes more than a few sentences. Note that in this case, all the content should be contained inside <p> elements.

If you are not the author of the note, you may need the attributes @resp and @source to declare someone else authorship. To give credit to another member of the project, add the @resp attribute, while the @source should be taken only if the note is adopted verbatim from a publication. Remember that according to DHARMA conventions, paraphrased notes are in this respect regarded as the product of the person doing the paraphrasing. The author of the original note should be declared using the regular bibliographic citation system, or depending on the context, simply by referring to the author's name.

<witDetail> – Witnesses Notes

Unlike the <note>, the element <witDetail> provides specific information related to a particular witness or witnesses to complete a variant reading. It works as a specialized version of the generic element <note>. It is used to comment on the <lem> or the <rdg> preceding it. You would have to add an attribute @wit to the <witDetail> with the sigil identifying the right witness or witnesses.

The <witDetail> can mention elements related to a specific aspect of the witness you would like to evoke or to record a specific information of the source texte, such as a note explaining that the variant readings might have two hands involved.

For instance the following apparatus entry given in the edition of Sañ Hyañ Siksa Kandañ Karesian, the initial comment on the shape of the script given in the diplomatic edition A could be retained in the critical edition through the element <witDetail>. Its placement is expected right after the variant reading of the witness A, here declared as <lem>. Any other variant <rdg> should then come after <witDetail>.

Example

```
Ulah <app>
   <lem wit="#A">mo</lem>
   <witDetail>The shape of <foreign>tarung</foreign> is unusual, it looks like <foreign>wulu</foreign> with the wavy stroke to the bottom, probably it was added later by the scribe.</witDetail>
   <!-- Potential <rdg> element -->
</app> turut sañ hyañ siksa
```

Since DHARMA uses <witDetail> as a solution to encode *ante* and *post correctionem*, we don't recommend using this element in other contexts to avoid confusion. However, if you feel like you might benefit from using it, contact the XML-TEI Data Manager.

Quotes

Generally, while encoding quotations, do not add any quotation marks, they will be produced but the transformation processes. However, if you don't encode it for whatever reasons, then add them.

When citing text from a publication, follow DHARMA transliteration conventions. When inside a <quote>, any transliterated words in the cited text shall be retained in their original form rather than being converted to DHARMA transliteration system. However, if quoting outside of the <quote> element, for instance in apparatus readings, translations and paraphrased/summarised opinions, the text should be converted to DHARMA transliteration system.

<cit> – Cited Quotation

To encode quotations from an external source, with accompanying information about its origin, use the element <cit> as a container and add a <quote> element in the <cit>. The second child element of <cit> depends on whether the author is referring to a work or if you, as editor, provide the information. For the first case, then use the element <ref> around the reference. If you want to normalize this reference, you can add the attribute @cRef to provide the canonical reference of this quoted text, see §XXX. On the other hand, you can choose the attribute

@target to give a reference of an edition of the text. In this case, see §Making a hyperlink to learn more.

Example
<cit>   <quote></quote>   <ref></ref> </cit>

For the second case, use the element <bibl>, if you are quoting this material from a published work, see §Bibliographic citations with Zotero on how to do use <bibl> with <ptr/>. Note that both <ref> and <bibl> elements could be used together to provide a specific edition to the reference made by the author of the text. In this case, use the element <ref> first followed by the <bibl> and its associated element.

Example
<cit>   <quote></quote>   <bibl><ptr/></bibl> </cit>

<q> – Quoted Material

Quoted text not attributed to a published source should be encoded in the element <q>, which is more neutral in its meaning than <quote>. We expect it to be applied mostly to quotations in the same language as the surrounding text such as translation of sentences inside the translation, or translation of direct speech given in the edition.

Example
<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>

Do not add quotation marks, they will be produced but the transformation processes. Nonetheless, if you need more control on the display, it is possible to omit the tag and directly write inside the file the desired characters. This method should be reduced to special cases as much as possible.

Bibliographic References

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> with the attribute @target, whose value shall be the Zotero Short Title of the cited publication, prefixed with the string “bib:”, see bib: — Referring to items in the DHARMA Zotero Group Library.

To give further specification regarding the part of the publication concerned, add the element <citedRange> as a second child of <bibl>. Its content should provide the citations of text units. To refer to page numbers, no need to add any other content: <citedRange>103</citedRange>. If you need to quote any other textual unit inside a publication, add the attribute @unit and fill it with one of the following values:

- “page”
- “part”
- “volume”
- “note”
- “item”
- “entry”
- “line”
- “figure”
- “plate”
- “table”
- “appendix”
- “section”

The quotation content of the textual units shall be given as Arabic numbers, any Roman or Devanagari will have to be converted by the editor. Any range should be recorded using a hyphen to separate the starting point from the ending one, both are expected 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
<bibl><ptr target="bib:Naerssen1976_01"/><citedRange>301-302</citedRange></bibl>
Display — van Naerssen 1976: pages 301–302.

If the name of the author(s) don't need to be display, use the attribute @rend with the value “omitname” to the <bibl> element, but if you expect to show “*ibid*”, use the value “ibid” for @rend. All this section is following recommendations established by the EDG, see §10.4.5.

Example
<bibl rend="omitname"><ptr target="bib:Naerssen1976_01"/></bibl>
Display — (1976)

Example
<bibl rend="ibid"><ptr target="bib:Naerssen1976_01"/><citedRange>301-302</citedRange></bibl>
Display — *ibid*.: pages 301–302

<title> – Encoding titles

Would you need to quote a title in any notes or commentaries, tag it with the element <title>. To determine what content you consider a title, follow the rules given in the EGD §10.4.2: all epigraphic and non-epigraphic primary sources, include secondary sources regardless of whether the title is cited in full, in abbreviated form or with a known acronym.

The rendering of those titles will be italic. When this is not desired, add the attribute @level with the value "a" (for “analytic”), this case will be displayed with quote marks around the title rather than italics, but use the attribute @rend with the value “plain” to obtain a title without any kind of typographic distinction from its surrounding text.

@resp – Attributing responsibility

It is possible to declare a responsibility of authorship with the attribute @resp. It can be added to any XML elements 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 mostly use it punctually. But when reviews will start, it will allow us to identify interventions made by any member of the project.

The attribute @resp should contain the personal identifier of the project member preceded by the prefix “part:”, see part: — Referring to other DHARMA team members. To credit more than one participant, separate each identifier with a blank space.

Encoding features available for apparatus and on their own

About non-alphabetical characters

<space> – Empty spaces

When a witness contains a blank space, you can use the empty element <space/> with the attributes @type for classification, @unit whose default value is “character” and @quantity to record the dimensions of the space, as stated in EGD §4.3.1.[56] However, we recommend you encode only the spaces relevant and meaningful enough to your edition, avoiding as much as possible any space used for semantic segmentation or layout purposes.[57]

Vacat space

Space left with the evident intent to be filled in later on, e.g. with a name, a place or a date, called “vacat” in the Western scholar tradition, should be encoded with the element <space/> with the attribute @type with the value “vacat”. In this case, you should always record the size of the space, e.g. <space type="vacat" quantity="3" unit="character"/>.

If the blank space has been partially filled, you can encode only remaining blank space if you are uncertain about the presence of an addition or if you can assert the exact extent. But if you are certain about both then encode the vacat for its entire length and use the element <add> for scribal additions as described in §Scribal additions, either after or before the space.[58]

Physical imposed space

The element <space> is also available to record a physical aspect with an impact on the edited witness, e.g. a defect in the material or a binding affecting the text forcing the scribe to skip . We keep the cases foreseen by the EGD: binding holes, defects surfaces and spaces resulting from other glyphs.

When a binding hole of the manuscript affects the text, you should encode it with the attribute @type="binding-hole" at the locus of the hole. You encode the impact on the text here rather than the existence of this hole on the object. So it means you won’t be encoding a binding hole if it is in the margins or between the text lines, but you will be encoded a binding hole interrupting one to several lines of texts. Note that you don’t need to provide information regarding the size of the space with the attribute @unit and @quantity, nor do you have to document the fact that a hole can affect more than one line. (See EGD §4.3.6)

---

[56] Please take into account also that DHARMA has a broad interpretation of the element <space/> compared to the regular TEI practice. The main intent is to encode space left for a segment of text to be filled in later or later. The <space> element shouldn’t be used for normal inter-word space or layout of the text.

[57] The critical editions guide doesn’t follow the sections §4.3.2 and §4.3.4 of the EGD about space for semantic segmentation and for visual layout

[58] See EGD §4.3.3, if you need to refer to it.

Example
<app> <lem type="emn">keṅətakna</lem> <rdg wit="#A">k<space type="binding-hole"/>eṅəttaknə</rdg> </app>

When a surface defect prevents the scribe to write on certain areas, you may encode it with the element <space> with the @ type="defect", @quantity and @unit. Encode only those elements that you deem relevant to understanding the history of the text, its creation or its transmission. (See EGD §4.3.7)

When a space is left blank because of another character used in the previous or following lines, you may deem necessary to encode it because it impacts the reading of your witness. In this case, you should encode it using the attribute @type with one of the two following values:

- “descender” when a character hangs down on the current line from previous ones.
- “ascender” when a character pops up on the current line form following ones

Spaces for ascenders and descenders don’t require any attributes @quantity and @unit. (See EGD §4.3.8)

Finally, if you are facing a space impacting your witness that you can’t seem to explain but still deem significant, you should encode it using the value “other” for the attribute @type and add attributes @quantity and @unit the same way as the cases above. (See EGD §4.3.9)

<g/> – recording a specific glyph

To encode all characters other than alphanumeric ones, you should use the element <g>. This TEI element underlines the fact that no equivalence can be found in the DHARMA transliteration system and by extension in the UTF-8 tables. Following rules already stated in the EGD, the element <g> should be used in two ways: as a text containing-element for any glyph expressing numerals, space fillers and punctuation marks, but as an empty element when no assumption of function can be made about the glyph, such as auspicious symbols at the beginning or end of a segment of text. With both ways, the element should be used with the attribute @type to record the specific shape of the glyph and eventually add a @subtype to provide more precision about it. The temporary values allowed as @type are gathered in the *Symbol Taxinomy Supplement*[59].

[59] See https://docs.google.com/document/d/1glfyQnFqPrbVOYzegfjKIOVrc-vMgznEQ1iNsFf7DE8/edit?usp=sharing. All encoders are requested to refer to that list before creating a token for a symbol and are encouraged to contribute to it by inserting clippings of symbols they encounter in their texts when no token already featured in the list matches their need. Note that those temporary values won’t be added into the validation schema. Some of the available values are: “dandaPlain”, “dandaHooked”, “dandaCross”, “dandaOrnate”, “ddandaPlain”, “ddandaHooked”, “ddandaCross”, “ddandaOrnate”, “comma”, “commaSmall”, “dashPlain”, “dashConcave”, “dashConvex”, “dashHooked”, “dashWavy”, “dashLong”, “dashDouble”, “dashDoubledot”, “dotMid”, “dotDouble”, “dotTriangle”, “circleSmall”, “circleLarge”, “circleHigh”, “circleLow”, “circleCross”, “circleDouble”, “circleFloret”, “circleHorned”, “circleConcentric”, “circleTarget”, “circleTriangle”, “spiralR”, “spiralL”, “crossPlus”, “crossX”, “floretQuatrefoil”, “floretComplex”, “gomutraInitial”, “gomutraFinal”, “gomutraFinalBars”, “gomutraFinalComplex”, “tennisBall”, “squiggelVertical”...

Remember that symbols can be described in detail and associated with a hand in the <handDesc>, see section §<handDesc> — Encoding the hand(s) with its/their script(s), and doing so is recommended for all symbols whose shape is not be evident to you and to any familiar reader with the corpora you are working with.

Several cases that have been prescribed by the EGD are still relevant for this guide. The following lines are a quick summary of those with the associated with the relevant reference to the EGD sections:

- Glyphs that are numeral characters not expressed by the means of digits and transliterated into Arabic numerals should be wrapped in <g type="numeral"> as described in EGD §4.2.2.
- Glyphs for punctuation marks are transliterated as the abstract punctuation character "." wrapped in <g> with an appropriate @type or a double ".." to represent a higher-level punctuation mark (e.g. a double daṇḍa), if a witness employs two levels of punctuation (see EGD §4.2.4)
- Glyphs identified as space fillers and transliterated with the dedicated character "§" are to be encoded with the element <g> and the appropriate value for @type (see EGD §4.2.5)

Example
<rdg wit="#L633">kṛmavidu cavanevayurde, <g type="filler">§</g><unclear reason="eccentric_ductus">E</unclear>da<add place="below">ta</add>yudena kaḻmsyurigrha<space type="binding-hole"/>s·taḥ</rdg>

- Miscellaneous non-alphabetic glyphs that are not transliterated with any character should be represented by the empty element <g/> with the appropriate @type (see EGD §4.2.6)
- Alphanumeric characters can occasionally be used as a symbol, transliterate the character normally, separated by the adjacent text by a space without encoding its function. In the same way, when a numeral sign works as a symbol, do not encode its function nor its semantic markup <num>. However, when a numeral is used as an alphabetic character, transliterate it as a digit, but without applying any semantic encoding (see EGD §4.2.7).

Indexables

It can be useful to encode names of people or places mentioned in the text of an edition, whether to facilitate the creation of an index or even to support search functionality. The editor can decide which names, if any, should be encoded. However, if those are to be encoded, the following section of the guidelines must be applied.

<persName> – Encoding Names

In the edited text, people's names should be encoded with the element <persName>. It is possible to provide a categorisation to those names using the @type attribute, as well as a subcategorization with @subtype.

The allowed values for @type are so far:

- divine
- human
- personification

The @subtype can take the following values. They can be used only if the attribute @type is already declared on the <persName> element.

- coronation (Rājarāja, Rājendra, …)
- sobriquet (biruda)
- title (pōttaraiyar, (kōp)parakēcarivarman̲ / (kō)rājakēcarivarman̲)
- other (pre-coronation name, e.g. Arumol̲i, Arumol̲ivarman̲)

To help establish the indexable, you can provide a normalized version of the name inside the attribute @key. Its content will be later transformed to a @ref attribute and linked with an authority list.

The element <persName> can be used within <lem> or <rdg> as needed to record variant readings of a name. If the variation provides proper nouns and common nouns, only the first should be encoded as a <persName> while the former isn't to be encoded.

<roleName> – Encoding Roles

The element <roleName> can be used inside the element <persName> to encode a position in society like a rank or status and a role. The rank and status are to be given using the attribute @type and the role with @subtype. Several values have been set up for the DHARMA project.

Propositions for the value of the attribute @type:

- King
- subordinateRuler (e.g. pallavaraiyan̲)
- landlord (e.g. uṭaiyar, kil̲avar)
- godLegalEntity (e.g. uṭaiyar)
- Priest
- Brahmin
- Monk
- merchant (e.g. nakarattār)
- Artisan
- brahminDelegate (e.g. sabhaiyār, sabhaiyōm)
- regionalDelegate (e.g. nāṭṭār, nāṭṭōm)
- officer (e.g. temple officer, royal officer)
- Dancer
- Singer
- Peasant
- shepherd (man̲r̲āṭi)

- unknown (this value is to be used when you want to encode a value for @subtype but can't provide a value for the @type.)

Propositions for the value of the attribute @subtype:

- Donor
- Donee
- founder (of a temple or a monastery)
- administrator (overseer of donation; e.g. the one who makes sure that the in-charge of a donation supplies what he has to supply).
- inChargeDonation (e.g. the one who has to supply oil every day)
- witness
- orderIssuer
- orderAddressee
- auditor (controller of transaction)
- beneficiaryMerit (e.g. transfer of merit; donation “on behalf of”, “in the name of”)
- commemoratedPerson (e.g. “in the honour of (a deceased warrior)”)
- scribe (exact role undetermined)
- composer (i.e. author of the text or part of the text; e.g. poet of the Sanskrit eulogy).
- handwriter (i.e. the one writing in chalk on the plate/stone for the engraver)
- engraver (i.e. the artisan who engraved the text on the support)
- sealer/solderer (i.e. the one who fabricated/sealed/soldered the seal)

Feel free to submit any new value to the XML-TEI Data Manager.

<placeName> – Encoding Places

The places can also be encoded if you feel the need. Use the element <placeName> to do so. Use the attribute @key to document a normalized version of the place’s name, or use the attribute @ref, if you already possessed a URI pointing to some reference (e.g., Pleiades) for the place name. The TF-A has worked on a temporary list for the attribute @type and @subtype.

Only two values are available for the attribute @type: “builtPlace” and “territorialDivision”.

For territorial and administrative divisions, the @subtype can be one of the following values:

- district (viṣaya, nāṭu, kūr̲r̲am)
- site (town, village)
- sitePart (e.g. quarter, hamlet, cēri)

While built places can be subcategorized with:

- temple
- shrine (e.g. for a secondary shrine in a temple complex)
- monastery (e.g. vihāra, maṭha)
- feedingHall (cālai, Skt. śālā, mess for devotee pilgrims)
- tank (artificial)
- pavillion (maṇḍapa)
- garden (nandavan̲am)

If any variant readings require <placeName>, it is possible to use it as needed in <lem> or <rdg>. But the same rules as for <persName> should be applied; only proper nouns should be considered while common nouns should be ignored.

<measure> –Encoding Measurements

When necessary, you can use the tag <measure> to encoding any references of quantity. An attribute @type should be used to record the kind, e.g. volume, weight, currency… and the quantity, the unit used as well as the commodity measure should be encoded using the attributes: @unit, @quantity and @commodity.
The first one indicates the unit used for the measurement expressed by its standard symbol, e.g. cm, m, ml, km, in …, while @quantity should contain a numeric value, and finally, @commodity for the measured substance. To know more about this, see EGD §7.4..4.

<num> – Encoding Numbers

All numbers in numeral signs as well as anything that has a numerical meaning can be encoded with the element <num>. It should be associated with the attribute @value which content records the machine-actionable form of the entire number. This guide follows the rules provided by the EDG §7.1.

Example
<num value="10">0</num>

Fractions will be given as decimal numbers. Never drop the 0 for fractions smaller than 1 and use the decimal point as marker, e.g. <num value="0.5">. When necessary, round the value to three digits after the decimal point, e.g. encode ⅓ as <num value="0.333">.

Words expressing a number should be encoded at once, e.g. tag the entire expression. Feel free to include words not in themselves expressing the number, but intermingled in the expression. Use it also for an expression combining words and numeral signs together.

Example
<lem wit="#A"><num value="1">sa</num></lem>

Encoding abbreviations

If your text includes abbreviations, two solutions are available for you. You can either flag the abbreviations or resolve them, as already stated in the EGD §7.3.

<abbr> – flagging abbreviations

To bring attention to any abbreviations used in your edition by one or several witnesses, wrap the element <abbr> around the abbreviation. It can be used on its own, outside any apparatus, or even inside the children of <app>.

Example

```
<app>
   <lem type="emn">
      <abbr>mā</abbr>
   </lem>
   <rdg wit="#L #K #M">ma:s·</rdg>
</app>
```

Resolving abbreviations

If you want to expand the abbreviations contained by your edition, add an <abbr> element around the abbreviated form and provide the resolved segment inside the element <expan>. Both should be wrapped by a parent element <choice>. This solution can be used also on its own when all the witnesses use the abbreviated form or inside the apparatus, when individual witnesses are involved. The application of this solution is left to the encoder's discretion, but we recommend it for cases where an abbreviation may be resolved in more than one way and you wish to indicate a specific resolution, either to avoid ambiguity or to emphasize your own interpretation.

Example

```
<choice>
   <abbr>mā</abbr>
   <expan>māṣa</expan>
</choice>
```

Expressing uncertainty

Attributes @cert and @precision

To express uncertainty about your interpretation or underline the fact that you are estimating some of the values provided in the attributes, you can use the attributes @cert or @precision. Both should be used only with the value “low”. The first attribute can be associated with the elements <unclear>, <supplied>, <num> and <seg>, while you will favor @precision with the element <gap/>.

<certainty/>

The empty element <certainty/> indicates the degree of certainty associated with some aspect of the text markup. It works with both attributes @locus and @match. The first one provides information about the subject of your uncertainty, thus the value “name” should be used when you have doubt about the choice of the element or the attribute, while the value “value” concerns the content of the element or the attribute value. The second attribute allows you to identify the ambiguous XML node using an XPath expression.

You can use <certainty/> with a stanza for which you can reasonably guess the metrical and to comment on an extent of a <gap reason="lost"/> you are unsure of, as prescribed by the EGD §2.3.4 and §5.4.6. With this first scenario, the encoding should following the opening part of the <lg> element and associated with @locus="value" and @match="../@met", while with the second, <certainty/> should be declared as a child element of <gap/> turning this empty element as a regular element formed with an opening and closing tags. Its attributes should be set on @locus="name" and @match="..".

Example

```
<lg met="anuṣṭubh">
   <certainty match="../@met" locus="value"/>
   <l>vyavahārān didṛkṣus tu</l>
   [...]
</lg>
```

Example

```
<gap reason="lost" quantity="30" unit="characters">
   <certainty match=".." locus="name"/>
</gap>
```

About formatting

<hi> – Typographical formatting

To format characters, you may find it useful to add simple instructions for the display without the semantic classification layer. You might use it inside your commentaries with the element <hi> with the attribute @rend with one to several of the following values:

- "italic"
- "bold"
- "subscript"
- "superscript"

Though, keep in mind that most of the all essential formatting should be handled globally by the XML tags you have used to encode your edition.

Example
Ed<hi rend="superscript">O</hi>

<list> – Encoding lists

If needed, you might format some text as a structured list using the elements <list> and <item>. We recommend keeping its use to translations and commentaries to avoid too much complexity in your edition. Indeed, ifa multi-level list is available, we don't recommend using it with the internal apparatus system of the edition.

Within an enclosing element, mainly <p> or <ab> , create a list with the element <list> with an element <item> to wrap each element of the list. Several styles are available if you add the attribute the attribute @rend: “bulleted” allows creating a bulleted style list and “numbered” will create an auto-numbered list. If you expect 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 nor <label>, the list will take the default plain behaviour with each item being displayed in a new line with an indent.

About languages

@xml:lang – identifying languages

The language of each TEI file forming a critical edition is to be set in English. To do so, an attribute @xml:lang shall be provided on the root element <TEI>. This default setting is to be understood as the language applied to all children elements. It means that every element in another language should be declared with the attribute @xml:lang. Its value should be conformant to the ISO standard 639-3. The main codes relevant for the project are given in the §Appendix C.

In this context, your critical edition should be given its own @xml:lang on the <text> element.[60] A text originally written in any Indic script and edited in Romanised transliteration should be suffixed with "-Latn", following the ISO 15924 code. 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

[60] The hasn't been kept for the present guide, see <text> – Structuring the text.

```
<text xml:space="preserve" xml:lang="osn-Latn">
```

Translations, Commentaries and bibliographies are by default set to English as well and so there is no need to declare the language at the <text> level. 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.

Tagging foreign language texts

Any foreign sections, being a word, a sentence or an even bigger portion of text, should be identified as such if the language shifts. At a word level, we recommend using the element <foreign> in combination with the attribute @xml:lang. It is not mandatory for the edition itself, since all the languages should be identified in the <teiHeader>, even though we recommend using it.

Adding the element <foreign> isn't mandatory, if the section of text concerned by this switch in language is already encoded with a structural element such as , <p>, <ab>, <lg> or <l>. If the part has a matching container, just add the @xml:lang to it with the relevant ISO code.

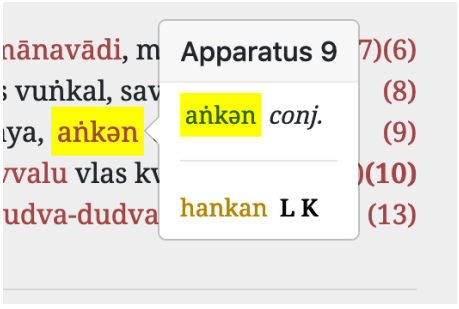
Note that some exceptions in these expectations are made for some editorial interventions. Indeed, the element <note> in the apparatus won't require that you declare the language each time.

Validation and review process

To help you in the reviewing process, you have the possibility to add the attribute @rend with the value "check". The resulting display of the content will be a yellow highlight.

Example

```
<app>
   <lem type="conj" rend="check">aṅkən</lem>
   <rdg wit="#L #K">hankan</rdg>
</app>
```



We recommend using the attribute @rend mainly on existing elements at the higher level of hierarchy in order to avoid needing to repeat the attribute several times on children elements.

To identify a precise part of the text which doesn't match any already existing tags, you can add the element <hi>. Note that any element <hi> only used with the @rend="check" will be deleted at some point. On other elements, only the @rend attribute will be deleted, except if several values are given in the @rend.

Additional contents in external files

Translation

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 source file. Every translation file should have a structure matching the structure chosen for the edition. Even if, for whatever reason, you decide to encode two distinct translations into the same language, 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.

A template is available in the project-documentation gitHub repository.

<teiHeader> for translations

The medata of your edition as a whole shall be declared in the <teiHeader> element of the main source file. See the section §<teiHeader> 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. Metadata will 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 §<publicationStmt> 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 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 prose.[61]

Example

```
<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;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;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>
```

[61] Note that authorship can also be sprecised through the use of attributes

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 §<encodingDesc> to identify all the elements at your disposal.

Finally, keeping track of your work is still recommended with the element <revisionDesc> and the use of the <change> elements. See the section §<revisionDesc>.

Structuring the translation

<text> and <body> elements

To structure the translation, follow the main rules set for the edition and explain under §<text>. Add both 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 C.

** – Sections of translations**

Then use the elements to reproduce the main sections of your edition. We expect you to keep the 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 some 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 ; 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".

Structural organisation of the translation

The main unit level for translated text will be the paragraph. The following elements <p>, <ab> and <lg> used in the edition will be replicated with <p> elements. It will offer you a versatile solution and allows you to organize the translated text as you see fit.

For paragraphs translating verses and stanzas, you can use the attribute @rend="stanza" to obtain a verse type display. In this case, also feel free to reproduce the attribute @n, if you have provided your own specific numeration of the stanza.

Identifying correspondence with editions

Parallelism with the edited text

To match translation parts with its corresponding edited text, we will make use of the declared @xml:id attributes on the elements , <p>, <ab>, <lg> and <quote>. On the paragraphs

used to structure the translation, add the attribute @corresp. Depending on how you have handled the translation process, you can use the attribute @corresp at several levels. Even so, we recommend trying as much as possible to use identifiers from the smaller semantic units such as <p>, <ab> and <lg> when possible. It might be useful to link the since being a higher block structure it will be easier to match between edited texts and translations.

The attribute @corresp shall contained a reference toward the semantic block @xml:id value preceding the an "#" and add a space if you need to declare more than one value. You need to provide every semantic @xml:id matching the translation, even if in case where the content is in fact non-contiguous. It means here that if you need to refer yourself to the paragraphs 1 to 3, you will need to add 3 @xm:l:id references separated by a blank space, while if you need to match your translation with the paragraph 1 and 3, you will add only the @xml:id for the paragraph 1 and 3 separated by a blank space.

Example

```
<div type="dyad" n="1" xml:id="svayambhu-02.01">
<quote xml:id="svayambhu-02.01.01" type="base-text"
xml:lang="san-Latn">
 <supplied reason="subaudible">
          <lg xml:id="svayambhu-02.01.01.01"
met="anuṣṭubh">
          <l>vyavahārān didṛkṣus tu</l> <l>brāhmaṇaiḥ
saha pārthivaḥ</l>
          <l>mantrajñair mantribhiś caiva</l> <l>vinītaḥ
praviśet sabhām ||</l>
          </lg>
    </supplied>
  </quote>
<p xml:id="svayambhu-02.01.02">kunaṅ hulaha saṅ
prabhu, uniṅānira taṅ vyavahāranikaṅ rāt,
harovaṅa ta sira brāhmaṇa vihikan maṅaji,
lavan <app><lem type="emn">mantri vruh</lem><rdg
wit="#L">mantri riṁ vruḥ</rdg><rdg wit="#K">mantri,
vruḥ</rdg></app>
<app><lem wit="#L">mavivekā</lem><rdg
wit="#K">mavikvekā</rdg></app>,
sulakṣaṇaha ta sira tumamaha riṅ sabhā.</p>
</div>
```

```
<div>
   <p n="1" corresp="#svayambhu-02.01"><quote
rend="stanza"><supplied reason="subaudible">When
the king is going to try a case, he should enter the court
modestly accompanied by Brahmins and counselors
who are experts in policy</supplied></quote> Now
should his Majesty undertake to attend to
litigation/affairs of this world, let him take as
companions Brahmins who are knowledgeable and
study Scripture, and likewise ministers with the wisdom
to judge with discernment; with good conduct, he
should enter the council.</p>
</div>
```

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 or <p>, for which you want to declare the author and its publication. Fill it 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's members are involved, it is possible to declare several references in those two attributes, as long as you add a space between each.

Encoding features for the translation

If you have encoded an edition, you do not need to replicate its encoded features in its translation. You may need however to mark up elements of the translation itself, like omitted text which are translating or the text of the translation reports. Note that miscellaneous encoding elements such as <foreign> or <hi> are available as well as editors’ remarks, not part of the translation per se, but encoded as <note>.[62]

For the sake of alignment between the edited texts and its translations, we will need to add some encoding features to reproduce some scribal interventions or damages made to the texts.

Additions into the translated text

Restoration of texts made 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. Its use isn't mandatory and its appreciation is left to the encoders depending on how literal the translation is.

The second case should be identified with the value “subaudible”. It must be used for words, not explicit in the edited texts, but that need to be made explicit to become 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” beared by 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 texts inside your translation.

[62] Most of this section will follow features already made available in DHARMA EGC from section §9.2.9 to §9.2.14)

Lacunae

Lacunae that can't be restored should be encoded with the <gap/> element with the same rules as explained in subsections of §Variant readings. You can keep all the attributes in cases where you need to explicitly represent information about the extent of the lacuna. However, if you don't need to, just keep the attribute @reason. However, if you attempt to restore it, use the <supplied> element with the attribute @reson="lost". We foresee in those cases that it will still be used in conjunction with a <gap/> element, since there is no way to know for larger lacunae all that have been lost.

If your intent is to represent a segment of text that has been skyped in the translation because it is not intelligible, then use the element <gap/> with an attribute @reason with the value "ellipsis". After such a phenomena, we expect either, for short segment of text, the <supplied reason="explanation"> with the child element <foreign> replicating the unintelligible text, or either, in case of larger segment of text, a <note> element shall provided an explanation of the missing part.

The <gap reason="ellipsis"/> can also be used to encode any extent of texts that you deem unnecessary to translate. We recommend in this case to also provide an explanatory <note> element right after the ellipsis.

Expressing doubts and incorrectness

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 uncertainty might arise from your own tentativeness or from the edited text whether being unclear, illegible, restored or by the language itself.

When text seems inappropriate as well as unexpected, tag the equivalent in the translation with the element <sic>.

Indicating bitextuality

It is possible to add 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.

Providing notes

You may add any necessary <note> in your translation as long as they are contained inside the <p> structuring translation and applying basic rules already stated in the section §<note> – Generic Notes.

You may add a credit note at the beginning of the translation with the @type="credit" as the first item within the first , or as the second item, immediately after the custom <head> if one is used as per § – Sections of translations. Its content shall be free text with one or more complete sentences, explaining the authorship of the translation in situations such as collaborative translation involving people outside DHARMA, a partial revision of a previously published translation by you or other DHARMA members, a use of an unpublished translation by a person outside DHARMA or any other circumstance that you might feel the need to clarify.

Commentary

Commentaries of critical editions as well as diplomatic editions are to be done in another file, to reduce the length of file and make it easier to work on it. As for translations, the edition file shall be understood as the main source, and the commentary file, if produced, shall be a secondary file linked to the first. It also means that you will have to follow the same recommendations regarding metadata.

The commentary on the edition isn't mandatory, but we offer this possibility if you feel that the <note> elements of the apparatus are not sufficient enough to cover all you have to say about your edition. It can be used to discuss readings for your edition with any details that could not be encoded within the apparatus, such as vague readings, precision about metrical phenomena or uncertainty about verse metres, your interpretation that impacts the emendations, conjectures, alternatives readings or the translation, and so on. Nonetheless, in case of doubts always favor filling in the apparatus itself rather than using the commentary file.

A template is provided in the project-documentation gitHub repository.

<teiHeader> for commentaries

As already stated in the section regarding translations, metadata about editions as a whole should be declared in the <teiHeader> of the main source file; see the section §<teiHeader>, and so the commentary <teiHeader> shall not contain any repetition of those metadata. Though, it is necessary to provide the basic information to insure you will on a long term still be able to identify it as the commentary of the edition file.

Concretely, the main parts of the <teiHeader> are kept; meaning, you will have a <fileDesc> gathering a <titleStmt>, a <publicationStmt> and a <sourceDesc>, followed by an <encodingDesc> as well as <revisionDesc>. This section will only give specific commentary and suppose you know the basics of encoding a <teiHeader>.

It is expected that you give a <title>. We don't expect any attributes in this case. Indeed, we recommend that you keep it simple and not ambiguous. But if you feel the need to add some attributes, to add a subtitle for instance, use those offered for editions adding the value "commentary" as @subtype. Then add <respStmt> elements to provide a list of people involved specifically with the commentary work and give a description of their role with the element <resp>. If the <editor> is the same, feel free to declare them as well.

You will also need to provide the <publicationStmt>; follow the section §<publicationStmt>, of this guide to fill it. The only difference shall be for the <idno type="filename”>. Indeed, it has to follow the file naming conventions set for the commentary 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. See the example XXX, in §Translation. The <encodingDesc> will at least contain the <projectDesc> element with its first mandatory <p> about the DHARMA project. If you have been providing another explanation about your own project, please copy and paste it

in your commentary file as well. Any other elements used for the main edition in <projectDesc> are available for the commentary, if necessary, but none are mandatory. Finally, keeping track of your work is still recommended with the element <revisionDesc> and the use of the <change> elements. See the section §<revisionDesc> to know more about it.

Encoding the commentary

<text> and <body> elements

To structure the commentary, follow the main rules set for the edition and explain under §<text>. Add both mandatory elements <text> with an attribute @xml:space="preserve" and <body>. Your commentary is expected in English and so you don't need to redeclare an attribute @xml:lang, since one is already declared on the <TEI> root of the file.

** – Sections of commentary**

Then use the elements to structure the main sections of your commentary. Each new subject should be encoded in its own and you can use a <head> element, as the first child of , to provide a title or use the attributes @n and @type on to specify its content.

Structural organisation of the commentary

The only unit level for commentary free-prose text will be the paragraph <p>, which may include globally permitted markup, but avoid non-XML markup such as brackets, asterisks and other signs, when possible.

Indicating correspondence between editions and commentaries

To match the commentary with its corresponding segment of text in the edition, we will make use of the declared @xml:id attributes on the elements , <p>, <ab>, <lg> and <quote>. Add the attribute @corresp on the <p> elements used for commentaries, using the same linking system as explained for translations, see §Indicating correspondence with the edition.

The attribute @corresp shall contained a reference to the @xml:id of the semantic block preceding the an "#". If you need to declare more than one value, add a blank to separate them. Each segment of text concerned by your commentary should be identified by its @xml:id.

Bibliography

Bibliography shall be declared in its own file connected to the main edition, following the same basic rules for commentaries and translations. The same way as for those, a template is given in the project-documentation repository. Your edition must have only one file of bibliography.

<teiHeader> for bibliography

Metadata about editions as a whole should be declared in the <teiHeader> of the edition; see the section §<teiHeader>, and it is not expected for you to repeat those inside the bibliography

file. Though, it is necessary to provide the basic information for long term identification and traceability of the file.

The main parts of the <teiHeader> are kept; meaning, you will have a <fileDesc> gathering a <titleStmt>, a <publicationStmt> and a <sourceDesc>, followed by an <encodingDesc> as well as <revisionDesc>. This section supposes you know the basics of encoding a <teiHeader>.

It is necessary for you to provide a <title>. We recommend that you keep it simple and not ambiguous such as “Bibliography” in association with by the name of your edition or of the text you are editing, e.g. *Saṅ Hyaṅ Siksa Kandaṅ Karəsian Bibliography*. Unlike for commentaries and translations, we don't foresee the work related to the bibliography to differ from people already involved in the main edition. So you won't have to provide any specific <respStmt> elements, declare only editors, if you have declared some in your main file. However, in theory if your bibliography involves a new person, you should give a list of people involved and describe their role with <resp>.

The <pubicationStmt> section will need to be filled in following rules given in the section §<publicationStmt>. The <idno type="filename”> will have to follow the file naming conventions set for the bibliographic file, see FNC, §2.2.4.

In the <sourceDesc> element, you will declare a <biblFull> element and copy and paste the <titleStmt> and <publicationStmt> sections of the edition file. While the <encodingDesc> will at least contain the <projectDesc> element with its first mandatory <p> about the DHARMA project. If you have been providing another explanation about your own project, please copy and paste it in your bibliography file as well. Finally, keeping track of your work is still recommended with the element <revisionDesc> and the use of the <change> elements. See the section §<revisionDesc> to know more about it.

Please note that only management metadata is necessary and expected for the bibliography file.

Encoding the bibliography

<text> and <body> elements

To structure the bibliographic file, you still need to follow the main rules set for the edition and explain under §<text>. Elements <text> with an attribute @xml:space="preserve”[63] and <body> will be already set up in the template. Eventual remarks are all expected in English, so you don't need to give an attribute @xml:lang, since one is declared by default on the root.

<listBibl> – Structural organisation of the bibliography

Any bibliographical item should be enclosed in a <listBibl> element. Unlike the inscription bibliography, we don't expect any @type attribute to distinguish between the primary or the secondary bibliography. Indeed, the primary is rather expected in the <sourceDesc> of the <teiHeader> in order to allow the sources to be identified and processed in the apparatus, see

[63] We are fully aware of the uselessness of this attribute in such a context, but we have kept it to help you with consistent habits.

§<sourceDesc> – Identifying and describing your sources. Should you need it to add primary bibliography, please contact the XML-TEI Data Manager of the project.

All secondary bibliographic items should be filled in the DHARMA Zotero group library and called from there using the referencing system described in the section §bib: — Referring to items in the DHARMA Zotero Group Library. To declare them, you will have to provide a <bibl> element for each bibliographic item and follow the rules stated in the section §Bibliographic citations with Zotero.

If at some point you need to declare an entry directly inside the XML file without using Zotero as a go-between, contact the XML-TEI Data Manager to establish a dedicated encoding model for your entry.

Example
<listBibl>   <bibl>   <ptr target="bib:Naerssen1976_01"/>   </bibl> </listBibl>
Display — Naerssen, Frits Herman van. 1976. “Tribute to the God and Tribute to the King.” *Southeast Asian history and historiography: essays presented to D G E Hall*, edited by Charles Donald Cowan and Oliver William Wolters, 296–303. Ithaca and London: Cornell University Press.

Sigla for the secondary bibliography

Declaring a siglum for primary sources should be done according to the rules in §<sourceDesc> – Identifying and describing your sources, using an attribute @xml:id to make it machine-actionable and an element <abbr> to display it according to your wishes. The secondary bibliographic sigla will follow the EGD logic, stated in the section §9.4.3. You will need to declare an attribute @n on the <bibl> element and it will be fetched to display your critical edition. Note that it is to be used only when you feel the need. So, unlike primary source sigla, they are in no way mandatory.

We recommend you try to keep your sigla short and simple as well as unique and reasonably straightforward to be understood by any reader. To ease processing, you should use ASCII characters and numbers in priority. Symbols and letters with diacritics are allowed but make sure to use the correct UTF-8 codes when using them to allow the machine-actionability. However, do not use &, < and >; those characters require to be escaped in the XML environment. You should note as well that the character + shall be transformed as a “&” in the display and is expected for the sigla of multi-author editions.

Generally, use only the initial of the surname of the author, e.g. “F” for “Fleet” and “C” for “Chhabra”. When the publication has more than one author, use the initial of the surname of each author separated with the character “+”, e.g. “S+G” for “Sircar and Gai”, displayed as “S&G”. To know more about this, refer yourself to the EGD §9.4.3.

Example
<listBibl>   <bibl n="vN">   <ptr target="bib:Naerssen1976_01"/>   </bibl> </listBibl>
Display — Naerssen, Frits Herman van. 1976. “Tribute to the God and Tribute to the King.” *Southeast Asian history and historiography: essays presented to D G E Hall*, edited by Charles Donald Cowan and Oliver William Wolters, 296–303. Ithaca and London: Cornell University Press. [siglum **vN**]

Encoding published books

If you are encoding an edition already published. 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> is put after it.

Two possibilities are offered to you:

1- you adapt the edition to match fully the rules stated for natively created 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. If it is the case, follow the rules stated here to organize the content. Then, all the content shall be kept in a single file using the elements <front> and <back> to structure it. Nonetheless, 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 §Encoding an external apparatus.

<front> – Front Matter

If you have chosen the second encoding strategy to encode an already published critical edition, you will store all the content coming before the edition itself inside the <front> element. It can contain a survey of witnesses and sources, bibliography and other content such as an introduction. Each subsection of the front matter should be given its own with a unique @xml:id identifying it.

In this case, you might be more comfortable reproducing a facsimile of this published critical edition and so your <front> will probably contain a survey of the witnesses and sources,

the bibliography, and other information that may be of interest. If you decide in favor of such a strategy, please follow rules to encode the witnesses as if they were declared in the <teiHeader>, see §<sourceDesc> – Identifying and describing your sources, especially about making machine-actionable sigla. In those cases, it means that no witness information shall be declared in the teiheader. 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

```
<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>
```

<back> – Back Matter

The content of the back of an already published edition shouldn't be encoded though. Indeed the edition being encoded, the traditional content found at the end of a publication can be generated automatically rather than being added manually. However, if necessary the <back> element, the last child of the element <text>, is available. Nonetheless, we recommend that you contact the authors and the XML-TEI Data manager, before adding notes, indices or any other kind of information.

If a <back> element is added, use the same kind of structure as the one used for the front matter. A shall enclose each subsection and need to have its own @xml:id attribute.

Example

```
<text>
   <front>
      <div></div>
   </front>
   <body>
      <div></div>
   </body>
```

```
   <back>
      <div></div>
   </back>
</text>
```

Appendix A: Editorial Conventions

In critical apparatus:
• copy/paste <rdg> from a file containing diplomatic edition, if you have one, and maintain full DHARMA transliteration. NB: this means you maintain use of : for tarung, ṃ or m̐ for ulu ricem, etc.
what about hyphenation?
what to do about punctuation?
we recommend use of , : and . in editing OJ text

In edited Old Javanese text

- no insertion of ə where its absence doesn't seem misleading
    - what about -ən forms of irr. pass. where mss. have -an or only -n after consonant and before vowel?
- nevertheless, in order to avoid needless variation of spelling, we take the liberty of simplifying combinations of -ḥ h- at word boundaries
- and do so also in cases of -m̐ ṅ- at word boundaries
- while we retain gemination word-internally, if it is supported by all witnesses
- interpret any ulu ricems in Sanskrit segments as plain m
- normalize vowel length of words borrowed from Sanskrit (and don't create an apparatus entry for any such normalized reading even if the "wrong" vowel length has unanimous support in mss. — see below)
- no variant reading recorded if
    - 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 m̐ on the wrong syllable without the misplacement of m̐ generating any possible meaning
    - the difference between lem and rdg would only concern ś / ṣ / s
    - the difference between lem and rdg would only concern a / 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 tinulis-akən)
    - the difference between lem and rdg would only concern dental/retroflex consonants
    - the difference between lem and rdg would only concern spelling e/ai or o/au

    - the difference between lem and rdg would only concern Cva/Co or Cya/Ce
    - the difference between lem and rdg would only concern C̥/Cra/Crə/Cru
    - the difference between lem and rdg would only concern the difference v/uv/h/Ø
        - we include under this heading mamuhaka :: mamvaka
        - we include under this heading sākṣyakna :: sakṣihakna
        - we include under this heading mavyavahāra :: mavyava:ra
    - the difference between lem and rdg would only concern the difference a/aha/ā
- The reading of any @wit inside <lem> is silently normalized on the same points
    - e.g., L had manavādi under stanza 3, but we quote L as supporting <lem> mānavādi
- word forms not recorded in OJED or words used in meanings not recorded there are indicated by *
- readings that involve emendation and/or a choice between two significantly divergent readings are in red

When the language is Sanskrit, we record variant readings in more detail.

Appendix B: Typology of variant readings and editorial normalizations deemed insignificant

Old Sundanese List

- The anusvāra (ṁ) followed by the velar nasal (ṅ) is merged with ṅ, for example
    - paṁṅaduA → paṅadua
- The visarga (ḥ) followed by the fricative (h) is merged into h, for example
    - maḥhala -> mahala
- The combination of layar followed by r is reduced to r, for example
    - kasampəṙran → kasampəran
- All vowel killers (paten/pangkon, virāma) are omitted in the edition and only shown if a reading is quoted in <rdg/>
- Vowel marker added to a ligature with subscript y, intended to be pronounced before the y (CyVi and CyVe) are rendered as CViya CVeya, for example syi → siya, rye → reya.

Three types of sibilants s, ś, and ṣ are merged into s, except part of ślokas intended as Sanskrit.

Appendix C: 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, 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	x-midkhmer
Khmer, modern	khm
Khmer, old	okz
Malay, modern (Bahasa Malaysia)	zlm
Malay, old	omy
Mon, old	omx
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

REMOVED FROM ELSEWHERE AND PROBABLY TO BE SCRAPPED ALTOGETHER

It may also happen that the lemma is supported by a published edition of your text, and in that case you may refer to it with the @source attribute. If you think or know that the previous edition may have had access to one or more of your witnesses in a physically less damaged state than the state in which you have access to it/them — in other words, if you suspect the editor(s) actually drew their reading from a manuscript witness — then use @source without using @type.

```
<app>
      <lem wit="#A" source="bib:Pollock2004_12">aṣṭādaśaitāni</lem>
       <rdg wit="#B">aṣṭadaśetani</rdg>
</app>
```

Of course it is also possible to reject a reading or an emendation proposed by a previous editor. All of this leads to a large range of possible scenarios which you may confront when dealing with one or more previous editions that we cannot foresee all of them and offer you specific rules. What we can do here is to offer you a variety of strategies that you may adopt as needed. It is important,

1. to describe what you know or infer about the witness(es) used in the previous edition (in a <note> as illustrated in §<sourceDesc>)
2. to always keep in mind the question whether a reading attributed to a previous edition is identifiable as representing one of the witnesses in your own <listWit> and encode accordingly
3. to always try to imagine how your encoding choice will be displayed and whether you need to apply ac/pc labels through <witDetail> (§Encoding readings ante and post correctionem) or can obtain their insertion by encoding an <rdg> with <choice><sic>XYZ</sic><corr>ABC</corr></choice>

1. The previous edition adopts another reading, which you find in one of your witnesses

```
<app>
      <lem wit="#A">aṣṭādaśaitāni</lem>
       <rdg wit="#B #EdP">caturdaśaitāni</rdg>
</app>
```

Display: *aṣṭādaśaitāni*] A, *caturdaśaitāni* B Ed^{P}

2. The previous edition adopts what you know or suspect to be an emendation but one that you actually find in one of your witnesses and that you adopt as <lem>.

```
<app>
      <lem wit="#A #EdP”>aṣṭādaśaitāni</lem>
       <witDetail wit="#EdP” type="pc”/>
       <rdg wit="#B”>aṣṭadaśetani</rdg>
</app>

          Display:       aṣṭādaśaitāni] A Ed Ppc, aṣṭadaśetani B
```

Observe again that this representation means you have no information as to what the previous editor actually read in his witness(es).

4. The previous edition reads one of your witnesses differently and adopts what you know or suspect to be an emendation different from your <lem>

```
<app>
      <lem wit="#A #B">ñovana</lem>
      <rdg wit="#EdA">ñevana</rdg>
       <witDetail wit="#EdA" type="pc"/>
      <rdg wit="#EdA">ñavana</rdg>
       <witDetail wit="#EdA" type="ac"/>
</app>

          Display:       ñovana] A B, ñavana EdAac, ñevana EdApc
```

5. The previous edition reads one of your witnesses differently and adopts what you know or suspect to be an emendation that agrees with your <lem>

```
 <app>
          <lem wit="#A #B">ñovana</lem>
          <rdg wit="#EdA">ñevana</rdg>
          <witDetail wit="#EdA" type="pc"/>
          <rdg wit="#EdA">ñavana</rdg>
          <witDetail wit="#EdA" type="ac"/>
  </app>

Display ñovana] A B, ñavana EdAac, ñevana EdApc
```

```
<app>
          <lem wit="#A #B">ñovana</lem>
          <rdg wit="#EdA">ñevana</rdg>
```

```
            <witDetail wit="#EdA" type="pc"/>
            <rdg wit="#EdA">ñavana</rdg>
            <witDetail wit="#EdA" type="ac"/>
            </app>

<app>
   <lem wit="#A #B">ñovana</lem>
   <rdg wit="#EdA">ñevana</rdg>
    <witDetail wit="#EdA" type="pc"/>
   <rdg wit="#EdA">ñavana</rdg>
    <witDetail wit="#EdA" type="ac"/>
</app>

   Display:     ñovana] A B, ñavana EdAac, ñevana EdApc
```

6. The previous edition reads one of your witnesses differently and adopts what you know or suspect to be an emendation that agrees with your <lem>

```
<app>
      <lem wit="#A #B">ñovana</lem>
      <rdg wit="#A" source="bib:Atja1981_03"><sic>ñavana</sic><corr>ñovana</corr></rdg>
</app>
```

Display: *ñovana*] **A B** em. **Ed**[A], *ñavana* **Ed**[A]

The previous edition reads one of your witnesses like you do but adopts what you know or suspect to be an emendation that disagrees with your <lem>

```
<app>
      <lem wit="#A">ñovana</lem>
      <rdg wit="#B">ñevana</lem>
      <rdg wit="#B" source="bib:Atja1981_03"><sic>ñevana</sic><corr>ñaivana</corr></rdg>
</app>

<app>
      <lem wit="#A">ñovana</lem>
      <rdg wit="#B"><sic>ñevana</sic><corr source="bib:Atja1981_03">ñaivana</corr></rdg>
  </app>


  </rdg>
      <rdg wit="#B" source="bib:Atja1981_03"><sic>ñevana</sic><corr>ñaivana</corr></rdg>
</app>


<witDetail type="ac" source="bib:Atja+SalehDanasasmita1981_03"/>
      <rdg wit="#B" source="bib:Atja+SalehDanasasmita1981_03">ñevana</lem>
```

Reporting a gap

If a witness or witnesses have a gap in coverage of the text, whatever the reason and larger extent, it might be noted in the apparatus with the elements <lacunaStart/> and <lacunaEnd/>. The term lacuna used to forge those two elements is confusing. Indeed, the TEI Guidelines (12.1.5 Fragmentary Witnesses)[64] use it to refer to a gap in a witness or a group of witnesses. If a gap occurs in the entire tradition, we recommend you use the section Lacunae.

<lacunaStart/> and <lacunaEnd/> are both empty elements. They are used as stand-off mark-up to avoid overlapping issues. So <lacunaStart> marks the beginning of the gap in a witness, while <lacunaEnd/> stands for its end. Using them within the <rdg> element means you will not need to repeat or comment on the absence of a witness among the variant readings each time an apparatus entry is opened in the length of the gap. Once declared <lacuneStart/> is valid until its end with <lacunaEnd/>.

Note that the use of those elements requires you to work with at least two witnesses.

Let's imagine two witnesses: the A is whole while the B contains a gap.

```
(The text is known by the witnesses A and B, but at this point a gap occurs in B)
<app>
    <lem wit="#A">pakən</lem>
    <rdg wit="#B"><lacunaStart/></rdg>
</app>
(After this point the witness B does not provide text while witness A continues)
[Some text]
(The text continues until the gap in the B witness ends. )
<app>
    <lem wit="#A">lemma</lem>
    <rdg wit="#B"><lacunaEnd/></rdg>
</app>
(Both witnesses are able to provide variants)
```

The elements <lacunaStart/> and <lacunaEnd/> can be used inside a <rdg> already containing a string of characters. The following encoding implies that the witness B has a large missing portion before "pak·kən", which in consequence becomes the first word of this section of its text.

```
(The text is only given by the witness A. The gap in witness B ends here allowing to provide a variant to the lemma)
```

[64] https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPMI

```
<app>
    <lem wit="#A">pakən</lem>
    <rdg wit="#B"><lacunaEnd/>pak·kən</rdg>
</app>
(Both witnesses are able to provide variants)
```

The elements can bear the attribute @wit which allows you to record several witnesses in one single <rdg> element while still being able to record the specific gap-related feature of the preserved portion of an incomplete witness. Note that the witness identifier declared in the <sourceDesc> must be repeated twice: in the @wit attribute of the element <rdg> and in the <lacunaStart/> or <lacunaEnd/> elements.

```
(The text before is only known by the witnesses A and C. The witness B gap ends here
allowing to provide a variant to the lemma A which is identical to the witness C)
<app>
  <lem wit="#A">pakən</lem>
  <rdg wit="#B #C"><lacunaEnd wit="#B"/>pak·kən</rdg>
</app>
(The text continues with the witnesses A, B and C)
```

If the witness B has a finishing gap occurring at the lemma “pakən”, however it only provides the variant in a partial form. In this case, you can use the element <gap/> to complete the first part of the word.

```
<app>
    <lem wit="#A">pakən</lem>
    <rdg wit="#B"><lacunaEnd/><gap reason="lost” quantity="” unit="”>k·kən</rdg>
</app>
```

Fragmentary witnesses

If a witness has only been partially preserved – whether a single fragment, a series of fragments, or a relatively complete text with one or more lacunae – it may be desirable to encode explicitly where its preserved portions begin and end.

The empty tags <witStart/>, which may occur within any <rdg> element, indicates the beginning, or resumption, of a fragmentary witness; while <witEnd/> indicates its end or interruption. It works as the empty elements <lacunaStart/> and <lacunaEnd/> to create a larger span with stand-off marking to avoid overlapping issues.

Note that those elements are to be used only if you have more than two witnesses.

Let's imagine three witnesses so that we can use elements from both sections “Reporting a gap” and “Fragmentary witnesses” – the A is whole, the B contains several gaps

and the C is a single fragment that preserved the text partially. For the former, we would use the <witStart/> to explain among other things the sudden apparition of the witness in the critical edition, while the <lacunaStart/> justify the witness B disappearance.

```
(The text is known by the witnesses A and B, but at this point a gap occurs in B, while the fragment C starts at the lemma "pakən")
<app>
    <lem wit="#A">pakən</lem>
    <rdg wit="#B"><lacunaStart/></rdg>
    <rdg wit="#C"><witStart/>pak·kən</rdg>
</app>
(After this point the witness B does not provide text while witnesses A and C continue)
[Some text]
(At this point the witness B still can't provide a variant, the fragment of witness C ends)
<app>
    <lem wit="#A">lemma</lem>
    <rdg wit="#C"><witEnd/></rdg>
</app>
[Some text]
(The witness C has ended but the witness B resumes)
<app>
    <lem wit="#A">lemma</lem>
    <rdg wit="#B"><lacunaEnd/></rdg>
</app>
```

Apparatus entries stored together in one <listApp>

If you have favored recording your apparatus entry in one single block, after the edited text for instance, or even in an external file, you might need to update some of the previous encoding pattern to underline the links between the edited text and the critical apparatus. The main structure should stay the same: you will declare one main <listApp> to contain all the <app> elements, one for each entry you wish to make. Any <app> should at least contain a <lem> element, and can come with one to several <rdg> elements and eventually a <note> element and the association between any reading with the witness in which it has been found should be given inside an attribute @wit.

The main difference with the previous system is to be found on the use of the @loc attribute on each <app> element. You should use this attribute to document the segment of text concerned by this apparatus entry. It means in this solution that a numbering of the segment of text is necessary at least under the form of @n attribute or a more machine-actionable linking system using @xml:id.[65]

[65] Note that if you favor using @xml:id, you adopt a double-endpoint solution rather than the usual location reference approach. To attribute ids to your elements, you should follow DHARMA's recommendation: numbering the hierarchical structure made by , <p>, <ab> and <lg> elements.

Example

Edited text –

```
<lg n="1.1" met="āryā">
<l n="ab">śaśadharasadṛśasaroje alake devākṣayānake caraṇau</l>
<l n="cd">sitarajabhāsvaravimalau śambhor nityaṃ [ ] pāntu janān</l>
</lg>
[...]
```

Apparatus entry after the edited text or in another file –

```
<listApp>
<app loc="1.1a">
    <lem wit="#msB #msD #msE #msF">alake</lem>
    <rdg wit="#msA">aUNCLlake/UNCL</rdg>
    <rdg wit="#msC">aUNCLla/UNCLke</rdg>
</app>
<app loc="1.1b">
    <lem wit="#msA #msB #msC #msE #msFpcorr">pāntu janān</lem>
    <rdg wit="#msD">pāntu janā×</rdg>
    <rdg wit="#msFacorr">pāta janān</rdg>
</app>
[...]
</listApp>
```

In this example, the stanza has been identified by chapter and by the position of the stanza in this chapter, separated by a dot, i.e. n="1.1" being the first stanza in the first chapter. Verse lines are encoded with an attribute @n as well with letters. All those values should be used to declare the location of the apparatus entry in @loc. To do so, add the reference to the stanza and then the reference to the verse line, i.e. loc="1.1a".

Eventually, you might run into a lack of precision regarding segments of text written in prose and then feel free to overcome this using <anchor/> elements.