

# Palm-Leaf Manuscripts Profiling Initiative (PLMPI) Encoding Guidelines

Giulia Buriola Meneghin  
in collaboration with Giovanni Ciotti and Marco Franceschini



<b>Version History</b>	<b>3</b>
Acknowledgments	3
<b>Preliminaries</b>	<b>4</b>
The PLMPI Encoding Guidelines	4
What is a colophon, and what is it made of?	5
Modules	5
<b>1. General Structure of a PLMPI Encoding</b>	<b>8</b>
<b>2. Encoding Metadata: The &lt;teiHeader&gt;</b>	<b>10</b>
2.1 The <fileDesc>	10
2.1.1 The <titleStmt>	10
2.1.2 The <publicationStmt>	11
2.1.3 The <sourceDesc>	13
2.2 The <encodingDesc>	14
2.2.1 The <projectDesc>	15
2.1.4 The <revisionDesc>	15
<b>3. Encoding Facsimiles</b>	<b>16</b>
3.1 Good Practices	16
3.1.1 Naming Patterns	16
3.1.2 Reporting Urls	16
3.1.2a	17
<b>4. Encoding Colophons and Their Translations</b>	<b>17</b>
<b>Appendix A: Language Codes</b>	<b>38</b>

## Version History

Author(s), Contributor(s)	Version	Changes	Dates
Giulia Buriola Meneghin, Giovanni Ciotti, Marco Franceschini	0.1	Redaction of the first version	2023-01 to 2023-12
Giulia Buriola Meneghin, Giovanni Ciotti, Marco Franceschini	1	Release of first version	2024-01

## Acknowledgments

The present Guide is part of the outcomes of the collaboration between Giulia Buriola Meneghin, Giovanni Ciotti, and Marco Franceschini within the scope of the activities of the Palm-Leaf Manuscript Profiling Initiative (PLMPI). The research of the project – funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2176 “Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures,” project no. 390893796 – was conducted within the scope of (1) the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg and (2) the project Texts Surrounding Texts (TST) Satellite Stanzas, Prefaces and Colophons in South-Indian Manuscripts (collections of the Paris BnF and Hamburg SUB) funded by ANR & DFG (2019–2023). We thank Syd Bauman, Piotr Bański, and Patrick McAllister for the insightful exchanges on the TEI-List. Our gratitude also goes to Peter Scharf, Emmanuel Francis, and Charles Li who shared with us not only their help and accurate opinions, but also various material, both XML-TEI related and non, for consultation.

## Preliminaries

The Palm Leaf Manuscript Profiling Initiative (PLMPI) is a multidisciplinary project presently conducted with the scope of the activities of the Cluster of Excellence “Understanding Written Artefacts” at the University of Hamburg, Germany. It combines input from the humanities (philology and codicology), computer sciences (mark-up language and pattern analysis and recognition) and natural sciences (chemistry and physics) with the aim of studying palm-leaf manuscripts from both material and cultural perspectives. Thus, although in the context of the PLMPI we could rely on – among others – the detailed guidelines for diplomatic editions compiled by D. Balogh and A. Griffiths (2020)<sup>1</sup> and A. Griffiths and A. Janiak (2023)<sup>2</sup> mostly in terms of metadata markup, we developed a new, project-specific strategy for encoding colophons. In this respect, a relevant contribution on such a kind of paratext and its encoding is the study by A. Ajotikar et al. (2016),<sup>3</sup> that provided some means of comparison with the encoding strategy developed for PLMPI. The purpose of the latter, beside recording the historical, geographical, and personal information transmitted in colophons, is that of retrieving stylistic, linguistic, and syntactic patterns that they might share through the identification of thematic modules, as described in section XY. Part of these guidelines appear in the forthcoming volume XY (TST Volume).

## The PLMPI Encoding Guidelines

The present guide is aimed at scholars who are working on a digital edition of colophonic material, and specifically at those working on identifying linguistic and stylistic patterns within different sets of source-materials. The mark up presented here follows the XML-TEI guidelines, established by the Text Encoding Initiative (TEI),<sup>4</sup> that cover the majority of the project’s encoding needs. The chosen and, where necessary, implemented strategies will be expounded in the coming chapters. Besides the TEI guidelines, as stated above, we partly follow the DHARMA Encoding Guide for Diplomatic and Critical Editions (Balogh and Griffiths, 2020; Griffiths and Janiak, 2023) – especially for the metadata storage and the shared, basic encoding approaches; therefore, we adhere to some of the markup strategies provided there.

---

<sup>1</sup> Balogh, Dániel, and Arlo Griffiths. 2020. “DHARMA Encoding Guide for Diplomatic Editions.” Technical Report. EFEO; Humboldt-Universität (Berlin); CEAIS-Centre d’Études del’Inde et del’Asie du Sud. <https://halshs.archives-ouvertes.fr/halshs-02888186>.

<sup>2</sup> Arlo Griffiths, and Axelle Janiak. 2023. “DHARMA Encoding Guide for Critical Editions.” EFEO - École française d’Extrême-Orient; CASE - Centre Asie du Sud-Est; CESA - Centre d’études sud asiatiques et himalayennes. hal-04085137.

<sup>3</sup> Ajotikar, Anuja P., Tanuja P. Ajotikar, and Peter M. Scharf. 2016. “Colophons in Sanskrit Manuscripts: A Study of the Sanskrit Library Manuscript Catalogue of Manuscripts at Harvard University, the University of Pennsylvania, and Brown University.” Banaras Hindu University.

<sup>4</sup> TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [4.6.0]. [2023-04-04]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> ([2023-08-09]).

Before delving into the project specific XML-TEI guidelines, however, we will present a brief review of the technical and/or philological concepts used in the guidelines, and of how we use them.

## **What is a colophon, and what is it made of?**

Throughout the guide, we use the term “colophon” to indicate any statement about the production, internal organisation, and usage of a given manuscript. By and large, as G. Ciotti and M. Franceschini put it,

being a text in itself, a colophon is composed according to a set of more or less fixed conventions, which translate, for instance, into a formulaic use of the language (e.g. use of a specific lexicon, fixed invocations, etc.) and a number of graphic devices (e.g. puṣpikās and piḷḷaiyār cuḷis) (2016, 60).

The main, conceptual distinction we make with regard to a colophon’s content is on what we call the “scribe’s statement” – preferred to “scribal statement” – that indicates that the author of the statement is the scribe (and not just that the statement is about the scribe; please note that this terminological choice improves upon Ciotti & Franceschini, 2016) and the “owner’s statements,” that includes the observations by the owner – who, note, might sometimes also be the scribe – generally found on the guard leaf. In practical terms, within the text of a colophon, we generally find information regarding the scribe and/or owner, their ancestors, their place of provenance, and the dates of the beginning and/or the end of copying – described with a varying level of precision and depth; such information, and how this comes to be shared in colophon(s), is fundamental to understand the potential relations and connections among scribes, owners, and manuscripts. Thanks to the work that Giovanni Ciotti and Marco Franceschini (2016; 2022) have been conducting in the past years on such kind of material, we can observe linguistic, syntactic, and structural patterns that might be of great use in retrieving historically and geographically relevant trajectories – not only of the manuscript sources themselves, but also of their scribes and owners.

## **Modules**

A specificity of the PLMPI, as stated above, is the focus on syntactic and linguistic patterns that appear in colophons. As a consequence, the main conceptual node of our material and subsequent markup is the definition of what we call thematic modules, namely sub-sentences that the editors grouped according to their internal theme and syntactic function. The aim of these macro-structures is that of establishing and highlighting the patterns present in the colophons of a given (colophonic) corpus. The following modules have, so far, been identified and are included in the present encoding guidelines:

1. [apology formula],<sup>5</sup> through which the scribe asks the readers for their understanding with regard to any mistakes made while copying;
2. [appraisal] encompasses a broad category of acknowledgements and appreciations for the act of copying, its benefits and consequences;
3. [borrowing/lending formula] includes borrowing formula put outside the colophon;
4. [copying formula]: indicates the beginning and/or end of the copying process;
5. [date]: provides variedly deep layers of information regarding the date of copying in a single colophon;
6. [invocation] refers to praises to the gods;
7. [item] includes text titles and/or words meaning “manuscript,” “book,” and so forth;
8. [owner] provides the name of the owner and, possibly, information about their family, provenance, and so forth;
9. [philology] indicates certain editorial activities carried out by the scribe, such as checking several extant copies of the same text before copying it into a new manuscript;
10. [scribe] provides the name of the scribe and possibly information about their family, provenance, and so forth;

To conceptually understand the function of modules, consider colophon SUB 35.3112 as an example:

[101v4] ❀ **dundubhināmasaṃvatsare**<sup>6</sup> **dakṣiṇāyane** [[kru]] **griṣṭmaṛtau**  
**āṣāḍamāse** **kṛṣṇapakṣe** **saptamyām** **saumyavāsare** **revatinakṣatre** [101v5]  
**sukarmmanāmayoge** **bavakaraṇe** **asmin** **dine** **nemmili**  
**kāśyapagotraśeṣādriyaputreṇa** **tiruma**[[.]]**lācālu** **kālaprakāśikajyotiśśāstram** ❀  
[102r1] ❀ **svahastalikhitam** ௨

In the year called Dundubhi, during the Southern course of the sun, in the Summer season, month of Āṣāḍa, dark fortnight, seventh [lunar day], Wednesday, constellation of Revati, *yoga* called Sukarmma, *karaṇa* of Bava – on that day, Tirumalācālu son of Nemmili Śeṣādri of the Kāśyapa *gotra* copied with his own hand the treatise *Kālaprakāśikajyoti*.

In terms of modules, we would represent the above data as follows:

❀ **dundubhināmasaṃvatsare** **dakṣiṇāyane** [[kru]] **griṣṭmaṛtau** **āṣāḍamāse**  
**kṛṣṇapakṣe** **saptamyām** **saumyavāsare** **revatinakṣatre** **sukarmmanāmayoge**

<sup>5</sup> We indicate modules by enclosing them within square brackets; this is currently also the way in which they appear in the static visualisations of the colophons (see § **XY**).

<sup>6</sup> Characters in bold correspond to those written in Grantha; non-marked are in Tamil script.

bavakaraṇe asmin dine nemmili kāśyapagotraśeṣādriyaputreṇa tiruma[.]lācālu  
kālaprakāśikajyotiśśāstram ❀ ❀ svahastalikhitam

[date][scribe][item][copying formula]

Some considerations starting from the above module [date] as an example. The fact that we are not concerned with recording the technical calendrical information – but rather with recording the syntactic and linguistic structure(s) through which such calendrical data is constructed and transmitted – is reflected by our encoding: we markup subsections of the module to indicate that the scribe recorded information on the year (i.e. not *which* year), the course of the Sun (not necessarily *which* course of the Sun), the season (not necessarily *which* season), the month (not necessarily *which* month), the *pakṣa* (fortnight) (not necessarily *which* *pakṣa*), the *tithi* (lunar day) (not necessarily *which* *pakṣa*), the weekday (not necessarily *which* weekday), the *nakṣatra* (constellation) (not necessarily *which* *nakṣatra*), the *yoga* (not necessarily *which* *yoga*), the *karaṇa* (not necessarily *which* *karaṇa*). Moreover, we want to mark up each of these information and focus on the specific patterns of expression: for instance, the year, *duṇḍubhināmasaṃvatsare*, is here built by adjoining a value (i.e. the proper name of the year, *duṇḍubhi*), an expletive (i.e. an unessential element, here *nāma*, "named"), and a marker for the term "year" (*saṃvatsara*); the *ayana* segment, *dakṣiṇayana*, is built here by adjoining a value (i.e. *dakṣiṇa*) and a marker for the term "course of the Sun" (i.e. *ayana*), without the presence of an expletive such as "*nāma*" or "*ākhyā*" and so on for each of the subsection of the [date] module. Similarly to this, we have developed a strategy to encode in different ways, to follow the project's need, each module and the information contained therein and deemed relevant.



# 1. General Structure of a PLMPI Encoding

Overall, the material of the project consists of mainly short pieces of text; therefore, we opted for encoding both colophonic material and its translation in two distinct <div> elements (divisions) within the same file. The following is the basic template for encoding, stored in the PLMPI GitHub Repository.<sup>7</sup>

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model
href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng"
type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>

<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="eng">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Repository number </title>
        <author xml:id="maro">
          <name>Mario</name>
          <surname>Rossi</surname>
        </author>
        <resp>Creation of the file</resp>
        <persName ref="part:robi">
          <forename>Roberta</forename>
          <surname>Bianchi</surname>
        </persName>
      </resp>
    </title>
    <publication>
      <authority>Palm-Leaf Manuscript Profiling Initiative (PLMPI)</authority>
      <pubPlace>Hamburg (Germany) and Pondicherry (India)</pubPlace>
      <idno type="filename"/>
      <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>
        </licence>
      </availability>
    </publication>
  </teiHeader>
```

<sup>7</sup>The GitHub for the project is available at <https://github.com/PLMPI>. In the project documentation folder (<https://github.com/PLMPI/project-documentation>), you can find the template files for the encoding.

```

    <p>Copyright © 2023-2024 by Mario Rossi and Roberta Bianchi.</p>
  </licence>
</availability>
<date from="2023" to="2024">2023-2024</date>
  </publicationStmt>
  <sourceDesc>
  <msDesc>
    <msIdentifier>
      <settlement> [...] </settlement>
      <repository> [...] </repository>
      <idno> [...] </idno>
    </msIdentifier>
    <msContents>
      <summary> [...] </summary>
    </msContents>
    <physDesc>
      <p> <!--The general notes on the mss or colophon (either on the physical
side or its content) can be stored here, and will be visualized as a note.--> </p>
    </physDesc>
  </msDesc>
</sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc>
    <!-- example from the DHARMA template for regex to retrieve external links such
as those for the authority lists of names, dates, etc.-->
    <p>The Palm-Leaf Manuscript Profiling Initiative (PLMPI). The research of the project
– funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
under Germany’s Excellence Strategy – EXC 2176 “Understanding Written Artefacts:
Material, Interaction and Transmission in Manuscript Cultures,” project no. 390893796 –
was conducted within the scope of (1) the Centre for the Study of Manuscript Cultures
(CSMC) at Universität Hamburg and (2) the project Texts Surrounding Texts (TST)
Satellite Stanzas, Prefaces and Colophons in South-Indian Manuscripts (collections of the
Paris BnF and Hamburg SUB) funded by ANR & DFG (2019–2023)</p>
  </projectDesc>
</encodingDesc>
  <revisionDesc>
    <change who="part:maro" when="2023-09-12" status="draft">Creation of the
template and basic encoding of the file</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. Add a <change> entry for each significant change to the file. -->
    </revisionDesc>
</teiHeader>
<facsimile>
    <surface xml:id="to-be-filled">
        <graphic url="to-be-filled" />
    </surface>
</facsimile>
<text>
    <body>
        <div type="scribeStmnt" subtype="prose" xml:lang="lan-code">
            <head>Edition</head>
            <ab><!-- Structure as necessary with <div>, <p>, <ab>, <lg> depending the
nature of the text -->
            </ab>
        </div>
        <div type="traslation">
            <head>Translation</head>
            <ab><!-- We do not markup the text within the translation division--></ab>
        </div>
        <div>
            <head>Modular Syntax</head>
            <anchor type="footer"/>
            <!-- This division contains, only in the visualization, the display of the modules
present in the colophon-->
        </div>
    </body>
</text>
</TEI>

```

Table 1

Please note that, in the main <TEI> element, we have set the main language as English, through the @xml:lang="eng".

## 2. Encoding Metadata: The <teiHeader>

As per the TEI Guidelines, “the teiHeader element supplies descriptive and declarative metadata associated with a digital resource.” Among the five possible sub-elements (fileDesc, encodingDesc, profileDesc, xenoData, revisionDesc), only <fileDesc> (file description) is mandatory, and should therefore always be included in the encoding; additionally, we include <encodingDesc> (encoding description), which documents the relationship between an electronic text and the source from which it is derived, and

<revisionDesc> (revision description), signalling all the revisions made within a single file.

## 2.1 The <fileDesc>

As stated above, in the file description one is expected to markup the description of the electronic file; within <fileDesc>, there are two mandatory sub-elements:

- the <titleStmt> (title statement), in which the title of the electronic file is marked up; additionally, we suggest to encode the file's authors/compiler as shown in the table below.
- the <publicationStmt>, which groups information concerning the publication or distribution of the text.

Additionally, we strongly suggest to include the <sourceDesc>, source description, in order to record relevant information regarding the source, such as the place where this is currently stored. In the paragraphs below, some specifications on these three elements and their usage are given.

### 2.1.1 The <titleStmt>

```
<fileDesc>
  <titleStmt>
    <title>Digital edition of the colophons in BN INDEN-2α</title>
    <editor xml:id="maro">
      <forename>Mario</name>
      <surname>Rossi</surname>
    </editor>
    <principal xml:id="robi">
      <forename>Roberta</name>
      <surname>Bianchi</surname>
    </principal>
  </titleStmt>
  [...]
</fileDesc>
```

Table 2

Within the <titleStmt>, we encode:

- in <title>, the title of the electronic work; with respect to this, we opted for using the catalogue number, rather than the original manuscript's title, since in a number of instances this is unknown or too generic. Moreover, since our

electronic work is derived from an existing source text, we follow the TEI suggestion to make the title clearly distinguishable from that of the source text through the sentence “Digital edition of the colophon in [identifier of the source].” This allows to distinguish the electronic work from the source text in citations and in catalogues which contain descriptions of both types of material.

- in <editor>, as many individuals, institutions, or organisations acting as editors, compilers, translators, and so forth as needed. Additionally, but not mandatorily, when the editor does not correspond to the encoder, one can add the element <principal>, where the person who has actually encoding the source can be named. As visible in the above table, each <editor> and/or <principal> has a unique @xml:id. These @xml:ids are created, following the practice of the DHARMA encoding guidelines (Griffiths & Janiak, 2023), by using the first two letters of the name and surname of the involved editors/encoders.<sup>8</sup> Within the parent elements <editor> or <principal>, you can encode the names by dividing them into <name> and <surname>.

### 2.1.2 The <publicationStmt>

```
<fileDesc>
<publicationStmt>
  <authority>Palm-Leaf Manuscript Profiling Initiative (PLMPI)</authority>
  <pubPlace>Hamburg (Germany) and Pondicherry (India)</pubPlace>
  <idno type="filename">BN-INDEN_2α</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 © 2023-2024 by Mario Rossi and Roberta Bianchi.</p>
    </licence>
  </availability>
  <date from="2023" to="2024">2023-2024</date>
</publicationStmt>
[...]
```

```
<fileDesc>
```

Table 3

<sup>8</sup> Please note that, at the time of the first release of the present guidelines and considering the relatively small team of this section of the PLMPI, we do not envisage the creation of an external authority list providing information on editors and compilers.

The <publicationStmnt> groups the metadata on the digital publication, supplying information about the publisher, the publication place, address, identifier, availability and date, and so forth. This information might be transmitted in a simple <p> (paragraph) or through more specific elements, as we propose for the project:

- <authority> supplies the name of a person or other agency responsible for making a work available: in PLMPI, the authority corresponds to the project itself and, therefore, the value of this element should always be “Palm-Leaf Manuscript Profiling Initiative (PLMPI)”;
- <pubPlace> supplies the place(s) of publication; in PLMPI, the publication place(s) should always be “Hamburg (Germany) and Pondicherry (India),” namely the places of the institutions that are working on the project;
- the <idno> (identifier), used to identify the entry in a standardised way, simply includes the ID number of the source as stored in the (digital or physical) repositories.<sup>9</sup> The <idno> should have an attribute @type with the value “filename.”

Within the <availability> element supplies information about the availability of a text, and therefore contains the licence that applies to each single file. As a standard, we use a Creative Commons licence, recorded in the element <licence> and targeting, through the attribute @target, the website where more information on the licence is provided. The type of licence, moreover, is further detailed within a free-text paragraph <p>: this sentence does not need any modification, which be however be required in the <p> following, where you will have to state the years for the copyright holding as well as the name(s) of the copyright holder(s). An additional element <date> might supply in machine-readable terms the dates of duration of the copyright, through the attributes @from and @to as well as the years (or more detailed dates) included in the element <date>.

### 2.1.3 The <sourceDesc>

```
<sourceDesc>
  <msDesc>
    <msIdentifier>
      <settlement>Paris, France</settlement>
      <repository>Bibliothèque nationale de France</repository>
```

---

<sup>9</sup>With regards to the naming patterns, please note that we add an en-dash (“-”) instead of the white space between the repository’s two- or three-letter siglum (i.e. “BN-”) and the other whitespaces – if any (as in BN-INDIEN, not to be given as BN INDIEN); moreover, between the letter singlum and the number(s), whitespaces are replaced by an underscore (“\_”) (“\_”, as in INDEN\_2α). Thus, if the original repository name is BN-INDIEN 2α becomes BN-INDIEN\_2α.

<pre> &lt;idno&gt;BN-INDEN 2α&lt;/idno&gt; &lt;/msIdentifier&gt; &lt;msContents&gt;   &lt;summary&gt; It seems reasonable to understand the word &lt;foreign xml:lang="tam-Latn" &gt;eḷut°&lt;/foreign&gt; in &lt;foreign xml:lang="tam-Latn" facs="#d1"&gt;eḷut° piṛiti&lt;/foreign&gt; as an abbreviation for eḷutappaṭṭa (“that was copied”). (2)   The manuscript is described in detail by Emmanuel Francis in an online   blog   post (<a href="https://tst.hypotheses.org/232">https://tst.hypotheses.org/232</a>). Francis points out that this   colophon is   at the end of the main text of the &lt;foreign xml:lang="tam-Latn"   &gt;Caṅkaṛpanirākaraṇam&lt;/foreign&gt;, although it was copied after its   commentary,   which occupies a later portion of the manuscript with a different   foliation.&lt;/summary&gt;   &lt;msItem&gt;&lt;p&gt;Manuscript of the &lt;foreign xml:lang="tam-Latn"   &gt;Caṅkaṛpanirākaraṇam&lt;/foreign&gt; with commentary.&lt;/p&gt;&lt;/msItem&gt; &lt;/msContents&gt; &lt;/msDesc&gt; &lt;bibl&gt;Cod. Palmbl. I 5 (see VOHD II,4, entry 1123)&lt;/bibl&gt; &lt;/sourceDesc&gt; </pre>	
Table 4	

After <titleStmt> and <publicationStmt>, we include the mandatory element <sourceDesc>, in which we store information regarding the encoded source. Within this element, we include the <msDesc>, manuscript description, that includes in its turn two sub-elements:

- <msIdentifier>, the manuscript identifier, which contains the information required to identify the manuscript. This element contains some further sub-elements:
  - <settlement> records the place of the repository in which the encoded manuscript is stored;<sup>10</sup>
  - <repository> simply supplies the name of the repository;

<sup>10</sup> Please note that this information can be given through different levels of specifications: for instance, although we have not included them in the present guidelines, one could also include the elements <country> and <region>, just before <settlement>.

- `<idno>` contains the name of the manuscript as it is (i.e. without altering dashes and/or whitespaces).<sup>11</sup>
- `<msContents>`, which describes the intellectual content of the manuscript. Within this manuscript's contents, we include two sub-elements in the following, mandatory order:
  - `<summary>`, where we supply possible notes on the colophons that are deemed useful by the editor-encoder: for example, specifications on the calculation of the date(s), on the scribe's hands, and so forth.
  - `<msItem>`, which provides – in a `<p>` – the title of the manuscript itself or of the manuscripts' collection, if available.

Be sure to add the `<summary>` before the `<msItem>`.

The `<summary>` will then be visualised as “Notes”; a possible, clickable output for the textual notes is the following:

► Notes

Manuscript of the *Caṅkaṣpanirākaraṇam* with commentary.

Fig. X: Visualisation of the `<summary>`'s clickable content and of `<msItem>`'s content.

▼ Notes  
It seems reasonable to understand the word *elur* in *elur piritti* as an abbreviation for *elutappatta* (“that was copied”). (2) The manuscript is described in detail by Emmanuel Francis in an online blog post (<https://tst.hypotheses.org/232>). Francis points out that this colophon is at the end of the main text of the *Caṅkaṣpanirākaraṇam*, although it was copied after its commentary, which occupies a later portion of the manuscript with a different foliation.  
Manuscript of the *Caṅkaṣpanirākaraṇam* with commentary.

Fig. X: Visualisation of the clicked `<summary>` and `<msItem>`'s content.

The `<fileDesc>` is, then, ended by a `<bibl>` (bibliographic citation) element, in which we add the reference to the relevant catalogue entries.

## 2.2 The `<encodingDesc>`

The encoding description specifies the methods and editorial principles used in the encoding of the text.

```
<encodingDesc>
  <projectDesc>
    <p>The Palm-Leaf Manuscript Profiling Initiative (PLMPI). The research of
the project –
  funded by the Deutsche Forschungsgemeinschaft (DFG, German Research
Foundation) under
  Germany's Excellence Strategy – EXC 2176 “Understanding Written Artefacts:
Material,
```

<sup>11</sup> This `<idno>` differs from the one in `<publicationStmt>` as the latter refers to the electronic file and is useful for the storing process, while `<idno>` this refers to the “official” name of the source.



<p>Interaction and Transmission in Manuscript Cultures,” project no. 390893796 – was conducted within the scope of (1) the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg and (2) the project Texts Surrounding Texts (TST) Satellite Stanzas, Prefaces and Colophons in South-Indian Manuscripts (collections of the Paris BnF and Hamburg SUB) funded by ANR &amp; DFG (2019–2023)&lt;/p&gt; &lt;/projectDesc&gt; &lt;!-- &lt;schemaRef type="guide" key="PLMPI_template" url="to-be-filled"/&gt; &lt;!--\- UPDATE the link once the release is made official -\-&gt; &lt;listPrefixDef&gt; &lt;prefixDef ident="to-be-filled" matchPattern="([a-zA-Z]+)" replacementPattern="to-be-filled"&gt; &lt;p&gt;Internal URIs using the part prefix to point to person elements in the &lt;ref&gt;persons&lt;/ref&gt; file.&lt;/p&gt; &lt;/prefixDef&gt; &lt;/listPrefixDef&gt;--&gt; &lt;/encodingDesc&gt;</p>
Table XY

### 2.2.1 The <projectDesc>

At the moment, the only mandatory sub-element is the following:

- <projectDesc>, which supplies in a <p> a free-text description of the project and the institution financing it; for the PLMPI, the standard information is:

The Palm-Leaf Manuscript Profiling Initiative (PLMPI). The research of the project –

funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under

Germany’s Excellence Strategy – EXC 2176 “Understanding Written Artefacts: Material,

Interaction and Transmission in Manuscript Cultures,” project no. 390893796 – was

conducted within the scope of (1) the Centre for the Study of Manuscript Cultures

(CSMC) at Universität Hamburg and (2) the project Texts Surrounding Texts

(TST)

Satellite Stanzas, Prefaces and Colophons in South-Indian Manuscripts  
(collections of  
the Paris BnF and Hamburg SUB) funded by ANR & DFG (2019–2023)

Although, at the time being, we use neither a project-specific schema nor external authority lists to store names of places, personal names, and so forth. However, we envisage the creation of such files for future purposes; therefore, we have included in the basic file a section including both of them: the `<schemaRef>` might contain the schema specification, and a specific link to the url to the schema itself.

To conclude, the `<listPrefixDef>` groups the definitions of the list of prefixes used within the encoding, such as “txt,” “bib,” “part” as used in the DHARMA encoding (for further details, see Griffiths & Janiak, 2023, 127f), which will then be retrieved through regular expressions.

#### 2.1.4 The `<revisionDesc>`

Although not mandatory, a good practice would be – as a conclusion to the `<teiHeader>` – having the revision description: here, one can record the changes made to the file in a `<change>` element; a new `<change>` should be added for each relevant change. It specifies in the attributes:

- `@who`, the person making the changes;
- `@when`, the date when the changes were made;
- `@status`, the status of the document (e.g. draft, published, unfinished, etc).

Within `<change>` one should write in words what the changes consist of.

```
<revisionDesc>
  <change who="part:maro" when="2023-09-12" status="draft">Creation of
the template and basic encoding of the file</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. Add a <change> entry for each significant change
to the file. -->
</revisionDesc>
```

### 3. Encoding Facsimiles

Due to the nature of the material, we have decided to include links to the manuscripts' images within each single entry. This can be done by using the <facsimile> element, that stands between the <teiHeader> and the <text>. The <facsimile> allows to create a link between a document of choice (generally images in jpg or pgn, but pdfs are also accepted) and a specific section of the encoded material, as is the case of a manuscript folio and its picture.

In PLMPI, the encoder should associate to each change of folio the picture of the folio itself. By and large, we have decided to keep the encoding of each figure within the single files rather than creating an external authority list; this choice might, however, change with time. The structure of the nesting should be the following:

```
<facsimile>
  <surface xml:id="id">
    <graphic url="to-be-filled"/>
  </surface>
</facsimile>
```

Table 5

In the PLMPI framework, an element <facsimile> should always contain at least one child element <surface>, within which one (or more) element <graphic> should be nested. A <facsimile> can have as many <surface> as the encoder needs, but s/he should bear in mind that each <surface> must have a unique @xml:id. As table XY shows, within each <surface>, the empty element <graphic> will provide information on the file to be linked. Thus, an attribute @url is required; its value will be the path to the source file. When the format is jpg or png, it is advised to include as attributes in <graphic> both @width and @height: they must contain the values, in pixels (px), respectively of the width and height of the image (which can be found in the source file's properties). Please note that the only valid format to indicate the number of pixels is the siglum 'px' preceded by the number of pixels themselves.

If you want to give more specifications or a brief description of the manuscript folio you are referring to, you can do so by adding a <note> within the relevant <surface>, below the <graphic/>. Here, you can simply add whatever information on the folio you might want to add; as for now, this information will not be displayed in the final visualisation.

```
<facsimile>
  <surface xml:id="id">
    <graphic url="to-be-filled"/>
  </surface>
  <note>
  </note>
</facsimile>
```

```
<note>Comment or information on the folio</note>
</surface>
</facsimile>
```

Table 6

While reporting the url path, one should be particularly careful: if the machine does not recognize it, the file you want to recall or display will not be found. You find some good practices both for the naming patterns for @xml:ids and some advice on reporting urls in the devoted section (§3.1 [Good Practices](#))

## 4. Encoding the Sources

Since the encoded material consists of colophons – i.e. generally short sections of text – we have decided to encode the original colophon and its translation in the same xml file. Thus, the main structure in <body/> will contain two parallel divisions <div/>, respectively with a @type="colophon" and a @type="translation", as in the example below.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model
href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng"
type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="eng">
<teiHeader>
<teiHeader> [...] </teiHeader>
<text>
<body>
<div type="colophon-type" subtype="prose" xml:lang="lan-code">
<head>Edition</head>
<ab>Here goes the colophon's text, in Roman script.</ab>
</div>
<div type="translation">
<head>Translation</head>
<ab>Here goes the translation of the colophon.</ab>
</div>
<div>
<head>Modular Syntax</head>
<anchor type="footer"/>
</div>
```

<pre> &lt;/body&gt; &lt;/text&gt; &lt;/TEI&gt; </pre>
Table 7

As Table **XY** shows, the edition and translation <div>s, besides containing each a <head>, whose titles are respectively “Edition,” “Translation,” and “Modular Syntax,” contain the colophonic material encoded in both the relevant division within a <ab>. The third <div>, with the <head> “Modular Syntax,” allows, through the <anchor> with @type “footer,” the correct display and visualisation in a separate layer of the modules (see [§4.1.3](#)).

## 4.1 Divisions

In the following sections, we present how to encode the three <div>s that we envisage in our encoding files.

### 4.1.1 Colophon’s Information in <div>

This first <div> element contains the three attributes @type, @subtype, and @xml:lang.

#### 4.1.1a Possible Values for @type

The attribute @type specifies who is the author of the colophon. The possible values for @type are:

- “scribeStmt” (scribe’s statement) that, as explained in the introduction, indicates that the author of the statement *is* the scribe – and not just that the statement includes information on the actions performed by the scribe (e.g. beginning of copying, etc.);
- “ownStmt” (owner’s statement) that, as explained in the introduction, is less frequent than the scribe’s statement and consists of observations made by the owner – who, note, might sometimes be the scribe – generally found on the guard leaf.

#### 4.1.1b Possible Values for @subtype

This attribute specifies whether the colophon is in prose, meter, or presents a mixture of the two.<sup>12</sup> Therefore, :

---

<sup>12</sup> Although, at this stage, we do not cover the strategies for encoding meters in verse lines, if one is interested in marking up this information, the *DHARMA Encoding Guide for Diplomatic Editions* (Balogh & Griffiths, 2020, 132–139) could be followed.

- If the colophon is completely in prose, we add a @subtype="prose" to the main division.
- If the colophon is completely in meter, we add a @subtype="met" to the main division.
- If the colophon has parts in prose as well as parts in meter, we add a @subtype="mix" to the main division.

#### 4.1.1c Possible Values for @xml:lang

The possible values of @xml:lang can be found in [Appendix A](#).

<div> <div subtype="prose" type="scribeStmt" xml:lang="tam-Latn"> <div>Edition</div> <div>[...]</div> </div> </div>
Table 8

As an example, the <div> in the above table marks up that the colophon consists of a scribe’s statement, written in prose and in Latinized Tamil.

Moreover, as you see, this division must include the <head> element “Edition,” followed by the <ab> (anonymous block), in which the colophon is encoded. In the section below, we introduce the strategy envisaged to encode the proper source material; as already explained in the introduction, the encoding strategy developed for this section has the specific aim of focusing on the modular syntax of each colophon; therefore, the encoding will be particularly detailed.

#### 4.1.1.1 Encoding Colophons in <ab><sup>13</sup>

##### 4.1.1.1a <date > and [date]

To begin with, we have preferred the element <ab> (anonymous block) to <p> (paragraph) due to the structural characterization of colophons, which –despite being somehow assimilable to paragraphs– are better defined as “phrase or inter level elements analogous to, but without the same constraints as, a paragraph” (TEI Consortium, 2022). The anonymous block itself has no specific attributes; in <ab>, the material can simply be encoded as prose, without the addition of other phrase-level elements, or as <lg> (line group) and <l> (lines) if it is in verse (see §XY). In both cases, the following elements should be then added:

<sup>13</sup> Most subchapters of this section are part of a forthcoming article for the volume *Texts Surrounding Texts* (Buriola Meneghin, Ciotti & Franceschini, forthcoming).

- when present, the <date>, used to encoded
- the <seg>s, used to encode sentence- and phrase-level text
- the <anchor>s

The module date is one of the most complex, and thus presents different layers of encoding.

The main hierarchical element, which contains as many <seg>s as needed, is <date>. As customary, in <date> we provide standardised information on the single dates through the following attributes:

- *@when*, if the exact corresponding day in the Gregorian calendar can be arrived at, either through the calendrical elements explicitly stated in the date or with the help of additional information available to the encoder (e.g. data found in another date recorded in the same colophon or in a another colophon to the same work); the value of this attribute should supply the value of the date in a standard form, e.g. yyyy-mm-dd;
- *@from* and *@to*, respectively indicating the starting and ending point of the period to be encoded; these attributes should be used only when a “from-to” range is explicitly mentioned in the text;
- *@notBefore* and *@notAfter*: these two attributes provide a time-range for the encoded date. The first reports, in a standard form, the earliest possible date for the encoded event; the second reports the latest possible date for the encoded event. They can also be used singularly.

Additionally, in recording Indic dates, there is a relevant element to take into consideration – namely the fact that a date might be a *single-match date* (i.e. convertible into a specific day in the Gregorian calendar), a *cyclical date* (i.e. recurring at regular intervals, typically – but not exclusively – every 60 years, being based on a cyclical so-called Jovian year), a *multiple-match date* (i.e. a date that can correspond to two or more days in the Gregorian calendar, but which do not recur cyclically), or *problematic* (i.e. a date that has no correspondence in the Gregorian calendar because it contains wrong calendrical elements). Thus, we opted for including this information in an attribute *@type* with “cyclical,”<sup>14</sup> “single-match,” “multiple-match”, or “problematic” as possible values (see the figures below for an application of <date>).

However, in our project, dates are also encoded as modules. The module [date] includes all the layers that a date might express, such as specifications on the year, the month, the day, the season, a festivity, the time, and so forth (see the figures for more practical

---

<sup>14</sup> Here, as in multiple-match dates, it is allowed to include a local <note>, in which possible date conversions, if hypothesised, can be added. At the moment, this information will be stored but probably not used for visualisation purposes.

examples), expressed through an attribute @type. The following values are accepted for @type:

- *date-koy*: this value indicates the fact that the encoded <seg> expresses the Kollam year in a date;
- *date-kay*: this value indicates the fact that the encoded <seg> expresses the Kali year in a date;
- *date-jy*: this value indicates the fact that the encoded <seg> expresses the Jovian year in a date;
- *date-chy*: this value indicates the fact that the encoded <seg> expresses the Christian year in a date;
- *date-śy*: this value indicates the fact that the encoded <seg> expresses the Śaka year in a date;
- *date-sm*: this value indicates the fact that the encoded <seg> expresses the solar month in a date;
- *date-lm*: this value indicates the fact that the encoded <seg> expresses the lunar month in a date;
- *date-wd*: this value indicates the fact that the encoded <seg> expresses the day of the week (*kiḷamai*, *vāram*) in a date;
- *date-t*: this value indicates the fact that the encoded <seg> expresses the day of the lunar month (*tithi*);
- *date-p*: this value indicates the fact that the encoded <seg> expresses the fortnight (*pakṣa*);
- *date-n*: this value indicates the fact that the encoded <seg> expresses the constellation (*nakṣatra*);
- *date-o*: this value indicates the fact that the encoded <seg> expresses other components – such as a part of the day (e.g. morning, evening), a season, a half of the year (i.e. the course of the sun, *ayana*, to the south or to the north of the equator), a festivity – in a date.<sup>15</sup>

These values allow the search of all instances of dates and their different layers. As briefly introduced above, a @subtype further describes the structure of the segment, providing information on the specific word-sequences that appear in each segment.

The values accepted for the @subtype are three:

- *v*: this abbreviation stands for “value” and is used to indicate any kind of specific name or value which identify a year, month, day, etc. See, for instance, the name “*dhātu*” in *dhātunāmasaṃvatsara*, “the year named Dhātu.”
- *e*: this abbreviation stands for “expletive” and is used to indicate any kind of specific expletive that might be attached to a value (v), often (but not always) for

---

<sup>15</sup> Apart from festivities, what is conceptually included in *date-o* is generally not useful for calculating the corresponding date in the Gregorian calendar.



metrical purposes. See for instance, the expletive “*nāma*” in *dhātuṇāmasaṃvatsara*, “the year named Dhātu.”

- *mk*: this abbreviation stands for “marker” and is used to indicate any word or symbol that might be attached to a value (v) or a value (v) cum expletive (e), whose purpose is to identify the category (year, month, day, etc.) the value belongs to. See for instance, the term “*saṃvatsara*” in *dhātuṇāmasaṃvatsara*, “the year named Dhātu”, or the symbol “⟨*Y*⟩” (standing for “Jovian year”) in *vila* ⟨*Y*⟩, “the Jovian year Vila.”

In the case that more than one marker is present in the segment (e.g. in word and in the form of symbol), the encoder can simply add as many “*mk*” as necessary as the value of the @subtype. These values should be assembled according to the order of appearance in the specific colophon’s <seg>. Additionally, to differentiate the type of numbers used (e.g. word, digit, ordinal, cardinal, uncertain), we encode the single numbers in an element <num> (number), defined by the attributes @type and @value. The four possible values of @type are: “digit-c” (cardinal in digit), “word-c” (cardinal in word), “digit-o” (ordinal in digit), or “word-o” (ordinal in word). We also take into consideration the possibility of having numerals in digit for which it is not explicit whether they are cardinal or ordinal: for them, we use the value “digit-u” (number in digit, unclear).<sup>16</sup>

```
<date type="cyclical" >
<seg type="date-jy" subtype="v-mk">
<hi rend="grantha">kharavarṣam</hi></seg>
<seg type="date-sm" subtype="v-mk">
<hi rend="grantha">māghamāsam</hi></seg>
<seg type="date-md" subtype="v-mk" n="digit-o">
<hi rend="grantha"> <num type="digit-c">27</num></hi> <g
facs="#d2">D2</g></seg>
</date>
```

Table 9: Encoding of the [date] module in SUB 35.3008: *kharavarṣam māghamāsam 27* ⟨D⟩, “year of Khara, month of Māgha, day 27”.

In the table above, the <seg>s within <date> mark up three different layers of the date. In the first segment, marking up *kharavarṣam* (“the year Khara”), the value of the segment’s @type indicates that we are dealing with a Jovian year in a date (date-jy); its @subtype indicates that the year is represented by the name of the year (marked up with the value “v”) and an additional marker (mk) for year, here *varṣa*. The same applies for the “date-sm” (solar month) segment, in which *māgha* is the month’s value and *māsa*

<sup>16</sup> The uncertainty concerns the numbers in Sanskrit words from 11 to 19, as cardinal and ordinal in the range 11-19 are identical. The numbers of days in Tamil, when expressed in words, are always ordinal.

is the marker term for month. Finally, the segment of the day encloses a day of the month (date-md), in which we find the numeric value of the day (27) as well as a symbol marking the day of the month (mk). The latter is additionally encoded with <g> (glyph),<sup>17</sup> while the digit is marked up by the element <num>, with a @type (digit-c) specifying that it appears as a digit and as an cardinal number.

Table 10 and 11 show slightly more complex examples, in which several @types are used.

```
<date type="multiple-match">
  <seg type="date-jy" subtype="v-mk">dhātuvarṣam</seg>
  <seg type="date-sm" subtype="v-mk">puṛattācī <abbr type="abbr" facs="#m"
    >m°</abbr></seg>
  <seg type="date-md" subtype="v-mk">31 <g facs="#d1">D1</g></seg>
  <seg type="date-wd" subtype="v-mk">tiṅkakkīlamai</seg>
  <seg type="date-ld" subtype="v"><num
type="word-o">pañcamī</num></seg>
  <seg type="date-n" subtype="v-mk">punarvasūnakṣatram</seg>
</date>
```

Table 10: Encoding of the [date] module in SUB 35.3363β: *dhātuvarṣam puṛattācī m° 31 (D) tiṅkakkīlamai pañcamī punarvasūnakṣatram*, “Year of Dhātu, month of Puṛattācī, 31st day, Monday, fifth [lunar day], constellation of Punarvasū”

```
<date type="single-match" when="1803-01-15">
  <seg type="date-jy" subtype="v-mk"><hi rend="grantha">
dundubhivarṣam</hi></seg>
  <seg type="date-sm" subtype="v-mk"><hi
rend="grantha">taimāsaṃ</hi></seg>
  <seg type="date-md" subtype="v-mk"><num type="word-o">ṇālā
  <lb type="2" n="118v1" facs="#to-be-filled"/>
  <hi rend="grantha">n</hi></num><hi rend="grantha">teti</hi></seg>
  <seg type="date-wd" subtype="v-mk"><hi rend="grantha">śa</hi>
  <damage extent="4" unit="character"><supplied resp="#mafr"><hi
rend="grantha">
```

<sup>17</sup> In our work, symbols used as markers are categorised on the basis of their graphic representation and labelled accordingly: e.g. the symbols for “day” can be classed into two groups, {D1} and {D2}. In the present article, though, this categorization has been preserved only in the text represented in the Figures; elsewhere it has been ignored and symbols are distinguished only on the basis of the calendrical element they mark: {Y} (symbol for “Jovian year”), {M} (symbol for “month”), {D} (symbol for “day”). All the graphic representations, reported as sigla in the edition’s text, are encoded as <surface>s and recalled through the attribute @facs. In the visualisation, this is displayed as a link that leads either to a pdf or a jpg file, where images of the various glyphs are collected.

```

        >niki</hi></supplied></damage>_lamai</seg>
    <seg type="date-n" subtype="v-mk" n="fd"><hi
rend="grantha">ci</hi>tt<supplied
    reason="omitted-in-original">r</supplied>ā<hi
rend="grantha">nakṣatraṃ</hi></seg> inta <hi
rend="grantha">śubhadina</hi>ttil
</date>

```

Table 11: Encoding of the [date] module in SUB 35.3028: **duṇḍubhivarṣaṃ taimāsaṃ nālān teti śa[niki]lamai citt<r>ānakṣatraṃ inta śubhadinattil**, “Dundubhi year, Tai month, fourth day, Saturday, constellation of Cittrā – on this auspicious day (...)”

For instance, in both the above tables, besides the already known @types and @subtypes,, we use in @type the value “date-n” (for date’s *nakṣatra*), which can be further specified through the attribute @n (number)<sup>18</sup> that is envisioned to allow the record of same day *nakṣatras* (“pd”, for “present day”) or following day *nakṣatras* (“fd”).<sup>19</sup>

#### 4.1.1.1b [apology formula]

The module is encoded in quite a simple way: the <seg> bears the attribute @type with value “apo” and, for the moment, no @subtype.

```

<seg type="apo">
    <hi rend="grantha">karakṛtam aparāḍham kṣantum arhanti santaḥ</hi>
</seg>

```

Table 12: Encoding of the [apology] module in SUB 35.3008: **karakṛtam aparāḍham kṣantum arhanti santaḥ**, “May the good ones forgive the mistake made by my hand”

#### 4.1.1.1c [copying formula]

The module includes any kind of copying formulas transmitting information on the beginning and/or the end of copying; more specifically, in our project, this consists of the exact verb or verbal constructs through which the act of copying is made explicit (e.g. *eḷuti muki-*, etc.). The used element is <seg>, with the @type indicating that the line includes an ending copying formula (expressed by the attribute’s value

<sup>18</sup> As per TEI guidelines, the attribute @n might contain as a value “any kind of label for an element, which is not necessarily unique within the document.” It does not need to be filled with actual numeric values.

<sup>19</sup> Conventionally, the *nakṣatra* associated with a day is the one which is current at sunrise on that day: this is what we call a “same day *nakṣatra*”. However, at some point during that day the following *nakṣatra* (i.e. the *nakṣatra* which will be current at sunrise on the following day and which, as such, is conventionally associated with the following day) becomes astronomically current; at times, this latter *nakṣatra* (called the “following day *nakṣatra*”) is recorded by the scribe, in place of the conventionally expected “same day *nakṣatra*”.

“copFor-end”) and the @subtype indicating the specific linguistic formulation to provide this information. The values accepted for @type are two:

- **copFor-beg**, which indicates that the copying formula provides information on the beginning of copying the manuscript (e.g. *āṛampam*);
- **copFor-end**, which indicates that the copying formula provides information on the end of copying the manuscript (e.g. [...] *likhitam*);

These two should be used at every occurrence of copying formulas, namely even if in a single colophon there is a copying formula dealing with the beginning of copying as well as one on the end of copying. The values accepted for @subtype are those listed by G. Ciotti in G. Ciotti and M. Franceschini (forthcoming) in the chapter devoted to activities described within the colophons. The chosen values are explicative of the segment’s content and, as visible below, make explicit mention of the grammatical structures and/or specific words used in the segment itself.

```
<seg type="copFor-end" subtype="likh_skt"><hi rend="grantha">likhitam</hi></seg>
```

Table 13: Encoding of the [copying formula] module in SUB 35.3008: *eḷuti mukiñcutu*, “was fully copied”

```
<seg type="copFor-end" subtype="eḷutiṇa">eḷutiṇa</seg>
```

Table 14: Encoding of the [copying formula] module in SUB 35.3028: *eḷutiṇa*, “was copied”

In both the above instances, the encoding marks up in @type the fact the scribe is here referring to the end of copying, while the @subtype specifies the main verbal category used to describe the act.

The values accepted for @subtype are various. We provide a list of values accepted so far, but in case additional ones need to be created, we suggest trying to be as consistent and clear as possible with the naming patterns. In particular, due to the grammatical specificities of the material, the values of this attribute must always include, as the last indicator, the language used for this specific segment, which might be Sanskrit (skt), Tamil (tam), or a mix (mix) of the two.

The chosen values are explanatory of the segment’s content and, as visible below, make explicit mention of the grammatical structures and/or specific words used in the segment itself. Thus, in copFor-beg, the accepted values for @subtype are:

*ārabh*  
*ārampi*

*tuvakku*  
*tuṛaṅku*

*tuṭaṅku*  
*eḷuti-varu*  
*ārampam*

In copFor-end, the accepted values for @subtype are:

<i>likh-fin_skt</i>	<i>eḷuti</i>	<i>ippaṭikku</i>
<i>likhitam_skt</i>	<i>eḷutiṇa</i>	<i>muḷu</i>
<i>vilekhya_skt</i>	<i>eḷuti-muki</i>	<i>muṭi</i>
<i>svahastalikitham_skt</i>	<i>eḷuti-muṭi</i>	<i>muki</i>
<i>vyālikh_skt</i>	<i>eḷuti-murṛ</i>	<i>murṛu</i>
<i>samāp-mix</i>	<i>eḷuti-muki-murṛum</i>	<i>muṭi-murṛu</i>
<i>samāpti-mix</i>	<i>eḷuti-muki-muḷutum</i>	<i>muki-murṛu</i>
<i>svahastalikhitam_tam</i>	<i>eḷuti-muṭi-murṛum</i>	<i>mukivu</i>
<i>hastalihitam_tam</i>	<i>eḷuti-muṭi-muḷutum</i>	<i>āccudu</i>
<i>likhitam_tam</i>	<i>eḷuti-niraiver</i>	<i>saṃpūrṇa_skt</i>
<i>svahastalikhitam-muḷuv</i>	<i>eriyaruḷappaṇṇu</i>	<i>pūrṇa_skt</i>
<i>atum</i>	<i>eḷuttu</i>	<i>lekhana_skt</i>
<i>eḷutu</i>	<i>kaiyeḷuttu</i>	<i>lipi_skt</i>

#### 4.1.1.1d [copied item]

This module aims at indicating the presence and position of the copied item – namely a fully explicit title of a work, any word for “book”, “manuscript”, and so forth, as well as any proleptic pronoun – within the colophonic material. The basic encoding should include the usual parent element <seg> and will depend on the level of complexity of the phrase referring to the copied item. For instance, if the phrase contains a full title and an adjectival pronoun as in table 15, the @type in <seg> might contain one of the following values:

- *copItem* to mark the sources that, as stated in the colophon, are copied by someone;
- *ownItem* to mark the sources that, as stated in the colophon, are owned by someone;
- *cop-ownItem* to mark the sources that, as stated in the colophon, are both copied and owned by the same or different persons.

The @subtype then specifies the form in which the colophon refers to the copied item. Its values can be: “title”, when the actual title of the work appears; “wordMs”(word for manuscript) when any Sanskrit or Tamil word for manuscript or book – such as *pustakam*, *cuvāṭi*, and so forth – is used; and “pron” (indicating either an adjectival pronoun or a pronoun), which implies the presence of any proleptic or deictic element that refers to either the non-explicitly mentioned manuscript or its explicit reference

through the mention of the title or a word for manuscript in the sentence. These values can be combined, as in table 15, if needed.

```
<seg type="copItem" subtype="pron-title">
  <hi rend="grantha">idam</hi>
  <seg type="title"><hi rend="grantha">āraṇam praśnāṣṭakam</hi></seg>
</seg>
```

Table 15: Encoding of the [copied item] module in SUB 35.3008: *idam āraṇam praśnāṣṭakam*, “This *Praśnāṣṭaka* of the *Āraṇa*”

As table 15 shows, within the main segment, we further mark up each part of the copied item phrase: the pronoun or adjectival pronoun can be marked by a <rs> (referencing string) and the title should be marked up as a <seg> with a @type=“title”. Additionally, if <rs> is used, this should be specified by an attribute such as @ref, pointing the title’s unique @xml:id: this would explicitly indicate the relation and connection between *idam* and *āraṇam praśnāṣṭaka*.<sup>20</sup>

In cases where a word for book or manuscript is also present, the phrase will be encoded as follows:

```
<seg type="copItem" subtype="title-wordMs">
  <seg type="title">tirumūḷi</seg>
  <seg type="wordMs"
  subtype="pustakam">po<hi rend="grantha">stakam</hi></seg>
</seg>
```

Table 16: Encoding of the [copied item] module in SUB 35.3170: *tirumūḷipostakam*, “the manuscript of the *Tirumūḷi*”

The main <seg> maintains “copitem” in the @type’s value and informs us that the title is followed by a word for manuscript through its @subtype “title-wordMs”. Within this main <seg>, then, the title will be marked as a <seg> with @type=“title”, while the word for manuscript, here *postakam*, will be encoded as <seg> with @type=“wordMs”. An additional @subtype might provide the normalized form of the word used for referring to the book/manuscript in that scenario, as shown in table 16.

There are, however, simpler cases in which only the title of the manuscript is mentioned, such as that in table 17.

```
<seg type="copItem" subtype="title">śrīhastigirimāhātmyam</seg>
```

<sup>20</sup>An external authority list containing all the texts from which colophons are taken is under construction. As expected, each of them will have a unique @xml:id and will be referred to where needed.

Table 17: Encoding of the [copied item] module in SUB 35.3028:  
*śrīhastigirimāhātmyaṃ*, “the illustrious *Hastigirimāhātmya*”.

#### 4.1.1.1e [scribe] and Personal Names

Encoding persons will appear as a mix of <seg> and <persName> elements. As tables 18 and 19 show, a parent <seg> encloses the whole name segment, bearing a @type attribute specifying that we are dealing with “person” (pers) and a @subtype specifying that the person enclosed in the <seg> is the scribe. Within that, a simple <persName> (personal name) will refer to an external authority list.<sup>21</sup> A <placeName> (name of place) might of course be added, where needed or deemed useful.

```
<seg type="pers" subtype="scribe">
  <persName role="scribe"> kiṭāmpi <hi
rend="grantha">kṛṣṇamācāluvāsale</hi>
  </persName>
</seg>
```

Table 18: Encoding of the [scribe] module in SUB 35.3028: *kiṭāmpi*  
*kṛṣṇamācāluvāsale*, “by Kiṭāmpi Kṛṣṇamācāluvāsa”

In the chance of more intricate references, as the one in figure 10, where we not only have the name of the scribe but also that of his *kula*, we proceed as before and add an attribute @type in <rs> that allows us to provide information on the structural pattern of the referring string. Here, “v” stands for value and refers to the name of the *kula*, namely *kaunḍanya*; *kula* is comparable to the “mk” (marker) used in [date], but is here context-dependent. This means that it should be changed according to the word used in the compound, which might be *vaṃśa*, *gotra*, *anvaya*, and so on; finally, we have “s” (suffix), that refers to any term used for indicating that one is “born from,” “sprung from” and so on, here *utbhavena*.

```
<seg type="pers" subtype="scribe" n="inst" >
  <rs ref="pers:0000X" type="v-kula-s"><hi
rend="grantha">kaunḍanyakulotbhavena</hi></rs>
  <persName role="scribe"><hi
rend="grantha">sūryyanārāyaṇena</hi></persName></seg>
```

<sup>21</sup>The external authority list of person names is still under construction.

Table 19: Encoding of the [scribe] module in SUB 35.3008: *kaṇḍanyakuloṭbhavena sūryyanārāyaṇena*, “by Sūryyanārāyaṇa, who was born in the Kaṇḍanya family”

Additionally, when deemed relevant, we add an attribute @n indicating the case used to indicate the person by whom the source is copied or to whom it will be donated – instrumental, in the above example<sup>22</sup> – within the <seg>. Adding this information might be useful for statistical purposes.

<abbr> and <g>

These two elements are used, within a <date> element, to mark up respectively:

- abbreviations or shorthands, such as those exposed in Conventions under the label Suffixes;
- glyphs or sigla for years, months, and other subgroups used for signalling sections of a date through symbols rather than words, as exposed in Conventions under the label Calendrical Elements.

In both cases, the encoder should include in the relevant element an attribute @facs, which will allow the linking – and the subsequent possibility for display – with the variants for the specific abbreviation or symbol in the tables.

For example:

```
<seg>dhātu <g facs="#yj1a">YJ1a</g> tai <g facs="#m" type="no">>m°</g> 26</seg>
```

The value of @facs should refer back to the value of the attribute @url provided in the element <graphic> in <facsimile> (for the encoding of <facsimile>, please see the section on facsimiles).

The proposed display for the above markup is the following:

dhātu {YJ1a} tai m° 26 mukivu |

YJ1a and m°, thus, will respectively be displayed as a link to the path defined in <facsimile>, reaching the proper files. The relevant documents, be them images or pdf, will be displayed in a separate page (please note that this display can be modified).

<sup>22</sup>Numeration follows, in roman numerals, the Sanskrit sequence of cases: nominative (I), accusative (II), instrumental (III), dative (IV), ablative (V), genitive (VI), locative (VII).



#### 4.1.1.1f [invocation]

This module indicates the presence, within the colophon, of any kind of invocation. The content of the invocation will not be encoded in-depth, since the aim of this module is that of marking the spot(s) where actual historical/geographical data are interspersed with invocations to find out possible patterns. Therefore, invocations should be encoded only when found in between other modules of a single scribe's statement: at this stage, since the project's aim is to mark up syntactic patterns, we do not envisage an in-depth encoding of the historical, geographical, or religious information that might be contained within invocations. Moreover, we do not distinguish between invocations in prose or in verse, as this has no implications in the study of the colophons' syntax and thematic patterns on which the interest of the project lies. Thus, we opted for a basic encoding, with only a @type and no @subtype:

```
<seg type="inv">tiyākāracar tuṇai cōkkaliṅkam tuṇai mīṇāṭci tuṇai cuppiṛamaṇiyar  
tuṇai &#x2013; rācentiramūr̥ti tuṇai aṛam vaḷattavaḷ tuṇai &#x2013; vetaṇāyaki tuṇai  
&#x2013; śrīrāma jeyam &#x2013; <lb type="inline" n="292v7"/> hari om kunaṇṛāka  
kuru vālka kuruve tuṇai keṇapatiye namā &#x2013;</seg>
```

Table 20: Invocation in BN-Indien 959

Such a module would, then, appear in a footer as a separator between other modules as '//'.<sup>23</sup>

#### 4.1.2 Use of <anchor>

We use empty anchor elements to mark up the spots of the colophon in which we have a module's shift, for instance when the content changes from [date] to [invocation] or [ending formula].

```
<anchor type="apo"/>
```

The @unit attribute is mandatory; its value should always be "section". As for the @type attribute, it should indicate which module is starting afterwards, without any other specification. Therefore, the accepted values<sup>23</sup> are:

- *copFor*, which will be displayed as "copying formula";
- *item*, which will be displayed as "item";
- *date*, which will be displayed as "date";
- *apo*, which will be displayed as "apology";

<sup>23</sup> Please note that, generally, for the naming patterns, if the word to be used as value has more than two syllables, we shorten it to its first three letters. However, there are some exceptions, as in the case of "philo", for which two syllables are employed instead of the first three letters, to avoid confusion.

- *own/own?*, which will be displayed either as “ownership” or “ownership,” depending on the certainty on the ownership;
- *copPlace*, which will be displayed as “copying place”;
- *ownScribe*, which will be displayed as “owner and scribe” and is to be used only when it is certain that the scribe is also the owner; this is generally the case for the colophons on guard leaves;
- *scribe*, which will be displayed as “scribe”;
- *phil*, which will be displayed as “philology”;
- *appr*, which will be displayed as “appraisal”;
- *rec*, which will be displayed as “reciter”;
- *inv*, which will be displayed as “//”

These elements will appear in a footer, in the same order in which they appear in the colophon. The modules will be encoded with `<seg>`, a general element that marks up any arbitrary segment below the “chunk” level that gives us the chance to specify which pattern appear in each colophonic `<seg>` through a `@type`, whose generic value relates the content of the segment (i.e. which part of a date it is, whether the year-section, the month-section, or the day-section), and a `@subtype`, that specifies the patterns used for giving the date (i.e. the presence of specific year/month/day’s names, suffixes such as -nāma and -ākhyā, and markers such as -saṃvatsara, -varṣa, etc). In the following paragraphs, we will show how to encode each module in-depth.

#### 4.1.2 Encoding Translations

As per Version 1.0, we have not developed encoding guidelines for the translation division. Thus, this section of your file consists only of the division itself, a title encoded through a `<head>`, and a `<ab>` containing the translation, without further elements or attributes.

```
<div type="traslation">
  <head>Translation</head>
  <ab>It was fully copied on the year called Dhātu, month of Māśi, 8th day, Sunday,
  three hours into the evening.</ab>
</div>
```

Table 21: translation `<div>` of colophon SUB 35.3005

Be reminded that all relevant elements and syntactical patterns are, then, encoded in the “Edition” division.

#### 4.1.3 Econding Modular Syntax

This `<div>` does not contain any free text, as it serves the purpose of creating an additional layer, which will be displayed in the visualisation, where to report the actual sequence of modules as they appear in the “Edition” division.

```
<div >
  <head>Modular Syntax</head>
  <anchor type="footer"/>
</div>
```

Table 22

As you see in the table above, this `<div>` includes the `<head>` “Modular Syntax” and an `<anchor>` with `@type=“footer”`. This `<div>` should not be modified, unless it is decided at the project level.

## Appendix A: Language Codes

The following are the language codes we use. When creating new codes, please consider using a pattern of three letters; although we are aware that this is not always possible and/or the best option, we encourage the employment of the first three characters of the language you are abbreviating.

Language	Code
English	eng
Broken Sanskrit <sup>24</sup>	brk-Latn
Mix of Tamil and Sanskrit	mix-Latn
Sanskrit	san-Latn
Tamil	tam-Latn

---

<sup>24</sup> For a definition of broken sanskrit, please see §[XY](#)

## Appendix B: Good Practices

### 3.1.1 Naming Patterns

Since each @xml:id should be unique, one should pay special attention when creating one.

#### 3.1.1a Figures

### 3.1.2 Reporting Urls

Reporting urls, especially when the files you want to recall are stored in our own pc, might be tricky: if one does not write the *exact* path, the url will not lead to the right folder or subfolder, and consequently image(s) will not be found during the transformation, leading to an empty page. Therefore, we envisage two ways for reporting the url path within the <graphic>:

#### 3.1.2a GitHub

You can use the links stored in the “Issues” section of our GitHub, following the instructions you find [here](#).<sup>25</sup>

#### 3.1.2b Files Stored in Personal Devices

---

<sup>25</sup> <https://github.com/PLMPI/material-all/issues/1>. The page will always be updated every time an addition/deletion will be made.

## References

Ajotikar, Anuja P., Tanuja P. Ajotikar, and Peter M. Scharf. 2016. "Colophons in Sanskrit Manuscripts: A Study of the Sanskrit Library Manuscript Catalogue of Manuscripts at Harvard University, the University of Pennsylvania, and Brown University." Banaras Hindu University: Banaras Hindu University.

Balogh, Dániel, and Arlo Griffiths. 2020. "DHARMA Encoding Guide for Diplomatic Editions." Technical Report. EFEO; Humboldt-Universität (Berlin); CEAIS-Centre d'Études del'Inde et del'Asie du Sud. <https://halshs.archives-ouvertes.fr/halshs-02888186>.

Buriola Meneghin, G., in collaboration with G. Ciotti & M. Franceschini. Forthcoming. "Marking up the Edges: XML TEI Encoding Strategies for Indic Colophons Collection" in E. Francis et al. *Texts Surrounding Texts*.

Ciotti, Giovanni. 2022. "Scribe, Owner, or Both? Some Ambiguities in the Interpretations of Personal Names in Colophons from Tamil Nadu," in Nalini Balbir and Giovanni Ciotti (eds), *The Syntax of Colophons: A Comparative Study across Pothi Manuscripts*, 149-170. Berlin: De Gruyter. <https://doi.org/10.1515/9783110795271-005>.

Ciotti, Giovanni and Franceschini, Marco. 2016. "Certain Times in Uncertain Places: A Study on Scribal Colophons of Manuscripts Written in Tamil and TAMILIAN Grantha Scripts." in Giovanni Ciotti and Hang Lin (eds), *Tracing Manuscripts in Time and Space through Paratexts*, 59-130. Berlin: De Gruyter. <https://doi.org/10.1515/9783110479010-004>.

Ciotti, Giovanni and Marco Franceschini. Forthcoming. *Vijayatetarām Etat Grandhaḥ* [sic!] (*This Manuscript Is most Victorious*).

Franceschini, Marco. 2022. "A Modular Framework for the Analysis of the Dates Found in Manuscripts Written in the Tamil and TAMILIAN Grantha Scripts," in Nalini Balbir and Giovanni Ciotti (eds) *The Syntax of Colophons: A Comparative Study across Pothi Manuscripts*, 171-208. Berlin: De Gruyter. <https://doi.org/10.1515/9783110795271-006>.

Arlo Griffiths, Axelle Janiak. DHARMA Encoding Guide for Critical Editions. EFEO - École française d'Extrême-Orient; CASE - Centre Asie du Sud-Est; CESA - Centre d'études sud asiatiques et himalayennes. 2023. hal-04085137.

TEI Consortium. 2007. "About these Guidelines" in: TEI Consortium (eds) *TEI P5: Guidelines for Electronic Text Encoding and Interchange [4.5.0]*, last modified 25-10-2022, xvii.

## Websites

École française d'Extrême-Orient (Paris, France), in collaboration with HiSoMA Research Centre (Lyon, France). "Early Inscriptions of Āndhradeśa." Accessed April 17, 2023. <http://hisoma.huma-num.fr/exist/apps/EIAD/>.

FOSTER consortium. (2018, November 26). *What is Open Science?*. Zenodo. Accessed April 17, 2023. <https://doi.org/10.5281/zenodo.2629946>

"SARIT: Search and Retrieval of Indic Texts (New)." Accessed April 17, 2023. <https://sarit.indology.info/sarit-pm/works/>.

"Siddham. The Asian Inscription Database." Accessed April 17, 2023. <https://siddham.network/>.