



dharma

Files Naming Conventions

Axelle Janiak & Adeline Levivier

Release Version 1, 2020-11-23

Versioning History

Author(s)	Version	Changes	Date
Axelle Janiak & Adeline Levivier	0.1	Creation of the file and redaction of the first draft	2019-08
Emmanuel Francis, Arlo Griffiths & Annette Schmiedchen	0.2	Review	2019-09
Emmanuel Francis, Axelle Janiak & Adeline Levivier	0.3	New draft, split in two documents FNC and Version Control	2020-04
Annette Schmiedchen & Arlo Griffiths	0.4	Review	2020-07
Dániel Balogh	0.5	Review	2020-11
Emmanuel Francis & Annette Schmiedchen	0.6	Review	2020-11
Axelle Janiak & Adeline Levivier	0.7	Changing some identifier pattern and deleting the appendices B and C	2020-11
Emmanuel Francis, Axelle Janiak & Adeline Levivier	0.8	Formatting and final validation	2020-11
Axelle Janiak & Adeline Levivier	1.0	Official release	2020-11

Contents

Summary	4
Introduction	5
1 Creating Identifiers	5
1.1 Identifiers for Inscriptions	5
1.2 Identifiers for Inscriptions from a Published Collection	6
1.3 Identifiers for Manuscripts	7
1.4 Identifiers for Inscribed Artefacts	7
1.5 Identifiers for Critical Editions	7
1.6 Identifiers for Diplomatic Editions of Manuscripts	8
1.7 Identifiers for Surrogates	8
1.8 Identifiers for Persons	8
1.9 Identifiers for Places	8
1.10 Identifiers for Collections	9
2 Naming Files	9
2.1 Project name	9
2.2 Designation	9
2.2.1 File name for textual data	9
2.2.2 File name for visual data	10
2.2.3 Metadata files	11
2.2.4 Files without identifier	11
2.2.5 Scholarly publications	11
2.3 Optional features	12
2.3.1 Version	12
2.3.2 Date	12
2.4 Extension	12
2.5 Some examples to help you name a file	12
2.5.1 Model for documentation files	12
2.5.2 Model for inscription files	12
2.5.3 Model for visual documentation files	12
2.5.4 Metadata files	13

2.5.5 Archives files	14
Appendix A: Corpora designation and corresponding GitHub repository	15
Task Force A	15
Task Force B	15
Task Force C	16

Summary

The identifier system is organized according to the following patterns. **See part 1 of the documentation for more details.**

1. Inscriptions: INSPallava00236 (**INS + first word designation or acronym of the corpus + 5-digits incremental number**)
2. Inscriptions in Published Collections: INSSII1400035 (**INS + acronym of the publication + volume number with 2-digits + item number with 5-digits**)
3. Manuscripts: LOr4530 (**Library's/repository's official abbreviation + collection if necessary + shelfmark**)
4. Artefacts: ART000001 (**ART + 6-digits incremental number**)
5. Critical Editions: CritEdSiksaGuru (**CritEd + workTitle**)
6. Diplomatic Editions: DiplEdSiksaGuruL633 (**DiplEd + workTitle + Manuscript ID**)
7. Surrogates: SUR000001 (**SUR + 6-digits incremental number**)
8. Persons: PSN00001 (**PSN + 5-digits incremental number**)
9. Places: PLC00001 (**PLC + 5-digits incremental number**)
10. Collections : COL00001 (**COL + 5-digits incremental number**)

The generic file naming system works with this model. **See part 2 of the documentation for more details.**

DHARMA_<designation>.<extension>

The file naming pattern for metadata file is:

DHARMA_mdt_<designation>.<extension>

The file naming for documentation should have a declared version:

DHARMA_<designation>_<version>.<extension>

Introduction

In the collaborative environment of DHARMA, we will exchange many files in many formats using various tools. We all have had the experience of losing time looking for a file because it was not in the right folder, or of wondering what could be inside a file laconically named “document.pdf” or “note.docx”.

A set of File Naming Conventions (FNC) is a framework for naming files and folders, in a way that describes what they contain and how they relate to other files in a consistent, logical and predictable way. If we can all agree to follow such FNC for the remainder of the project, this will help us to save time and maintain peace of mind, while also helping to ensure successful data retrieval and facilitating the management of access rights at folder level.

This guide formulates rules on how to name your files for the best management of all your data that are meant to enter the “DHARMA-base”. But the Guide does not concern your work when you are writing your own articles or books, or are creating/editing any other file as long as it is not meant to be shared with other team members.

Be aware of the difference between an “identifier” and a “filename”. A filename is a name given to an electronically stored computer file, conforming to limitations imposed by the operating system, as in length or restricted choice of characters. While an identifier can be used to form a filename, it serves as a means to identify a unique object through a string with its own specifications.

1 Creating Identifiers

1.1 Identifiers for Inscriptions

Concerning a single inscription, the identifier will be composed of three elements: **INS + corpus designation + item No.**

It must always start with the three letters that form the abbreviation of “inscription”, written in uppercase letters: **INS**.

The **corpus designation** is a one-word designation or acronym of the corpus in question. We retain a number of acronyms from pre-existing projects, notably CIC, CIK, IDENK and EIAD, but changing its form to follow DHARMA format (e.g. the inscription

from Campa C0152 will become INSCIC00152). See Appendix A for the project's corpus designations.

Finally, the **item No** is the number given to identify the inscription in the corpus itself. Format the number with leading zeros so that it always forms a 5-digit number, e.g. 00001.

Some inscriptions are not clearly attributable to any corpus, yet clearly attributable to an area covered by DHARMA. In such rare cases, use Varia as corpus designation, e.g. INSVaria00001. Such a designation can also be used if the provenance/area of origin of the inscription is not clearly known. The EpiDoc files of such inscriptions will be stored in a GitHub repository common to the whole project, that is not related to a specific task-force. For these inscriptions only, the identifier can be changed during the project, in case new knowledge makes it possible to connect them to a more specific corpus.

A single inscription can thus be recorded as belonging to several different corpora, e.g. INSPallava0001, stored in the Github repository *tfa-pallava-epigraphy*, will also be identified as INSEIAD00139 with an element `<altIdentifier>`.¹ However, there should be only one single EpiDoc file for this specific inscription in the entire DHARMA framework. Such instances should thus be recorded in order to keep track of them and avoid duplicate EpiDoc editions. You can use the readme file of the repository on Github to point out that the inscription belongs to a secondary corpus. You can also add a link toward the main edition in the repository to which the "alternative identifier" belongs.

1.2 Identifiers for Inscriptions from a Published Collection

This section is an exception to the identifier pattern given in §1.1. It concerns inscriptions published in the *South-Indian Inscriptions* (SII) collection as well as, if this is found more convenient for other collections such as, e.g., the *Epigraphia Carnatica* (EC and ECNS).²

For corpora of inscriptions defined by such publications --- that is, inscriptions, not associated with any of the DHARMA regional/dynastic corpora and named according to an existing compendium --- it is possible, but not compulsory, to attribute, with the approval of your task-force leader, an ID recording the position in the collection rather than giving an ID related to a corpus defined on other grounds.³

¹ It will be dealt with in the metadata spreadsheets.

² The volumes of the old series of *Epigraphia Carnatica* will be identified as EC, while the ones published in the new series should be abbreviated with ECNS.

³ This option is particularly meant for TF-A, in which a large portion of the SII will be just digitally encoded (without revision by want of access to visual documentation). The IDs following the pattern "INSSII#####" will enable easy location of the files named after it.

In such cases the pattern must be: **INS + abbreviated title + volume No + item No.**

We keep the abbreviation of “inscription” in uppercase letters: **INS**, already used to identify inscriptions in §1.1.

The **abbreviated title** is an acronym for the collection as a whole, e.g. SII, EC, or ECNS.

The **volume No** must be given with a 2-digit number, e.g. 01. The number is used to identify the volume of the collection. For SII, it must be contained between 01 and 41, while *Epigraphia Carnatica* old series will be numbered from 01 to 12 and 14 to 18, and the new series from 01 to 13.

Finally, the **item No** is the number given to identify the inscription in the volume itself. Format the number with leading zeros so that it always forms a 5-digit number, e.g. 00001.

1.3 Identifiers for Manuscripts

For a manuscript, keep the official pattern from the library/repository: **Library's/repository's official abbreviation + collection if necessary + shelfmark.**

1.4 Identifiers for Inscribed Artefacts

The data model of the DHARMA project requires us to treat artefacts, i.e. text supports, as different entities from inscriptions, i.e. inscribed texts, and to assign to each artefact its own identifying number: **ART + item No.**

Use the abbreviation **ART** (in uppercase) to determine the identifier as related to an artefact.

For artefacts the **item No** and is to be formatted with leading zeros so that it always forms a **6-digit number**, e.g. 000001. These item numbers will be pregenerated for you.

1.5 Identifiers for Critical Editions

To identify a critical edition, the pattern will be composed of the following two elements: **CritEd + WorkTitle.**

Start with the abbreviation of “critical edition” with an uppercase letter at the beginning of each part: **CritEd.**

The **WorkTitle** starts with an uppercase letter. Don't use any diacritics. Please consider shortening long titles, to **obtain a string of no more than 15 characters.**

1.6 Identifiers for Diplomatic Editions of Manuscripts

To identify a diplomatic edition of a manuscript, the pattern will be composed of the following three elements: **DiplEd + WorkTitle + manuscriptID**.

Start with the abbreviation of “diplomatic edition”: **DiplEd**.

The **WorkTitle** starts with an uppercase letter and is without diacritics. Please consider shortening long titles, in order to **obtain a string of no more than 15 characters**.

The **manuscriptID** should be established following the pattern given in § 1.3: **Library’s/repository’s official abbreviation + collection if necessary + shelfmark**.

1.7 Identifiers for Surrogates

To identify a surrogate (i.e. estampage, plaster cast, drawing, analog photograph), use **SUR + item No**.

Use the three uppercase letters that form the abbreviation of “surrogate” to start the ID: **SUR**.

The **item No** is to be formatted with leading zeros so that it always forms a **6-digit** number, e.g. 000001.

1.8 Identifiers for Persons

Persons have to be declared as a canonical reference inside an authority file.⁴ An identifier must be provided for each entry. To build the identifier, use **PSN + item No**.

Start with the three uppercase letters that form the abbreviation of “person”: **PSN**.

The **item No** is to be formatted with leading zeros so that it always forms a 5-digit number, e.g. 00001. These numbers will be pregenerated for you.

1.9 Identifiers for Places

Places have to be declared as a canonical reference inside an authority file.⁵ An identifier must be provided for each entry. To build the identifier, **PLC + item No**.

Start with the three uppercase letters that form the abbreviation of “place”: **PLC**.

⁴ The workflow regarding authority file for persons will be described in a dedicated guide (workflow under construction).

⁵ The workflow regarding authority files for places is managed first at the project level and will be imported in the [Pleiades](#) gazetteer with a batch upload. It will help to avoid manually typing each entry. The workflow will be described in a dedicated guide (workflow under construction).

The **item No** is to be formatted with leading zeros so that it always forms a 5-digit number, e.g. 00001. These numbers will be pregenerated for you.

1.10 Identifiers for Collections

To identify a collection or an institution, use **COL + item No**.

Start with the three uppercase letters that form the abbreviation of “collection”: **COL**.

The **item No** is to be formatted with leading zeros so that it always forms a 5-digit number, e.g. 00001. These numbers will be pregenerated for you.

2 Naming Files

Most files related to DHARMA should be named following this model: **DHARMA_<designation>.<extension>**.⁶ However, in some special cases, e.g. documentation files, you will need a specific pattern containing a version statement, so **DHARMA_<designation>_<version>.<extension>**.

Additional date or version can be added, wherever necessary. See § 2.3.

Metadata will require the addition of the prefix “**mdt**” between underscores after the project name and before the <designation>: **DHARMA_mdt_<designation>.<extension>**.

2.1 Project name

The acronym of the project should always be included in the name of a file created in the framework of the project. Always start with it and spell it with capital letters: **DHARMA**. It should always be followed by an underscore.

2.2 Designation

2.2.1 File name for textual data

Always use the text identifier in the <designation> segment.

- For a file containing the edition of an inscription:
DHARMA_<InscriptionID>.<extension>
- For a file containing the edition of an inscription reproduced from a published collection or compendium (see § 1.2):

⁶ In this part, every element in angle brackets should be replaced depending on the context and on your usage.

DHARMA_<InscriptionIDwithinCorpus>.<extension>

- For a file containing a critical edition: **DHARMA_<CritEdID>.<extension>**
- For a file containing a diplomatic edition: **DHARMA_<DiplEdID>.<extension>**

2.2.2 File name for visual data

For visual data (i.e. photos and 3D models), we need the possibility to quickly identify the content of the image, which can be an inscription, an artefact, a manuscript, a surrogate (e.g. an estampage), or the view of a place. The model for the visual data is the same as for any file: it contains successively the project name, the designation (designation of inscription, artefact, manuscript, place), the <view No> with leading zero, when necessary, so that it forms a 3-digit number and, finally the extension defining the format of the file (jpg, tif, raw, nef, ply, obj): **DHARMA_<designation>_<view No>.<extension>**

- For a photo file concerning an inscription: **DHARMA_<InscriptionID>_<view No>.<extension>**
- For a photo file concerning an artefact: **DHARMA_<ArtefactID>_<view No>.<extension>**
- For a photo file concerning a manuscript: **DHARMA_<ManuscriptID>_<view No>.<extension>**
- For a photo file concerning a surrogate: **DHARMA_<SurrogateID>_<view No>.<extension>**
- For a photo file concerning a view of a place (general views, archaeological sites): **DHARMA_<PlaceID>_<view No>.<extension>**

The 3D and the RTI models are managed in a different way given that 3D and RTI processing produce various sets of files named by the software itself, that can not be changed. However, the photo files used to create those models are to be named following the DHARMA convention. Each set is to be stored in a main folder named according to this model:

DHARMA_rti_<InscriptionORArtefactID>

OR

DHARMA_3d_<InscriptionORArtefactID>

It has then to be compressed (the extension for a compressed file is .zip or .rar). The compressed folder includes all the photos needed to process RTI and 3D models.

2.2.3 Metadata files

Metadata require the addition of the prefix “**mdt**” between underscores after the project name and before the <designation>: **DHARMA_mdt_<designation>.<extension>**.

The <designation> should always be the identifier regarding the inscription, artefact, surrogate, person, place or collection described in the file.

2.2.4 Files without identifier

For entities without an identifier, you are free to choose the <designation> segment of the filename. We just ask you to follow some basic rules:

- Avoid blank spaces and underscores
 - use hyphens and “camelCase” in separating parts of file names
- Avoid the following special characters / \ : * ? “ ” < > | [] { } \$ + %
 - note that only the special character & is allowed
- Create meaningful but brief names
 - your entire file path, including folder names, should never contain more than 256 characters
- Avoid diacritics

2.2.5 Scholarly publications

The scholarly publications collected and stored in ShareDocs to create the DHARMA Library have their own naming system:

<authorName>_<year>_<abbreviatedTitle>.<extension>.

If there are two authors, the pattern is:

<authorName1&authorName2>_<year>_<abbreviatedTitle>.<extension>.

If there are more than two authors, the pattern is:

<authorName1&al>_<year>_<abbreviatedTitle>.<extension>.

Finally, in the case of multi-volume works referred to by their abbreviated title (e.g. ARIE), follow the following pattern: **<acronym>_<volume>_<year>.<extension>**.

Presentations from the project’s workshops and conferences can be referenced like this:

<authorName>_<year><eventName>_<abbreviatedTitle>.<extension>,

e.g. Morlock_2019KickOff_IntroductionXmlTeiEpidoc.pptx

Consult the *ShareDocs Memo* for more information or contact the Visual Data Manager.

7

⁷ Guidelines for using ShareDocs are available in the presentation made during the project’s Kick-off meeting (Berlin, 2019). See <https://bit.ly/3nOb9n0>.

2.3 Optional features

The optional features can be used after an underscore for any file format except txt and xml since for those two the versioning is taken care of by git.

2.3.1 Version

The version is to be indicated with the lowercase letter “v” followed by a 2-digit number, e.g. **v01**.

2.3.2 Date

The date in this block has to follow the pattern **YYYYMMDD**. It is mandatory only for archiving purposes but, if it seems useful, it is possible to record a date associated with a version. Note that you must type the version first followed by an underscore and then the date according to this pattern: **v00_YYYYMMDD**.

2.4 Extension

The extension defining the format of the file is mandatory. This extension starts with a full stop and is followed by a code of 3 or 4 letters (e.g. .xml, .jpg, ...).

2.5 Some examples to help you name a file

2.5.1 Model for documentation files

DHARMA_<designation>_<version>.<extension>

DHARMA_ZoteroGuide_v01.pdf

DHARMA_TransliterationGuide_v01.md

DHARMA_TransliterationGuide_v01.pdf

2.5.2 Model for inscription files

DHARMA_<InscriptionID>.<extension>

DHARMA_INSGangaEast00001.xml

DHARMA_INSUttiramerur00007.txt

DHARMA_INSSII1400035.xml

2.5.3 Model for visual documentation files

DHARMA_<InscriptionID>_<view No>.<extension>

DHARMA_INSUttiramerur00007_002.jpg
DHARMA_INSBengalCharters00032_001.ply

DHARMA_<ArtefactID>_<view No>.<extension>

DHARMA_ART000243_001.tif
DHARMA_ART000167_001.nef

DHARMA_<ManuscriptID>_<view No>.<extension>

DHARMA_LOr4530_002.jpg

DHARMA_<SurrogateID>_<view No>.<extension>

DHARMA_SUR000007_002.jpg
DHARMA_SUR000032_001.raw

DHARMA_<PlaceID>_<view No>.<extension>

DHARMA_PLC00065_023.jpg
DHARMA_PLC00178_003.tif

DHARMA_rti_<InscriptionID>.<extension>

DHARMA_rti_INSUttiramerur00007.zip
DHARMA_rti_INSBengalCharters00032.rar

2.5.4 Metadata files

DHARMA_mdt_<InscriptionID>.<extension>

DHARMA_mdt_INSGangaEast00001.xml
DHARMA_mdt_INSSII1400035.xml

DHARMA_mdt_<artefactID>.<extension>

DHARMA_mdt_ART000243.xml
DHARMA_mdt_ART000167.xml

DHARMA_mdt_<SurrogateID>.<extension>

DHARMA_mdt_SUR000007.xml
DHARMA_mdt_SUR000032.rdf

DHARMA_mdt_<PlaceID>.<extension>

DHARMA_mdt_PLC00065.xml
DHARMA_mdt_PLC00178.xml

DHARMA_mdt_<PersonID>.<extension>

DHARMA_mdt_PSN00065.xml

DHARMA_mdt_PSN00178.xml

DHARMA_mdt_<CollectionID>.<extension>

DHARMA_mdt_COL00065.xml

DHARMA_mdt_COL00178.xml

2.5.5 Archives files

DHARMA_<designation>_<version>_<date>.<extension>

DHARMA_FNC_v01_20201106.pdf

Appendix A: Corpora designation and GitHub repository

Task Force A

Name of the corpus	Corpus Designation	GitHub repository
Accāḷpuram	Accalpuram	tfa-accalpuram-epigraphy
Cempiyaṇ Mahādevī	CempiyanMahadevi	tfa-cempiyan-mahadevi-epigraphy
Cīrkāḷi	Cirkali	tfa-cirkali-epigraphy
Koṭumpāḷūr	Kotumpalur	tfa-kotumpalur-epigraphy
Mēlappaḷuvūr & Kīḷappaḷuvūr	MelKil	tfa-melappaluvur-kilappaluvur-epigraphy
Pallava	Pallava	tfa-pallava-epigraphy
Pāṇḍya	Pandya	tfa-pandya-epigraphy
South Indian Inscriptions	SII	tfa-sii-epigraphy
Tamil Nadu (varia)	TamilNadu	tfa-tamilnadu-epigraphy
Uttiramērūr	Uttiramerur	tfa-uttiramerur-epigraphy

Task Force B

Name of the corpus	Corpus Designation	GitHub repository
Arakan	Arakan	tfb-arakan-epigraphy
Bādāmi Cālukya	BadamiCalukya	tfb-badamicalukya-epigraphy
Bengal Charters on Copper Plates, Stones and Other Supports	BengalCharters	tfb-bengalcharters-epigraphy
Bengal Dedication Inscriptions	BengalDED	tfb-bengalded-epigraphy

Bhaumakara	Bhaumakara	tfb-bhaumakara-epigraphy
Dakṣiṇa Kosala	DaksinaKosala	tfb-daksinakosala-epigraphy
Early Andhra	EIAD	tfb-eiad-epigraphy
Eastern Gaṅga	GangaEast	tfb-gangaeast-epigraphy
Epigraphia Carnatica	EC	tfb-ec-epigraphy
Kalyāṇa Cālukya	KalyanaCalukya	tfb-kalyanacalukya-epigraphy
Licchavi	Licchavi	tfb-licchavi-epigraphy
Maitraka	Maitraka	tfb-maitraka-epigraphy
Rāṣṭrakūṭa	Rastrakuta	tfb-rastrakuta-epigraphy
Śailodbhava	Sailodbhava	tfb-sailodbhava-epigraphy
Sātavāhana	Satavahana	tfb-satavahana-epigraphy
Telugu Inscriptions	Telugu	tfb-telugu-epigraphy
Veṅgi Cālukya	VengiCalukya	tfb-vengicalukya-epigraphy
Viṣṇukunḍin	Visnukundin	tfb-visnukundin-epigraphy

Task Force C

Name of the corpus	Corpus Designation	GitHub repository
Campā	CIC	tfc-campa-epigraphy
Khmer	CIK	tfc-khmer-epigraphy
Nusantara	IDENK	tfc-nusantara-epigraphy