Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

CULTURE LETTERARIE E FILOLOGICHE

Ciclo 36°

Settore Concorsuale: 11/A4

Settore Scientifico Disciplinare: M-STO/08

Modelling the infrastructure of the "Śivadharma Database"

Presentata da: Martina Dello Buono

Coordinatore Dottorato
Prof. Marco Antonio Bazzocchi

Supervisore
Prof.ssa Francesca Tomasi

Co-supervisori
Prof.ssa Florinda De Simini
Prof. Aldo Gangemi

Esame finale anno 2024

Che la fiducia in te stesso sia con te.

Abstract

A *Digital Scholarly Edition* is a conceptually and structurally sophisticated entity. Throughout the centuries, diverse methodologies have been employed to reconstruct a text transmitted through one or multiple sources, resulting in various edition types. With the advent of digital technology in philology, these practices have undergone a significant transformation, compelling scholars to reconsider their approach in light of the web. Thus, in the digital age, philologists are expected to possess (too) advanced technical skills to prepare interactive and enriched editions, even though, in most cases, only mechanical or documentary editions are eventually published online.

The *Śivadharma Database* is a web Content Management System (CMS) designed to facilitate the preparation, publication, and updating of Digital Scholarly Editions. By providing scholars with a user-friendly CRUD (Create Read Update Delete) web application to reconstruct and annotate a text, they can prepare their *textus* with additional components such as *apparatus*, notes, translations, citations, and parallels. It is possible by leveraging an annotation system based on HTML and graph data structure. This choice is made because the text entity is multidimensional and multifaceted, even if its sequential presentation constrains it. In particular, editions of South Asian texts of the *Śivadharma corpus*, the case study of this research, contain a series of phenomena that are difficult to manage formally, such as overlapping hierarchies. Hence, it becomes necessary to establish the data structure best suited to represent this complexity.

In *Śivadharma Database*, the *textus* can be viewed as an HTML file that is readily displayable. Textual fragments, annotated via an interface without requiring philologists to write code and saved in the backend in HTML, form the atomic unit of multiple relationships organised in a graph database. This approach enables the formal representation of complex and overlapping textual phenomena.

The methodology and practical approaches adopted by the *Śivadharma Database* allow for good annotation expressiveness with minimal effort required to learn the relevant technologies during the editing workflow.

Keywords: Digital Scholarly Editing; Digital Philology; Data Modelling; Web Design; Web Development.

Table of contents

Introduction

The *Śivadharma Project — Translocal Identities: The Śivadharma and the Making of Regional Religious Traditions in Premodern South Asia*, is a research initiative aimed at exploring the influence of the *Śivadharma* tradition on the propagation of the *Śaiva* religion in South Asia during the Middle Ages and early modern times. The project entails studying the *Śivadharma corpus*, which comprises manuscripts in Sanskrit and Dravidian languages, inscriptions, and icons. This research project is being carried out by a consortium of several institutions, including the L'Orientale University of Naples, the École française d'Extrême-Orient, the DH.arc Research Centre, the Department of Classical Philology and Italian Studies (FICLIT), and the Department of History and Cultures of Alma Mater Studiorum University of Bologna (DISCI).

The team at the central host institution, L'Orientale, is dedicated to preparing Scholarly Editions of *Śivadharma* resources, while a part of the team at DH.arc and FICLIT are developing a web CMS called *Śivadharma Database*. This application offers a user-friendly interface for producing Scholarly Digital Editions from scratch, including storage, publication, cataloguing, updating, and tools for visualisation and browsing. Specifically, it allows annotating texts to provide a *textus* with its complementary components, e.g., *apparatus*, notes, translations, citations, and parallels. In this research, the significance of providing scholars with a *user-friendly* environment to prepare Scholarly Digital Editions without coding emerges. This approach can save significant time that would otherwise be spent learning coding concepts and practice, enables better error handling, and allows for homogeneous scholarly data on which to perform operations such as filtering. Besides, scholars can prepare and publish their editions directly in their final visualisation, which significantly streamlines the entire process.

Starting from a specific case study relating to Sanskrit literature, we identify common patterns in the field of Digital Scholarly Editing and present a *product* that can be reused in other contexts. Indeed, one of the primary objectives of the *Śivadharma Database* is to achieve *flexibility*, making the environment easily customisable and reusable regardless of the application domain. Hence, we present a discussion on the design and development practices employed for creating this resource, aimed at fulfilling the crucial requirement of *flexibility* and *reusability*.

Commencing with the identification of issues and practices relevant to philology, and specifically to ecdotics, this study delves into the techniques of preparation and critical features of *traditional*, i.e., printed, Scholarly Editions (Ch. 1). This specific analysis, which is part of the broader scope of the *Śivadharma* project design stage, is intended to investigate the origins of the *edition* and subsequently explore its evolution in the *digital* landscape. The underlying objective is to gain a

comprehensive understanding of the context in which the research project at hand operates. Subsequently, we debate the essential proficiencies that a *digital* philologist must possess and why such competencies are deemed too advanced from a technical point of view. One of the corollaries of this situation is the paucity of online editions that present a *textus*. It is necessary to foster increased cooperation between Philology and Digital Humanities to address these issues, thereby designing and developing a *digital* environment conducive to resolving the gap between Philology and the digital domain (Ch. 2).

Then, the benchmark stage of the project, specifically related to the context of Digital Scholarly Editing, is presented in Ch. 3. The workflow for preparing digital editions is analysed, presenting the possible practices and their implications at each step. It is essential to establish which tasks to cover in the *Śivadharma Database*, in which sequence, and which technologies best support these features that can be very hard to represent formally and handle technically.

In Ch. 4, we systematically examine the concept of CMS as a viable solution to address the gaps identified in the preceding chapters. Moreover, the study presents an overview of the current state of the art of comparable technologies employed in Digital Scholarly Editing. It is also discussed why *digital* philologists are *no less digital* if they use user-friendly tools.

Ch. 5 constitutes a comprehensive examination of the case study, i.e., the *corpus* of *Śivadharma* texts. This analysis is conducted with an emphasis on complexity at the levels of tradition, textuality, and language. The study identifies phenomena of particular significance, including the relationships between texts within the corpus and their intra- and intertextual connections, which could give rise to a highly complex formal structure.

The last two chapters are devoted to the culmination of these reflections, namely, the *Śivadharma Database*. Ch. 6 focuses on the design choices made for the project, including discussions on project missions and design technical aspects. This includes the presentation of sketches and wireframes for the program interface, strategies for achieving a high level of intuitive usability, and methods for accessing and utilising resources. The chapter concludes with an overview of the look-and-feel of the CMS.

In conclusion, Ch. 7 deliberates on a recurring theme that has been highlighted throughout this entire research: the technical challenges associated with formally representing *overlapping structures*. Hence, the *Śivadharma* technical choices are presented and contextualised from the modelling and implementation points of view. Specifically, the project leverages an annotation system that is based on HTML and a graph data structure. This approach facilitates the formal representation of complex, multidimensional and multifaceted overlapping textual phenomena

while also providing an interface that allows for the annotation of textual fragments without requiring philologists to learn and write code.

Chapter 1

Preparing an edition: the *traditional* workflow

Let us start from the basics. The first lesson to learn, or better, the first lesson learned, is that it is necessary and compelling to comprehend the origin of a product before delving into its development and evolution.

Devoting time and effort to formulating an approximately precise definition of *edition* is the primary task of a digital humanist involved in Digital Scholarly Editing in general and especially in the case of who is going to develop a web platform for preparing and publishing Digital Scholarly Editions. We will call this project *Śivadharma Database*. The definition of *edition* is only an approximately precise definition since scholars disagree on a single, indisputable, and infallible definition (Franzini, Terras and Mahoni 2015, 1:2). Although it is currently the subject of ongoing discussion, understanding the concept behind and identifying an edition's core features serves as a means of laying the groundwork for its evolution.

The creation of an edition results from a lengthy process encompassing numerous elements and variables. Born and raised as *printed*, editions are evolving to *digital*. Nevertheless, the philological and ecdotic processes entailed in their preparation remain unchanged.

1.1 Taking over the complexity of the edition's concept: Sahle's definition

The definition of *edition* carries a certain level of complexity, as the process involved in preparing it is multifaceted. To provide a structured framework for this complexity, we refer to Sahle's definition (2016, 23), which effectively translates the concept of *edition* into written form:

> *A scholarly edition is the critical representation of historical documents.*

The term *edition* refers to critically[1] examining and re-coding a text and its transmission using various methods, tools, and practices from disciplines such as philology and textual criticism to make it worthwhile for research in the Humanities (Sahle 2016, 22). This process may involve

[1] According to Sahle (2016, 22), critical reflection by the philologist responsible for editing a text is a fundamental attitude. The adjective *critical* implies a series of activities in which scholarly knowledge and reasoning are applied to transform a text into an edition. These activities include, for instance, identifying structures or objects of interest, establishing rules for transcription, making judgments regarding punctuation, orthography, emendations, corrections, etc.

materially oriented reproduction of documents and the constitution of new text readings. The text reconstruction may be based on one source, i.e., *codex unicus*, or more — with up to five thousand in the case of the Greek New Testament (Haugen 2020, 359). If the edition is based on a single witness, we say *mono-testimonial* edition. On the contrary, *multi-testimonial* editions are based on more than one witness. In any case, this task aims to reconstitute a reliable version of the text that is as close as possible to its original form, i.e., the *Urtext*, author's intention, or editor's intent (Rizzo 1973).

In the event of more witnesses, text reconstruction assumes that the transmission of the text is rarely linear and often includes contamination through mutual exchange between texts, known as *horizontal contamination* (Italia 2023, 324–50).

1.2 Types of Edition

The structure of an edition reflects the philologist's examination of the text witnesses. Different types of editions exist, depending on the editor's aim and the available scholarly material. Stoppelli's Italian Philology handbook (Stoppelli 2019) and Haugen's classification (Haugen 2014; 2020) outline the main edition types. They are presented below in order of complexity[2] and offer valuable guidance for scholars engaged in preparing editions.

0 ***Mechanical edition***. A mechanical edition is a specific type of textual reproduction method that uses technology, such as photography or scanning, to create a facsimile of the original document. Unlike other types of editions, in this case, the philologist does not interfere with the text in a critical sense.[3] Instead, the technological process acts as the mediator between the original and its reproduction. The primary objective of a mechanical edition is to produce an accurate and reliable version of the original text without any interpretation or modification. This method is particularly beneficial for scholars who require access to the original document but cannot view it in person.

1 ***Documentary edition***. A documentary edition is an edition of a single source, so it does not contain variants from other witnesses. Italia (2020, 65) defines this type of edition *hyper diplomatic* since the philologist has no room to intervene in the text.

[2] Complexity (from 0 to 4.1) in two directions, one dependent on the other. (1) Complexity of the philological work necessary for the preparation of the edition and, consequently, (2) availability of edition supporting and ancillary material, e.g., *apparatus*.

[3] Dahlström defines such scholarly editing as *non-critical*, as it requires no intervention on the text (Dahlström, 78).

2 ***Monotypic/monoptic edition***. This type of edition allows the editor to base their work on either the entire tradition or a single witness, depending on the number of available sources. In contrast to other types of editions, the monotypic/monoptic edition allows the editor to intervene in the text in a critical sense, thereby adding an interpretive layer to the reproduction.

2.1 ***Diplomatic edition***. The diplomatic edition represents a hybrid approach that incorporates both mechanical and critical editing methodologies. In this type of edition, the philologist assumes an active role in the reproduction of the manuscript's content, utilising modern typefaces to convey its meaning.

2.2 ***Diplomatic-interpretative edition***. The diplomatic-interpretative edition strives to reproduce the text of a particular manuscript while respecting its original handwriting. This approach involves minimal interventions by the philologist, such as distinguishing between *u* and *v*. The goal is to maintain the integrity of the manuscript while simultaneously providing a minimal critical interpretation.

3 ***Synoptic edition***. There are two significant types of synoptic editions. One consists of juxtaposing different versions of the text of the same work. The other corresponds to the comparison between the witnesses of the same work. These two types may overlap.

4 ***Eclectic edition***. Eclectic editions always rely on more than one witness. Producing an eclectic edition involves an attempt to approximate an earlier stage of the textual transmission by selecting readings from multiple manuscripts. This approach aims to identify and select the most reliable readings, particularly in instances where conflicting readings are present in the manuscripts. In contrast to the pursuit of the original text, the editor of an eclectic edition typically seeks to discern the author's intention (Roelli *et al.* 2015).

4.1 ***Critical edition***. In the case of critical editions, philologists intervene by restoring the text by critically examining the entire tradition, i.e., all the witnesses transmitted over time. The philologists' work thus consists of a series of attempts to report the author's final intentions, when discernible, or the version of the text deemed most reliable. Thus, they lends their interpretation to variant readings, copying errors, corruptions, cancellations, contaminations, alterations, and other phenomena with the aim of achieving the highest level of textual accuracy possible.

As outlined by Haugen (2020, 361), we can define many editions as *critical* since they offer structured information on some of the aspects of the text restoration. However, such restoration is not as expected in a 'truly critical edition', e.g., not offering a genealogical *recensio* or basing the text on it (Ch. 1.3.2). In reality, this distinction is not clear-cut and well-defined. As the Fig. 1 shows, editions can be placed in multiple groups for their methodological approach to the text.

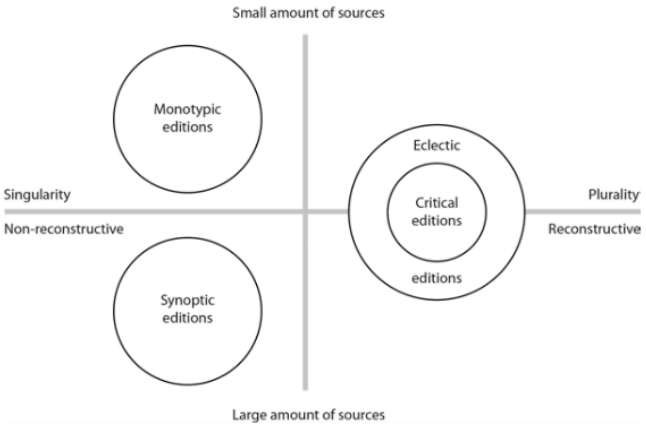


Fig. 1. Editions classification (Haugen 2020, 362).

Philologists choose what type of edition to prepare depending, above all, on the witnesses available. In the case of a *codex unicus*, the only possible types are the *mechanical*, *documentary*, *monotypic*, *diplomatic*, and *diplomatic-interpretative*, but not without exceptions. For instance, multiple versions of the same text may exist, but not all of them can be included in a strict *genealogical recensio*. However, they can clarify some textual aspects that would otherwise not be resolvable.

The choice is not immediate when working with multiple witnesses, i.e., multi-testimonial edition. It is necessary, explains Haugen (2020, 363), to consider multiple variables. The size and complexity of the tradition, the contamination, the degree of fragmentation, the availability of previous editions, and so forth, may influence the choice. So the editor can decide to work with the *archetype*, i.e., an approximation of the original text (Haugen 2020, 363) — *reconstructive editing* — or with the best or most typical manuscript and a selection of manuscripts in a synoptic approach — *non-reconstructive editing*.

1.3 Critical editions

A critical edition aims to establish a reliable version of a text by comparing several of its witnesses in the case of texts in multiple versions or analysing a single manuscript. The result of the editing process is the *textus constitutus*. In addition, specific ancillary material supports the philologist's choices in rebuilding the text and provides evidence.

1.3.1 Evidence to support the text reconstruction in critical editions

The *note to the text* is the first tool readers can find in printed editions to know about the philological work which led to the *textus constitutus*. It documents the tradition, justifies the editor's decisions regarding the most reliable witnesses, and explains the edition criteria (Stoppelli 2019, 31). It can be the preface to the edition and ideally should include the *stemma* (Fisher 2020, 406) (Ch. 1.3.2).

The *apparatus* represents the entire elaboration process and corresponds to the core of the philologist's work (Ch. 1.3.4, 1.3.5). As defined by Italia (2021), the *apparatus* provides documentary evidence of the concrete application of the hypothesis related to the text. This evidence reflects each critical and methodological decision by the philologist, i.e., the readings accepted and those rejected, any corrections, and alternative hypotheses (Stoppelli 2019, 31), which may or may not be explicitly declared (Fischer 2019, 207), uncertainties, changes of witnesses, and brief justifications of editorial decisions (Fischer 2020, 406). In addition, explanatory and commentary *notes* can act as a corollary to the *apparatus* since their consultation concerns specific textual passages (Ch. 1.3.7). Finally, *appendices* can specify other technical information related to the text at the end of an edition (Hunter 2019, 92).

1.3.2 Ecdotic methods in the field

The critical edition of a text is the process of reconstructing a text (Robinson 2016, 196), and its documentation branches off into the previously described components (Ch. 1.3.1). However, how does the philologist work in the field? A critical edition's preparatory work consists of iterative research based on witness(es) and aims to restore a text close to the *original*. The most complex case is the one in which more than one witness is available, where each witness corresponds to a copy of a given text, probably dating back to different periods. In this case, the philologist has to identify the *variant readings*, i.e., the textual alternatives between the witnesses, and choose the most reliable, the one that corresponds to the original version or is the closest to it. Naturally, the selection criteria, e.g., statistics, grammar, semantics, logic, and identifying errors, vary from case to case (Most 2016, 170).

The philological methodologies have refined over time until their formalisation. The first systematisation by the German philologist Karl Lachmann (1793-1851), dating back to the second half of '800, dictates still valid principles related to the treatment of multiple sources (Stoppelli

2019, 71). Lachmann's method[4] involves a series of subsequent operations to build the so-called *stemma codicum*, a hypothesis representing the potential copying process through witnesses (Haugen 2020, 365). The first is the *recensio*, i.e., collecting the available witnesses of a given text. It implies the analysis of the *direct*, i.e., partial or complete copies of the text, and *indirect transmission*, e.g., translations and quotations. The *collatio* compares the available witnesses word by word to identify the occurring differences, i.e., *variant readings*. According to the philologist's interpretation, only a set of readings corresponds to *significant errors*. Errors are essential to establish the genetic relationship between witnesses by eliminating derivative manuscripts and reconstructing lost ancestors (Buzzoni 2020, 382). Principles of logic govern the construction of the *stemma codicum*, whose tree structure summarises such relations. At the top of this model, the *stemma* represents the original, while the *archetype* is immediately beneath it. The leaves of the tree correspond to the preserved copies. A sample *stemma codicum* (Li 2020, 384–385) is presented below (Fig. 2), with five manuscripts, i.e., A, B, C, D, and E. The hypothetical common ancestor of all these manuscripts is α. However, the relationship between α and the original text is unclear. The hypothesis suggested in the *stemma* is that A and B have a common source, as do C and D. E is a direct copy of D. However, it is not helpful for reconstruction since it only contains additional errors.

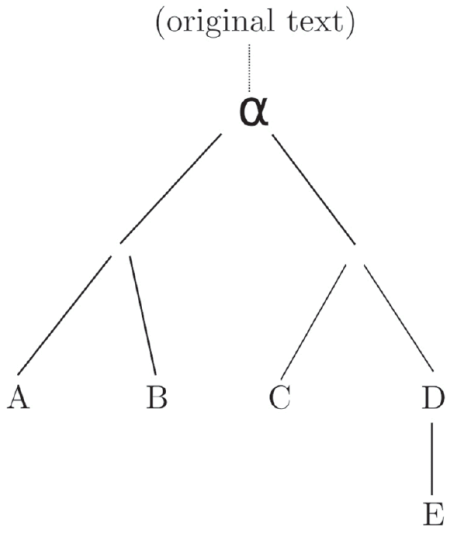


Fig. 2. Sample *stemma codicum* (Li 2020, 284–285).

[4] The expression "Lachmann's method" is conventional. This method, developed during the 19th century and well established in the 1830s, is not the result of the study of a single philologist but of the reasoning of many scholars prior to Lachmann, extension by contemporaries, and refining by later ones (Palumbo 2020, 88).

The *stemma* directs the reconstructive process since it constitutes the first criterion of choice between variants. An in-depth analysis of the sources according to heterogeneous variables, i.e., *examinatio*, leads to the *constitutio textus*, i.e., the establishment of the text as performed by the editor, whose choices in this regard are explained in the *apparatus* to the text (Haugen 2020, 359). During the *examinatio*, the philologist examines the significant errors and the all-out variants. Particular attention is on identifying the corruptions, which can be corrected in the *emendatio* (Buzzoni 2020, 393).

1.3.3 From *stemma codicum* to *apparatus*

Even if it is not a common practice to offer the *stemma codicum* to the readers, such an interrelationship between manuscripts is the entrance door to the edited texts (Haugen 2020, 359). When *examinatio* is not sufficient to make mechanical choices between variants, the *emendatio* is the necessary procedure to judge the variants according to internal criteria, i.e., *selectio*, or conjecturally, i.e., *divinatio*, when the results from *selectio* are not decisive. In general, the rule to apply is the *criterion of the majority of the families* (Bausi 2008, 13–46), i.e., if most direct witnesses report the same reading, it most likely represents that present in the original. It is not applicable in the following cases: (1) the *stemma* are bipartite; (2) all the readings from every descendant differ; (3) the *contaminatio* is present in the transmission. In one of these cases, the philologist may apply the *lectio difficilior*, according to which the more complex the reading, the more preferable it is. Also, the author's *usus scribendi* and *combination,* i.e., combining variants that are partially correct, help make philological choices.

On the other hand, if philologists dispose of a *codex unicus*, their judgement has a decisive role. The final stage of the editorial work is the *dispositio*. Once the text is established, it is accompanied by the *apparatus criticus* (Buzzoni 2020, 398–399). The *apparatus criticus* is the core of a scholarly critical edition as it transparently shows the editorial decisions (Fisher 2020, 412). Specifically, it provides the necessary information that certifies the philologist's work on the collated witnesses (André 1972, 14). It abridges the analysis of the primary sources conducted by the philologist. The comparison line by line of the witnesses or the analysis on a single copy led the philologist to decide whether to accept or reject the readings. Such analysis is complex, and, as a result, choosing what is relevant to include in an apparatus is not an easy task (Moureau 2015, 348).

1.3.4 *Apparatus* formatting and basic components

There is no standard regarding the structure of the *apparatus*. It depends on the information considered relevant by the philologist. Regarding its formatting, as Moureau (2015, 348) explains, the editor has to take into account four different options: (1) *language*, (2) *layout*, (3) *format of the data*, and (3) *syntax*.

1 **Language**. The language of the *apparatus* is a problem to address when starting to compile the *apparatus*. Specific abbreviations or symbols can replace a whole word, e.g., *add.* or + for *addidit*. The *conspectus siglorum* or the edition introduction usually lists the abbreviations and symbols.

2 **Layout**. The layout may be different in each case. Editors can choose to number the *apparatus* entries or to refer to their position in the paragraph or stanza line. Their position may be at the bottom of the page, at the end of the edition, or less often in the margins as in the mediaeval system of *marginalia*.

3 **Format of the data**. We can distinguish three diverse kinds of data: (a) readings, (b) witnesses, and (c) editorial information.

(a) **Readings**. Readings are usually introduced by referring to the chapter, verse, or internal text subdivision. Their form generally corresponds to the one found in the witnesses, i.e., without homogenising or formally polishing them. Two different kinds of readings may be available. The *lemma* corresponds to the adopted reading. It is at the beginning of each critical unit, i.e., *apparatus* entry, and its form is the same as present in the *textus*. The *variant readings* follow the *lemma*. The level of detail changes depending on whether the *apparatus* is positive or negative. The positive style records the variants of all the available witnesses, both accepted and rejected, allowing for an in-depth reflection on the history of the tradition. Instead, the negative format notes only the witnesses attesting to a variant reading. Thus, it does not mention the witnesses containing the same reading the editor chose for readability reasons. The choice between these two styles is up to the philologist (André 1972, 16–18).

(b) **Witnesses**. Specific *sigla* designate the witnesses, i.e., abbreviations, names, or symbols, attesting each reading.

(c) Editorial information. Editorial notes may add specific information about the readings.

A specific syntax regulates the composition of the *apparatus*. In general, the entry always contains a reference to the position of the reading in the text, e.g., the number of lines or paragraphs. At least in classical studies, the *lemma* is in the first place, and then the other readings. A colon separates them in a positive *apparatus* and a square bracket in a *negative* one. The readings, instead, are divided by commas or semicolons. Even with the various schools of practice, the definitive aim is to reach a certain level of readability, uniformity, and avoid ambiguities (André 1972, 14).
Finally, the order of witnesses should follow philological criteria, e.g., witnesses' families or other reasons, such as the proximity between the variant readings and the chosen reading (Moureau 2015, 351).

The following example identifies the main components of an *apparatus* entry from (Fig. 3).

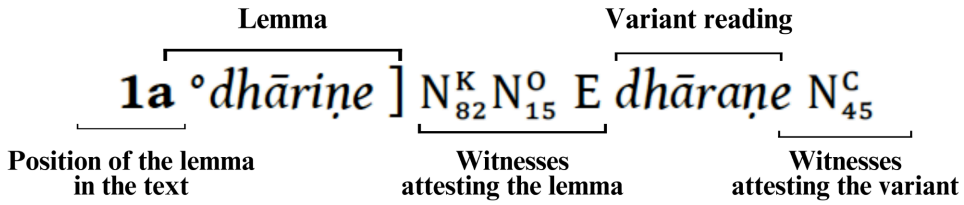


Fig. 3. Sample *apparatus entry*.[5]

1.3.5 Additional phenomena in the *apparatus*

The *apparatus* includes the representation of specific modifications to the text. André (1972, 27–30) lists such phenomena. In particular, it concerns (1) *additions* and *suppressions*, (2) *lacunae*, (3) *transpositions*, (4) *conjectures*, and (5) *omissions*.

1 ***Additions* and *suppressions***. The terms *addition* and *suppression* refer to a text added or deleted from a witness by the scribe or a person handling the manuscript before the edition. In suppressions, textual strings may be stroked through, erased, or scraped off (*ante ras*). In other cases, dots may be next to the characters to be deleted (*exp.* = *expunction*) (Griffith and Janiak 2023, 100).

[5] *Apparatus* entry from *Śivadharmottara*, a Sanskrit work currently edited by Florinda De Simini (Ch. 5).

2 ***Lacunae***. *Lacunae* are gaps of missing text in manuscripts. There are various reasons why manuscripts have missing sections. Since manuscripts are physical objects, loss and damage to their content are frequent. Material damages may be due to fire, fungi, insects, rodents, water, or bad binding in case of lost or misplaced quires (O'Sullivan 2020, 19).

3 ***Transpositions***. When a reading has been moved in the text relative to other manuscripts, we speak of *transposition*. Philologists refer to the term *inversion* when the transposition involves two contiguous words (Andrews 2020, 172).

4 ***Conjectures***. The term *conjecture* is strictly related to the concept of *emendatio* (Ch. 1.3.3). In passages where text is difficult to restore, the presence of conjectures is so frequent that it is almost impossible to list them all (André 1972, 29–30).

5 ***Omissions***. An *omission* is a fragment of text that a copyist does not reproduce in the copied text but is present in the exemplar. As in the case of additions, no assertion on whether an omission is secondary is implied (Conti 2020, 245).

1.3.6 Complementary apparatuses

Supplementary material can be fundamental for the *constitutio textus* and a better appreciation of the text, especially from the point of view of its composition technique and literary impact (Giannouli 2015, 16). While *fontes* and *parallela* are regarding the content of the *textus*, *testimonia* and *imitationes* refer to its impact on other writers.

Bidez and Drachmann (1938, 30) define *fontes* as a critical edition component complementary to the *apparatus criticus*. Specifically, they consider *fontes* (1) the sources of the *textus*, i.e., passages from earlier authors that the author of the *textus* re-adapted; (2) parallel passages, i.e., passages from other authors on the same subject referring to the same sources.

Revising the guidelines for the critical editions of the Association Guillaume Budé published by Louis Havet (1925) in 1925, Irigoin (1972) defines the *testimonia* as citations and quoted excerpts of the *textus* by other authors. He considers them a proper part of a critical edition since they constitute an indirect tradition. Finally, *imitationes* are passages from contemporary or later authors (Giannouli 2015, 23).

As suggested by Giannouli (2015, 23), it is recommended to dispose of these complementary materials separately from the *apparatus criticus* and each other.

1.3.7 Other tools to comprehend the *textus*: translation and notes

Recommending good practices in the field of editions, André (1972, 32–33) includes the *translation* and the *notes* in the traditional set of edition components.

Regarding the *translation*, he proposes to align the paragraph of the original text with those of the translation. In addition, it is a good practice to report the reference to the numbering of book, chapter, paragraph, and verse and follow the specific syntax as regards the additions, e.g., between square brackets.

The philologist often includes free-text notes to give readers a more comprehensive and detailed explanation of her choices or the *textus*. The notes may discuss philological, historical, or literary issues.

1.3.8 Limitations of *reconstructive editing*

The *reconstructive editing*, i.e., based on the reconstruction of a *stemma*, is only sometimes possible for several reasons attributable to the text variability. Following Haugen's discussion (Haugen 2020, 379–380), we list the textual features that affect the effectiveness of this kind of editing: (1) *textual dynamics*, (2) *linguistic diversity*, and (3) *ravages of time*.

Textual dynamics refers to texts not faithfully copied by the scribes but revised, added to, or erased according to their tastes or the audience. In this case, the number of variants could be too high to be recorded in an *apparatus* of variants.

At the same time, *linguistic diversity* could result when texts were copied over time and space, thus acquiring different orthographies and linguistic forms from the original.

Finally, the *ravages of time* could hinder *reconstructive editing* in case of irrevocably lost, fragmented, or too complex manuscript families.

1.3.9 Beyond Lachmann's method: Gaston Paris and Joseph Bédier

In the 19th century, the German Karl Lachmann (1793–1851) and the French Gaston Paris' schools were considered the foundation for the genealogical method (Palumbo 2020, 88–109) (Ch. 1.4, 1.5).

Gaston Paris (1839–1903) sets out Lachmann's method in the introduction of the *Vie de Saint Alexis* edition in 1872 (Paris and Pannier 1872, 7–15), but at the same time, he introduces specific corrections to that method (Formisano 2005, 5–22).

In the following years, scholars gradually discussed the predominance of the Lachmannian school. The first to criticise the method of common errors was Joseph Bédier, who considered it ineffective for various reasons. First of all, the reconstruction of a *stemma* is never certain, as it is never the only one possible. In addition, editors could find situations where the *stemma* is bipartite, thus forcing them to choose between different variants only by relying on subjective criteria, e.g., taste or intuition. As a result, the editions produced are composite and highly arbitrary. The solution proposed by Bédier was to base the critical edition on the best manuscript, i.e., *bon manuscrit*, possibly providing an *apparatus* of the variants from other witnesses.

Thus, the scope of editions shifted from reconstructing the authorial version to publishing one or more scribal versions.

Despite the intense criticism surrounding Lachmann's method, it was never abandoned. Instead, it was applied more consciously, considering its fragilities and the great weight of philologists' judgement (Palumbo 2020, 88–109).

1.3.10 Problems uncovered by Bédier:[6] Giorgio Pasquali and the Neo-Lachmannism

According to Trovato's discourse (2020, 109–138), improvements to Bédier's method were made from 1929 to the present, and they have been demonstrated to be effective in cases of very complicated textual tradition. With the term *Neo-Lachmannism* coined by Contini, we refer to the scholars contributing to the refinement of Lachmann's method triggered by Bédier. We discuss below some of the corrections by Giorgio Pasquali (1885–1952).

First of all, Pasquali distinguished the *scribal* from the *authorial* variants. Thanks to this innovation, the task of the philologist dealing with multiple authorial versions is no longer to choose the variant that best represents the work but the one that represents the author's intention in that precise stage of the work's evolution (Ch. 1.4).

Another fundamental practice introduced by Pasquali (1934, 126) is the *open recensio*, according to which, where it is impossible to establish the reading of the archetype mechanically, i.e., no reading represents the majority of primary branches, the philologist should use *iudicium*, i.e., choosing based on internal criteria (Ch. 1.3.3). He also adds that scholars should keep recent witnesses, i.e., *recentiores non deteriores*. Frank (1976, 135) extends Pasquali's discourse by explaining that, in peculiar cases, the transmission conditions may differ from those described by Lachmman. Specifically, he points out that (1) multiple originals may exist (real, virtual, or possible); (2) variations and contaminations may originate from oral transmission; (3) copyists may use several

[6] According to (Roelli 2020, 3).

divergent sources, and this may cause contamination; (4) previous editions employing conjectures may influence the transmission.

Trovato (2020, 117) continues reporting Pasquali's thought that other factors that affect the transmission are errors, or rather, the genesis of errors. Depending on their genesis, errors made by the copyists could be *monogenetic*, i.e., the variants could have been produced independently of one another, or, vice versa, *polygenetic*.

Finally, Pasquali assigns a fundamental role importance in the reconstruction of the *stemma*, especially in an overabundance of resources, to the geographical area where the manuscripts were written. The so-called *criterion of peripheral areas* claims that if the readings of manuscripts written in peripheral areas distant from each other and the centre of culture coincide, they are probably reliable. As Pasquali (1952, xvii–xviii) explains concerning ancient and mediaeval texts, a Vulgate began to take shape over time. It progressed from the centre to the suburb without necessarily reaching it, as it was a fashion.

1.4 Authorial philology

The authorial philology concerns the edition of the *original*, specifically (1) the *autograph*, i.e., the manuscript written directly by the author; or (2) the *idiograph*, i.e., the manuscript written by a copyist under the author's supervision; and (3) prints edited by the author. Authorial philology may be helpful when dealing with editions of *in fieri*, i.e., an autograph or idiograph presenting traces of a reworking process, and *in multiple versions*, i.e., the work is preserved in more than one authorial manuscript or printed texts (Italia 2021, 29–69). Although it differs from genetic philology, it is not its opposite but a complementary part. The notion of *variant* is central to both. While in the former, variants are considered *deviations* that must be corrected, in the latter, they are indicated by the term *réécritures*, meaning *rewriting*, as part of an invention (Van Hulle 2016, 35), where the text is approached as a process rather than a fixed and finished entity (Katajamäki and Pulkkinen 2023, 7).

In the 1980s, Dante Isella, working on the complete Scholarly Edition of the *Opere* by Carlo Emilio Gadda, preempted by his mentor Contini, gave birth to a new method of representing the variants. He defined the peculiarities of authorial philology through a diachronic and systemic *apparatus*. In the context of authorial philology, the author's will may be investigated through her manuscripts and the corrections made by the author. The author's will to archive and preserve her notes takes on a brand new role. The main peculiarity of authorial philology is the focus on the author's drafts, i.e., the creative process that led to the definitive version of a work (Italia 2023, 324–50). Another peculiarity is binomial philology/criticism. Contini (1974, 232–241) explains the value of such a study. Studying a work from a *dynamic* perspective through the author's corrections means getting

closer to her point of view (Contini and Di Meana 1989, 161). Therefore, the variant has no longer only an *instrumental* value for reconstructing a text but a value in itself for profoundly understanding an author's creative process over time. Hence, authorial philology does not undermine the necessity to reconstruct a text but allows the reader to know incomplete texts not corresponding to the author's last will (Italia 2023, 324–50).

Chapter 2

Traditional philology, Digital philology, and Digital Humanities: in between

In the preceding chapter (Ch. 1), we embarked on a quest for the authors' final truth that may have never existed. Various complex paths were traversed, yet never in parallel. The final objective of the philologist has always been to reconstruct the authors' final intentions and deliver them to the reader in an authoritative form (Ch. 2.1). All the roads crossing each other were directed towards this common destination, and philologists have debated over the best road for centuries. However, what has changed in modern times? How has the digital age impacted this field (Ch. 2.2)? Throughout this chapter, we delve into the evolution of philology in the digital age, examining the profound impact of digital technology on the field, particularly in preparing Digital Scholarly Editions. The chapter highlights a shift from *traditional philology* to *digital philology*, which shares the traditional philology missions but also leverages digital tools to present texts with interactive features and broader audience (Ch. 2.3). Furthermore, we briefly shed light on the challenges and complexities of producing Digital Scholarly Editions, focusing on the need for collaboration between philologists and digital humanists (Ch. 2.4, 2.5). Then, we discuss the trend of reproducing primary sources online — and not *editions* — and the technological and technical barriers that hinder the production of advanced digital editions (Ch. 2.6). In conclusion, a theoretically plausible answer to the issues identified throughout this chapter is proposed. It suggests that philologists strive to reduce *ad unum* the text, while digital humanists should explore technological solutions in close collaboration with philologists to counter this prevailing trend.

2.1 Centuries spent searching for the author's intentions: a recap

For centuries, the field of philology, particularly ecdotics, has long debated the most effective method for uncovering textual truth. It refers to the authors' intended will for their work and delivering it to their audience. One of the most prominent figures in this field was Lachmann (Ch. 1.3.2), whose method underwent refinement, debate, rejection, and revival over time (Ch. 1.3.9; 1.3.10). Regardless of the approach, philologists aimed to produce a scientifically valid and authoritative text close to the original work. The end goal was to create a printed edition that was considered to be an inviolable and indisputable entity. This black-and-white text was primarily accessible to a select few.

2.2 Where is philology now?

In the current era, we find ourselves in a transitional phase where the digital ecosystem has become a significant player. This middle ground represents a merging of a paradigm blend, where digital media has yet to wholly replace analogue media in producing, using, and transmitting textual sources (Italia 2020, 7). According to Italia (2020, 56–61), in light of the ever-expanding digital landscape, some have speculated that the field of philology may have met its demise. However, not all share this perspective, as many view digital technology, first of all, as a means to overcome the physical limitations of paper. Over the centuries, philologists have dedicated themselves to concisely presenting their *apparatus*, often resorting to increasingly abbreviated forms to save space. However, with the vast digital expanse at their disposal, they started to explore an abundance of resources, reproducing all available witnesses and placing greater emphasis on the digital reproduction of the *textus certus*, rather than the ideal reconstruction of an uncertain archetype (Ch. 2.4). The matter of space is not the sole factor that validates the shift from analogue to digital. The digital realm offers the opportunity to establish a connection between different constituents of an edition, such as *textus* and *apparatus*, which cannot be accomplished in the analogue environment, where the *apparatus* is frequently relegated to the concluding section of the book, in the appendix, or at the bottom of each page. Additionally, diverse typographical markers can be utilised to indicate the varied components and attributes of the text.

So, if in the past traditional philologists held a degree of scepticism towards the web, viewing it solely as a more *user-friendly* means of reconstructing and editing texts, we have since progressed to a point where the web is no longer merely perceived as a *tool*. Instead, it forces us to *rethink* philology (Ch. 2.3), also to provide users-readers with space and tools to develop their own hypotheses and ways of reading texts (Fisher 2020, 407). Integrating digital methodologies in philology necessitates a more rigorous assessment of the edition's characteristics to be represented. The outcome is particularly effective when one captures and portrays these characteristics in a dynamic and interactive perspective (Ch. 2.3.1). As we shall observe subsequently (Ch. 2.3.4, 2.6), the technical proficiencies in modelling and implementing editions have advanced to such a degree that they necessarily require cooperation between philology and digital humanities.

2.3 Digital philology exists and it is multifaceted as its outcomes

According to Tomasin (2019, 20–22), philologists' responsibility entails preserving and curating texts[7] in a physical or virtual repository. It involves transmitting such texts and imparting the

[7] Actually, Tomasin does not say *texts*, but *discourses*, referring to the work *La linguistica romanza* by Heinrich Lausberg (1971).

necessary knowledge to interpret them accurately. Over time, their literal meaning and historical context may become obscured or difficult to comprehend, hence the need for proper interpretation. The initial step in any philological inquiry is, without a doubt, the *text* itself. Philology has a central role in the preservation (M-1), comprehension (M-2), and perpetuation (M-3) of textual sources. M-1, M-2, and M-3 summarise the pillars of the *mission* of philology. The text also determines the best philological *method*(s) for dealing with it. The *competencies* necessary *a priori* to deal with the text are (C-1) historical competence and (C-2) linguistic competence.

For the sake of organisation, let us proceed with a summary of the fundamental principles of *traditional philology* identified by Tomasin (2019, 20–22) in Table 1.

Traditional philology	
Starting point	(S-1) Text
Mission	(M-1) preservation
	(M-2) comprehension
	(M-3) perpetuation of texts
Method(s)	(MT-1) Not definable *a priori*, but determined by the text.
Necessary competence	(C-1) historical competence
	(C-2) linguistic competence

Table 1. Fundamental principles of *traditional philology* identified by Tomasin (2019, 20–22).

Thus, according to Tomasin's argument, the existence of *digital philology* is questionable as its primary objective of *rendering a text* contradicts the previously identified missions M-1, M-2, and M-3. A systematic refutation of his thesis will be presented below.

2.3.1 Digital philology missions

Contrary to what Tomasin claims, *digital philology* is primarily concerned with the preservation, i.e., M-1, comprehension, i.e., M-2, and perpetuation, i.e., M-3, of texts. The thesis above has been entirely confirmed by Italia (2020, 13):

> In un sistema culturale che consideri i testi come beni comuni, parte fondativa del nostro patrimonio culturale, proteggerli, curarli – nel senso di "prendersene cura" –, garantirne la qualità e favorirne la diffusione, diventa un vero e proprio compito sociale, prima ancora che culturale, compito che ogni studioso dovrebbe prefiggersi per non fare sopravvivere solo l'*editing* creativo, quello di cui beneficiano (e ancora molto) i bestseller impilati nelle librerie, ma anche l'*editing* meno creativo, quello che si prende cura della "verità testuale" dei nostri classici. Grandi e piccoli.[8]

The fact that the texts are digital makes these missions even more effective since their digital support makes them accessible to a broader audience beyond the few who originally intended to read them (Pierazzo 2019, 210; Italia 2020, 11).

So why create a digital edition? The reasons are manifold and correspond to the same reasons that drive *traditional philology*: to ensure that the text will not be lost, to offer an interpretation that takes into account its historical context and original meaning that may be subject to the passage of time, and to pass it on to future generations like a relay race baton, no longer restricted to professional runners.

The production of the final output of an edition, which is the *rendering of the text*, or better, to *discover and enhance the multiple facets and levels of a text on the user side*, is an objective that aligns with the ones mentioned above. As Pierazzo (2020, 74) claims, digital texts are not merely presented in accordance with their contents, but also incorporate interactive elements. The interface serves as a crucial intermediary between the data pertaining to the editions, the software to manipulate them, and the users who will ultimately read the content. In this regard, unlike printed editions, digital editions offer tools that can effectively represent and allow taking full advantage of the richness of the text and its various intra and intertextual relationships. For instance, one can link an *apparatus* entry to the *lemma* in the *textus* by clicking on a specific text string, which will immediately retrieve the corresponding entry without the need to search at the bottom of the page or even the book. Similarly, one can associate textual strings to their diachronic variants, accessible just by clicking to understand how a text changed over time.

These seemingly trivial operations significantly transform the readers' use of the text and enhance their experience. These tools' creation strongly depends on the philologist's work and their decisions on what is relevant to make explicit (Ch. 3.2, 3.3, 3.4). Essentially, it is the philologist's

[8] «In a cultural system that considers texts as common goods, a founding part of our cultural heritage, protecting them, caring for them – in the sense of "taking care" of them –, guaranteeing their quality and promoting their diffusion, becomes a real social task, even before than cultural, a task that every scholar should set for herself in order not only to make creative editing survive, the one from which the bestsellers stacked in bookstores benefit (and still a lot), but also the less creative editing, the one that takes care of the "textual truth" of our classics. Big and small».

responsibility to make the connection between the *lemma* and *textus* explicit and determine where that *apparatus* entry must be linked to the text. They also have to declare the variants behind textual strings in the *textus*. In this regard, *digital philology* deviates from the *traditional* one. The philologist has an additional objective: to identify and make explicit those connections visible that only experts can see. It follows that the traditional objectives are supplemented by a constant search for a good level of expressiveness of the various facets of the text that, in a second step, can be reached through an interface by the reader-user. This forces the philologists to *rethink* their work: not only to systematically organise the information related to the text, but also to envision which (and how) elements are noteworthy and hold scientific value for visualisation. Of course, flanked by a digital humanist to overcome technological and technical issues (Ch. 2.6).

2.3.2 Digital philology starting point

The *text* (S-1) is the starting point of the philological investigation also in a digital environment, as it is always the subject or object of all the missions (Ch. 2.3.1). Furthermore, it is necessary for philologists to work first of all on the text and, above all, to have a deep knowledge of the main features of the text before designing tools that allow exploring it. For instance, consider designing a tool that allows readers to trace the source of each citation in a text. Such a tool would only make sense in the case of a text with citations. So, first of all, the citations should be identified within the text. Secondly, the sources should be explicit, e.g., the cited author's name, work title, year, and edition. It is a philologist's task to make explicit the information necessary to make the connection possible, and it is their knowledge of the text that leads to the design of such a tool.

2.3.3 Digital philology methods

According to the fundamental principles of *traditional philology*, as outlined in Table 1, it is asserted that the methods for treating a text cannot be predetermined, as they are contingent upon the qualities of the text in question (MT-1). Undoubtedly, this principle is equally relevant in the digital context, where the philologist's work initially aligns with traditional analogical workflow (Ch. 1), as will be further explored by an example in Ch. 5.11. While integrating any text into a digital environment, and so, while defining the relevant information to be displayed and how to organise, systematically model, arrange it in the mirror of the digital page, and make it accessible and navigable through an interface (Ch. 2.3.4), it is crucial to work on recovering the text as described in Ch. 1. Following the *traditional philology* starting point (2.3.2), it is necessary to

commence the text recovering process from the manuscript(s)[9] and subsequently proceed with a series of deliberate choices[10] by the philologist, culminating in the attainment of a comprehensive version of the text. For instance, when dealing with a *critical edition* (Ch. 1.2), the philologist must make crucial decisions among the variants they intend to put into the *textus*. The question that the philologist is called upon to answer is, first of all, Which version of a specific work to restore is the closest to the original the author intended to convey? Either by mechanically demonstrable reasons or by *divinatio*, they are required to provide an answer, i.e., a *textus*. Hence, while venturing into the digital, the philologist needs to surmount all the obstacles of *traditional philology*, choose a precise direction at every crossroad, reconstruct the entire journey, and consider the digital aspect. However, as we shall observe later (Ch. 2.4), this occurs only sometimes.

To provide a concise overview of the methodologies associated with Digital Scholarly Editions, we reference those outlined by Pierazzo (2020, 47–71). This topic will subsequently be thoroughly examined and discussed in Ch. 3. As noted by Pierazzo (2020, 47), it is evident that digital editions lack a standard approach in terms of presenting their contents and making them accessible through user-friendly tools. Nevertheless, a specific standard concerning the methodologies for preparing Scholarly Digital Editions has been established, which is known as the *Text Encoding Initiative* (TEI).[11] The reference conceptual model is the *source-output* model. The source corresponds to a collection of one or more files that contain an appropriately annotated text that then will be transformed into an output, i.e., visualisation of the edition (Ch. 3.4). These annotations may pertain to the structure of the text, e.g., chapters, stanzas, and paragraphs, or supplementary constituents, e.g., *apparatus* and its subcomponents. Typically, the source material is encoded in XML format in adherence to the TEI guidelines (Ch. 3.3.1). Before commencing the annotation process, it is necessary to devise an *editorial model* (Ch. 3.2). Essentially, the philologist must determine the pertinent elements of the text that are crucial to visualise an edition. In other words, which aspects of a text require emphasis? In the case of a critical edition, for instance, the *apparatus* is of utmost significance since it summarises the reconstruction process of a text. This phase is pivotal as the visualisation of the final edition hinges on the editorial model (Ch. 3.4.2).

The digital methodologies employed in Scholarly Editing also include digitising the text into a sequence of characters or facsimile reproductions in digital format (Ch. 3.1).

---

[9] We are specifically discussing *manuscripts*, as we are not alluding to *digitally produced texts*.

[10] The selection of different options to prepare an edition (Ch. 1) is contingent upon the philologist's objectives. These options may differ depending on the accessibility of copies of the work in question (Ch. 1.1), as well as the nature of the edition being produced (Ch. 1.2).

[11] TEI: https://tei-c.org/.

2.3.4 Necessary competencies in the digital philology field

Here we come to a burning issue: the skills required of a *digital* philologist. For a level playing field and to be fair, in the end, we will reverse the situation by analysing which skills are required of a digital humanist working in Digital Scholarly Editing (Ch. 2.3.5). Because the exchange here is mutual, even if the subject has only been treated subtly for now.

Thus far, we have observed how the editorial process envisioned by *digital philology* aligns with that of *traditional philology*, at least in its initial stages (Ch. 2.3.2, 2.3.3). These fields, (once again) share common ground regarding the necessary competencies for handling an edition of a text, be it analogue or digital. In order to accomplish the tasks outlined in M-1, M-2, and M-3, the historical and linguistic abilities identified by Tomasin (2019, 22) are essential. The importance of the philologist's historical skills is inherent in Sahle's (2016) definition of edition (Ch. 1.1), where he defines the concept of *edition* as the critical representation of *historical* documents. Subsequently, he proceeds to furnish a comprehensive elucidation of the *historical* notion:

> Editions [...] explain what is not evident to the present-day reader. In short, they bridge a distance in time, a historical difference. Texts that are created today do not need to be critically edited. They can speak for themselves. Only historic documents and texts need an editor to make them speak clearly.

It is that *historical difference* that the philologist is supposed to grasp. History also influences the language in which the philologist is specialised, as Tomasin (2019, 22) explains.

Switching to *digital*, the digital domain poses a complex set of challenges that require a high degree of technical proficiency from philologists. From edition designing, involving the creation of an accurate edition model, to development, philologists need to be adept at various technical aspects. Encoding, for instance, plays a crucial role in creating digital editions, whether using TEI or other vocabularies or methodologies (Ch. 3.3). However, this often distracts philologists from focusing on the preparatory steps required for an edition (Ch. 1). These issues will be addressed in Ch. 4. For now, it is essential to note that traditional philological skills must dovetail with (too advanced) technical competencies in Digital Scholarly Editing.

2.3.5 Competencies required of a digital humanist in Digital Scholarly Editing field: a mutual exchange

In order to establish a clear understanding of the role of a digital humanist, it is pertinent to reference the remarks by Tomasi, which aptly describe the work of experts in the field of Digital Humanities (DH). Tomasi (2022, 7), in her discourse on the Semantic Web, expounds upon this topic in the following manner:

> Il Web semantico ha saputo fornire una nuova prospettiva di approccio al sapere [...] perché capace di tradurre ciò che ha qualificato, fin dalle origini, il lavoro degli esperti di DH: riflettere sui sistemi computazionali per acquisire (nuova) conoscenza da dati e documenti. Esplorare prima di tutto una testualità, esito della sedimentazione degli oggetti analogici della tradizione culturale, per lanciare una sfida ermeneutica alla macchina.[12]

The role of a digital humanist is to analytically and critically reflect on the most effective methodologies for manipulating information by its unique characteristics and expressing the hermeneutics of that information. It is important to note that this process is not simply a technical or mechanical exercise but rather a deliberate effort to design and implement systems that enable the organisation of knowledge through the lens of observation or, more specifically, the interpretation of a humanist perspective on data. The digital humanist seeks to create meaningful frameworks for information grounded in a deep understanding of its underlying context and significance. Nevertheless, the interpretation alluded to in this particular scenario is exclusively formal. The digital humanist endeavours not only to construct a precise model of a particular domain but also to ensure that this conceptualisation is both understandable and capable of being processed by a machine, thereby rendering it computable. The scope of this endeavour extends beyond the mere conceptualisation of a domain, encompassing the entire life cycle of data, from their initial production to their manipulation, preservation, dissemination, and discovery in relation to the end user.

In philology, the *datum* concept carries a multidimensional connotation when dealing with *text*. *Text* is a complex entity that conveys implicit semantics through strings of characters. As a result, the data associated with text can no longer be restricted to a single, well-defined structure and is instead

[12] «The Semantic Web has been able to provide a new perspective of approach to knowledge [...] because it is capable of translating what has qualified, from the outset, the work of DH experts: reflecting on computational systems to acquire (new) knowledge from data and documents. First of all, to explore textuality, the result of the sedimentation of the analogical objects of the cultural tradition, to launch a hermeneutical challenge to the machine.»

subject to logic that is not linked to predefined schemes. This is why character strings are referred to as *unstructured data*. An additional step is required in the cycle presented earlier to move from *unstructured* to *structured* data. In addition, formalising the text to make it *computable* is complex also because it involves various aspects, which the digital humanities and philologists work together to address. Of particular importance is the history of the tradition of the text, as well as the interpretive acts that are usually summarised in the *apparatus* (Tomasi 2022, 43–74).
However, it holds that the selection of a model is contingent upon the nature of the subject data. In that case, the digital humanist must possess knowledge of the corresponding domain, precisely one associated with textuality. In Digital Scholarly Editing, it becomes essential for them to comprehend the issues of philology to facilitate philologists with their contribution towards the modelling and formalisation of their research object, which is sometimes very hard for a non-expert.

2.3.6 Traditional vs. Digital philology: a final comparison of their fundamental principles

The preceding reflections demonstrate that *digital philology* is a legitimate field of study. Moreover, it shares many fundamental tenets, methodological approaches, and core competencies with *traditional philology*. Indeed, the principles that underpin both disciplines are mainly congruent, with only minor deviations. However, what is required in a digital environment is broadening the philologist's methods and skill set to accommodate the challenges and opportunities presented by this new context and to respond effectively to the evolving demands of the discipline's research and scholarly missions. This conclusion, summarised in Table 2, coincides with Rosselli Del Turco's reflection (Rosselli Del Turco 2016, 227) that producing a digital edition is inherently more resource-intensive than its traditional counterpart. The traditional workflow is decidedly more straightforward, as the philologist's labour is complete upon delivery of the work to the publisher, who handles all aspects of publication and distribution. Conversely, the digital workflow presents significant challenges, as the philologist must navigate a largely undefined process, compounded by the lack of established standards to guide their efforts (Ch. 3).

	Traditional philology	***Digital philology***
Starting point	(S-1) Text	
Mission	(M-1) preservation	
	(M-2) comprehension	
	(M-3) perpetuation of texts	(M-3) perpetuation of texts

		transmitted not only to specialists, but to general public
		(M-4) rendering of the text, or better, express the multiple facets and levels of a text on the user-side
Method(s)	(MT-1) Not definable *a priori*, but determined by the text.	
		(MT-2) Digitising the text
		(MT-3) Modelling the edition
		(MT-4) Encoding the edition
Necessary competence	(C-1) historical competence	
	(C-2) linguistic competence	
		(C-3) modelling competence
		(C-4) technical competence
		(C-5) designing competence

Table 2. Fundamental principles of *traditional philology* identified by Tomasin (2019, 20–22) and of *digital philology*.

To put a seal on this discussion, the field of philology has undergone a significant transformation of its paradigm in the digital age. Undeniably, the obligations demanded of a digital philologist undertaking a digital edition project are more quantitatively outstanding. While it is undoubtedly true that the philologist now has access to a more expansive workspace than that of costly paper, they are also much more occupied in their digital laboratory, where they must meticulously model and formally articulate each aspect of their intellectual labour, and play more imaginative with the design of their editions, flanked by a digital humanist, in order to effectively convey it to their audience effectively.

2.4 Scarcity of Digital Scholarly Editions and philologists like Bédier in the digital age

Two decades ago, Robinson (2003) asked (rhetorically) whether a Digital Scholarly[13] Edition should solely feature archived images or a collated text, emphasising the editor's interpretation. A decade later, when addressing the subject of digital images of manuscripts and books accessible

[13] To be exact, Robinson says *electronic* edition.

online, he posits that this availability of millions of images lacks revolutionary implications concerning textual scholarship (Robinson 2016, 182–183):

> [...] providing access changes nothing, of itself. If people actually use that access to make new editions, new scholarship, of a kind never seen before, which readers may use in ways never known before, then that would indeed be a revolution. [...] One presumes in editions at least a minimum level of scholarly intervention, in selection of material, in the provision of transcriptions, annotations and commentary.

According to Robinson, the distinguishing factor of an edition lies in its *interpretation* – specifically, the editor's attempt to comprehend how diverse primary documents are interconnected. The mere production of a text does not suffice. Therefore, sharing digital images of manuscripts online falls short of the mark. Instead, it is necessary to trace back to the works' creation, detailing their conception and reception to achieve a Digital *Scholarly* Edition.

Schmidt (2018) validates the inclination of digital philologists to refrain from finalising a text and emphasises that the archival phase takes precedence over the editorial phase, rendering it discretionary. Zaccarello (2017, 118) confirms the same attitude of digital philology to accurately represent the historical materiality of testimonies rather than proposing their reconstructive hypothesis. Monella (2018), instead, emphasises the absence of Scholarly Digital Editions of classical[14] texts with a manuscript-based multi-testimonial tradition.[15]

It is worth wondering whether this trend has shifted in a different direction during the present days.

[14] Monella (2018, 142) refers specifically to Greek and Latin texts belonging to the so-called canon of Classical literatures.

[15] According to Monella (2018, 147), this can be attributed to the so-called *canonisation* of classical texts, i.e., the process of transforming literary texts into a *Canon*. On the one hand, it implies a tendency toward standardisation, as it is difficult for authorial textual variance to survive for millennia. Furthermore, the reverence that scribes and philologists felt towards such texts discourages the introduction of variations, often unintentionally introduced due to errors or distractions. For this reason, textual variance is not considered significant at a cultural level but merely a *tool* to reconstruct a *good* text, i.e., one as close as possible to the oldest development of that text. Variants are relegated to the *apparatus* solely to demonstrate the process that led to such construction. Therefore, since digital editions are appealing precisely for their ability to represent textual variance, they are not a necessity for classicists.

Henceforth, we refer to the analysis by Italia (2020, 41–55) as a means to delineate the current state of this trend.[16] Italia speaks precisely of *proliferation of documents*,[17] attesting to a bédierian attitude on the net:

> La prima conseguenza del prevalere del documento nelle edizioni digitali è costituita dal fatto che il filologo, che prima ricostruiva/sceglieva un'edizione, ora è sempre meno spinto a ricostruire/scegliere, preferendo la presentazione di *tutti i testimoni* piuttosto che una loro selezione, e preferendo la fisicità del documento, in parallelo con la sua edizione diplomatica, piuttosto che l'astrattezza di un testo ricostruito, e, dal punto di vista materiale, storicamente mai esistito. Una posizione sempre più béderiana e sempre meno lachmanniana.[18]

Italia explains why digital philologists *do not choose*, i.e., do not reconstruct a text a few lines later: they can publish *all* the editions (Italia 2020, 53). The digital medium offers unlimited space, freeing the philologist from the constraints of summarising the variants of each witness in a limited *apparatus* confined to a corner of the book. The philologist can now make available a comprehensive range of textual features and information, often retaining the freedom to select *nothing*. This approach has faced significant criticism from Italia (2020, 56), who argues that reproducing all the available sources fails to fulfil the requirements of philology in restoring a text. However, the availability of numerous technological tools for producing a scientifically accurate and rich Digital Scholarly Edition (Ch. 2.5), and not only mechanical or documentary editions (Ch. 1.2), raises a question — Why do we have more online *facsimiles* and *transcriptions* than *editions* if the technology allows us to create *any* content or tool to browse the contents? It is necessary to understand why philologists have yet to leverage this abundance of technological options.

---

[16] Robinson and Schmidt analyse the trend of document availability on the web, as opposed to editions, from a digital humanist perspective, thus referring to purely digital themes, such as making tools available for accessing and browsing editions. In contrast, Italia approaches the subject predominantly from a philological lens, and explains why the tendency to publish only documents, and not editions, is a problem on a philological level.

[17] «proliferazione dei documenti» (Italia 2020, 51).

[18] «The first consequence of the prevalence of documents in digital editions is that the philologist, who previously reconstructed/chose an edition, is now less and less likely to reconstruct/choose, preferring the presentation of *all the witnesses* rather than a selection of them, and the physicality of the document, in parallel to its diplomatic edition, rather than the abstractness of a reconstructed text, and, from a material point of view, historically never existed» (Italia 2020, 52).

2.5 Blame for the scarcity of Digital Scholarly Editions and philologists like Bédier in the digital age

According to Italia (2020, 56), we are witnesses of a resurgence of *bédierism*, which refers to the philologists' inclination to present the reproduction of primary sources rather than a *textus*. Fisher (2020, 406) confirms the tendency to no longer consider a single established critical text as a necessary final product. The underlying reasons can be attributed to a series of consequences of introducing the digital medium in Digital Scholarly Editing, which push but hinder the development of both philologically and technologically complex editions. While technology offers philologists ample space to publish and a wide range of tools to enrich their editions with various components, e.g., *apparatus*, parallels, etc., and tools, it also creates an issue of *horror vacui*, leading philologists to the publication of *any* text,[19] witness, facsimile, or transcription (Ch. 2.5.1). Moreover, it presents technical difficulties as there is a lack of user-friendly tools for preparing editions at present (Ch. 2.5.2). The philologists' current solution lies in learning and leveraging the writing of code that encodes the characteristics of the text, along with the support of IT staff (Rosselli Del Turco 2016, 227) to make their results public and maintain them over time.

It is easy to see that the DH community must bridge these gaps to create sophisticated editions beyond simply displaying documents (Ch. 2.6).

In the subsequent sections, we shall delve into a comprehensive review of the issues above.

2.5.1 Boundless sense of digital space

The trend that we call *bédierism* is attributed to digital philologists by Italia (2020, 56). She explains that they have traditionally had to employ ingenuity to summarise and conserve space in their apparatuses. They are now liberated from these constraints due to the virtually limitless space afforded by digital media. As a result — as the worst consequence — an *apparatus* seems no longer necessary as it is now possible to reproduce all primary sources. Consequently, the predilection online is for the *primary source* over the *editions*.

This approach can be seen as a response to the philologists' realisation that they have an unprecedented digital space. Still, there could be other factors driving this trend.

[19] See, for instance, (Italia, 2016; 2020), and (Robinson 2016).

2.5.2 Lack of user-friendly tools to create Digital Scholarly Editions

The tendency to publish *facsimiles* instead of *editions* is not solely attributable to the philologists' sense of freedom. On the contrary, it is more likely a technological issue that hinders the production of Digital Scholarly Editions. Rosselli Del Turco (2016, 226) observes that the lack of *easy-to-use* tools for creating Digital Scholarly Editions is accountable for their scarcity on the web. Pierazzo (2019, 210) confirms this lack. Monella (2018, 144) also points out the inadequate digital literacy among traditional philologists.

Let us start from a basic level. It is evident that publishing a facsimile online, such as a picture or a scan, is significantly *easier* than developing a web application that offers an interactive *textus* connected to its *apparatus*, aligned with its translation, and traceable to all parallel sources with requisite metadata. The technical process for achieving these two operations' outcomes is strikingly different and complex. The final result depends not solely on philological efforts but also on technical expertise (Pierazzo 2019, 2012).[20] Even so, scanning and transcribing manuscripts can be accomplished by a philologist working *independently*. Developing a highly interactive application requires advanced technical skills beyond the scope of philologists' expertise. It is not feasible to expect a philologist to acquire these technical skills and become a developer. This is why the need for *user-friendly* tools to create Digital Scholarly Editions is compelling.

Some methodologies, such as TEI, are relatively easy to learn and use, as will be discussed later (Ch. 3), but they are not immune to objective difficulties in their application. As Italia (2020, 68–69) and Pierazzo (Pierazzo 2019, 2015) point out, such methodologies are *too* normative and flexible, presenting challenges in their implementation. And do not forget that a TEI file is primarily an XML file, and XML is not intended to display data.[21]

2.5.3 Missing publisher

In print publishing, the role of the *publisher* was pivotal. The responsibilities encompassed by this position were broad in scope, ranging from overseeing the processing and management of the

[20] Pierazzo (2019, 212) identifies the acquisition of technical knowledge as a crucial requirement: XML, TEI, XSLT, HTML, CSS, web design. In actuality, these could undergo a complete transformation if the intention is to pursue a *graph* Scholarly Digital Edition (Ch. 3.4.1).

[21] See, for instance, (Rosselli Del Turco 2016, 233): «visualisation tools: one could maintain that, if you correctly encode your text, the edition is already there and only needs to be 'extracted' from the XML document base, but of course this phase is at least as delicate and complex as the previous one [...] This is also an area where tool development is particularly intense».

edition data to ensuring its final presentation met the desired standards. However, in the case of digital editions, this role remains unfulfilled, as the uncontrollable nature of the web and the possibility of self-publishing renders it increasingly common (Rosselli Del Turco 2016, 228). The absence of a publisher in Digital Scholarly Editing (Pierazzo 2019, 2012) connotes a missing set of editorial protocols. Consequently, the philologist is not obligated to publish a *textus* but may be free to opt for alternative approaches.

2.5.4 Lack of models

Another reason that would justify the scarce presence of scholarly editions on the web is that no accepted and validated publishing model for Digital Scholarly Editions exists for now. It presents a significant challenge for digital philologists. As Pierazzo (2019, 210) explains, each edition follows its model, presenting the text differently. Some presentations lead back to traditional types of editions (Ch. 1.2), while in other cases, the models of textuality are unprecedented, and consequently, the possible interactions disorientate readers. A standard model exists, namely TEI, but the related issues (Ch. 3.3.1) again spark debate.

Thus, the absence of a model places a heavy burden on the philologist: Which tool to choose to visualise and browse the edition, e.g., facets, buttons, or more complex tools? And so, what data modelling is required for such a tool to work? It is a strictly technical topic and inextricably linked to modelling and web development techniques since the visualisation strictly depends on these two domains (Ch. 3.4). For this reason, Buzzoni (2016, 60) stresses the importance of establishing a *protocol*, which serves as a guiding framework for structuring the critical components of Digital Scholarly Editions.

2.6 A reasonable solution in Digital Scholarly Editing: *reductio ad unum* and multiple forms with the DH help

Digital Scholarly Editing is marked by two conflicting forces, as observed in the previous analysis in Ch. 2.4 and 2.5. Firstly, technology provides a wide range of sophisticated tools to perform complex operations on texts. These tools enrich digital editions from essential search functions to complex widgets, such as automated alignments between text, *apparatus*, parallels, and other components, as well as facets and data visualisations. However, on the other hand, editions are often documentary or even facsimile, lacking scholarly interpretation by the philologist. Consequently, the adjective *scholarly* accompanying editions is often missing, and the text is not actually reconstructed to its original form (Ch. 1.3) or the last will of the author (Ch. 1.4).

Furthermore, the tools for preparing Digital Scholarly Editions available to philologists need to improve usability (Ch. 2.5.2, 4), leading to the heart of the problem: if a philologist does not reconstruct a text and, therefore, does not propose a unique version approximately close to the original or the author's intentions, the primary missions of philology (Ch. 2.3.1), whether *traditional* or *digital*, remain incomplete. In Ch. 2.5, four main reasons corroborate this problem: (1) philologists tend to utilise the vast space provided by the network as much as possible as a reaction to the fact that they have always been forced into tiny spaces; (2) the lack of easy-to-use digital tools for preparing Digital Scholarly Editions forces them to learn and implement very advanced techniques and technologies; (3) there is not necessarily a publisher on the web who imposes guidelines; and (4) the lack of established models of Digital Scholarly Editing leads to disorientation. As a result, the network is filled with *documents* but lacks *editions*. The issues above are inherently linked. Without easy-to-use digital tools and solid models for preparing complex editions and without the guidance of a publisher, philologists tend to resort to technically more straightforward tasks where they can work independently. Hence, they often limit their efforts to placing facsimiles online, possibly accompanied by their transcription.

The resolution to address the problem is straightforward — design and develop digital tools based on solid models that cater to the fundamental requirement of inducing philologists to propose a single version of the text with ease but also provide readers with the textual multiple forms and variants, and tools to explore them. This venture, as debated by Van Zundert (2016, 13), calls for a symbiotic relationship between DH and philology.

Chapter 3

Digital Editions: how to (technically) right now

Once the existence of *digital philology* is assured, it has been argued that the skills required of digital philologists are too advanced. Through the reflection in Ch. 1 and 2, we investigated the origins of the *edition*. Now we move to its evolution in *digital*, presenting a benchmark of Digital Scholarly Editing common methodologies, practices, issues, and implications. In practical terms, it demonstrates that preparing digital editions is a complex undertaking that requires a significant level of technical proficiency.

3.1 Digitising the primary sources

According to Pierazzo (2020, 50), every edition commences with digitising primary resources, either manuscripts or printed.[22] Digitisation[23] techniques include typing the text as a sequence of characters in a machine-readable form (MRF) or scanning/digital photography. The selection of digitisation techniques depends on several factors, including the medium, printed vs. handwriting text, language, and more. To convert digital images into electronic text, popular approaches such as manual text entry, the Optical Character Recognition (OCR) software, and semi-automatic methods are commonly used (Nguyen *et al.* 2021, 123:1–124:2), and each has its strengths and weaknesses, as discussed below.

3.1.1 Automatic transcription

OCR is the most commonly employed automated transcription method for printed texts. It examines digital images of a text and attempts to interpret its signs. However, achieving 100% accuracy is

[22] It cannot be assumed that *every* digital edition project commences with the digitisation of its sources through scanning or digital photography unless it is intended to prepare a *mechanical edition* (Ch. 1). There are instances where the source facsimiles have already been digitised, published online, and are freely accessible and can be included in external digital editions. The collections offered by the Cambridge Digital Library serve as an example of such sources: https://cudl.lib.cam.ac.uk/collections/. In addition, an increasing number of editions focusing on modern and contemporary authors do not begin with digitisation efforts, primarily due to the fact that the material is originally produced in digital format (Carbè, 2023).

[23] Digitisation is defined as a transformation process of paper-based materials into electronic texts (Nguyen *et al.* 2021, 123:1).

often challenging, as the text needs to be as readable as possible, which is only sometimes the case with manuscripts and, generally, with handwritten documents.[24] Thus, despite the continuous improvements in OCR engines, their optimal performance on historical documents remains challenging due to inadequate training data from past documents. The physical quality of the original materials, complicated layouts, and old fonts further complicate the matter, causing significant difficulties for the current OCR software. As a result, the OCR outputs are often noisy and may affect any downstream applications that rely on these textual materials as their input (Nguyen *et al.* 2021, 124:2). Correction methods are available, such as *crowdsourcing* systems. Unfortunately, these methods can be costly, often leading to their neglect (Pierazzo 2020, 50–52).

3.1.2 Manual transcription

Manual transcription is an alternate approach that is more time-consuming but more accurate. Not to be underestimated, the process is very expensive (Nguyen *et al.* 2021, 124:2).[25] Three different entities can carry out such a process: (1) a philologist or a scholarly expert who possesses in-depth knowledge of the language and the characteristics of the text to transcribe, (2) non-expert staff in cases of *outsourcing* solutions, and (3) individuals from the general public in the case of *crowdsourcing* systems. It is important to note that in both case 2 and case 3, the transcribers may not necessarily possess expertise in the domain (Pierazzo 2020, 52–53), and there may be security issues since information is shared with a third party (Nguyen *et al.* 2021, 124:2).

3.1.3 Semi-automatic transcription

There are intermediate solutions available which offer a collaborative approach for transcribing paper documents into digital data. This data is then utilised to train OCR models for generating automatic transcripts. *Transkribus*[26] is one such computer-assisted transcription tool that is widely used for this purpose (Nguyen *et al.* 2021, 124:2). It is based on Artificial Intelligence (AI) and, starting with setting the language and the type, i.e., handwritten or printed, of the document to transcribe, it can recognise the writing and layout, and generate its transcription. Eventually, it is possible to correct the errors that occurred. *eScriptorium* represents an alternative solution to

[24] See (Memon *et al.* 2020) for a review of OCR performance on handwritten documents.
[25] The IMPACT project (http://www.impact-project.eu/about-the-project/concept/, accessed August 25, 2023) reports around 1 EUR per page (Nguyen *et al.* 2021, 124:2).
[26] Transkribus: https://readcoop.eu/it/transkribus/.

accomplish the same task. It relies on Natural Language Processing (NLP) or Named Entity Recognition (NER) line-based techniques.[27]

3.1.4 Facsimiles

The proliferation of photographic reproduction of manuscripts and texts in print or on other media is ubiquitous and has significantly aided the work of philologists. The availability of an enormous amount of online material has facilitated their research and the publication of Digital Scholarly Editions, which present the facsimiles of the text next to their transcription. However, this method of digitisation has its drawbacks. Primarily, it is worth noting that the unit of measurement in photographic reproduction is typically restricted to the individual *page*, leading to the potential loss of valuable codicological, iconographic, and *mise en page* information. Moreover, photography fails to provide an accurate sense of scale for the object captured, regardless of whether it is a manuscript or a print (Pierazzo 2020, 53–56).

3.1.5 IIIF to describe and share images over the web

The International Image Interoperability Framework (IIIF) comprises a collection of open standards that facilitate the large-scale annotation and dissemination of high-quality images and audio/visual files[28] in web environments, enabling users to view and interact with them. The fundamental premise of this project is to augment the resources owned by institutions and enable their accessibility to the public. These standards are mainly adopted by galleries, libraries, archives, and museums (GLAM) organisations worldwide (Haynes 2023, 50). The IIIF Application Programming Interfaces (APIs) facilitate the interaction between various image repositories globally to overcome their fragmentation. This approach counters the prevalent issue of isolation that has traditionally impeded private and public initiatives aimed at preserving and valorising cultural resources (Bernardo *et al.* 2023, 124). Hence, researchers whose work is rooted in the analysis of visual sources benefit from this initiative by accessing relevant materials for conducting their research (Cohoner and Hueber 2023, 101).

[27] Differing from OCR-based tools, which perform character-level recognition, eScriptorium executes line-by-line recognition (Stokes et al., 2021).

[28] While the IIIF community devised technical solutions to simplify sharing and reusing 2D images and audiovisual (AV) corpora, 3D objects are still studied by the IIIF 3D Technical Specification Group. See (Haynes 2023).

3.1.6 IIIF, Mirador and the study of manuscripts

The utilisation of IIIF technologies in the study of manuscripts has proven to be particularly beneficial: (1) as manuscripts are preserved in specialised institutions, often far from the researchers' study places, technology provides remote access to these valuable resources;[29] (2) manuscripts often exist in the form of fragmented *codices*. The application of IIIF technologies simplifies their reconstruction and eliminates the problem of image siloing (Fagin Davis 2023).[30] (3) the comprehensibility of manuscripts is a fundamental aspect of their study, not only from a philological perspective. Tools such as *Mirador*,[31] an open source, web-based, general-purpose image viewer, aid scholars in recognising and comprehending the content of manuscripts by providing functionalities such as deep zoom (Fig. 4) and rotation.[32] In addition, Mirador's capacity to compare images is of paramount importance. Considering that juxtaposing codices and manuscripts constitutes the cornerstone of textual scholarship, the capability to simultaneously display two, three, or more codices on a single screen proves highly advantageous. Ultimately, Mirador is not merely an image viewer but also facilitates their annotation. Indeed, selecting a specific area of the image enables the addition of arbitrary annotation text (Van Zundert 2018, 2–7).

[29] See also Ch. 5.11 to check how these resources could be used in practice.

[30] See, for instance, the projects by *Fragmentarium* — Laboratory for Medieval Manuscript Fragments: https://fragmentarium.ms/.

[31] Mirador: https://projectmirador.org/.

[32] As IIIF, Mirador lacks functionalities for 3D visualisation and rendering (Van Zundert 2018, 3).

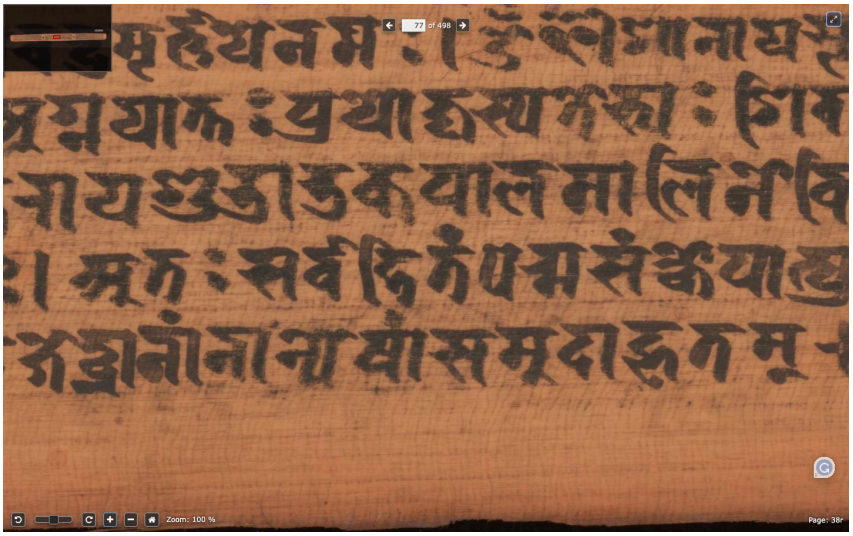


Fig. 4. Maximum zoom by Mirador of *Śivadharma* palm-leaf manuscript MS Add.1645. Page: 38r. Contents: *Śivadharmottara* (Ch. 5.2).[33]

3.1.7 IIIF and Mirador for the reconceptualisation of DSEs as distributed resources

One of the most significant challenges in Digital Scholarly Editing is *interoperability*. Despite the existence of tools attempting to overcome this problem, such as TEI (Ch. 3.3.1), what often occurs is the creation of editions that correspond to silos unable to communicate with other projects due to differing implementation methods. Van Zundert defines such editions as *data silos* and notes that the majority of digital editions are indeed so, being locally-integrated server applications. He argues that the underlying issue behind this situation is not purely technological but rather attributable to an institutional attitude. Specifically, he speaks of an 'institutionalised tradition of walled-in local digital resources and specific local methods of working with those resources' (Van Zundert 2018, 11). *Mirador* could be a solution enabling the widespread openness and distribution of editions due to its architectural component, which, through lightweight protocols, informs it of where to find the data and what to do with it, for instance, rotating an image, zooming in, etc. On the other hand, more advanced features, e.g., linking the manuscript transcription or clicking on a verse of the transcription and reaching the exact image area, are challenging to implement from a technical standpoint as it would require direct modification of Mirador's source code. And these would only represent some of the basic functionalities of digital editions (Van Zundert 2018, 33).

[33] The Sanskrit manuscript shown is preserved in the Cambridge University Library. The image of the page 38r is accessible at the following web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/77. Licence: CC BY-NC 3.0.

3.2 Modelling the edition

A Scholarly Edition is a complex entity both conceptually and structurally, as it represents a critical representation of historical documents, as per the definition (Sahle 2016) (Ch. 1.1). The objective of re-coding a text in an edition is twofold: first, to reproduce a historically transmitted document reliably, and second, to constitute a new reading of the text. Editors perform two synchronous actions, striving to reestablish the original form of the text as accurately as possible while providing their point of view and interpretation. This intrinsic sophistication makes editing a text inherently complex, including reflecting on the text's transmission, identification and coherent analysis of the structure and entities of interest. The reconstructed text is constrained by its sequential presentation but is also characterised by a multidimensional complexity (Neill and Kuczera 2019). Formal modelling of such complexity necessitates starting with the reading and analysing the text, its relevant layers, and inter-textual and intra-textual relations. Hence, modelling a digital edition is a multifaceted process that bears significant implications for every subsequent step in the production pipeline. Specifically, during the technical phases such as *encoding* and *development* (Ch. 3.3, 3.4), philologists should be supported by technicians who help them as much as possible to consider the fundamental variables, i.e., textual features, for the edition's presentation.

3.2.1 What to model

In the process of preparing a Digital Scholarly Edition, *modelling* stands as a critical step. The digital edition's structure, contents, visualisation, and usage depend on this aspect. Italia (2020, 63) deems modelling as the core component of digital philology since the edition's birth and development stem from it. At its essence, modelling entails laying the edition's foundation by deciding the end goal beforehand. This phase requires a thorough text analysis to identify pertinent characteristics deemed relevant for the reader-user. It demands meticulous attention to detail. Pierazzo (2020, 48) notes that well-defined, consistent, and formalised rules must be established when entrusting operations to a computer. This is why setting a list in bullet points of the edition's features to make available serves as guidelines in the design and development phase. For instance, as we will discuss further (Ch. 3.2.2), when seeking to furnish an edition with an *Index of Persons*, the model must incorporate a precise method for identifying and formalising every individual referenced in the text to make it computer-processable. Without this information in the model, it cannot be incorporated into the digital edition. This, in turn, renders it inaccessible for the user to navigate.

But what exactly to model? If all the characteristics of the edition depend on the model, the model depends on the editor's objectives. The initial determinant that significantly shapes the modelling decisions is the type of edition the editor intends to create (Ch. 1.2). In the case of a *mechanical edition*, the primary emphasis would be on selecting a witness to exhibit, if multiple exist, and highlighting any relevant characteristics to make accessible, such as the division into chapters. In the most complex scenario, i.e., *critical edition*, the edition is expected to comprise a selection of an extensive range of information, including the *apparatus* and the phenomena it can report (Ch. 1.3.4, 1.3.5), complementary apparatuses (Ch. 1.3.6), translation, and notes (Ch. 1.3.7). It is implied that each choice depends on the characteristics of the text.

To summarise the discussion, the fundamental principle of modelling an edition requires establishing a specific perspective on the object to be modelled, which entails including or excluding relevant details. In particular, (1) it is based on a selection of witnesses; (2) reflects the editor's critical point of view on the text; (3) sets the text according to standardised editorial models, e.g., diplomatic edition, critical edition etc.; and (4) includes a selection of details related to the text, e.g., variants in the *apparatus* (Pierazzo 2020, 49).

The scope of textual information to be represented in the model is limitless. It can encompass any information deemed relevant by the editor, ranging from traditional data contained in the *apparatus* to indices of individuals, dates, and places of interest mentioned in the text, as exemplified in *Vespasiano da Bisticci, Lettere* project[34] (Tomasi 2020). The model can also incorporate very specific details, e.g., corrective categories related to variants, such as graphic variants, punctual variants, idiomatic phrases, and adiaphoric variants, as demonstrated in the *PhiloEditor* project.[35]

3.2.2 What happens if the model is not correct

In an optimal scenario, the closure of each step in a workflow would signify its permanent completion. However, in practical planning, such a notion is often disrupted by what is commonly known as the *ripple effect*, i.e., selecting an out-of-bounds option at an advanced stage in the workflow will entail reconsidering decisions made at the previous stage (Garrett 2011, 23). Hence, if any issues arise during the encoding or development phases, it necessitates the revision of the model. It cannot be guaranteed that the model is entirely accurate once it is perceived to be *completed*. Conversely, it is often during the development phase, with the aid of visualisation, that existing errors are identified (Tomasi 2022, 133). This event is not uncommon: it occurs *very*

[34] Vespasiano da Bisticci, Lettere: http://projects.dharc.unibo.it/vespasiano/.

[35] PhiloEditor: https://projects.dharc.unibo.it/philoeditor/.

frequently. The modelling process follows an iterative approach, which suggests that the model will be deemed complete only after several iterations. Consequently, each iteration, i.e., modification to the model, refines the analysis method. Therefore, the editor is unlikely to prepare the *textus* in the first attempt and finalise all the edition's features at once.

Notwithstanding the possibility and likelihood of modifications to the model, editors should maintain high precision and focus, considering as many variables as possible. Each alteration to the model carries significant implications, which may necessitate restarting the work and extending the project timeline (Pierazzo 2020, 49–50). Modifications of such nature can prove to be considerably costly, particularly when they require alterations to the software (Kaur and Singh 2015, 6). In order to optimise the modelling process of an edition, it would be advisable to enlist the assistance of professionals who possess a comprehensive technical understanding of the variables that significantly impact the presentation and management software of the edition.

3.3 Encoding the text

Once the relevant textual features and relationships are identified, it becomes necessary to develop a formal system that can aptly represent and make them computer-readable and processable. Annotations are instrumental in establishing specific relations between annotated data in a given context (Oren *et al.* 2006). Each relation embodies a particular perspective on the text according to the editor's interpretation (Barzaghi 2021).

The annotations are conveyed via precise *markup* languages, facilitating the formal representation of the identified textual phenomena. Two primary markup languages are utilised for text annotation: (1) the eXtensible Markup Language (XML), which is based on the Standard Generalized Markup Language (SGML), and (2) LaTeX, which is based on TeX.

The conventional encoding of digital editions is typically carried out following the Text Encoding Initiative (TEI) standard guidelines, founded on XML. Thus, the TEI standard employs a hierarchical tree structure where elements are nested inside each other. As we will observe in the subsequent sections (Ch. 3.3.1), this process entails various challenges, particularly concerning the expressiveness of the formal representation of not strictly hierarchical structures. Further analysis will be conducted to explore alternative solutions to the standard. These alternatives include graph-based technologies and frameworks (Ch. 3.3.3, 3.4.5), extending the encoding systems to complex and overlapping structures, but not without some drawbacks.

As discussed by Del Gratta *et al.* (2022, 6), the existence of an extensive range of document models[36] indicates the necessity of comprehending their inherent characteristics. Simultaneously, it highlights the challenge of discovering a cross-model that can be applied across all reference domains in a particular field of study or application.
Finally, we reflect on the impact of the choice of a text encoding system on the final display of the edition (Ch. 3.3.7).

3.3.1 *Tree* editions, TEI and related issues

XML/TEI has been a long-standing standard in the Humanities field for years[37] and continues to hold its position even now (Pierazzo 2020, 61; Pierazzo and Leclerc 2016, 186; Tomasi 2012, 266). While XML serves as a *metalanguage*, providing a set of rules that aid in creating an infinite number of languages (Fiormonte *et al.* 2015, 138; Holzner 2001, 1), the TEI offer a standardised set of semantic annotations corresponding to tag labels that indicate the function of each textual element (Buzzetti 2003, 182). TEI tags facilitate translating the text interpretation by philologists, e.g., humanistic concepts as *person*, *place*, and *organisation*, into a formal language that computers can comprehend and process, modelling a given source as a result.
TEI is defined by the XML syntax, which conforms to the paradigm of embedded markup in a single rooted tree (Vogeler 2021, 77). Acquiring proficiency in XML is relatively straightforward (Segers 2021, 340), as it is a very flexible metalanguage. However, scholars must learn it and its schemas, e.g., TEI, in-depth (Bonsi *et al.* 2020, 52). In addition, the data structure of XML may not be optimal from an efficiency standpoint in specific contexts of use, i.e., *overlapping structures* (Ch. 3.4.3).
Upon analysis of TEI, it can be observed that its most notable feature is the extensive range of its lexicon. The many tags that represent identical textual phenomena can be a mixed blessing. While it offers philologists a broad *spectrum* of terminologies to semantically define the object of annotation, it can also pose a challenge (Schmidt 2019, 5). Searching for the most appropriate definition and tag to represent a specific textual feature in extensive documentation spanning over

[36] The edition model classification presented in this discussion is also in Del Gratta *et al.* 2022 where they identify three main models for representing documents: the sequence of characters (*strings*), generalised markup (*hierarchical representation*), and the Resource Description Language (RDF - *network model*).

[37] Since its inception in 1994, the TEI Guidelines have garnered immense popularity and widespread usage among libraries, museums, publishers, and scholars.

2033 pages[38] can be labour-intensive. Moreover, differences between various tags are often subtle nuances of meaning, making the selection process even more daunting.[39] It is noteworthy that TEI Lite[40] came into existence as a deliberate measure. It is a more concise and comprehensible subset of TEI designed to enhance user-friendliness. However, TEI Lite's scope is relatively limited compared to its counterpart and may not effectively cater to all the textual complexities that may arise in a given text.

A related additional issue is the TEI interoperability capabilities (Schmidt 2021). Since there is not only one way to code for each specific textual phenomenon[41] included in the system and its terminology is not exempt from synonymy, the interoperability of documents is consequently difficult to guarantee.

Finally, the semantic expressiveness of TEI is debated. The analysis below delves into specific inadequacies in expressing semantic relations in TEI, as observed and reported by Renear, Dubin, and Sperberg-McQueen (2002, 122).

1 ***Relationships between elements, attributes, or attributes values***. No general constructs to formally classify specific relations are available in the guidelines. For instance, the attribute `@ref` allows setting a link to the `@id` of any element, but it does not specify the semantics related to that link.

2 ***Context where an element may be used with different meanings***. It is challenging to articulate the context in TEI formally. For instance, the `<title>` element marks a textual string as a title, but it can have a different meaning depending on its position. `<title>` within `<head>` means *title of the document*, while it means *title of the enclosing chapter* within the `<chapter>` element. There is no mechanism to express the *title of* kind of relation.

---

[38] We refer to the latest English version of the TEI guidelines, specifically the *P5: Guidelines for Electronic Text Encoding and Interchange* (TEI Consortium 2023). The page count is subject to variation based on the language preference selected.

[39] A significant example can be found in the Epidoc specifications (https://epidoc.stoa.org/), which offer a subset of TEI for encoding scholarly editions. Initially designed for ancient inscriptions, an extension of these guidelines covered papyrus and manuscripts. For every markup tag, they include instructions to guide its proper usage. For instance, in the case of the `<gap>` element, there is potential semantic confusion with the `<unclear>` and `<del>` elements. So, they provide further clarifications: https://epidoc.stoa.org/gl/latest/ref-gap.html.

[40] TEI Lite: https://tei-c.org/guidelines/customization/lite/.

[41] For instance, consider the diverse approaches that exist for linking the *apparatus* to the text in TEI: https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPLK.

Based on the examples above, it can be asserted that XML/TEI semantic value and consistency depend largely on human interpretation and control. Therefore, any markup restriction or semantic role must be expressed in natural language as instructions or documentation for human users (Ciotti and Tomasi 2016, 2). It suggests that XML markup language users are forced to guess about the semantic relationships the encoder intended but had no way to express formally. Providing additional documentation in natural language to express the semantics behind a text may be a partial solution since such documentation would still not be machine-readable (Renear, Dubin and Sperberg-McQueen 2002, 119–120). Many proposals have been put forth to address the integration of XML/TEI with semantic technologies (Ch. 3.3.3) to overcome the TEI's issue in semantic expression.[42]

3.3.2 TEI in *tree* editions: some examples

In the field of Humanities, the annotation systems predominantly rely upon using XML/TEI (Ch. 3.3.1). Therefore, TEI-based editions are widely popular.[43] The following example of a critical edition marked in TEI, *Edizione Logica Avicennae*,[44] was selected for its complexity, owing to the abundance of components included in addition to the *textus*, i.e., *apparatus criticus*, *fontes*, and *parallels* (Ch. 1.3.4, 1.3.5, 1.3.6). A snippet of the TEI markup (Listing 1) and its visualisation on the edition's interface (Fig. 5) are presented below to provide a comprehensive understanding of the encoding behind the edition.

[42] See, for instance, Ciotti 2018; Ciotti et al. 2016; Ciotti, Daquino and Tomasi 2016; Tomasi *et al.* 2013.

[43] Their strong adoption is evident from the analysis of two digital editions catalogue: (1) the *Catalogue of Digital Editions* by Franzini (https://dig-ed-cat.acdh.oeaw.ac.at/), and (2) *A catalog of: Digital Scholarly Editions* by Sahle (https://v3.digitale-edition.de/).

[44] Edizione Logica Avicennae: http://evt.labcd.unipi.it/demo/evt2-beta2/avicenna/index.html#/readingTxt?d=doc_1&p=C-112v&s=text-body-div&e=critical&app=text-body-div-head-app.

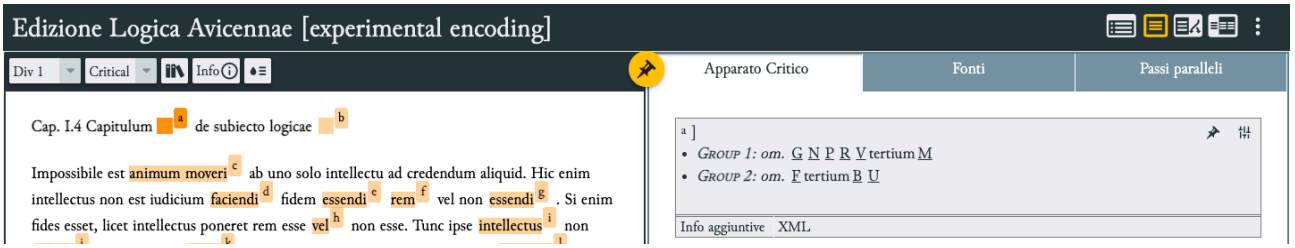


Fig. 5. *Editione Logica Avicennae*. Focus on an *apparatus* entry.

```
<head>
      <supplied resp="#silvia">Cap. I.4</supplied>
      Capitulum
      <app resp="#silvia">
            <lem wit="#F #G #N #P #R #V" />
            <rdg cause="add" wit="#M #B #U">tertium</rdg>
      </app>
      de subiecto logicae
      ...
</head>
```

Listing 1. *Editione Logica Avicennae*. Focus on an *apparatus* entry.

The example above, corresponding to a formalised *apparatus* entry in TEI, exhibits a strictly hierarchical and nested structure in XML format. Notably, the annotated strings, such as *lemma* and *variant reading*, are associated with specific tags that determine their semantic meaning, while attributes are assigned to these tags to specify additional information. In contrast, no tags include strings that are not objects of annotation.

3.3.3 *Graph* editions, semantic technologies and related issues

In most cases, the textual entity-relation structure is non-linear since it mirrors the multidimensionality of the text and its interpretation. As discussed by Neill and Schmidt (2021, 45), annotations are inherently *overlapping*. It implies that any given range of text may simultaneously be the subject or object of more than one annotation and is often *out-of-order*, as defined by Peroni and Vitali (2009). As a result, the traditional text encoding, which is based on SGML/XML syntax such as the standard XML/TEI, needs to be revised when annotations do not fit into a linear and hierarchical tree structure. As cited by Marinelli, Vitali, and Zacchiroli (2008, 1), an instance of overlapping markup is presented below (Listing 2, Fig. 6).

```
< doc > <b > John <i > likes </ b > Mary </ i > </ doc >
```

Listing 2. Instance of overlapping markup (Marinelli, Vitali, and Zacchiroli 2008, 1).

John *likes* *Mary*

Fig. 6. Expected result of the markup of Listing 2 (Marinelli, Vitali, and Zacchiroli 2008, 1).

The XML markup fragment above is not *well-formed*. Even if its syntax looks like XML, the nesting of the `<i>` element with the `<b>` element and vice-versa is improper because a portion of both elements overlaps. As shown in Fig, the word *likes* is both bold and cursive.

This inadequacy leads to a need for more semantic expressiveness. The literature has widely discussed the limitations of the tree representation of text, which is based on SGML/XML-based syntax (Peroni and Vitali 2009; Renear, Dubin and Sperberg-McQueen 2002; Schmidt 2010). Several solutions have been proposed to overcome these limitations. Some research demonstrates that the tree hierarchies may be extended to solve problems such as overlapping while maintaining the tree as the primary data structure, i.e., by introducing techniques such as TEI *milestones* and *fragmentation*[45] (Di Iorio, Peroni and Vitali 2011, 1698).

The *graph* data structure is widely regarded as a more practical approach for conveying texts' structural and semantic multidimensionality. Indeed, the emergence of Web 3.0 and semantic technologies has enabled graph-based solutions to address the challenges associated with hierarchical tree data structures. Applying such technologies resulted in a reconsideration of the *information resources* notion (Daquino, Giovannetti and Tomasi 2019, 50).

While the global network was typically conceptualised as a network of documents, it currently comprises a collection of entities uniquely identified and connected through typed links. Consequently, the previous notion of a document as a *full-text entity* has been replaced by an entirely different model. Now, a document is perceived as a subject or object that can potentially be associated with a collection of other entities, subjects, or objects, which, in turn, may have other relationships with documents or entities, such as people, places, or events. This intricate network of relationships is what generates the *graph*.

The same applies to the specific context of Digital Scholarly Editing. The prevalent conceptual and technological framework employed in most digital editions revolves around the primacy of the document, interconnections between closed databases, and hierarchical textual encoding.

[45] TEI Non-hierarchical Structures: https://tei-c.org/release/doc/tei-p5-doc/it/html/NH.html.

Conversely, the *knowledge graph* paradigm is gaining traction by facilitating the categorisation of relationships amongst data in an explicit, machine-readable semantic manner (Spadini and Tomasi 2021, 1). Hence, *graph* text modelling facilitates a comprehensive approach wherein a text is viewed not merely as a surface-level construct but as a multidimensional entity characterised by diverse dimensions and relationships (Bonsi *et al.* 2015, 92). The predominant language for technical and formal graph development is the *Resource Description Framework* (RDF).[46] It enables data organisation as *nodes* connected by typed links in a graph. It allows for resolution of issues such as *overlapping hierarchies*, as it is possible to assign any annotation to the same range of text (Peroni and Vitali 2009).

A viable solution in Digital Scholarly Editing could be based on the Open Annotation Model[47] together with the clients built on top of it, such as *Recogito*[48] and *Hypothesis.is*.[49] Specifically, the Open Annotation Model is an extensible and interoperable framework for expressing annotations according to an RDF graph, allowing sharing between different platforms. The structure behind annotations is basic. There are three primary resources describing the annotation: (1) the *Annotation*; (2) the *Body*, i.e., any content regarding a *Target*; and (2) the *Target*, i.e., the subject of the annotation (Sanderson, Ciccarese and Van de Sompel 2013, 366). This model offers sufficient flexibility and depth to cater to complex requirements while still being simple enough to accommodate everyday use cases, such as attaching a piece of text to a single web resource. However, these technical solutions come with certain limitations. For instance, they do not allow annotating arbitrary textual ranges. Additionally, they assume that the annotated text is immutable, meaning that any changes made to the original text automatically cancel the annotations (Neill and Schmidt 2021, 46).

Through techniques of this kind, except Resource Description Framework in Attributes (RDFa) RDF serialisation,[50] annotations are stored separately from the text and may be overlapped freely, thus solving the inherent problems of an XML-based tree structure (Ch. 3.3.1). However, RDF is an expensive solution regarding learnability: no formal models, standards and infrastructures are

[46] RDF: https://www.w3.org/RDF/.

[47] Open Annotation Model community: https://www.w3.org/community/openannotation/.

[48] Recogito: https://recogito.pelagios.org/.

[49] Hypothesis.is: https://web.hypothes.is/.

[50] The RDFa approach falls short in effectively addressing the issue of overlapping, as its structure adheres to that of HTML incorporating specific attributes. This suggests that the structure remains bound to a tree-like hierarchy (Neill and Schmidt 2021, 46).

RDFa: https://www.w3.org/TR/rdfa-primer/.

available for now (Spadini and Tomasi 2021, 1–2). In addition, visualisation is outside the scope of RDF (Daquino, Giovannetti and Tomasi 2019, 56).

3.3.4 Semantic technologies in *graph* editions: some examples

In recent times, Semantic Web technologies applied to the Humanities have resulted in noteworthy benefits, particularly the enhancement of digital editions of texts (Tomasi 2012, 265). With the growing popularity of these technologies, digital editions of scientific and non-scientific texts showing that the graph is flexible enough to represent the discontinuity of a text's interpretation are now available. However, it is still premature to assert the widespread adoption of shared standards that leverage the potential of Semantic Web technologies (Bolioli, Tasso and Rosselli del Turco 2017, 167).

In the *Edizione Nazionale delle Opere di Aldo Moro* (Moro 2021), the RDF serialisation RDFa provides HTML documents with structured data, such as people and places mentioned in a given text. The examples below show the display (Fig. 7) and the underlying formalisation of the person *Manzini, Vincenzo* mentioned in a text (Listing. 3).

Fig. 7. *Edizione Nazionale delle Opere di Aldo Moro*. Focus on mentioned people in the text.

```
<p class="paragraph" data-counter="2" id="p-2">
      In Italia il problema è stato proposto dal <span
      about="https://w3id.org/moro/enoam/data/210001/v1/mention-0"
      class="mention person" id="mention-0" property="dcterms:references"
```

```
      resource="https://w3id.org/moro/enoam/data/vincenzo-manzini"
      typeof="deo:Reference">Manzini</span>
      ...
</p>
```

Listing 3. *Edizione Nazionale delle Opere di Aldo Moro*. Formalisation of a person mentioned in the text in RDFa.

In the *Semantic Scholarly Digital Edition of Paolo Bufalini, Appunti (1981-1991)*[51] (Tomasi *et al.* 2019), RDF triples extracted from the XML/TEI document of the edition form the graph (Daquino, Giovannetti and Tomasi 2019). The examples below show the display (Fig. 8) and the underlying formalisation of the person *Publio Tacito* mentioned in the notebook (Listing 4, 5).[52]

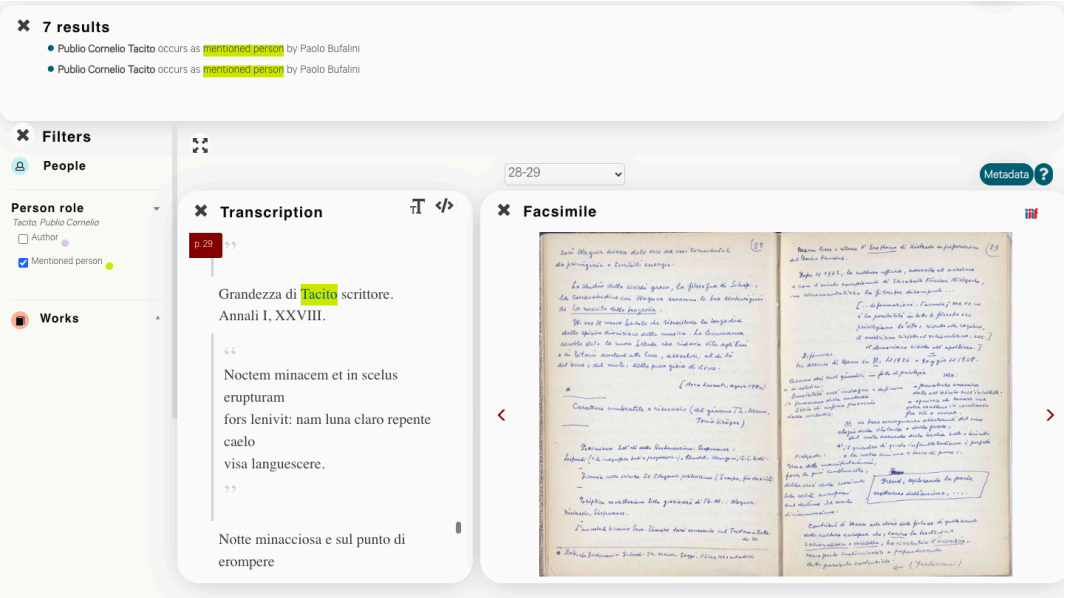


Fig. 8. *Semantic Scholarly Digital Edition of Paolo Bufalini, Appunti (1981-1991)*. Focus on mentioned people in the text.

```
<note type="comment" xml:id="comm-125a" corresp="#quot-181" resp="#PB">Grandezza
di
<persName ref="#CT">Tacito</persName> scrittore.
</note>
```

Listing 4. *Semantic Scholarly Digital Edition of Paolo Bufalini, Appunti (1981-1991)*. Formalisation of a person mentioned in the text in XML/TEI.

```
<rdf:Description rdf:about="https://w3id.org/bufalinis-notebook/person/ct">
```

[51] Semantic Scholarly Digital Edition of Paolo Bufalini, Appunti (1981-1991): https://projects.dharc.unibo.it/bufalini-notebook/.

[52] The workflow is summarised in Dello Buono 2023.

```
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person">
      <rdfs:label xml:lang="it">Publio Cornelio Tacito</rdfs:label>
      <rdfs:label xml:lang="en">Tacitus</rdfs:label>
      <owl:sameAs rdf:resource="http://viaf.org/viaf/100226923">
      <owl:sameAs
rdf:resource="http://it.dbpedia.org/resource/Publio_Cornelio_Tacito"/>
</rdf:Description>
```

Listing 5. *Semantic Scholarly Digital Edition of Paolo Bufalini, Appunti (1981-1991)*. Formalisation of a person mentioned in the text in XML/RDF.

In Paolo Bufalini's notebook edition, we observe a significant shift in the digital edition paradigm from the conventional XML/TEI tree model to a more advanced graph model. This novel approach represents the text not as a mere data sequence but as a complex network of relationships between various entities. Each markup element assumes the role of either a subject, object or property connected to the source document via typed links. This approach enables the construction of a knowledge graph consisting of a set of RDF triples extracted from the TEI document.

3.3.5 *Graph* editions, graph databases and frameworks, and related issues

Graph databases, such as Neo4j,[53] represent a viable alternative to semantic technologies. Its ability to incorporate *stand-off* textual annotations that remain entirely detached from the text allows for circumventing the problem of overlapping, resulting in a highly effective approach to semantic representation. Specifically, the text can be saved in a database as plain text and the markup as an array of properties pointing to the text. However, NoSQL databases, such as graph databases,[54] lack standardisation than relational databases,[55] where query results can be visualised in standardised environments. Unlike relational databases, every NoSQL implementation has a unique user interface which needs to be comprehended and learned (Carro 2014).

Drawing a comparison with semantic technologies, which utilise specific databases known as *triplestores*, graph databases such as Neo4j offer advanced graph metrics and weighted edges.

[53] Neo4j: https://neo4j.com/.

[54] Graph databases, such as Neo4j, are often categorised as *native* graph databases and are among the many NoSQL stores available. Other NoSQL systems include wide-column, document, and key-value (KV) stores (Besta *et al.* 2023).

[55] According to Vogeler (2021, 75), relational databases are ubiquitous in software engineering, yet their usage in scholarly editing remains limited. An instance of a digital edition that employs a relational database is the *Catullus Online* project (Kiss 2020): http://www.catullusonline.org/CatullusOnline/index.php.

Moreover, they are remarkably straightforward for developers to set up and maintain (Sippl, Burghardt and Wolff 2021, 191).

3.3.6 Graph databases and frameworks in *graph* editions: some examples

Analysing the main digital editions catalogues by Franzini (Franzini, Mahoni and Terras 2016) and Sahle (2020), it appears that digital editions relying on graph databases and frameworks are currently unavailable. However, the *Codex* (Neill and Kuczera 2019) project deserves significant attention even if it is not featured in the catalogues, as it is not classified as an edition. The primary goal of *Codex* is implementing a user-friendly textual annotator that formalises annotations according to a stand-off architecture to create a history atlas. The formalisation is accomplished by representing the annotated strings and annotations as nodes and relations in a graph database, i.e., Neo4j (Fig. 9). The annotations are diverse, not limited to text presentation features like pages, lines, and paragraphs, but also include semantic aspects. The *Codex* interface allows for marking individuals, places, and natural and artificial objects. This list can be customised by developers who choose to adopt this solution.

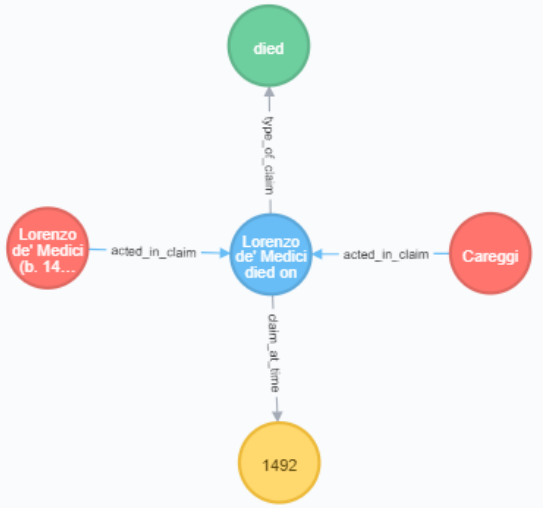


Fig. 9. Example of a graph representation of entities identified in the text along with their corresponding relationships in Neo4j.

3.3.7 Direct dependency of the edition's model on the encoding

As Buzzetti argues (1999, 128–130), all forms of text representation involve certain assumptions about a model, either implicitly or explicitly. It means that the conventional representation of a text is itself a model. The same applies to digital representations of text, where the markup is used to

express the form of representation and impose precise constraints on the automatic processing of the document's information content. It is necessary to apply processing procedures to formal information that support different investigation methods of critical and interpretive practice to achieve an appropriate digital representation of text. The effectiveness of the digital representation is determined by the structure assigned to the information and the applicable operations (Buzzetti 2000). In Digital Scholarly Editions, markup defines the model of the text in question. Consequently, annotations translate interpretations of the text into a formal format by making explicit declarations about the logical functions of the text strings.

3.4 Visualising the edition

The modelling, design, and development phases result in the digital edition interface. Such a final phase, i.e., the *visualisation*, demands the highest technical expertise (Ch. 3.4.1). As debated by Segers (2021, 340) and Turska, Cummings, and Rahtz (2016, 2), it is often necessary to rely on the IT staff at this point. In addition, the visualisation phase is a test case for validating the entire workflow of preparing a digital edition. It encompasses and concretely shows to the reader-user eyes all the design choices made earlier, ranging from the edition modelling and component selections (Ch. 1.3.4, 1.3.5, 1.3.6, 1.3.7, 3.2) to the quality of encoding (Ch. 3.3) and the resilience of user application development. Considering that the interface is essentially a byproduct of modelling (Ch. 3.4.2), we shall elaborate on the interface design features already during the modelling phase of the edition. This planning ahead strategy is necessary to prevent a scenario where a model and encoding that lacks comprehensiveness fall short of fulfilling the functions of the digital edition as required by the philologist-stakeholders. Other challenges associated with designing an interface should be considered. In particular, the limited available space in a computer layout (Rosselli Del Turco 2019, 104) and usability concerns (Rosselli Del Turco *et al.* 2014, 3) must be taken into account while accommodating the various components of a digital edition.

3.4.1 Visualisation techniques

Multiple data visualisation techniques exist. The choice depends on the underlying data structure. In the context of Digital Scholarly Editions, the final output file is typically HTML since a digital edition's hosting platform is usually a web application.

Following the conventional workflow (Ch. 3.3.1), editions' data are encoded in XML/TEI. Thus, it is necessary to employ dedicated scripts in languages such as XSLT, XPath, or XQuery to transform XML into HTML. Conversely, in the case of a graph-based edition (Ch. 3.3.3, 3.3.5), the process of

transforming data into HTML becomes more intricate. Regardless of whether RDF is the underlying framework, the data must be stored in a database. Consequently, extracting the data from the backend using specific queries becomes necessary to transmit it from the backend to the frontend and present it to the final user.

Due to the significant number of languages to learn and the intricate processes of data extraction and visualisation (Table 3), philologists face the most critical challenges in managing the visualisation phase, whatever the starting data structure. At this phase, external intervention becomes indispensable.

Examples of conversion in HTML				
Data structure	**Data hosting**	**Data extraction**	**Data visualisation**	**Reference example**
XML(/TEI)	XML file	XSLT, XPath, XQuery, or similar XML converting languages into HTML	XSLT, XPath, XQuery, or other converting language processors, such as Saxonica processors[56]	*shivadharma-xslt-base-txt*[57]
Graph in RDF(/any serialisation)	Graph database supporting RDF/SPARQL APIs, such as Blazegraph[58]	SPARQL queries	Conversion in JSON in Python via RDFLib/ SPARQLWrapper [59] or similar libraries Visualisation in Jinjia2[60] or similar templating engines	*RDFpyApp*[61]
Not semantic graph	Graph database, such as Neo4j	Cypher queries	Visualisation in EJS or similar	*Śivadharma Database* [62](Ch.

[56] Saxonica: https://www.saxonica.com/products/feature-matrix-12.xml.

[57] shivadharma-xslt-base-txt: https://github.com/martinadellobuono/shivadharma-xslt-base-txt.

[58] Blazegraph: https://blazegraph.com/.

[59] RDFLib/SPARQLWrapper: https://github.com/RDFLib/sparqlwrapper.

[60] Jinjia2: https://pypi.org/project/Jinja2/.

[61] RDFpyApp: https://github.com/martinadellobuono/rdfpyapp-web-app-template.

[62] Śivadharma Database: https://github.com/martinadellobuono/shivadharma-database.

			templating engines	7)

Table 3. Examples of technical workflows to convert an edition encoding into HTML.

As visualising editions involves significant technological complexities, numerous tools have been developed to ease this phase, eliminating the need for coding scripts for conversion. The Edition Visualization Technology (EVT)[63] (Ch. 4.6.1) and TEI Publisher[64] are two of the most employed XML/TEI edition viewers.

3.4.2 Edition's interface as result of modelling

The role of the interface in DH research projects has been gaining increasing importance recently. Caviglia, Ciuccarelli, and Coleman (2012) highlight this trend. According to Manovich (2010, 100–101), the interface is a crucial component that corresponds to the underlying data structure and algorithms. By providing users with tools to create or access content and facilitating them to utilise algorithms that process and generate data without direct contact, the interface plays a pivotal role in user experience. The representation in this context comprises two interrelated components: data structured by specific prerequisites and an interface with tools to navigate and manipulate such information. The correlation between data modelling and the interface is integral and indissoluble.

In the context of the increasing focus on visualisation tools during the modelling process, a more open-minded approach to visualisations is required that considers the real world, as opposed to a purely analytical approach. While various disciplines, particularly Information Visualization, have extensively studied visualisations from a strictly analytical perspective, focusing on specific types of visualisations used primarily in limited, specialised contexts, the Humanities require consideration of how visualisations can be applied to real-world contexts of use (Masud *et al.* 2010, 445). Therefore, visualisation is not solely a technical matter. As a *thinking process* more than an *end product* (Hinrichs *et al.* 2023, 43), it requires critical and methodological reflection concerning the practical application of the final products.

Regarding the Digital Scholarly Editing workflow, it is crucial to understand *precisely* what the user wants to see in a digital edition and what operations they want to allow *already* during the modelling stage. The edition's model is the basis for its development and ultimately determines its

[63] EVT: http://evt.labcd.unipi.it/.

[64] TEI Publisher: https://teipublisher.com/index.html.

presentation[65] to the end reader-user. As previously mentioned (Ch. 3.3.7), the editor's definition of the model guides the edition's development process. According to Tomasi (2022, 138–147), the visualisation of the edition, i.e., *presentation*, results from data *interpretation*, summarised in a *model*, leading to services for end-users and the development of modes and tools for accessing content. Then, selecting a suitable visualisation methodology is contingent upon the characteristics of the data and its relevance to the research objectives. As debated by Jänicke *et al.* (2017), any DH project should commence with comprehensive deliberations regarding research questions and perspectives that may prove advantageous for visualisations. These discussions should specifically focus on analysing the data's properties, which govern the development of the model. The following illustration (Fig. 10) depicts the interdependency relationships among the various phases of preparing digital editions, i.e., *modelling*, *encoding*, and *development*.

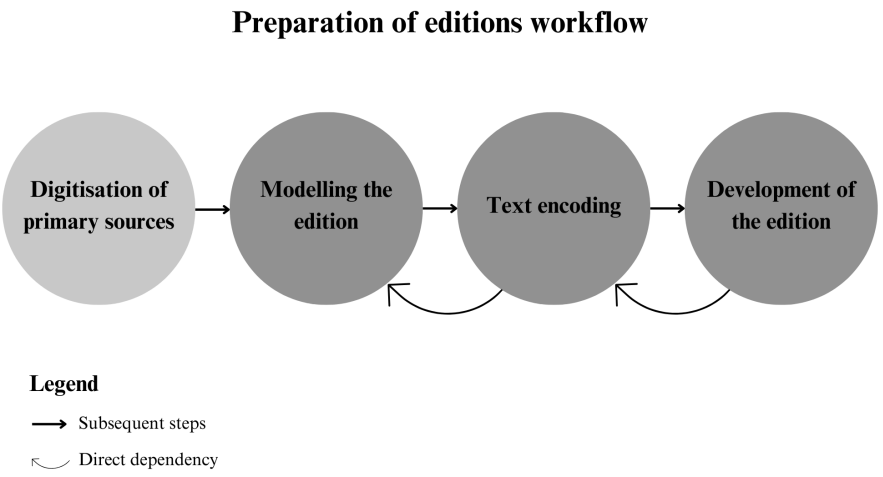


Fig. 10. Preparation of editions workflow. Direct dependency between the modelling, encoding, and development phases.

Data can only be presented after formalising them according to a predetermined model. At the same time, supporting a specific data format determines the following choices regarding the processing and visualisation tools already during the modelling phase (Rosselli Del Turco 2019, 105) due to this interdependency.

[65] We say *presentation* meaning the final view of the edition, usually a web page, referring to Barabucci, Spadini and Turska 2017, where they distinguish the edition's *data*, i.e., file(s) usually in XML/TEI, but not only (Ch. 3.4), by their *presentation*, usually in HTML.

3.4.3 Interdependence between modelling, encoding and visualisation: a scenario

The following hypothetical scenario[66] concerning the design and development of an *Index of Persons* enabling users to access a comprehensive list of individuals mentioned within a specific text shows the interdependence between the editions' modelling, encoding, and visualisation stages in practice.

Scenario:

The developer is requested to extract all the names of individuals mentioned in a given text.

Possible plots and end of the stories:

1 ***Correct model***. The editor anticipated the relevance of this information in the earlier stages of the edition preparation workflow, included it in the model, and formalised it according to technical specifications that enable the computer to process it.
=> It is possible to provide a comprehensive index of all the individuals mentioned in the text.

2 ***Incorrect model***. The model does not comprise such data; consequently, its visualisation would not have been deliberated during the edition's design phase, and there is no formalisation of such data.
=> It is NOT possible to provide a comprehensive index of all the individuals mentioned in the text.

The scenario demonstrates that the inadequate or missing information in the model weakens its formal representation and, thus, hinders its development and, consequently, visualisation. Vice versa, modifications to the final presentation necessitate alterations to the foundational data or improvements to the rendering software (Barabucci, Spadini and Turska 2017, 43).

[66] Utilising scenario-based methods, which involve the use of scenarios, i.e., detailed depictions of hypothetical events that may occur during a certain phase of a product's life cycle, can prove to be beneficial not only for software, interaction, and product design, but also for User Experience (UX) design (Michailidou, Haid and Lindemann 2015, 609–610).

Chapter 4
All the reasons why we need a Digital Scholarly Editing CMS

The process of preparing digital editions is fraught with technical challenges that go beyond the workflow methodology. As discussed in Ch. 3.3, the encoding stage demands a high level of technical expertise, including learning and mastering specific languages (Pierazzo 2019, 212; Robinson 2005; Schmidt 2018). Similarly, modelling and visualising digital editions require considerable technical skills (Ch. 3.2, 3.4). These steps must ensure a comprehensive representation of the textual complexity while being technically easy to handle by scholars to allow them to focus on *scholarly* rather than *technical* issues. Therefore, it is crucial to have *user-friendly* software, tools, and services that can streamline scholars' work and make it more accessible. Since it is philologists' responsibility not only to undertake philological work to restore a *textus* (Ch. 1) but also to learn and master some of the technical aspects of implementing digital edition editions (Ch. 2), the technological choices made in this regard have an impact not only on the expressiveness of the edition model (Ch. 3.2), encoding (Ch. 3.3), and interface (Ch. 3.4) but also on their effort in learning how to prepare an edition in practice. A potential solution to make Digital Scholarly Editing technologies *user-friendly* is a CMS, which has already been tested in this context (Ch. 4.6). A CMS allows editors to create and release content via an interface (Ch. 4.1), without requiring them to write code (Ch. 4.2) and preview the eventual edition's visualisation (Ch. 4.3). Finally, it is possible to publish content on the same platform where it was created, resulting in a reduction of costs associated with the publishing process (Ch. 4.4).

4.1 What is a CMS

Using a CMS can facilitate various tasks such as collecting, managing, distributing, and updating content across various media, as noted by Jagamogan *et al.* (2021, 1). The architecture of a CMS consists of four interaction layers, as defined by McKeever (2003, 688): *content*, *activity*, *outlet*, and *audience* (Fig. 11).

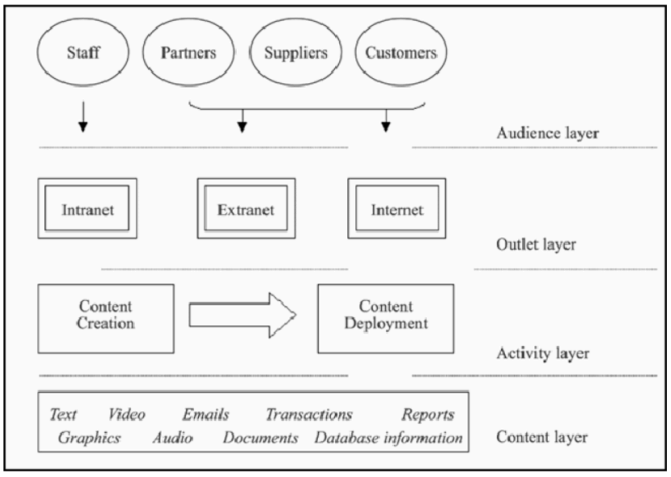


Fig. 11. CMS layers (McKeever 2003, 688).

The first layer of the hierarchy, i.e., *content*, is responsible for managing different types of content, such as text, video, documents, and more, which require a system or process for their efficient management. The *activity* layer represents the tasks in handling the *content* layer, including its creation and deployment to an appropriate outlet. The *outlet* layer identifies the web outlets that allow users to access the content, such as intranet, extranet, or internet. Lastly, the *audience* layer defines the groups of individuals who are expected to interact with the web outlet.

The CMS lifecycle follows a *collection-release iteration*, as depicted in the figure provided by McKeever (2003) (Fig. 12). The *collection* phase involves creating or acquiring content from an existing source, while the *release* involves delivering the content to the intended audience.

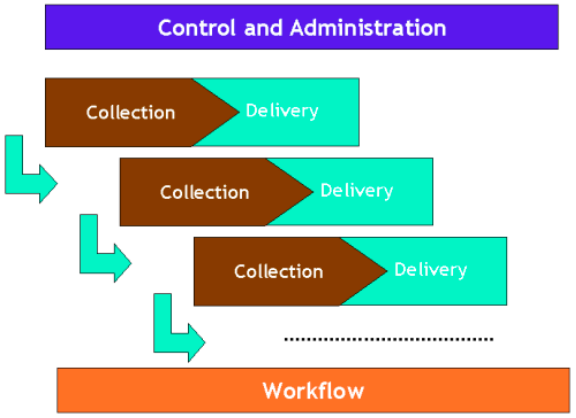


Fig. 12. CMS lifecycle (McKeever 2003, 688).

Content creation often involves multiple individuals with unique responsibilities, such as editing or reviewing documents. Most users involved in content creation may not possess technical expertise, and thus require user-friendly interfaces. Typically, an interface is interposed between the content

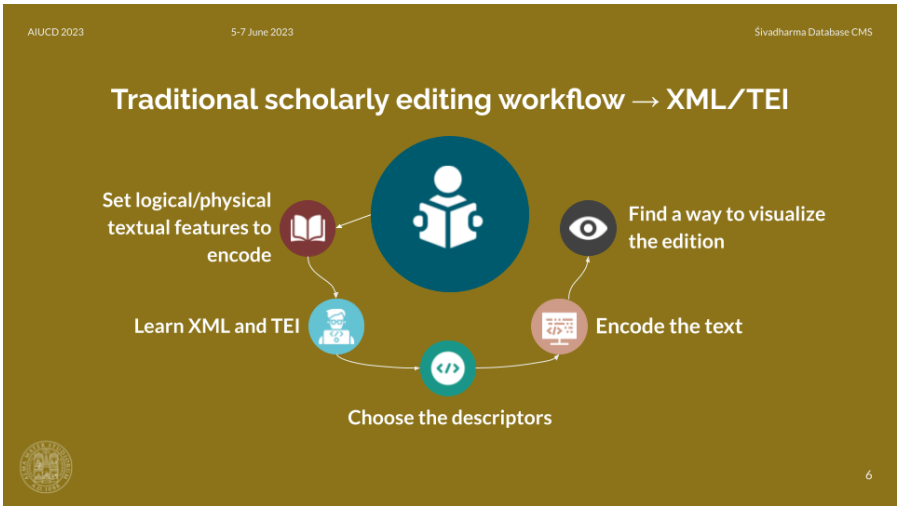
creator and the software that manages the content. Delivering content to websites in a non-automated environment can be tedious and time-consuming, but with the aid of a CMS, web pages can be dynamically generated from the content repository, significantly streamlining the process.

4.2 A Digital Scholarly Editing CMS as a solution to streamline philologists' work (and perhaps make it more rewarding)

Digital Scholarly Editing tools' inadequacy has been widely debated in the literature. Robinson (2005) argues that scholars should be able to create digital editions. However, the available tools demand high dedication, discouraging them from undertaking this task and diverting from their scholarly pursuits to focus on technical aspects instead. Schmidt (2018) and Pierazzo (2019, 212) also share the same concerns about the technical expertise required in the editing workflow. Zenzaro *et al.* (2022, 20) add that philologists frequently regard the shift from *traditional* to *digital* as a burden and the existing technological resources as not only insufficient, but also exasperating. It impedes the integration of knowledge, methodologies, and tools. To address this, they advocate for software, tools, and services that require minimal technical skills.

The Digital Scholarly Editing technological challenges should not be overlooked. In the editions' preparatory phase, acquiring knowledge of modelling patterns and coding languages can be a significant time and effort investment, especially for non-technical scholars. During the editing process, this leads to additional expenses to choose the best formal method of annotating textual features and apply it to texts, and distracts from addressing philological questions (Hagel 2007, 79). One may contend that mandating a subset of the TEI guidelines they can quickly grasp to achieve the desired edition visualisation would suffice to reduce the technological expertise. However, is it rewarding for a philologist to spend hours, days, and months marking texts by hand?

To address these issues, a single environment runnable on the web allowing editors to create editions via an interface without installations and coding, such as a CMS, would streamline the editing process and make it more rewarding (Fig. 13).



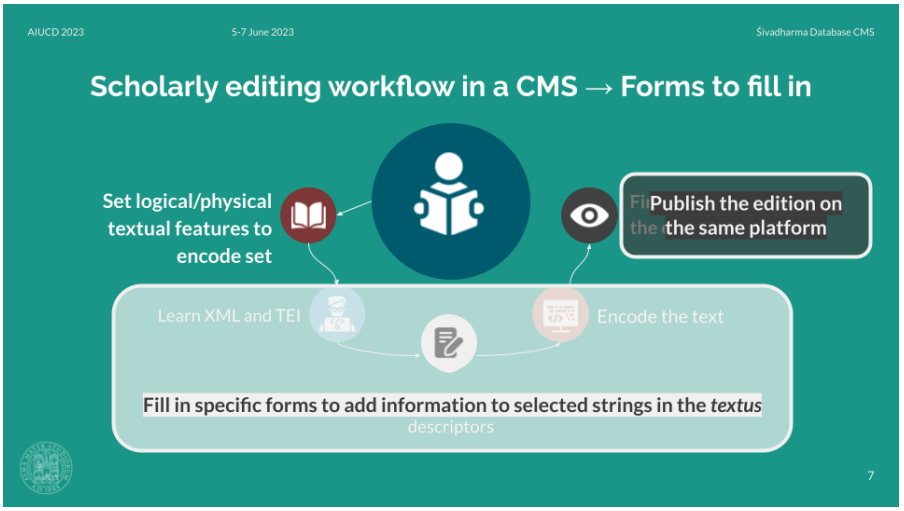


Fig. 13. Digital Scholarly Editing traditional vs. CMS workflow.[67]

4.3 CMS solution to fill the gap between the input and the output sources

Another gap related to Digital Scholarly Editing technologies is the mismatch between input (e.g., the XML/TEI file) and output (e.g., the HTML rendering), which can lead to misunderstandings between the editors and the users (Hagel 2007, 79).

Consider the XML/TEI-based tree edition development workflow as an example.[68] Initially, the edition is encoded in XML/TEI (Listing. 6). Subsequently, the XML/TEI file is processed by an XSLT script to obtain a conversion from XML to HTML (Listing. 7), eventually leading to a visualisation (Fig. 14).

[67] Excerpt from the AIUCD 2023 presentation *Śivadharma Database CMS. HTML and graph as a starting point for digital editions* by Martina Dello Buono and Francesca Tomasi: https://doi.org/10.6084/m9.figshare.24433486.v1.

[68] The example in Fig. is from the sample edition of *Śivadharmottara Digital Scholarly Edition*, currently edited by Florinda De Simini: https://github.com/martinadellobuono/shivadharma-xslt-base-txt.

```
<!-- stanza 1 -->
<div type="section" n="1">
  <lg met="Vaṃśastha">
    <gap>
      <app xml:id="app-1-1-1-1" resp="#flodes">
        <lem wit="#NK28 #NKo77 #P" rend="">
          <gap></gap>
        </lem>
        <rdg wit="#PT75" rend="">
          <lg met="">
            <l n="a">śivam astu sarvajagatāṃ</l>
            <l n="b">parahitaniratā bhavantu bhūtaguṇāḥ</l>
            <l n="c">doṣāḥ prayāntu śāntiṃ</l>
            <l n="d">sarvatra sukhībhavantu sakalalokāḥ</l>
          </lg>
        </rdg>
      </app>
    </gap>
    <l n="a">
      namo &#8217;stu tasmai śakalendu<app xml:id="app-1-1-1a-1" resp="#flodes">
      <lem wit="#NK82 #NO15 #E" rend="circlefront">dhāriṇe</lem>
      <rdg wit="#NC45" rend="">dhāraṇe</rdg>
    </app>
  </l>
  <l n="b">
      phaṇīndraratnadyutikaṇṭha<app xml:id="app-1-1-1b-1" resp="#flodes">
    <lem wit="#NC45 #NO15 #E" rend="circlefront">rāgiṇe</lem>
    <rdg wit="#NK82" rend="">dhāriṇeṃ</rdg>
  </app>
```

Listing 6. Sample XML/TEI encoding of an edition. Focus on `<section>` element.

```
<!-- section -->
<xsl:template match="tei:div[@type='section']">
    <div class="row px-4 py-3" data-type="{@type}">
        <xsl:apply-templates select="@* | node()[not(self::tei:translation)]"/>
    </div>
</xsl:template>
```

Listing 7. Sample XSLT script to convert the edition's XML/TEI file into HTML. Focus on `<section>` element.

namo 'stu tasmai śakalendudhāriṇe
phaṇīndraratnadyutikaṇṭharāgiṇe |
harāya śubhrābhrakapālamāline
vibhinnadaityasphuritaikaśūline || 1. 1 || om. app. ▾

Fig. 14. Final visualisation of the `<section>` element.

It is evident from Fig. 14 that the final visualisation of the edition is graphically distinct from the encoding and the XSLT script. As a result, it becomes challenging to maintain control over both the code and visualisation simultaneously, despite the possibility of obtaining an in-progress preview. It requires experience and skill to ensure that the results returned by the code files are in sync with the edition's desired visualisation.

4.4 CMS solution to publish editions without extra investments

Schmidt (2018) identifies a CMS-like environment as one of the possible solutions to overcome technically inexpert scholars' challenges in creating but also *sharing* editions. Pierazzo (2019, 209) notes that a framework for publishing and not only creating Digital Scholarly Editions would help address the significant investments required for their production.

A CMS comprehensively addresses this requirement by facilitating the creation and publishing of content on a unified platform (Ch. 4.1).

4.5 CMS solution to obtain consistent data and better error handling

A CMS software accomplishes multiple tasks related to data production, reading, and processing (Cinquilli *et al.* 2012, 1). Developing a CMS software that accepts a uniform data format based on a singular model and a singular tool for data entry, such as a *form*, can lead to the acquisition of structurally homogeneous data, thereby offering greater control in the event of errors. If editors were allowed to structure the data autonomously, for instance, through the TEI standard, there would be less control over the consistency of the data as it provides different encoding methods and terminology not exempt from synonymy (Ch. 3.3.1). Hence, it increases the likelihood of errors.

4.6 Digital Scholarly Editing CMS state-of-the-art

The academic community is increasingly searching for scholarly tools that offer user-friendly options for text encoding, which are actually rare. While some of these tools conform to conventional XML/TEI annotation systems to ensure technology popularity and reusability (Ch. 3.3.1), others explore novel technologies like RDF and noSQL databases (Ch. 3.3.3, 3.3.5). A brief analysis of their pros and cons is reported below.

4.6.1 Annotations in XML/TEI: Edition Visualization Technology (EVT)

The Edition Visualization Technology (EVT)[69] is currently one of the most commonly employed tools in the Digital Scholarly Editing domain.[70] While it finds its foundation in XML/TEI, it functions primarily as a digital edition *viewer* rather than an *editor*. Upon input of an XML/TEI file, EVT returns a visual interface of the edition, including two edition levels, i.e., diplomatic and normalised transcriptions, links between the text and its facsimiles, hot-spots of specific

[69] EVT: http://evt.labcd.unipi.it/.

[70] At least, in Italy. See RIDE Issue 15 for an updated list of Digital Scholarly Editing tools: https://ride.i-d-e.de/issues/issue-15/.

manuscripts areas, manuscripts structure, search engine, and more (Rosselli Del Turco 2019, 92). The basic EVT functionalities find their expression in the pilot EVT project, namely *Digital Vercelli Book*,[71] a diplomatic/diplomatic-interpretative digital edition (Fig. 15) (Ch. 1.2).

Fig. 15. *Digital Vercelli Book* Digital Edition. EVT visualisation.

Advanced features like critical *apparatus* and parallel texts, are available in *Edizione Logica Avicennae* [experimental encoding] (Fig. 16).[72] The availability of diverse features is a response to the requirement for a certain degree of program flexibility, which is essential for accomplishing customised editorial work.

[71] Digital Vercelli Book: http://evt.labcd.unipi.it/demo/evt_v1-3/dotr/.

[72] Edizione Logica Avicennae [experimental encoding]: http://evt.labcd.unipi.it/demo/evt2-beta2/avicenna/index.html#/readingTxt?d=doc_1&p=C-112v&s=text-body-div&e=critical.

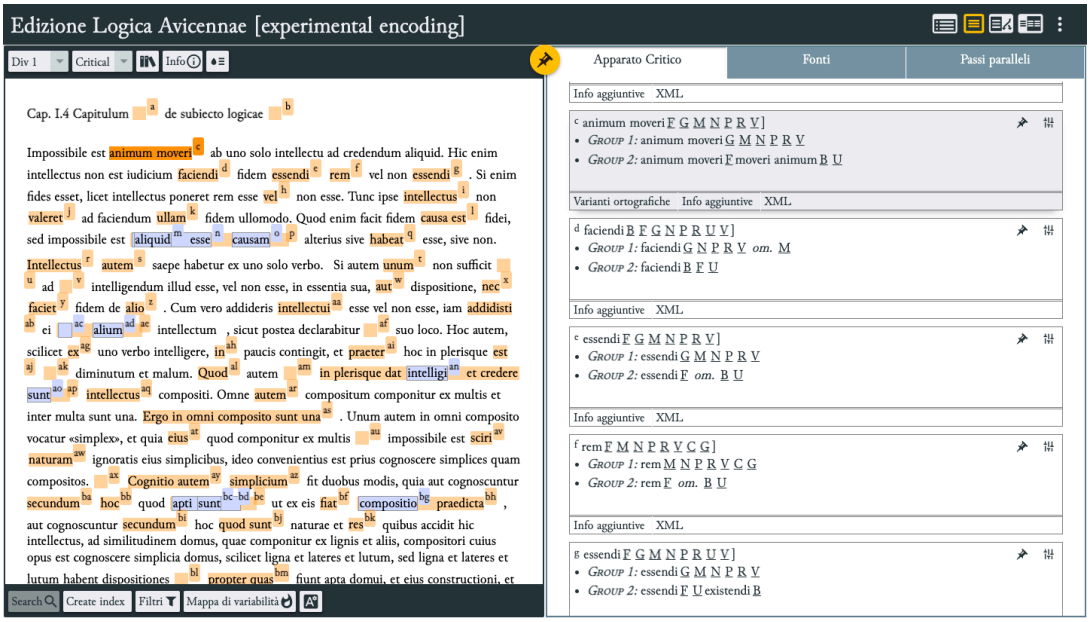


Fig. 16. *Edizione Logica Avicennae* [experimental encoding]. EVT visualisation.

Gaining an understanding of why EVT was developed is of utmost interest. It provides insight into how the community perceives the shortage of tools available for digital editions. This includes issues such as the reliance on proprietary software, limited compatibility with certain platforms, complicated installation and usage procedures, and inability to adapt the available tools to projects beyond their original context. EVT is a practical solution to this insufficiency. However, the main drawback of this system is that scholars need to possess prior knowledge of the specific XML/TEI schema to visualise their editions.

4.6.2 Annotations in semantic technologies: KeyWords In Context, KeyWord Out of Context, KeyWord Alongside Context (KWIC KWOC KWAC)

RDF is the underlying framework of rare annotation systems. One system leveraging RDFa is KWIC KWOC KWAC (KeyWords In Context, KeyWord Out of Context, KeyWord Alongside Context)[73] (Fig. 17), which offers a user-friendly annotation mechanism.

[73] KWIC KWOC KWAC software: https://github.com/sanofrank/KwicKwocKwac.
KWIC KWOC KWAC documentation: https://aldomorodigitale.unibo.it/about/docs/processing#structure-section.

Fig. 17. KWIC KWOC KWAC editor.[74]

The *Edizione Nazionale delle Opere di Aldo Moro* project demonstrates the effectiveness of KWIC KWOC KWAC in enabling the annotation of entities like individuals and geographic locations in full-text documents. Nonetheless, the incorporation of annotations into texts may compromise their legibility.

4.6.3 Stand-off annotations stored in a graph database: SPEEDy

SPEEDy (Fig. 18) is a web-based editor that allows editors to annotate textual strings corresponding to people, places and events in a user-friendly environment.

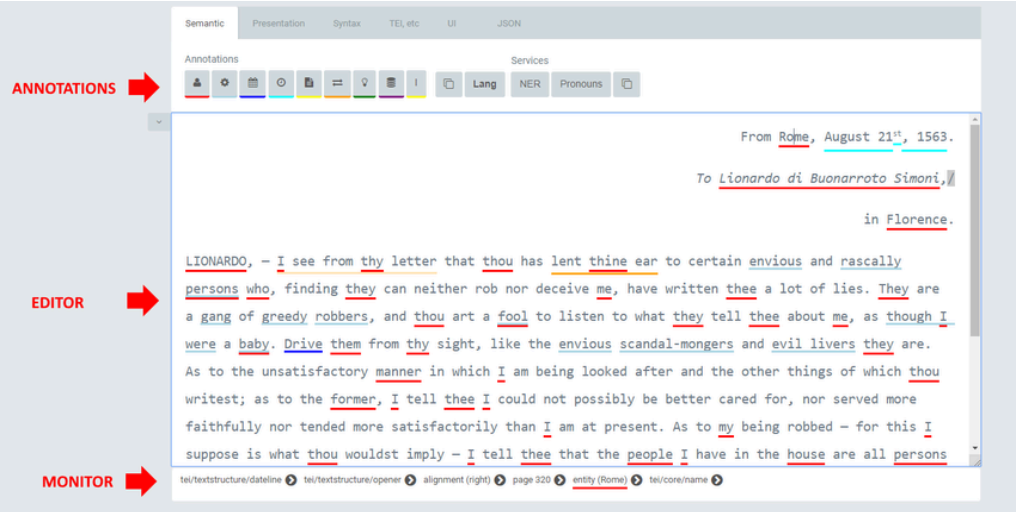


Fig. 18. SPEEDy editor (Niell and Schmidt 2021, 50).

[74] The image is shown for illustrative purposes only. It is taken from the documentation of the *Edizione Nazionale delle Opere di Aldo Moro* project: https://aldomorodigitale.unibo.it/about/docs/processing#addmetadata-section.

It differs from KWIC KWOC KWAC in the technological choices behind its implementation. Specifically, it records the textual annotations as *stand-off* properties in Neo4j, a graph database, so that it can handle complex and overlapped structures elegantly. The annotation system is characterised by its clarity and transparency. In particular, the annotated sequences of characters are encapsulated within `<span>` elements, with attributes and assigned values corresponding to their properties (Neill and Schmidt 2021, 47).

Despite considerable flexibility, this annotation system has limitations as it does not support the annotation of more complex structures such as the *apparatus*.

4.6.4 Other types of annotations: Pundit, Classical Text Editor and CoPhiEditor

Much like KWIC KWOC KWAC and SPEEDy, Pundit[75] provides a user-friendly interface for highlighting and commenting on web pages (Fig. 19). Even though it can be highly functional in reducing the technical skills required, its licence is not open and therefore not freely usable.

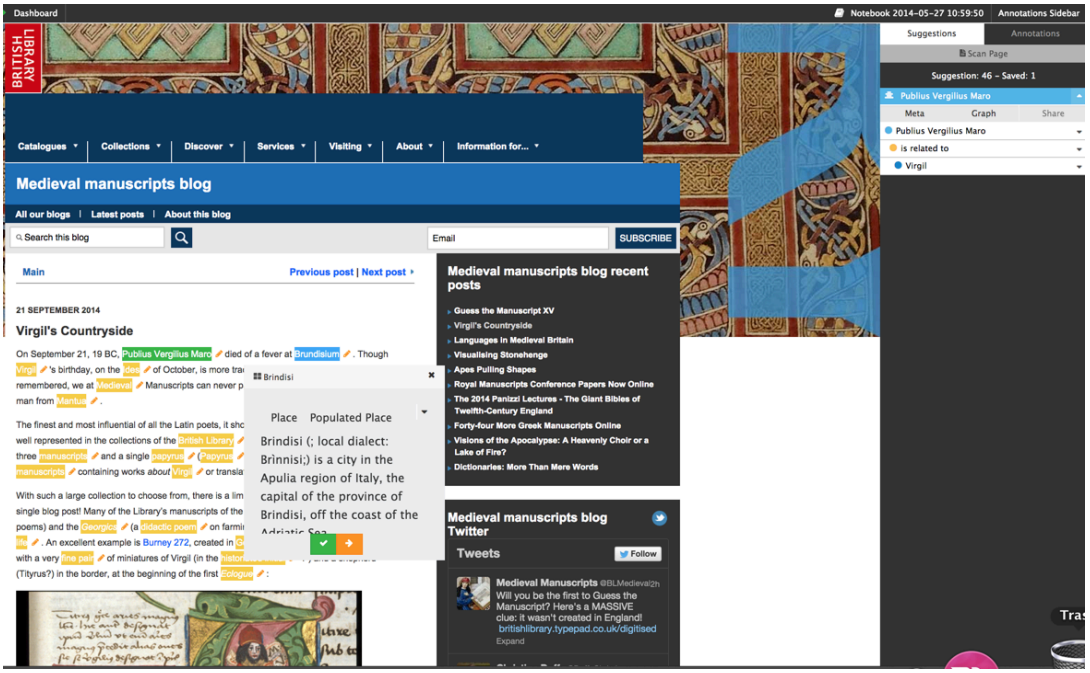


Fig. 19. Pundit editor.[76]

[75] Pundit: https://www.netseven.it/pundit/.

[76] The image shown for illustrative purposes only. Is taken from: https://www.digitalmeetsculture.net/article/pundit-for-digital-humanities-by-net7/. All rights reserved.

An example of using Pundit can be found in the *Burckhardt Source*[77] project, which is a digital collection of letters written by and to Jacob Burckhardt. The project utilises annotations to reference individuals, places, artwork, and bibliographies mentioned in the letters (Di Donato and Müller 2014). In addition, it employs semantic technologies to produce semantic annotations and link them to external resources such as Wikipedia entities. Although KWIC KWOC KWAC, SPEEDy, and Pundit do not require coding skills, their annotation systems do not support the creation of complex edition components like the *apparatus*. The Classical Text Editor[78] environment (Fig. 20), instead, fills this gap. It provides advanced features such as functionalities to create the *apparatus*, notes, and parallels. Additionally, it enables XML/TEI export for interoperability.

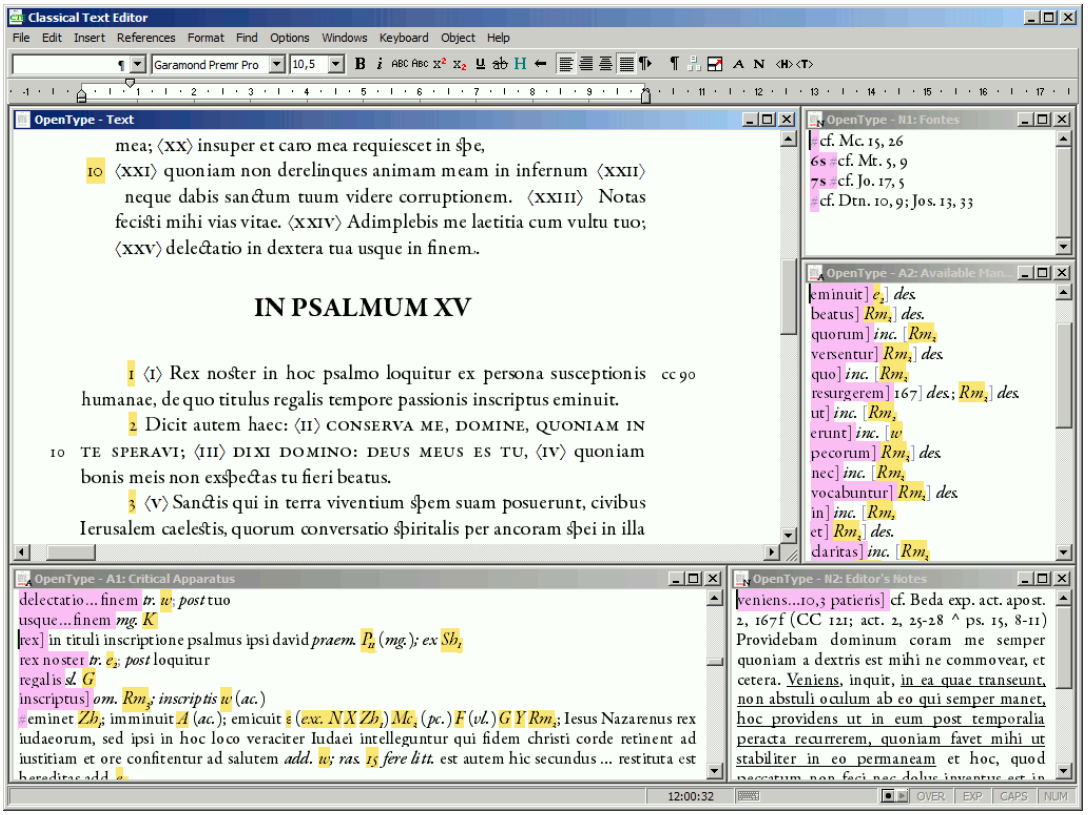


Fig. 20. Classical Text Editor.[79]

Nevertheless, just like Pundit, the licence for the Classical Text Editor does not permit its widespread usage in the community. Moreover, it is primarily focused on printed editions.

Finally, CoPhiEditor (Fig. 21) is a web platform that employs a distinct set of technological tools for accomplishing similar tasks, i.e., creating papyrological Scholarly Digital Editions.

[77] Burckhardt Source: https://burckhardtsource.org/.

[78] Classical Text Editor: https://cte.oeaw.ac.at/.

[79] The image shown for illustrative purposes only. Is taken from: https://cte.oeaw.ac.at/. All rights reserved.

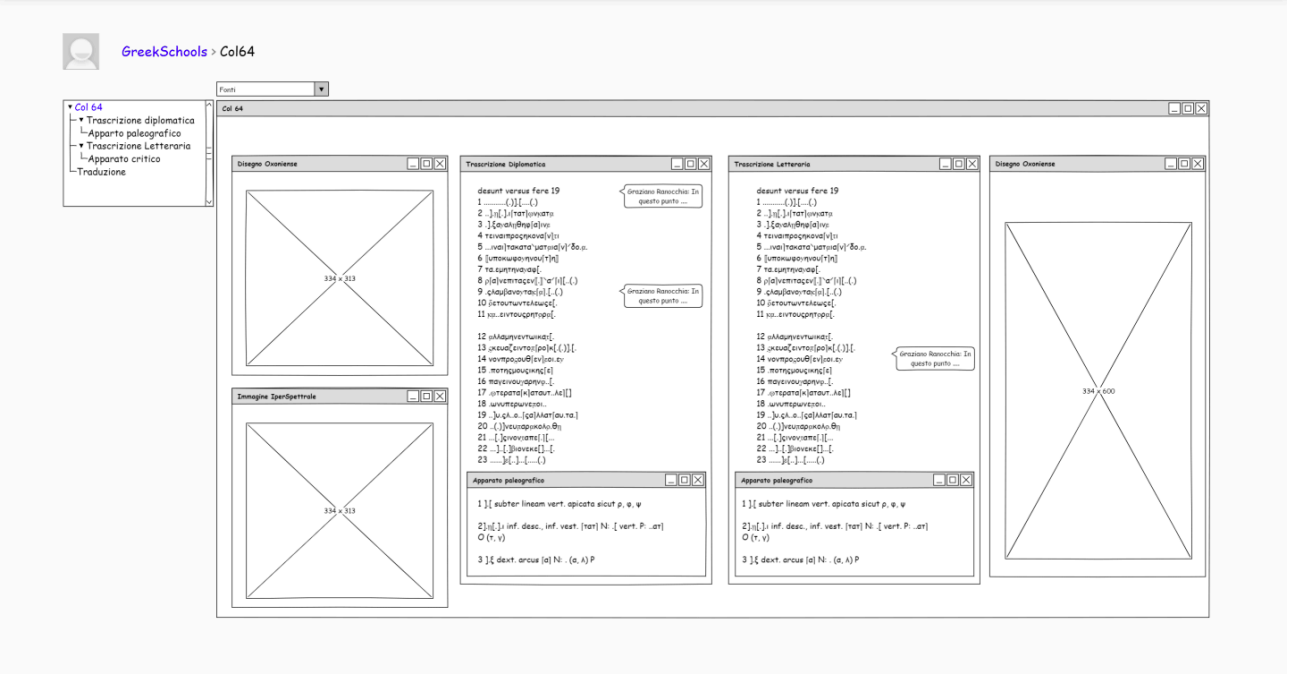


Fig. 21. CoPhiEditor wireframe (Zenzaro *et al.* 2022, 24).

The platform is built on a Domain-Specific Language (DSL) that is specifically designed for a limited scope of knowledge or activity. This technological choice is motivated by the fact that general-purpose metalanguages such as XML or languages like Python are not familiar to specialists in the Humanities. These language constructs and lexicons distinctly differ from natural language or formalisms that are familiar to these specialists. As a result, encoding digital editions significantly differs from traditional practices, e.g., preparing an *apparatus* in XML/TEI or writing it step-by-step in a document. One of the most significant benefits of a DSL is the familiarity with the formalisms used within a domain that has already optimised the representation of the information related to their objects of study. However, a significant drawback of using a DSL is its limited applicability outside the specific domain for which it was designed. In the case of CoPhiEditor, the DSL is tailored to the requirements of papyrological Scholarly Digital Editions, which means that it may only be suitable for other areas with significant redesign and development efforts.

4.7 Does digital philologists remain *digital philologists* if they use a CMS to prepare editions?

Preparing digital editions can be challenging for philologists, especially regarding the technical aspects of encoding and visualising the text. To alleviate this workload, a framework that makes it easier to create digital editions by reducing the required technical skill set would be beneficial. However, this framework should not oversimplify the process. Such a framework can only partially

resolve the challenges of philology, be it traditional or digital. It can serve as an *instrument* for philologists to streamline and improve their work in a digital environment. However, at the same time and in any way, they must undertake the traditional process of preparing an edition, starting from manuscripts and culminating in a *textus*. According to Leonardi (2021, 79), digital philology and its tools must effectively accommodate both digital and traditional paradigms. While traditional paradigms are embodied in the tools of analogue philology, the digital context can enhance these paradigms even further. Therefore, digital philology must balance conforming to digital paradigms while still retaining the best practices of traditional philology.

It is important to note that a CMS does not negate the need for digital skills that a digital philologist must possess. These skills include modelling, technical, and designing competencies. They are in a *digital* environment, even if it enables the content creation without coding expertise. Philologists must model the edition's components and contents and determine the operations to make available to the user. A CMS could present numerous options, and philologists must choose the appropriate solutions for their research objectives. In other words, a CMS relieves them of purely technical tasks, such as learning coding languages, guidelines, visualisation methods, server setups, and library installations, allowing them to focus on their work as digital philologists rather than developers. On the other hand, philologists must learn how to use this technology appropriately and functionally to achieve their research objectives. This is where the *digital* philologist comes into play, bridging the gap between traditional philology and digital tools. In addition, the customisation and extension of the CMS functionalities may initiate from the specific requirements of philologists for their digital projects.[80]

[80] This is what occurred in the case of EVT (Ch. 4.6.1): «Requests for modification of existing features, and implementation of new features hitherto not taken into account because they were not useful or relevant [...] have been the main source of innovation and evolution of this tool [...] The exchange of ideas and views with other scholars engaged in the preparation of editions of different kinds of works has helped us enormously in the development of EVT resulting in a much better product.» (Rosselli Del Turco 2019, 94).

Chapter 5

Textual scholarship in South-Asian tradition and the *Śivadharma corpus*: a case study

The discipline of textual scholarship in the South-Asian tradition is incredibly rich and multifaceted. South Asia boasts an extensive literary heritage, from ancient oral traditions to preserved written texts (Ch. 5.1). Even if, as Bausi (2008, 22) notes, there is no specific contribution to the history of the editorial technique of Oriental texts, there is no doubt that the philological practices described in Ch. 1, developed and mainly referred to the Classical Studies, including the New Testament and Romance Studies, have been adapted and modelled to the specific features of Oriental texts. Designing Digital Scholarly Editions of South-Asian texts and a system to manage the workflow of their creation could be a challenge and a test case due to their complexity. By first examining significant phenomena and textual features of a collection of Sanskrit texts commonly referred to as the *Śivadharma corpus* as a case study, along with the philological methodologies employed in this domain, we can facilitate the comprehension of how to design a Digital Scholarly Editing CMS, i.e., *Śivadharma Database*, enabling scholars to produce, publish, and update Digital Scholarly Editions directly via an interface (Ch. 6).

5.1 Between oral tradition and written texts

Since religion holds a position of absolute prominence in India, many literary works are not only *literature* but *sacred writings* often revealed by divine entities and transmitted to men by other men endowed with supernatural powers. Until the classical period,[81] three significant religions spread in India: (1) *Brahmanism*, from which Hinduism originated; (2) *Jainism*; and (3) *Buddhism*. These three religious currents gave rise to a rich literature that stands out for its antiquity, extent, and importance (Boccali, Piano and Sani 2000, 5).

The most archaic literature is the Vedic literature, at the foundation of *Brahmanism*. As far as the writing language is concerned, we cannot speak about *Sanskrit* yet. The Vedic texts are in *ancient Indian*, often referred to by the term *Vedic*. Sanskrit refers to a later stage of the language compared to the period in which these texts fall, i.e., the chronological arch from the 15th to the 5th BC. It is an *artificial* language since it did not undergo a natural evolution but resulted from an extended

[81] In Indian Literature, the classical period is not a phase delimited in a specific chronological span but a literary movement born around the 1st century BC and developed until the 13th century AD.

elaboration of the *Vedic* by Indian grammarians. We cannot even speak about *writing* yet. Let us start with the composition of the Vedic texts. The *Veda* is the sacred knowledge par excellence from which all Indian religious thoughts depart. The Vedic texts originate from a divine revelation to the *ṛṣi*, i.e., mythical seers considered their authors. In later years, tradition no longer envisages an actual *composition*. They began to believe that these texts were not *composed* but *seen* supernaturally following a series of penitences inflicted on men and then communicated to other men in the form of words. It follows that no actual *originals* of these texts exist since (1) we do not know exactly their sources, (2) there are no specific authors, and (3) they were handed down orally. The content of such texts leads to a kind of oral rather than written transmission. It mainly consists of religious precepts related to legends and mythical figures, whose most natural means of transmission is word of mouth. Not for nothing, Indians preserved the sacred texts through the memory of ministers-poets who recounted them from generation to generation for a long time. Like anything that depends solely on memory, it is impossible to trace back to the *original*, but several versions rise due to constant variations caused by memory. The texts began to be written down for practical legal purposes around the 7th BC. However, we do not know the related practices (Boccali, Piano and Sani 2000, 7–9).

Alongside the *Vedas*, an immense heritage of works, on the one hand, transmit the ancient Brahamanic truth professed in the *Veda*, i.e., the texts illustrating the *dharma*,[82] including the *Itihāsa*, the *Purāna*, and the *Śastra*, while, on the other hand, they disseminate a new *revelation*, i.e., *Tantra* texts. Even many of the pillar works of such a current are not attributable to specific authors due to their *divine* origin. For instance, the authorship of the epic work *Mahābhārata*[83] included in the *Itihāsa*, i.e., the main Brahmanic works portraying the behaviour of the righteous and the wicked containing numerous didactic passages, is uncertain. Tradition says that the redactor is Vyāsa, who, according to Hindu beliefs, will later become the incarnation of the deity Viṣṇu to rearrange the *Vedas* and other sacred writings (Sutton 2020, ix). Some scholars think instead that there is not a single author of this enormous work. It resulted from a composite collection of pieces dating back to diverse times generically put together in an epic form (Hopkins 1901, 398–402). Still, others claim that a redactor did this final rearrangement (Hacker 1978, 170; Van Buitenen and

[82] As defined by Boccali, Piano, and Sani (2000, 125), the *dharma* is to be intended as the «ordine cosmico e norma dell'agire» [cosmic order and law regulating human activity].

[83] Western writers designed the *Mahābhārata*, together with the *Rāmāyaṇa*, as *epic* referring to a work narrating heroic gestures. In the Hindu tradition, instead, especially *Mahābhārata* is considered primarily as a religious text with the status of scripture (Sutton 2020, ix). The text of *Mahābhārata* itself defines its character as extremely sacred: «Those who learn even one verse of the holy recitation of *Bhārata* are purified from all their sins» (*MBh* 1, 1, 254) (Boccali, Piano and Sani 2000, 166).

Fitzgerald 1973, xvi; Biardeau 1986, 27). It is also impossible to date this work. Boccali, Piano, and Sani state that the most archaic parts date back to no earlier than 400 BC, while they probably reached their current form around 400 AD (Boccali, Piano and Sani 2000, 166).

It is the same fate for the *Purāṇa*s, texts of which it is impossible to trace their originals. They were a series of heterogeneous materials adapted according to the reference community[84] and their geographical location. They were handed down orally by the so-called *sūta*, through which women and *śūdra*, i.e., servants and men doing menial jobs, were educated. Their *divine* nature and target audience were well suited to an oral transmission: they told about myths and legends regarding gods and seers, histories of the origin of sacred places, and more, and often had their autonomy. Only around the 4th and 5th centuries did Brahmins start to write them down (Boccali, Piano and Sani 2000, 219–220).

In some cases, instead, texts were written down but lost, such as the Canon of the 24th *tīrthaṃkara*,[85] a text belonging to the so-called *canonical literature*, reference for the Jainist religion. Also, thanks to the genres, i.e., doctrinal essays, biographies, allegorical tales, lyrical poems, and scientific treatises, texts written in Sanskrit belonging to the so-called *exegetical literature* developed around the Canon began to appear around the 7th–12th centuries. Proof that they were in written form, in addition to the use of Sanskrit in some cases, is the fact that we know their authors, e.g., Umāsvāti and Siddhasena Divākara in the 7th century, Haribhadra in the 8th century, and Hemacandra in the 12th century (Boccali, Piano and Sani 2000, 66–77).

In still other cases, texts transmitted orally and, more recently, in written form come down to the present day through translations, e.g., Buddhist literature[86] (Boccali, Piano and Sani 2000, 81).

Regarding classical literature, we observe a significant shift in perspective. The genesis of texts in the *kāvya*, term used to define the classical literature, laid in their written form, which was subsequently intended for public consumption through the means of reading, i.e., *śrāvya*, or recitation, i.e., *dṛśya*. Such literary works may exist in standalone form or as part of an anthology. Frequently, the authorship of literary works is discernible, albeit occasionally, the attributions presented by the anthologists are not entirely trustworthy (Boccali, Piano and Sani 2000, 399–400).

[84] The audience consisted exclusively of these two social categories, i.e., women and *śūdra*.

[85] According to the Jain religion, a *tīrthaṃkara* was the one who preached the doctrine in a specific age.

[86] Although Buddhism originated in India, it spread throughout most Southeast Asia, including Nepal, Tibet, Indochina, Indonesia, China, and Japan. However, India went in the opposite direction. Buddhism gradually disappeared, absorbed by other religions, i.e., Śivaism and Vaishnavism. The flourishing of Buddhism in these regions is why written texts developed in southeastern Asia rather than in India.

5.2 *Śivadharma corpus*

The study of the *Śivadharma corpus* is relevant for defining a standard methodology of textual criticism to the critical editing of texts transmitted in different South-Asian regions whose manuscripts were carriers of texts and supports of worship.

Such *corpus* is a collection of ca. 85 specimens from the second half of the 11th century onwards. Translocality is one of the notable *Śivadharma corpus* features.[87] Although the attestations are predominantly Nepalese, the *Śivadharma* tradition is widely attested in several regions, e.g., Kashmir, Bengal, and Tamil Nadu. It implies that such texts were studied and transmitted in areas of different languages and manuscript traditions (De Simini 2017, 506). The corpus is comprised of eight Sanskrit anonymous texts which describe in the form of dialogues practices, rituals, and behaviours of lay Śaiva householders (De Simini and Mirnig 2017, 587) addressed to Śaiva devotees (De Simini 2016, 234). The texts are the following: (1) *Śivadharmaśāstra*,[88] (2) *Śivadharmottara*,[89] (3) *Śivadharmasaṃgraha*,[90] (4) *Śivopaniṣad*,[91] (5) *Umāmaheśvarasaṃvāda*,[92] (6) *Uttarottaramahāsaṃvāda*,[93] (7) *Vṛṣasārasaṃgraha*,[94] and (8) *Dharmaputrikā*.[95] The *corpus* probably included a ninth text, the so-called *Lalitavistara*, but was so unsuccessful that it was not transmitted in the later tradition (De Simini and Mirnig 2017, 588). According to Kafle (2021, 235), it consisted of an early draft of the *Umāmaheśvarasaṃvāda*, even if a specific manuscript, i.e., N_{77}^{Ko}, reports them both.

The earliest works are *Śivadharmaśāstra* and *Śivadharmottara*, which reached Nepal between the 7th and 9th centuries, but we do not have detailed information about their author(s), and their composition time. Direct and indirect evidence, i.e., references and quotations in later literary and epigraphic sources, allows us to affirm that these two works were well-known in India in the mediaeval time, probably contributing to the early development of Śaivism (Bisschop, Kafle and

[87] As De Simini (2017, 506) points out, this statement is valid for two works in particular, i.e., the earliest *Śivadharmaśāstra* and *Śivadharmottara*, which were studied also outside Nepal. The remaining six works were found exclusively in Nepal with rare exceptions.

[88] «Treatise on Śaiva Religion».

[89] «Continuation [of the Treatise] on Śaiva Religion».

[90] «Compendium of Śaiva Religion».

[91] «Essential Teachings of Śaiva».

[92] «Dialogue Between Umā and the Great Lord».

[93] «Great Dialogue [Made of] questions and answers».

[94] «Compendium of the Essence of the Bull [of Dharma]».

[95] «Daughter of Dharma».

Lubin 2021, 4).[96] On the other hand, only one case attests to the two works together, i.e., Bengali manuscripts. However, other attestations quoting them both make us think that they were culturally and traditionally connected (De Simini 2016, 236–237).

It follows the transmission of the other works exclusively in Nepal, probably as a response to the popularity of Śaivism in this region during the Middle Ages (De Simini 2016, 239). However, we do not know where they were composed. Following De Simini and Mirnig's hypothesis (2017, 590), such different works related to the same topics took the shape of a homogeneous *corpus* starting from the 12th century and transmitted in Nepalese multiple-text manuscripts (MTMs) (De Simini 2016, 233).

5.3 *Śivadharma* context of production

The manuscripts of the *Śivadharma collection* originated in the Nepalese religious environment. The purposes of their production were various. First, it was intended to transmit the rules of conduct in the context of rituals and religious institutions. As part of the literature on *dharma* (Ch. 5.1), they share its normative character (Boccali, Piano and Sani 2000, 276). They constituted a doctrinal system in practice for those who adhered to the Śivaist religion (De Simini and Mirnig 2017, 587–588). On the other hand, copying and gifting the manuscripts by the devotees allowed them to accumulate merits. In addition, testimonies report that they were recited during the rites of appeasement and read publicly in celebrations. It testifies to how alive was Śaivism in Nepal during the Middle Ages (De Simini 2016, 268–270).

5.4 Nepalese MTMs of the *Śivadharma corpus*

The *Śivadharma corpus*, as described in Ch. 5.2, corresponds to a fixed set of eight works attested as a *corpus* in Nepalese ancient palm-leaf (Fig. 22) and more recent paper MTMs. The most attested format is the so-called *pothi*, palm-leaf manuscripts where the minimal codicological units are loose leaves. There is no binding, but leaves are piled on each other, and removable strings and upper and lower covers keep them together (Fig. 23, 24).

[96] For a definition of *Śaivism*, see (Sanderson 1988, 660): "The term *Śaivism* here refers to a number of distinct but historically related systems comprising theology, ritual, observance and yoga, which have been propagated in India as the teachings of the Hindu deity Śiva. A *Śaiva* is one who practices such a system.".

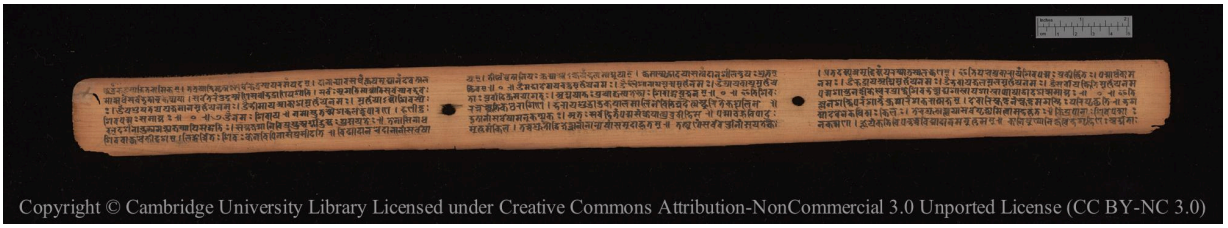


Fig. 22. *Śivadharma* palm-leaf manuscript MS Add.1645. Page: 38r. Contents: *Śivadharmottara*.[97]

Fig. 23. *Śivadharma* manuscript MS Add.1645 wooden cover outside and inside.[98]

Fig. 24. *Śivadharma* manuscript MS Add.1645 wooden back cover inside and outside.[99]

[97] The physical location of MS Add.1645 is the Cambridge University Library. The credit for this image goes to the Cambridge University Library. The image of the page 38r is accessible at the following web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/77. Licence: CC BY-NC 3.0.

[98] The physical location of MS Add.1645 is the Cambridge University Library. The credit for this image goes to the Cambridge University Library. The image of the cover outside can be accessed at the following web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/1. The image of the cover inside is instead available at this web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/2. Licence: CC BY-NC 3.0.

[99] The physical location of MS Add.1645 is the Cambridge University Library. The credit for this image goes to the Cambridge University Library. The image of the cover inside can be accessed at the following web address:

These kinds of *blocks* do not correspond to complete works, but there are hints that the various parts of the manuscripts were independent and sometimes used separately. However, the internal subdivision does not hinder the manuscripts' cohesion and the intent of creating and preserving the texts as a *corpus*. We follow De Simini's examination of the palm-leaf manuscript Add. 1645 held at the University Library of Cambridge (De Simini 2016a, 243–245) (Fig. 22, 23, 24). It dates back to NS 259 (1139–40 CE). It comprises 247 folios attesting all eight canonical works of the *Śivadharma corpus*. They are copied one after the other without blank pages or spaces as dividers. They are divided only by an initial and final heading, except the first work, i.e., *Śivadharmaśāstra*. At the end of *Śivadharmaśāstra*, we find not only the traditional final statement but also *mantras*, i.e., invocations, to Śiva, standing as a barrier between the earliest text and all the others. In this regard, we can define Add. 1645 as a *homogeneous monomerous* (Gumbert 2004, 25), i.e., an unarticulated codicological unit with some internal boundaries. This structure is common to several manuscripts belonging to the *Śivadharma* collection.

5.5 *Śivadharma* tradition complexity

The large number of primary sources is a challenge for the philologists working on the *Śivadharma* witnesses, thus forcing a meticulous *recensio* in order to clearly define the genealogical links between manuscripts (Ch. 1.3.2). This work turns out to be particularly difficult as the entire tradition of the *corpus* is not known. Therefore, as it is impossible to access all the available copies, the current works focus on a selection of manuscripts, which, on the other hand, allows us to make deductions regarding the *Śivadharma* history of tradition. Referring to the analysis conducted by De Simini (2017, 505–547), the 12th and last chapter of *Śivadharmaśāstra* shows the various regional variants in the transmission, i.e., inclusion or omission of specific groups of stanzas and different internal arrangements. The observed alterations are attributable to diverse factors, including (1) subgroups transmitting a specific variant; and (2) scribes or other parties involved in the transmission deliberately modifying the texts. Furthermore, the identifiable links between manuscripts are unstable when the philological analysis extends to other parts of the *corpus*. Depending on the section where the philologist moves, diverse genealogical connections may be found and established. It may hint that there are probably cases of internal horizontal contamination. This suggestion comes from the analysis of the manuscript copying process. Although philologists are currently unable to define the practices related to the copying process, it can be said that scribes' attitude was that of alternating mechanical and non-mechanical practices, thus introducing

https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/497. The image of the cover outside is instead available at this web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/498. Licence: CC BY-NC 3.0.

alterations on purpose. In addition, they also copied works from various manuscripts, making it impossible to transmit a single text.

We must also consider another critical factor which makes it difficult to reconstruct a definite *stemma*. The South-Asian regions' hot and humid climate undermined the preservation of the manuscripts. In this context, Lachmann's method (Ch. 1.3.2) contributes to clarifying specific issues but, on the other hand, it cannot be applied rigorously. What puts it in crisis are the temporal and spatial extension of the diffusion of the *corpus*, the horizontal contamination, and the damages due to the climate. As a result, the choices by the philologist for the reconstruction of a *stemma* are mainly *open* and non-mechanical.

5.6 How does a philologist work when the transmission is complex, as in *Śivadharma* tradition?

The history of the transmission of the *Śivadharma corpus* is not linear and influenced by several factors (Ch. 5.5). So how do philologists behave in a situation of such complexity? De Simini (2017, 540) reports that the predominant philological approach is closer to the post-Lachmannian philologist Pasquali (Ch. 1.3.10). According to Pasquali's point of view, the transmission history strongly affects the reconstruction of *stemma*. Starting from the assumption that in the event of contaminated tradition, as in *Śivadharma*, it is impossible to build a definitive *stemma*, he appeals to what he defines as *open recensio*. Thus, the Sanskritist philologist editing the *Dharma* of Śiva moves among the manuscripts as described by Pasquali. In the *recensio* phase, they collect all the available witnesses.[100] During the *editio* phase, instead, the choice of the most reliable reading is not mechanical or based on fixed criteria, e.g., genealogical *stemma* (Ch. 1.3.2), majority rule (Ch. 1.3.3), Bédier's concept of *bon manuscrit* (Ch. 1.3.9). The *iudicium* is pivotal: the assessment of each reading follows the principles established by the editor, which are based on the history of the tradition. Therefore, a previous step in the workflow is necessary. Before assessing the readings, each manuscript has to be assigned a specific place, which certainly influences the forms of the text at the different stages of its transmission. The combination of the genealogical method and the *open recensio* method is thus the key to studying the complex transmission of the *Śivadharma corpus*.

5.7 *Śivadharma* textual complexity

The texts of the *Śivadharma collection* are rich in parallels, quotations, and influences from other sources. To explain the complexity of such phenomena, we take five researches by Kiss (2021,

[100] It is not always easy to access *Śivadharma* witnesses. In addition, it is uncertain whether all the manuscripts found are all the existing ones (Ch. 5.5).

183–201), Kafle (2021, 233–253), Bisschop (2021, 157–171), Battistini (2021, 233–253), and Trento (2021, 233–253), comparing some *Śivadharma* works with other works,[101] internal or external to the *corpus*, respectively (1) the *Vṛṣasārasaṃgraha* with the *Mahābhārata* (Ch. 5.7.1), (2) the *Umāmaheśvarasaṃvāda* with the *Uttarottaramahāsaṃvāda* (Ch. 5.7.2), (3) the *Śivadharmaśāstra* with the *Bhaviṣyapurāña* (Ch. 5.7.3), (4) the *Bhikṣāṭanakāvya* and its commentary *Bhāvadīpikā* (Ch. 5.7.4), and (5) the *Śivadharmottara* and its Tamil translation *Civatarumōttaram* (Ch. 5.7.5). They serve as examples to outline the network of connections between the *Śivadharma* and Hindu literature. These examples demonstrate, the works we refer to with the umbrella term *Śivadharma* are not *silos*. On the contrary, they are often subject or object to parallelisms and influence, more or less marked, from several works, internal and external to the *corpus*. It outlines that the tradition evolved around Śiva is never one-sided but composite and open to influences, manipulations, and alterations. Such a tendency to absorb cults, deities, and principles (Bisschop 2018, 1), remarkable at a textual level, places it within an extended and complex network of references.

5.7.1 Influence between works not belonging to the same *corpus*: the *Vṛṣasārasaṃgraha* and the *Mahābhārata*

The *Vṛṣasārasaṃgraha* is a Sanskrit work with twenty-four chapters included in the *Śivadharma corpus* (Ch. 5.2). Later than the cornerstones of the *Śivadharma*, i.e., *Śivadharmaśāstra* and *Śivadharmottara*, its composition can be dated around the 10th century. As the other works in the same collection, the *Vṛṣasārasaṃgraha* concerns *Dharma*, i.e., religious duties.

This work is placed in a position of direct continuation with the *Mahābhārata* starting from its first three chapters. Precisely, the *Vṛṣasārasaṃgraha* opens with a dialogue between Janamejaya and Vaiśampāyana, which takes up the starting scene of the *Mahābhārata* where the same two characters are present.

The narration of *Mahābhārata* starts in a forest frequented by crowds of ascetics and *ṛṣi*, i.e., seers (Ch. 5.1), where Sauti Ugraśravas arrived. At that moment, the Wise Men, particularly Janamejaya, performed a sacrificial rite of twelve divine days,[102] during which they sacrificed some snakes. During this celebration, he heard an extraordinary story about a war for the possession of the whole world narrated by Vaiśampāyana. In turn, Vaiśampāyana had heard it from another great Wise Man

[101] The works mentioned here are described briefly in Ch. 5.1.

[102] Corresponding to twelve human years.

of ancient times, Vyāsa (Ch. 5.1). So, Sauti Ugraśravas decided to tell the story in turn to the *ṛṣi* present there (Boccali, Piano and Sani 2000, 164–166).

The introductory setting of the *Vṛṣasārasaṃgraha* is the same as *Mahābhārata*. Vaiśampāyana recited the whole *Mahābhārata* during a snake sacrifice headed by Janamejaya. However, Janamejaya was not satisfied at the end of the story. He longed to hear a higher teaching on *Dharma* to lead to liberation. That is when Vaiśampāyana related a dialogue between Vigatarāga, who was Viṣṇu in disguise, and Anarthayajña, an ascetic.

This connection triggers a will or, perhaps better, a particular need, that of providing more specific Dharmic teachings than those in the *Mahābhārata*.

5.7.2 Literary parallels between works belonging to the same *corpus*: the *Umāmaheśvarasaṃvāda* and the *Uttarottaramahāsaṃvāda*

The *Umāmaheśvarasaṃvāda* is a Sanskrit work in twenty-one or twenty-two chapters[103] included in the *Śivadharma corpus* (Ch. 5.2).

While in the example mentioned above the *Mahābhārata* inspires the composition of the *Vṛṣasārasaṃgraha*, greater textual parallels link the *Umāmaheśvarasaṃvāda* and the *Uttarottaramahāsaṃvāda*, both belonging to the *Śivadharma corpus* (Ch. 5.2). The example below shows *verbatim* parallelism between the chapter 10 of *Umāmaheśvarasaṃvāda*[104] and the chapter 8 of *Uttarottaramahāsaṃvāda*.[105] We can observe that the parallel is literary, and discrepancies, marked in grey Table 4, are minimal.

Umāmaheśvarasaṃvāda *10.1*	***Uttarottarasaṃvāda*** *8.5cd–6ab* (fol. 21r6–21v1)
devy uvāca |	*umovāca* |
yady evaṃ yugadoṣeṇa	*yady evaṃ yugadoṣeṇa*
naranārījanās tathā |	*naranārījanās tathā* ||
luptalajjā bhaviṣyanti	*luptalajjā bhaviṣyanti*

[103] The transmission of the *Umāmaheśvarasaṃvāda* is disrupted. While one cluster of manuscripts reports its text divided into twenty-one chapters, another cluster in twenty-two (De Simini 2017, 529).

[104] Edited by Nirajan Kafle (2021, 233–253).

[105] *Ibid.*

dharmo yāsyati saṃkṣayam ||	*dharmā yāsyanti saṃkṣayam* |

Table 4. Parallel between two works belonging to the same *corpus*, i.e., chapter 10 of the *Umāmaheśvarasaṃvāda* and chapter 8 of the *Uttarottarasaṃvāda* edited by Nirajan Kafle (2021, 233–253).

The borrowings usually go from *Umāmaheśvarasaṃvāda* to *Uttarottarasaṃvāda*. What happens is a recovery, refinement, or reworking of the main themes of the former work by the latter.

5.7.3 Literary parallels between works belonging to different *corpora*: the *Śivadharmaśāstra* and the *Bhaviṣyapurāña*

The *Śivadharmaśāstra* is the most archaic Sanskrit work included in the *Śivadharma corpus* (Bisschop 2018, 6) (Ch. 5.2). It is divided into twelve chapters. As in the previously analysed case (Ch. 5.7.2), this work shows literary parallels with other works but external to the *corpus*. Specifically, chapter 171 of *Bhaviṣyapurāña*[106] shows a large-scale textual reuse of chapter 11 of *Śivadharmaśāstra*,[107] the so-called *Śivāsramādhyāya*, albeit with some revisions. The most evident revision concerns replacing the name of the god Śiva with the name of the deity Sūrya. This operation is due to the shift of the work's audience. The reference audience was not the Śiva community anymore but the Sūrya devotees to whom Saura teachings were addressed. Revisions are even more complex, concerning more significant retelling of parts of the work and additions without parallels in either text. An example of these revisions[108] is reported below (Table 5).

Śivadharmaśāstra *11*	***Bhaviṣyapurāña*** *1.171*
	ya eṣa dharmaḥ sūryeti *tavākhyāto mayānagha* | *magadharmaḥ sa evoktaḥ* *sarvapāpabhayāpahaḥ* || 2 ||
sarveṣām eva varñānāṃ	*sarveṣām eva varñānāṃ*
śivāśramaniṣevināṃ |	*magadharmaniṣevañam* |
śivadharmāḥ śivenoktāḥ	*magadharmaś ca saṃprokta*

[106] Edited by Bisshop (2021, 157–171).

[107] *Ibid.*

[108] For a more extended version of the parallels from the *Śivadharmaśāstra* 11 to the *Bhaviṣyapurāña* 1.171, see (Bisschop 2021, 157–171).

dharmakāmārthamuktaye || 1 ||	*eteṣāṃ bhavamuktaye* || 3 ||

Table 5. Parallels shared with two works not belonging to the same *corpus*, i.e., chapter 11 of the *Śivadharmaśāstra* and chapter 171 of the *Bhaviṣyapurāña* edited by Peter Bisschop.

As observed in Table 2, the parallel in the *Bhaviṣyapurāña* 1.171 of the *Śivadharmaśāstra* 11 contains several revisions. In particular, we focus on two diverse kinds of discrepancies: (1) additions, e.g., the entire stanza in the *Bhaviṣyapurāña* 1.171, absent in the *Śivadharmaśāstra* 11, and (2) reformulations where the epithet *Śiva* changes to *Magas*, i.e., Sūrya-worshipping Brahmins, e.g., *śivadharmāḥ* > *magadharmaś*.

5.7.4 A base text and a commentary as dependent text: the *Bhikṣāṭanakāvya* and its commentary *Bhāvadīpikā*

The *Bhikṣāṭanakāvya*,[109] also referred to as *Śivabhikṣāṭana* or *Parameśvarabhikṣāṭana*, is a poem authored by Utprekṣāvallabha dated back around the early 13th century.[110] Diverging from the *Śivadharma* works, it is classified under the genre of classical literature, particularly as a lyrical *kāvya* production (Ch. 5.1). While it may not be considered an integral component of the *Śivadharma corpus*, this work, along with its commentary, serves as a valuable supplementary source for understanding certain aspects of the religious and societal developments surrounding the spread of Śaivism throughout South India.

Among the copies of commentary within the *New Catalogus Catalogorum*,[111] it appears that this particular work has been analysed and discussed in an incomplete commentary entitled *Bhāvadīpikā*[112] by a king named Maṅgapa during the 15th or 16th century. This commentary's contents can be attributed to a recurring schematic format within a dense and, at times, interrupted prose, which includes quotations. The structure of each gloss is articulated in (1) an *avatārikā*, i.e., premise and subject of the stanza; (2) a proper commentary to the stanza, including the marked words, their syntactic constructions, synonyms, details about their composition, and explanation of grammatical rules; (3) the *pratīkas*, i.e., «words from the text commented upon» (McAllister 2021, 31). The present commentary is intriguing as it offers a glimpse into the author's personal narrative.

[109] «Poem on the Mendicancy [of Śiva]».

[110] Battistini (2021, 287–291) presents compelling evidence that supports this thesis and delves into the authenticity of the author's name.

[111] Volume 12, 147.

[112] «Lantern on emotions».

Notably, the author starts his commentary with a lengthy verse preamble, delving into his lineage, effectively spinning a thread of human connections rather than literary works.

5.7.5 A base text and a translation as dependent text: the *Śivadharmottara* and its Tamil translation *Civatarumōttaram*

Maṟaiñāṉa Campantar, also known as Vedajñāna or Nigamajñāna, was a teacher and prolific author in Tamil. His existence is believed to have occurred during the 16th century in Chidambaram, South India. He was central in restructuring and organising a collection of ritualistic, social, and theological knowledge regarding the Caiva Cittānam, an early modern religious tradition in Tamil-speaking South India. His masterpiece was the *Civatarumōttaram*, a poetic translation of the *Śivadharmottara* (Ch. 5.2). This literary work was relevant in early modern and contemporary Tamil Śaivism. It was extensively circulated and even cited in other theological and devotional works in Tamil. However, it was not just a mere translation. As Trento noted (2021, 104), the translation process involved a simultaneous synthesis and reorganisation of the Caiva Cittānam tradition, skillfully blended with the lay Śaivism tradition in an organic manner. Additionally, from a stylistic perspective, it was characterised by a fusion of different elements. On one hand, the pedagogical and doctrinal objectives were discernible stylistically. However, on the other hand, a distinct pursuit of literary enjoyment was also evident.

5.8 *Śivadharma* language complexity

Research on *Śivadharma* is still ongoing. For this reason, it has not yet been feasible to conduct a comprehensive linguistic analysis of the *corpus*. Kiss has put forth some linguistic considerations regarding a specific work, namely the *Vṛṣasārasaṃgraha* (Ch. 3.2) (Kiss 2021, 187). In particular, it has been observed that the language used in the *Vṛṣasārasaṃgraha* is peculiar and often deviates from standard norms. The text is distinguished by its irregularities, even at the prosodic level. However, these features cannot be generalised to the entire *corpus*.

Nonetheless, even at a basic level, the intricacies of the Sanskrit language pose a challenge in handling texts, starting with the script. The *Śivadharma* texts are written in *Nāgarī*, i.e., «of the city», also known as *Devanāgarī*, i.e., «divine *Nāgarī*», which is a native script of South Asia that is written from left to right (Fig. 25). Technically, it is an *alphasyllabary*, where each consonant-vowel sequence is considered a unit or *akṣara*, with the vowel corresponding to a mandatory *diacritic* to the consonant. It should be noted that this is not the only script used for Sanskrit. Although the

Devanāgarī is used today for rendering the printed Sanskrit, early manuscripts were typically transcribed in local scripts (Bright 1996, 384).

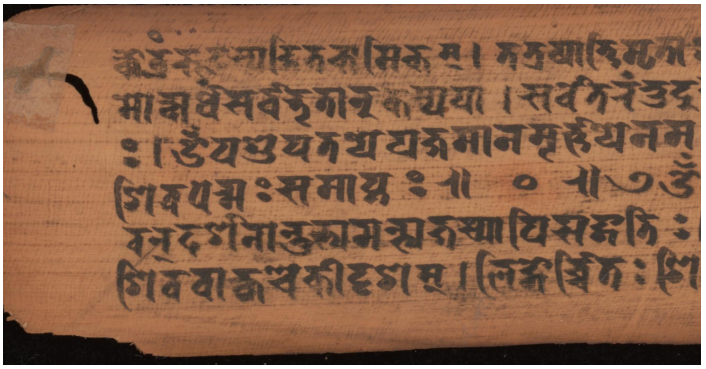


Fig. 25. *Śivadharma* palm-leaf manuscript MS Add.1645. Page: 38r. Contents: *Śivadharmottara*. Zoom on the *Devanāgarī* script.[113]

It is customary to transcribe the *Devanāgarī* using the Latin alphabetical system, which is referred to as *romanised Devanāgarī*, to make Sanskrit texts accessible to the Western audience and vice versa for the South-Asian audience. Table 6 provides an example of transcription of stanza 1 of chapter 1 of *Śivadharmottara*, currently edited by Florinda De Simini (2016a; 2016b; 2017), into *Devanāgarī* and *romanised Devanāgarī*.[114]

Śivadharmottara *1.1 – Devanāgarī*	***Śivadharmottara*** *1.1 – Romanised Devanāgarī*
namo 'stu tasmai śakalendudhāriṇe	नमो स्तु तस्मै शकलेन्दुधारिणे
phaṇīndraratnadyutikaṇṭharāgiṇe |	फणीन्द्ररत्नद्युतिकण्ठरागिणे |
harāya śubhrābhrakapālamāline	हराय शुभ्राभ्रकपालमालिने
vibhinnadaityasphuritaikaśūline || 1 ||	विभिन्नदैत्यस्फुरितैकशूलिने || 1 ||

Table 6. Transcription of stanza 1 of chapter 1 of *Śivadharmottara*, edited by Florinda De Simini, in *Devanāgarī* (on the left) and in *romanised Devanāgarī* (on the right).

[113] The physical location of MS Add.1645 is the Cambridge University Library. The credit for this image goes to the Cambridge University Library. The image of the entire page 38r is accessible at the following web address: https://cudl.lib.cam.ac.uk/view/MS-ADD-01645/77. Licence: CC BY-NC 3.0. See also Ch. 5.4.

[114] The process of converting *romanised Devanāgarī* into *Devanāgarī* has been successfully accomplished with the help of the software tool known as *Sanskrit C(s)onverter*. The credit for developing this application goes to Csaba Kiss, and it can be accessed at the following web address: https://filedn.com/lFSw9FGgUBpyrpsGtImyUHh/converter.html.

Furthermore, there exist certain linguistic phenomena that deserve particular attention, particularly those that relate to specific word combinations. One of the most well-known phonological phenomena is *sandhi*. It involves the assimilation of a sound segment at the end of a word to the first segment in the following word, irrespective of the syntactic phrasing (Kessler 1994, 35). In this way, small words, i.e., *morphemes*, are combined to form compound words (Aralikatte *et al.* 2019). To comprehend the mechanism of this process, an example of *sandhi splitting*, i.e., breaking a compound word into its constituent morphemes (Aralikatte *et al.* 2019), is reported below (Hellwig 2015, 289).

gardabhaścāśvaśca > *gardabha*(*ḥ-c*)(*a-a*)*śva*(*ḥ-c*)*a* > *gardabhaḥ ca aśvaḥ ca*[115]

5.9 *Śivadharma* prosody

The *Śivadharma* works are composed of stanzas containing verses. Each stanza generally consists of four *quarters*, i.e., *pādas*, which are commonly referred to as *pāda* a, *pāda* b, *pāda* c, and *pāda* d. Printed versions of these texts typically present the *pādas* on two or four lines, depending on their length. A mark is placed at the end of each line, i.e., *daṇḍa*, while a double mark, i.e., *daṇḍas*, is used to indicate the end of each stanza to facilitate readability.[116] For an example of *pādas* presented on both four and two lines, reference can be made to chapter 1 of the Scholarly Edition of *Śivadharmottara*, currently edited by Florinda De Simini (2016a; 2016b; 2017) (Table 7).

Śivadharmottara *1.1*	***Śivadharmottara*** *1.2*
namo 'stu tasmai śakalendudhāriṇe *phaṇīndraratnadyutikaṇṭharāgiṇe* | *harāya śubhrābhrakapālamāline* *vibhinnadaityasphuritaikaśūline* || 1 ||	jñānaśaktidharaṃ śāntaṃ kumāraṃ śaṅkarātmajam | devārisūdanaṃ skandam agastiḥ paripṛcchati || 2 ||

Table 7. Stanza 1 and 2 of chapter 1 of *Śivadharmottara*, edited by Florinda De Simini, displayed respectively on four (on the left) and two lines (on the right).

The division into *pādas* of stanza 1 shows that each displayed line corresponds to a *pāda* (Table 8).

[115] «the donkey and the horse and».

[116] It is interesting to note that the *daṇḍa*(s) is the sole punctuation utilised in Sanskrit texts (Ruppel 2017, 33).

Śivadharmottara *1.1*	
namo 'stu tasmai śakalendudhāriṇe	***pāda*** **a**
phaṇīndraratnadyutikaṇṭharāgiṇe |	***pāda*** **b**
harāya śubhrābhrakapālamāline	***pāda*** **c**
vibhinnadaityasphuritaikaśūline || 1 ||	***pāda*** **d**

Table 8. Stanza 1 of chapter 1 of *Śivadharmottara* edited by Florinda De Simini divided into *pādas*.

Instead, the division into *pādas* of stanza 2 shows that each displayed line includes two *pādas* (Table 9). The differentiating factor lies in the line length, with lines exceeding the display limit divided into two equal parts. It is important to note that the ending of a *pāda* may not necessarily coincide with the conclusion of a word.

Śivadharmottara *1.2*	
pāda **a**	***pāda*** **b**
jñānaśaktidharaṃ śāntaṃ	*kumāraṃ śaṅkarātmajam* |
pāda **c**	***pāda*** **d**
devārisūdanaṃ skandam	*agastiḥ paripṛcchati* || 2 ||

Table 9. Stanza 2 of chapter 1 of *Śivadharmottara* edited by Florinda De Simini divided into *pādas*.

Regarding the metre in Sanskrit literature, also known as *chandas*, a categorisation based on the unit of measurement discerns two types of metres: (1) the *akṣaravṛttas*, counted by syllables, and (2) the *mātrāvṛttas*, counted by *morae*. As the metre of *Śivadharma* works falls under the *akṣaravṛttas* category, we shall examine the structure it establishes below.

1 ***Samavṛttas***. All the four *pādas* share the same structure: **a** = **b** = **c** = **d**.
2 ***Ardhasamavṛttas***. The four *pādas* share the same structure alternately: **a** = **c**; **b** = **d**.
3 ***Viṣamavṛttas***. The four *pādas* have a different structure: **a** ≠ **b** ≠ **c** ≠ **d**.

The *samavṛttas* is then further divided into different categories depending on the number of syllables per *pāda*. In the case of *Śivadharma*, the metre is the *anuṣṭubh*, also called *śloka*, which is a Vedic metre consisting of eight syllables per *pāda*, totalling thirty-two syllables (Hahn, 1–3).

Although its system is complex due to its freer form and several subtypes (D'Avella 2019, 1), the metre in *Śivadharma* works is generally consistent. However, a break of metrical structures occurs when reporting clauses mark the change of interlocutor in the dialogue. An example of a break is reported below (Table 10).

***Śivadharmottara** 1.2*
jñānaśaktidharaṃ śāntaṃ kumāraṃ śaṅkarātmajam | *devārisūdanaṃ skandam agastiḥ paripṛcchati* || 2 ||
<agastir uvāca>
bhagavan darśanāt tubhyam antyajasyāpi sadgatiḥ | *saptajanmāni vipraś ca svargād bhraṣṭaḥ prajāyate* || 3 ||

Table 10. Stanza 2 and 3 of chapter 1 of *Śivadharmottara* edited by Florinda De Simini. Focus on the break of the metre.

5.10 Traditional methods and digital tools in South-Asian textual scholarship

In Sanskrit textual scholarship, technological methodologies and advancements are progressively complementing and enhancing traditional practices. Although some scholars still adhere to traditional methods, there is a growing trend towards alternative automatic methods. Per the edition preparation workflow outlined in Ch. 1, we analyse the digital tools used in South-Asian Digital Scholarly Editing.

5.10.1 Tools for building the *stemma codicum*

The initial phase of the editing process involves the creation of a *stemma codicum* when multiple witnesses of a work are available. Scholars utilise a combination of digital tools and on-site visits to institutions that house the manuscripts. The acquisition of facsimiles is a crucial aspect of the research. In the South-Asian domain, the Cambridge Digital Library[117] is a prominent resource. This library houses a vast collection of manuscripts curated from the late 19th century.

After the witnesses have been gathered, the subsequent step involves the *collatio*, performed manually or availing of digital tools. Alongside the widely used *CollateX*,[118] there exist other

[117] Sanskrit manuscripts collection of Cambridge Digital Library: https://cudl.lib.cam.ac.uk/collections/sanskrit/1.

[118] CollateX: https://collatex.net/.

specialised tools that aid philologists in the collation of Sanskrit texts. One such tool is *helayo*,[119] developed by Li (2022). It automatically aligns multiple Sanskrit text sequences starting from pre-processed and tokenized texts. It returns an XML/TEI file reporting the alignment and a phylogenetic tree as output. In addition, a web-based matrix editor[120] is provided to edit the alignment. As Maas (2013, 31) highlights, these tools assist scholars studying Sanskrit literature and linguistics in producing and processing large amounts of complex data.

5.10.2 Tools for building the *apparatus*

The *textus* and the critical *apparatus* are the core of the philologists' work. These elements form the foundation of the philologists' craft in critical editions, allowing for thorough analysis and interpretation of texts (Ch. 1.3.3, 1.3.4, 1.3.5). Typically, Sanskrit philologists manually prepare the *apparatus*. It involves recording the necessary information in a document that can be utilised for a print version of the edition (Fig. 26) or utilising markup languages for a digital version (Ch. 3.3).

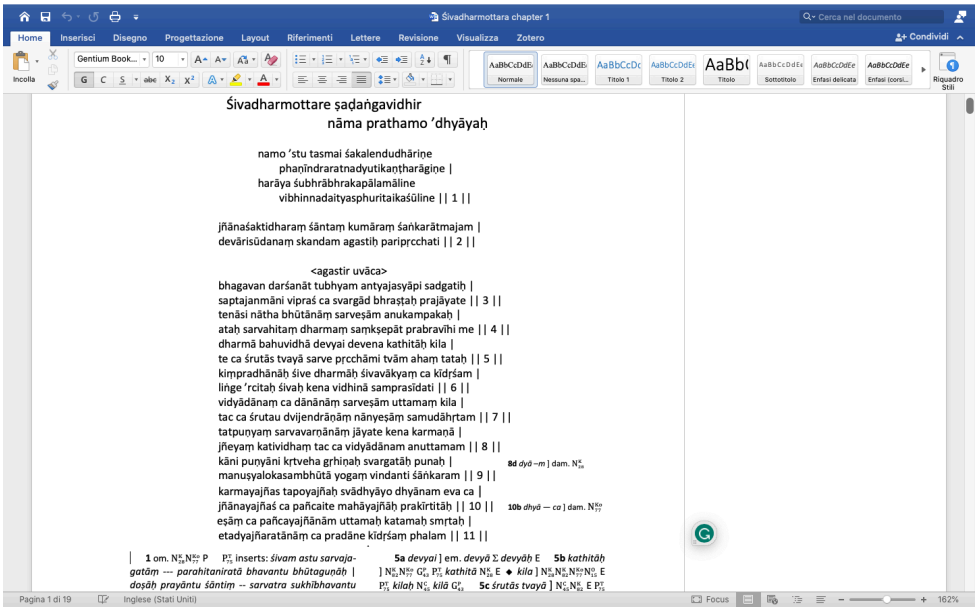


Fig. 26. .docx document of *Śivadharmottara* chapter 1, edited by Florinda De Simini.

In the South-Asian context, specific guidelines based on the TEI have been developed for encoding editions[121] in the DHARMA project.[122] These guidelines are beneficial when encoding South-Asian

[119] helayo: https://github.com/chchch/sanskrit-alignment/tree/master/helayo.

[120] Matrix editor: https://github.com/chchch/sanskrit-alignment/tree/master/matrix-editor.

[121] DHARMA Guides: https://dharma.hypotheses.org/guides.

[122] DHARMA project: https://dharma.hypotheses.org/. https://dharma.hypotheses.org/guides

critical[123] (Griffiths and Janiak 2023) and diplomatic editions[124] (Balogh and Griffiths 2020). In due course, we shall delineate this approach as *XML/TEI methodology*. Upon analysing an encoded instance outlined in the DHARMA guidelines, it is evident that these guidelines are congruent with TEI standards (Table 11).[125]

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

Display –

lemma] K L, variant reading M • optional observation

Table 11. DHARMA Encoding Guide for Critical Editions. Critical apparatus encoding following the TEI *Parallel Segmentation* method.[126]

Other encoding systems are concurrently utilised. Specifically, a distinct system has been devised for edition encoding purposes within the *Śivadharma* project.[127] It employs XML and offers an intuitive approach to marking the *apparatus*. We shall henceforth refer to this approach as *Generic XML methodology*. An example is reported below (Table 12).

```
<RMTEXT> sadbhāvaḥ satyam ity āhur dr̥ṣṭapratyayam eva vā |</RMTEXT>

<RMAPP>
    <VSNUMPADA>1a</VSNUMPADA>
    <LEM>sadbhāvaḥ</LEM>
    <span class="ms"><b>C</b></span>
    <span class="ms">K<sub>82</sub></span>
    <span class="ms">K<sub>07</sub></span>;
    sadbhāva° <span class="ms">K<sub>10</sub></span>
```

[123] DHARMA Encoding Guide for Critical Editions: https://hal.science/hal-04085137.

[124] DHARMA Encoding Guide for Diplomatic Editions: https://shs.hal.science/halshs-02888186.

[125] TEI Consortium, eds. "12 Critical Apparatus" TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Version 4.6.0]. [Last updated on 4th April 2023]. TEI Consortium. https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html ([10th August 2023]).

[126] The TEI Parallel Segmentation method implies an inline *apparatus* where each segment of text on which there is variation is marked by an <app> element directly in the text.

https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPDE.

[127] This system is used strictly privately as a drafting tool.

`    <span class="ms">E</span> ` `</RMAPP>`
Display – sadbhāvaḥ satyam ity āhur dṛṣṭapratyayam eva vā | 1a sadbhāvaḥ] CK_{82} K_{07}; sadbhāva° K_{10} E

Table 12. *Śivadharma* critical *apparatus* encoding.

The *textus* is encapsulated into the `<RMTEXT>` tag, while the *apparatus* entry in `<RMAPP>`. The *apparatus* entry components (Ch. 1.3.4) are identified as follows: (1) `<VSNUMPADA>` containing the location of the lemma in the stanza and pada; (2) `<LEM>` for referring to the location of the lemma in the stanza and *pāda*; (3) `<span class="ms">` to associate a witness attesting to the *lemma* or a variant. Conversely, no tags enclose variants. Unlike the TEI *Parallel Segmentation* method, the *lemma* is not directly included in the *textus* for reasons of visualisation.

More advanced technical methods to prepare the *apparatus* have been developed as well. Specifically, we refer to automatic methods for compiling the *apparatus* from scratch. According to Li (2017, 305), no originals exist within the framework of Sanskrit textuality (Ch. 5.1). Instead, there are numerous manuscript copies of literary works, many of which are highly corrupted. Nevertheless, automated techniques enable an alignment that would otherwise be unattainable. Consequently, the *apparatus* can be automatically reconstructed, thereby rendering all frequently unmentioned emendations explicit, including common spelling variations. The collation algorithm may be configured with a set of crucial principles to ensure accurate and reliable results. These principles are essential for sifting through and discerning relevant information while discarding any irrelevant or uncritical content in the *apparatus*.

5.11 Prepare and finally visualise editions in the South-Asian domain: the workflow

A comprehensive summary of the steps necessary for preparing South-Asian critical editions is outlined below.[128] Table 13 highlights the diverse techniques discussed in the preceding section (Ch. 5.10).

[128] We analyse the workflow of preparing *specifically* critical editions since it is considered the most intricate scenario (Ch. 1.2).

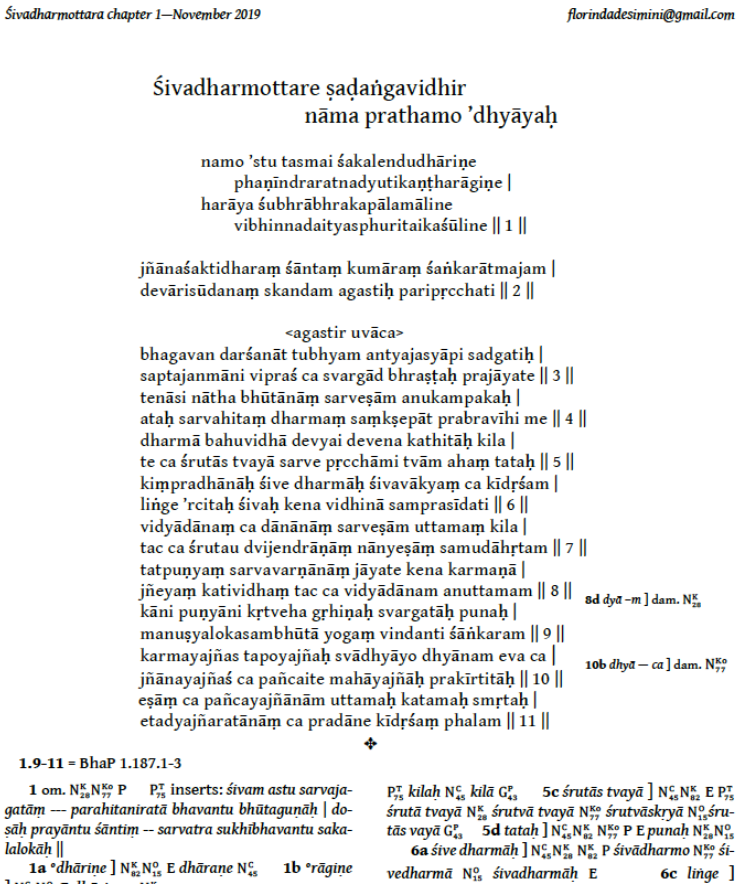
Workflow to prepare a critical edition					
		Methodologies			
		Manual	XML/TEI based	Generic XML based	Automatic
1	*Recensio* (Ch. 1.3.2)				
	Collection of all the witnesses of a given text.	The philologist goes to institutions that preserve manuscripts, or utilises digital repositories.			
					Following the initial acquisition of manuscripts, each of them is transcribed and tokenized to facilitate the alignment process.
2	*Collatio* (Ch. 1.3.2)				
	Comparison of the available witnesses to identify the variants.	The philologist compares the witnesses word by word to identify the variant readings.			A specific software, i. e., *helayo*, aligns the transcriptions of the manuscripts and returns a phylogenetic tree.
3	*Stemma codicum* (Ch. 1.3.2)				
	Construction of a genealogical tree establishing the relationships between witnesses. It represents a possible copying process	Drawing from a comparison of the manuscripts and a thorough analysis of significant errors, the philologist reconstructs the *stemma codicum*.			The *stemma codicum* is reconstructed automatically by using a specific

	through witnesses.				software, i.e., *helayo*.
4	*Examinatio* (Ch. 1.3.2; 1.3.3)				
	Analysis of the sources.	The philologist critically analyses the sources.			The philologist automatically obtains a critical *apparatus*. It can be refined by specific parameters according to critical principles set by the philologist.
5	*Emendatio* (Ch. 1.3.3)				
	When *examinatio* is not sufficient to make mechanical choices between variants, the *emendatio* is the necessary procedure to judge the variants according to internal criteria, i.e., *selectio*, or conjecturally, i.e., *divinatio*.	The philologist judges the variant and chooses the most reliable.			In the case of *helayo*, the philologist uses a web interface to filter the information collected in the *apparatus*.
6	*Constitutio textus* and *apparatus*				
	The editor proposes a specific version of the text and summarises the editing process in the *apparatus*.	Based on the previous analysis, the philologist proceeds with reconstructing the text while documenting each of the choices made in the *apparatus*.	Based on the previous analysis, the philologist proceeds with reconstructing the text while documenting each of the choices made in the *apparatus*. Both the textual content and the *apparatus* information are encapsulated in distinct tags, providing them with a precise semantic meaning.		The *textus* and the *apparatus* have already been prepared in the previous phases.
7	*Visualisation*				
	The edition is paginated.	The	The	In the digital	In the layout

		presentation of the edition follows a traditional layout. The *textus* takes centre stage and has an absolute prominence on the page. The *apparatus* is situated at the foot of the page and, to some extent, alongside the reference verse. Lastly, the references to the parallels are positioned between the text and the *apparatus* (Fig. 27).	XML/TEI-based edition's display only partially reflects the conventional printed version. The text takes centre stage, accompanied by a list of witnesses which attest to it and various metadata such as the title, publisher, and copyright. Each stanza is followed by its respective parallels and translation. The *apparatus* is situated beside the text, corresponding with the reference verses, and is accessible by hovering the mouse over the reference numbers. The *lemma* is denoted in red within the text and is highlighted when the corresponding device entry is opened (Fig. 28).	edition based on generic XML, the text is not the only protagonist of the page because it is directly accompanied by its translation. Each of these components takes up one half of the page. The *apparatus* and the parallels are found under each reference line (Fig. 29).	of the automated edition, the focal point is the text, which is introduced by metadata such as the title, author, and commentator. The text is situated in the centre of the page and is alternated with stanzas of the base text and its commentary. On the right-hand side, the *apparatus*, automatically generated, parallels and sources are aligned with the stanzas/paragraphs of the commentary. On the same side, users can access various tools to refine the information contained in the *apparatus*, such as witnesses. Once a selection has been made from the available options, the *apparatus* can

					be regenerated (Fig. 30).

Table 13. Workflow to prepare a critical edition with multiple witnesses. Different methods.



Śivadharmottara chapter 1—November 2019 florindadesimini@gmail.com

Śivadharmottare ṣaḍaṅgavidhir
nāma prathamo 'dhyāyaḥ

namo 'stu tasmai śakalendudhāriṇe
phaṇīndraratnadyutikaṇṭharāgiṇe |
harāya śubhrābhrakapālamāline
vibhinnadaityasphuritaikaśūline || 1 ||

jñānaśaktidharaṃ śāntaṃ kumāraṃ śaṅkarātmajam |
devārisūdanaṃ skandam agastiḥ paripṛcchati || 2 ||

<agastir uvāca>
bhagavan darśanāt tubhyam antyajasyāpi sadgatiḥ |
saptajanmāni vipraś ca svargād bhraṣṭaḥ prajāyate || 3 ||
tenāsi nātha bhūtānāṃ sarveṣām anukampakaḥ |
ataḥ sarvahitaṃ dharmaṃ saṃkṣepāt prabravīhi me || 4 ||
dharmā bahuvidhā devyai devena kathitāḥ kila |
te ca śrutās tvayā sarve pṛcchāmi tvām ahaṃ tataḥ || 5 ||
kiṃpradhānāḥ śive dharmāḥ śivavākyaṃ ca kīdṛśam |
liṅge 'rcitaḥ śivaḥ kena vidhinā samprasīdati || 6 ||
vidyādānaṃ ca dānānāṃ sarveṣām uttamaṃ kila |
tac ca śrutau dvijendrāṇāṃ nānyeṣāṃ samudāhṛtam || 7 ||
tatpuṇyaṃ sarvavarṇānāṃ jāyate kena karmaṇā |
jñeyaṃ katividhaṃ tac ca vidyādānam anuttamam || 8 ||
kāni puṇyāni kṛtveha gṛhiṇaḥ svargatāḥ punaḥ |
manuṣyalokasambhūtā yogaṃ vindanti śāṅkaram || 9 ||
karmayajñas tapoyajñaḥ svādhyāyo dhyānam eva ca |
jñānayajñaś ca pañcaite mahāyajñāḥ prakīrtitāḥ || 10 ||
eṣāṃ ca pañcayajñānām uttamaḥ katamaḥ smṛtaḥ |
etadyajñaratānāṃ ca pradāne kīdṛśaṃ phalam || 11 ||

8d *dyā –m*] dam. N^{K}_{28}

10b *dhyā — ca*] dam. N^{Ko}_{77}

1.9-11 = BhaP 1.187.1-3

1 om. $N^{K}_{28}N^{Ko}_{77}$ P P^{T}_{75} inserts: *śivam astu sarvajagatāṃ --- parahitaniratā bhavantu bhūtaguṇāḥ | doṣāḥ prayāntu śāntiṃ -- sarvatra sukhībhavantu sakalalokāḥ* ||

1a °*dhāriṇe*] $N^{K}_{82}N^{O}_{15}$ E *dhāraṇe* N^{C}_{45} **1b** °*rāgiṇe* P^{T}_{75} *kilaḥ* N^{C}_{45} *kilā* G^{P}_{43} **5c** *śrutās tvayā*] $N^{C}_{45}N^{K}_{82}$ E P^{T}_{75} *śrutā tvayā* N^{K}_{28} *śrutvā tvayā* N^{Ko}_{77} *śrutvāskṛyā* N^{O}_{15} *śrutās vayā* G^{P}_{43} **5d** *tataḥ*] $N^{C}_{45}N^{K}_{82}N^{Ko}_{77}$ P E *punaḥ* $N^{K}_{28}N^{O}_{15}$ **6a** *śive dharmāḥ*] $N^{C}_{45}N^{K}_{28}N^{K}_{82}$ P *śivādharmo* N^{Ko}_{77} *śivedharmā* N^{O}_{15} *śivadharmāḥ* E **6c** *liṅge*]

Fig. 27. Printed edition of *Śivadharmottara*, edited by Florinda De Simini. Draft version.

Fig. 28. Digital Scholarly Edition of *The Svayambhu*, edited by Arlo Griffiths and Timothy Lubin. Draft version.

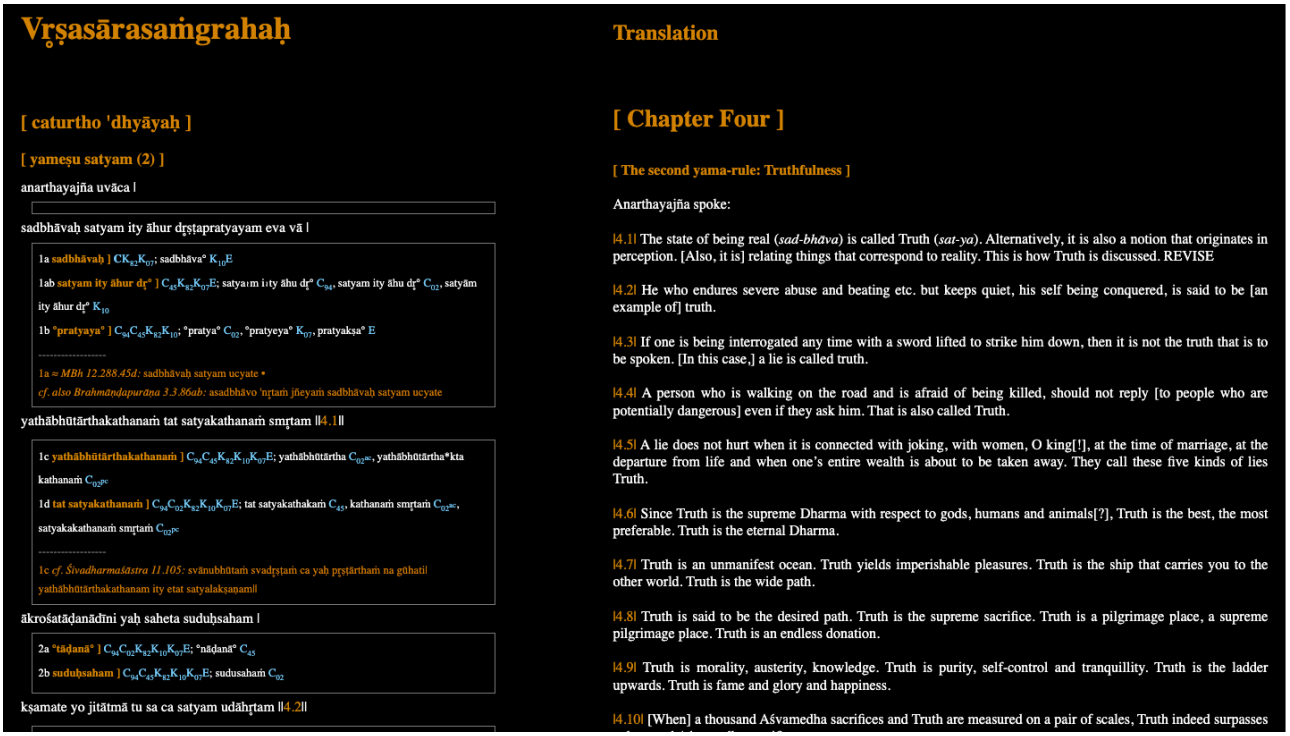


Fig. 29. Digital Scholarly Edition of the *Vṛṣasārasaṃgraha*, edited by Csaba Kiss. Draft version.

Fig. 30. Digital Scholarly Edition of the *Dravyasamuddeśa*, edited by Charles Li. Draft version.

Upon analysing the data presented in Table 13, it can be observed that the editorial workflow involved in preparing printed and digital versions[129] of an edition shows a negligible level of variation. As expected, a significant discrepancy lies in the visualisation of the editions. Philologists have the potential to showcase a vast array of possibilities, from a *reductio ad unum* to a multitude of versions (Ch. 2.6). The work by Li (2016, 307–309) serves as a paradigmatic illustration of the multiplicity of text versions that can be provided. Indeed, the alteration of particular parameters through the facets enables the acquisition of diverse versions of the *textus*, which, for instance, may include specific punctuation (Fig. 31) or orthographic variations (Fig. 32).

[129] It is interesting to acknowledge that XML holds a crucial position in the development of digital editions, even in the South Asian domain. Indeed, it is utilised in automated methodologies as well (Ch. 5.10.1; 5.10.2)

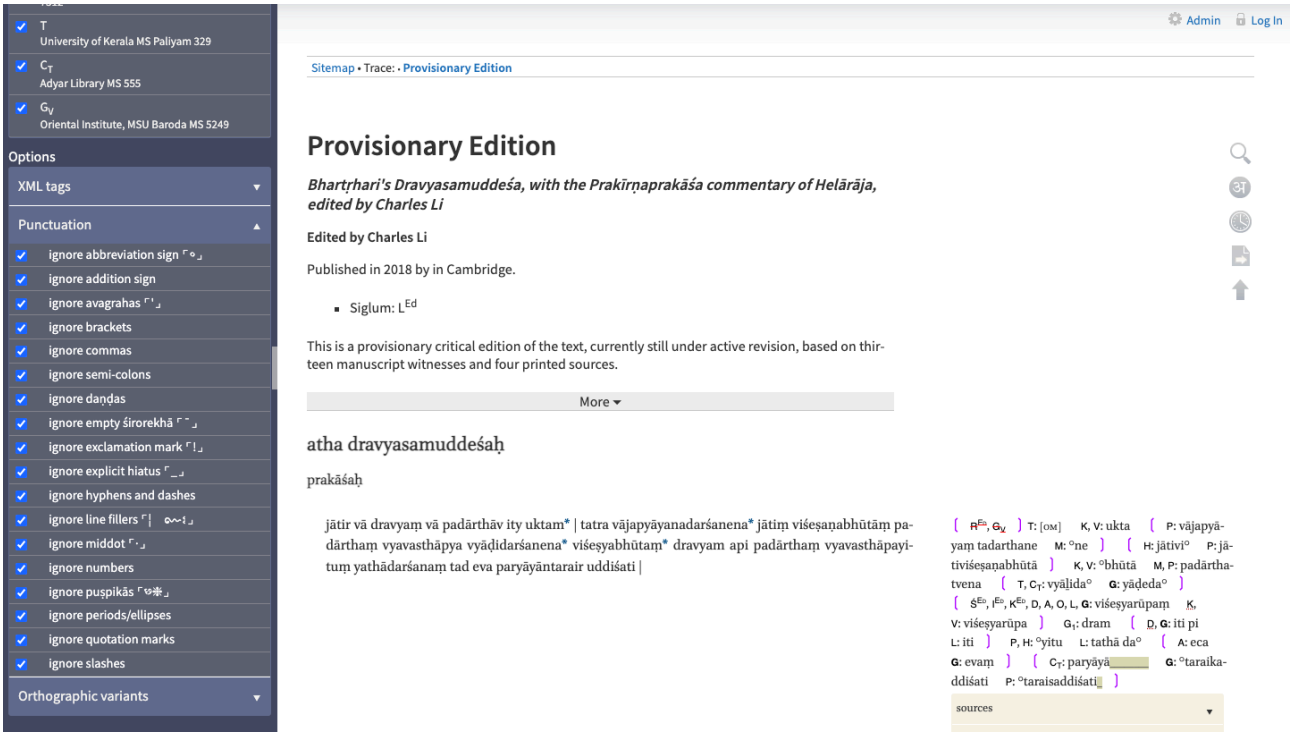


Fig. 31. Digital Scholarly Edition of the *Dravyasamuddeśa*, edited by Charles Li. Draft version. Focus on widgets to filter the punctuation.

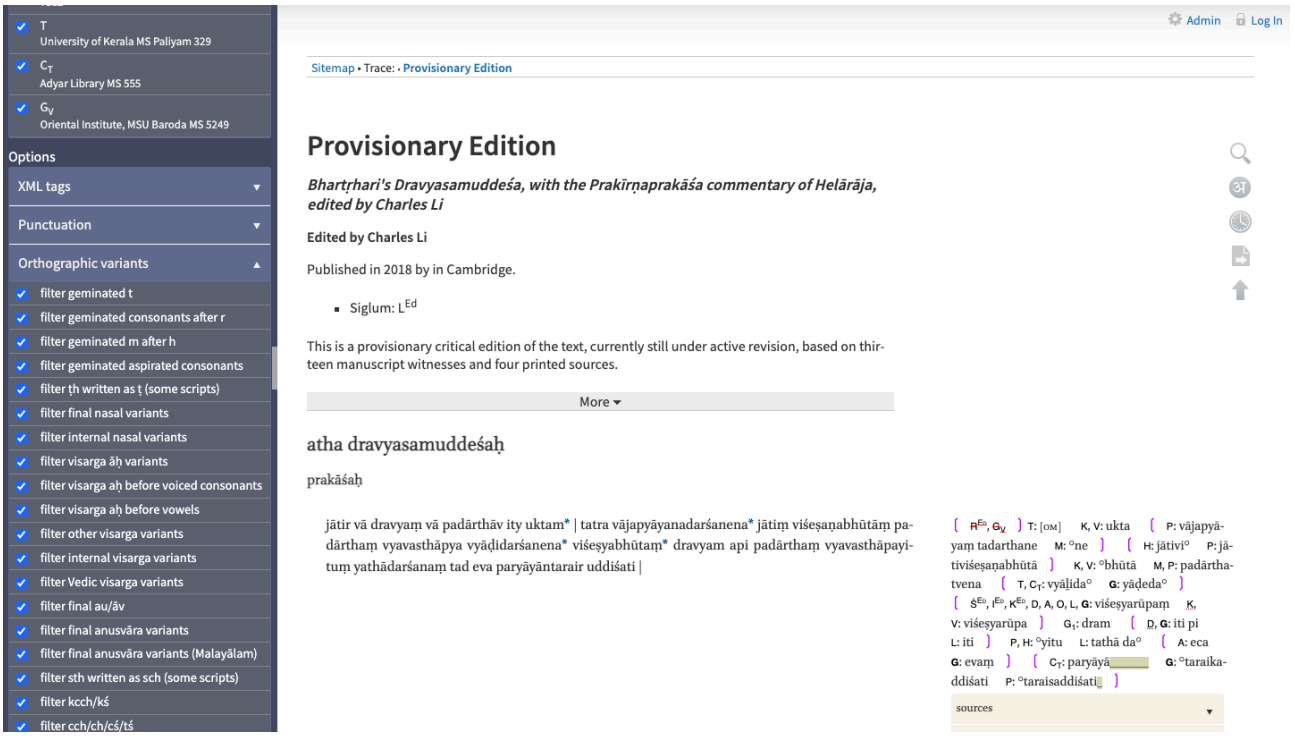


Fig. 32. Digital Scholarly Edition of the *Dravyasamuddeśa*, edited by Charles Li. Draft version. Focus on widgets to filter the orthographic variants.

Chapter 6
Digital Scholarly Editing CMS: *Śivadharma Database* design

The preceding chapters served as a prelude to the current design and implementation stages of the *Śivadharma Database*, a CMS for creating, publishing, and updating Digital Scholarly Editions in a user-friendly web environment. As Rosselli Del Turco (2019, 107) explains, one of the primary and often overlooked issues in Digital Scholarly Editing is that non-philologists often design tools for philology. The collaboration of philologists and digital humanists could solve this gap (Ch. 2.3.4, 2.3.5). Over two years of engagement in specialist readings, frequent meetings and discussions with experts in the domain, and field trips, we learnt about the philological issues and methodologies to consider as variables when designing the *Śivadharma Database*. Retracing the various topics addressed in the previous chapters, we now review such variables and, for each variable, we design a corresponding solution in terms of User Interface and User eXperience (IU/UX). These include:

1 **Definition of the concept of edition**.
2 **Different types of editions**.
3 **Main components of an edition**.
4 **Traditional and digital approaches to philology**.
5 **Differences and influences between the two areas**.
6 **Problems that arise in both areas**.
7 **Competencies required of philologists**.
8 **Assessment of the currently available Digital Scholarly Editing tools, their pros and cons**.

Through a detailed analysis of these variables on a real case study, i.e., South-Asian *Śivadharma corpus* (Ch. 5), we step-by-step design each of the *Śivadharma Database* functionalities and features and finally show the look and feel of the developed web application.

6.1 Reference design model: James Garrett's goal-oriented model

UX design is pivotal in designing the *Śivadharma Database* CMS since the project aim is to provide an experience covering a *workflow*, i.e., preparing a Digital Scholarly Edition (Ch. 3). It implies providing them with easy access to a Digital Scholarly Editing environment where they can find *user-friendly* tools to prepare their editions and guide them through the flow.

Hence, during this stage, the focus is not primarily on refining the features from an algorithmic perspective but on *studying and improving the user's overall experience dealing with the overall product*, i.e., the *Śivadharma Database*. As Garrett (2011, 8) puts it:

> when it comes to the parts of a product that are user-facing – the buttons, displays, labels, and so forth – the 'correct' form isn't dictated by functionality at all. […] User experience makes sure the aesthetic and functional aspects of the button work in the context of the rest of the product, asking questions like, […] "Is the button in the right place relative to the other controls the user would be using at the same time?"

James Garrett's *goal-oriented model* (Garrett 2011) has been adopted as a reference model to design the *Śivadharma Database* CMS interface and functionalities. This design model has been chosen since it focuses on website and web application design. It comprises a linear process that starts by studying what the stakeholders intend to achieve through the product and what they aim to deliver to their users. It defines a series of functional specifications and content requirements based on them. In conclusion, such a model provides a more practical stage, where the final product is designed on sketches of its layout, UI widgets and related functionalities.
Hence, following a comprehensive analysis of the variables arising from the preceding chapters, which significantly impact the design of the *Śivadharma Database* application, we describe its functional and content requisites, furnishing, where necessary, a design sketch of the corresponding widget to accomplish the defined tasks.

6.2 *Śivadharma Database* missions

The *Śivadharma Database* is a web application designed to do *digital philology*. So, it shares the digital philology's missions outlined in Ch. 2.3.1. Traditional philology aims to preserve, comprehend, and perpetuate texts. The *Śivadharma Database* should facilitate the digital workflow to achieve the same objectives in a digital environment. Its purpose is to ensure that the texts are not lost and are duly interpreted to bring their historical context and meaning back to light and to pass them on to future generations. The *user-friendly* nature of this platform emphasises the importance of perpetuating texts and making editions accessible to a broader audience. Allowing scholars who may not have the coding and programming skills and the necessary knowledge to create Digital Scholarly Editions (Ch. 3) provides an additional opportunity for philology to carry out its missions. The advantages of possessing easy-to-use tools for digital edition preparation become even more apparent in settings where philological practices are notably intricate, as in *Śivadharma* use case (Table 14) (Ch. 5).

Reasons why philological practices might be notably intricate: *Śivadharma* use case		
1	*It is difficult to reconstruct a text.*	The tradition is mostly oral.
		In most cases, no original exists.
		In most cases, there are no specific authors.
		Many of the written texts have been lost.
		A large number of primary sources is available.
		Witnesses can be highly damaged.
		There are many variations between witnesses.
		In most cases it is not possible to make mechanical choices.
2	*The textual structure is not linear.*	The text is full of influences.
		The text has relations with other types of texts, e.g., commentary and translation.
3	*The script and the language are not standard.*	The language differs from the norm.
		Witnesses are transcribed in local scripts.

Table 14. *Śivadharma* use case summarisation.

In such cases, philologists must prioritise strictly philological and non-technical queries, considering the intricacy of the philological tasks at hand. A CMS could respond to these needs.

6.3 Starting from the *Śivadharma* case study to reach *many*: flexibility as one of the primary project objectives

Whilst the *Śivadharma Database* project is inherently linked to the *Śivadharma Project*, one of its main objectives is to establish a user-friendly and adaptable platform that caters to the requirements of editing and publishing workflows without being restricted to any particular application domain. This decision aligns with our commitment to the principles of *reuse*, ensuring that the community has access to a valuable tool that can facilitate research activities.

Identifying the consistent patterns in the editing workflow (Ch. 1, 3), irrespective of the reference field, is crucial in developing a valuable toolkit for effectively carrying out the associated tasks. While the primary objective of this study is to address a wide range of use cases, it is necessary to

note that customisation may be necessary to suit specific project requirements outside the scope of *Śivadharma* (Ch. 4.7).

6.4 *Śivadharma Database* in practice: both editors and readers online

According to Pierazzo (2019, 209), a Digital Scholarly Editing framework can serve two purposes — it can provide scholars with adequate tools and cover the functions related to publication, resulting in cost savings. By adopting a UX perspective, such a solution could offer additional benefits as editors would be able to view their finished work in real time and make necessary changes. Furthermore, if the platform offers both creation and publishing capabilities, it could function as an *editing tool* and a *reading platform*.

Inspired by this, the *Śivadharma Database* is designed to serve a dual purpose. Once users log in, they are offered the option of assuming the roles of either *editors* or *readers*.

To access the editing section and undertake the role of an *editor*, one must possess a password provided upon request by the *Śivadharma* team. It allows specialists to verify the texts published online. Once logged into the editing section, editors can publish their material or save it in a draft, thus ensuring that no external user can access its contents. The editor can choose to allow the access to *co-editors*.

On the other hand, users in the role of *readers* are granted access to a specific section where all the published editions are listed and readily available. In this way, the platform provides a seamless experience for editors and readers, thus making Digital Scholarly Editing more accessible and cost-effective.

6.5 Editors in the *Śivadharma Database*: tools to allow scholars to reconstruct a text

The primary aim of Scholarly Editing, whether digital or otherwise, is to restore a *text*. Philologists employ one of a variety of models to perform this task, selecting an appropriate approach based primarily on the number of witnesses available for the text under consideration (Ch. 1). Thus, in the context of designing the *Śivadharma Database*, it is necessary to offer a comprehensive set of tools that cater to all types of methodologies (Ch. 1.3.2, 1.3.8, 1.3.9, 1.3.10, 1.4) and editions, ranging from *mechanical* to *critical* (Ch. 1.2). This approach ensures that philologists have the flexibility to opt for any method. In addition, the software in a CMS allows them to modify their approach during the construction phase easily. So, in-depth scrutiny of the preparatory stages in creating traditional and digital editions, as outlined in Ch. 1 and 3, guides the platform's design.

An analysis of the design preferences of the *Śivadharma Database* in reconstructing a text considering various possible methodologies and types of editions is reported below.

6.5.1 Ecdotic methods in the *Śivadharma Database*

Regarding applying a specific ecdotic method (Ch. 1.3.2, 1.3.8, 1.3.9, 1.3.10, 1.4), there are no restrictions on the *Śivadharma Database*. While the philologists of the *Śivadharma* team utilise a methodology similar to the one suggested by Pasquali (Ch. 1.3.10, 5.6), the platform does not dictate the decisions philologists make in the ecdotic domain. As explained in Ch. 2, digital technology does not eliminate the need for *traditional* philological analysis and contemplation; instead, it is a supplementary tool that imports its paradigm and extends the traditional philological paradigm.

6.5.2 Tools for digitising primary sources: *reuse* of external services

Starting with the most basic type of edition (Ch. 1.2), i.e., *mechanical* edition, it is necessary to set up a platform that enables uploading manuscript images and corresponding metadata for efficient cataloguing. In this regard, several frameworks and services offering such functionalities are already available, obviating the need for a complete system overhaul. The *Śivadharma Database* leverages the *AMS Historica*[130] service the University of Bologna provides.[131]

Several factors drive the decision to depend on an external service. Reusing pre-existing services enables significant savings in design, development, testing, costs, and human resources. Additionally, uploading images necessitates considerable server space, mainly when dealing with an extensive collection of images. Separating the images from the actual CMS enhances the system's architecture and organisation in server spaces.

The specific selection of *AMS Historica* can be attributed to several reasons, primarily due to its web-based functionality, which aligns with the web environment of the CMS *Śivadharma Database*. Furthermore, it leverages advanced technologies for image description and sharing, specifically IIIF and its API, such as Mirador, whose advantages have been widely acknowledged by the community, as highlighted in Ch. 3.1.5 and 3.1.6.

An example of a manuscript visualised in *AMS Historica* is presented below (Fig. 33).

[130] AMS Historica: https://amshistorica.unibo.it/.

[131] The uploading and description of the images is in progress.

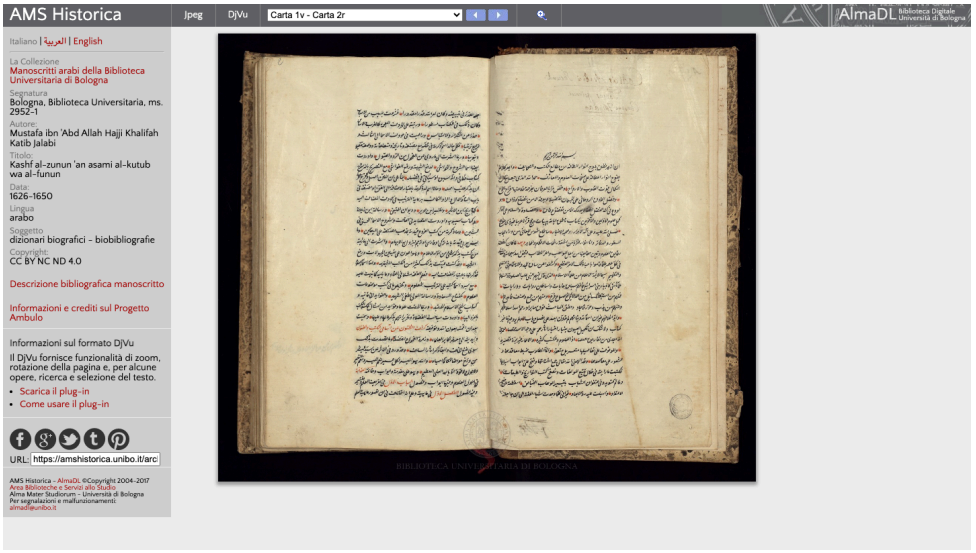


Fig. 33. Example of manuscript visualised in *AMS Historica*. Facsimile on the right and related metadata on the left.[132]

In the event of using the *Śivadharma Database* for projects other than the one it was initially created for, namely, the *Śivadharma Project*, it is mandatory to host the images in external environments separate from the *Śivadharma* system. It will ensure that the structure of the CMS remains distinct from the image management system.

Considering the functionalities of the CMS, it should implement a system at the interface level for importing images and describing them via metadata. As illustrated in Fig. 34, the proposed solution is based on the *form* interface component, enabling data input and saving (Ch. 6.5.5). Hence, specific forms are created for importing images into the *Śivadharma* system via URI and assigning metadata to them, including manuscripts *sigla*, institutions that preserve them, and more.

[132] Manuscript of the *Manoscritti arabi della Biblioteca Universitaria di Bologna* collection on *AMS Historica*: https://amshistorica.unibo.it/1668. Copyright CC BY NC ND 4.0.

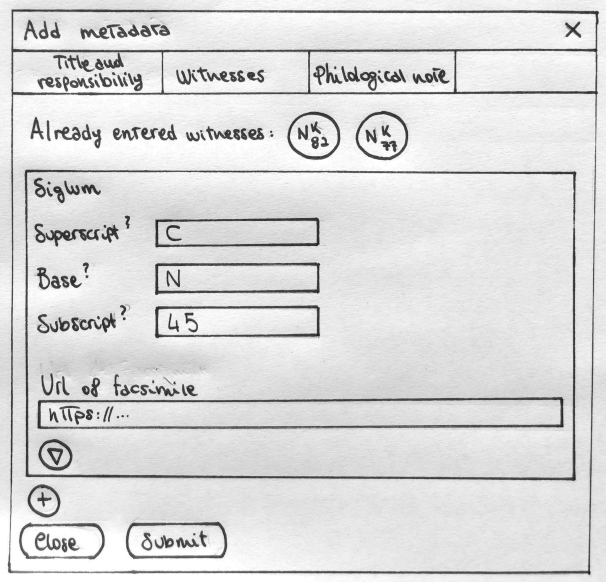


Fig. 34. Sketch of the form to declare and import the witnesses from an external repository into the *Śivadharma* system.

6.5.3 Tools for transcribing the text

Transcribing texts is a time-consuming and intricate process, especially when they are handwritten and almost illegible. While automatic (Ch. 3.1.1) and semi-automatic (Ch. 3.1.3) techniques are available, manual transcription remains the most precise method (Ch. 3.1.2). In the *Śivadharma Project*, philologists agreed not to transcribe every single manuscript but to propose a reconstructed text directly. However, the *Śivadharma Database* provides a textarea allowing users to manually transcribe texts during their reconstruction or at the end of the process or paste transcriptions generated by automatic or semi-automatic programs. It also offers tools for editing and formatting (Fig. 35).

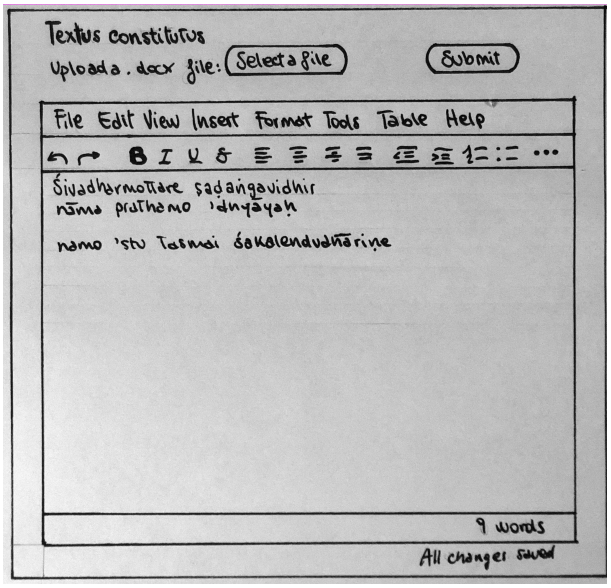


Fig. 35. Sketch of the textarea to type the *textus constitutus* in the *Śivadharma Database*.

6.5.4 Tools for modelling the edition: *modularity* as the key

Modelling editions is crucial, as this stage is the cornerstone for their structure, contents, visual representation, and usage. It is the pivot in the entire workflow, as all subsequent steps, including encoding and visualisation of the edition, strictly depend on it. Any errors or inconsistencies in the editions' model can adversely affect all subsequent project steps, requiring the editor to revisit and make necessary changes (Ch. 3.2.2). It is advisable to create a comprehensive list of all the components required in an edition, based on the type of edition the editor intends to create, to mitigate such risks. This approach helps establish a clear and concise framework for modelling an edition and ensures that all critical components are included for a successful outcome.

In the *Śivadharma Database* modelling, our focus lies in the development of a systematic infrastructure equipped with a user-friendly interface, specifically designed for the preparation of *critical editions* (Ch. 1.3). This approach is driven by two key factors: (1) the project's primary objective necessitates the creation of *critical editions*, and secondly, (2) aiming to make the CMS versatile and flexible, we provide a comprehensive suite of tools that can cater to even the most complex editions. The key is achieving *modularity*: editors are provided with a *range* of tools, each corresponding to a specific edition's component (Ch. 1.3), that they can selectively employ based on their research objectives and desired outcomes (Ch. 4.7).

Henceforth, below is presented a list of the components editors can use to create their editions on the *Śivadharma Database*. It results from extensive deliberations with *Śivadharma* stakeholders, who established their editorial guidelines,[133] and the research study documented in Ch. 1, 2, and 5.

1 ***Textus constitutus***. The *textus constitutus*, which represents the final edited version of a work or document, is the main objective of philology, whether traditional or digital. This pursuit is often regarded as a moral imperative from which philologists cannot stray. As debated in Ch. 6.2, the *Śivadharma Database* must furnish tools that enable text reconstruction, including a dedicated textarea for transcription and a platform to document reference sources (Fig. 35).

2 ***Note to the text***. Philologists typically use free-text notes to expose their editing criteria. A textarea is provided to accomplish this task, also enabling text formatting (Fig. 36).

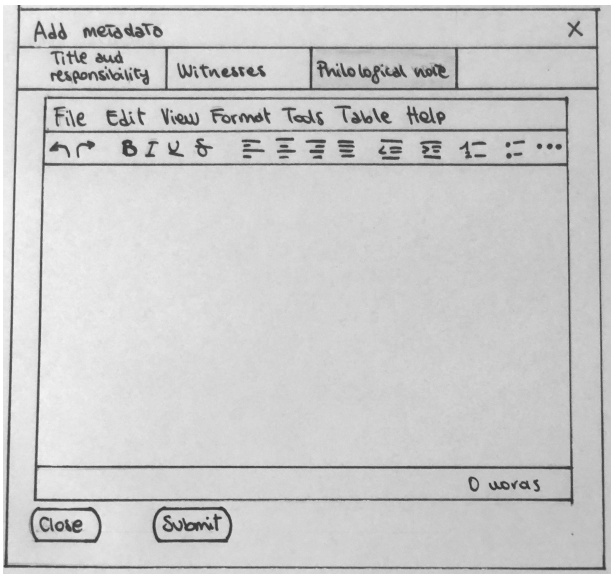


Fig. 36. Sketch of the textarea to type the *note to the text* in the *Śivadharma Database*.

3 ***Apparatus***. The *apparatus* is the core of a critical edition, wherein philologists record their choices, variant by variant, and even correction by correction. In the *Śivadharma Database*, from a design point of view, the *apparatus* is represented as follows. Regarding the *language* used for the composition of the *apparatus*, philologists are granted the liberty to choose any language. However, the interface does not permit modification of the abbreviations and symbols that may appear in the

[133] One of the primary reasons philologists refrain from publishing *scholarly* editions online is the absence of a dedicated publisher to oversee their projects and provide guidelines (Ch. 2.5.3). In order to address this issue, *Śivadharma* has implemented a set of comprehensive editorial guidelines to ensure the production of scientifically valid and uniform results.

entries, e.g., *om.* denoting an omission, unless a developer intervenes. As for the layout, the *apparatus* will be positioned at the bottom of the *textus*, and beneath each section. Finally, concerning the *format*, the system must facilitate inserting the following information via forms: the position of the *lemma* in the text, the *lemma*, the witnesses attesting to the *lemma*, and the variants along with the witnesses attesting to them, separated from the *lemma* by a space.

To fully implement the *apparatus* module (Fig. 37), it is necessary to incorporate functions that can effectively handle the various phenomena that can be observed in an *apparatus*, as described in Ch. 1.3.5: (1) *additions* and *suppressions*, (2) *lacunae*, (3) *transpositions*, (4) *conjectures*, and (5) *omissions*. Other functionalities are described below.

(a) **Live check**. The *Live check* functionality, shown in Fig. 37, allows a preview of the result of filling out the forms. This feature is exclusively accessible for the *apparatus* module. For other components, it is necessary to click on *Check all* in the module (Fig. 45) or save the edition as a draft to preview the changes.

(b) **Save**. The *Save* button allows sending the data to the database.

(c) **Check all**. When moving to the *Check all* tab, the list of *apparatus* entries already created is available in the preview. This functionality is identical in each module for creating components, together with the form for indicating the location of the selected fragment.

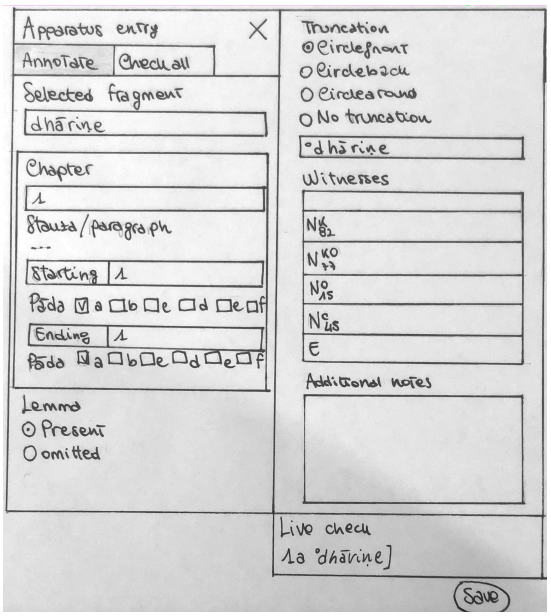


Fig. 37. Sketch of the form to associate an *apparatus entry* to the *textus* in the *Śivadharma Database*.[134]

[134] The same scheme is replied for the variants. A button allows cloning the form to create a variant.

4 ***Complementary apparatuses***. The *Śivadharma Database* offers additional features supporting supplementary materials, such as *parallels* and *citations* (Ch. 1.3.6). These features facilitate the specification of passages from other authors and citations of entities, such as places and people, through dedicated forms.

(a) **Parallels**. An instance that exemplifies the concept of *parallel* is the parallel described in Ch. 5.7.2. In this specific case, the *Shivadharma corpus*'s work *Umāmaheśvarasaṃvāda* shows parallelism with another work, the *Uttarattaramahāsaṃvāda*. A structure that encompasses all the relevant information about this textual phenomenon is created based on this example (Fig. 38).

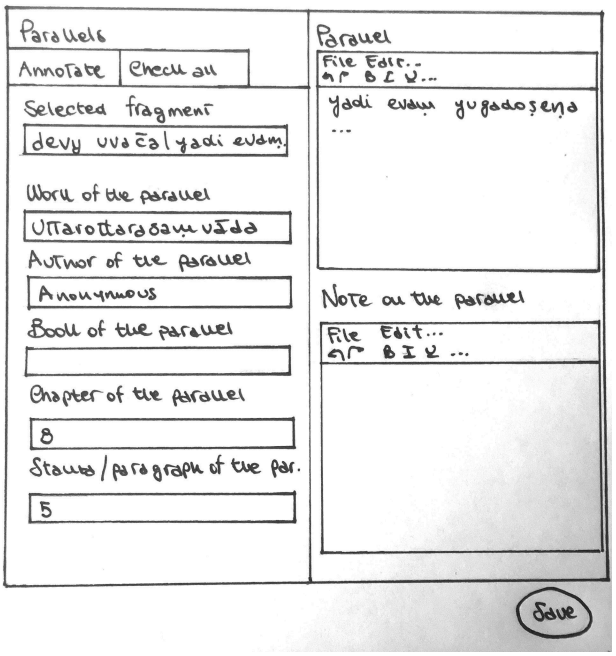


Fig. 38. Sketch of the form to associate a *parallel* to the *textus* in the *Śivadharma Database*.[135]

It is possible to incorporate the following information to establish a parallel connection with the *textus*: (1) *work of the parallel*, (2) *author of the parallel*, (3) *book of the parallel*, (4) *chapter of the parallel*, (5) *stanza/paragraph of the parallel*, (4) *parallel*, and (5) *note on the parallel*.

(b) **Citations**. An instance concerning the citation of an entity, particularly a *person* entity, is expounded in Ch. 5.7.1. The work in question is the *Vṛṣasārasaṃgraha*, belonging to the

[135] Owing to its uniformity across all modules, the form for declaring the location of the selected fragment has been excluded from this sketch and the following ones. Please refer to Fig. to check its structure.

Śivadharma corpus, which mentions several characters. A form is devised to facilitate the input of such information drawing from this instance (Fig. 39).

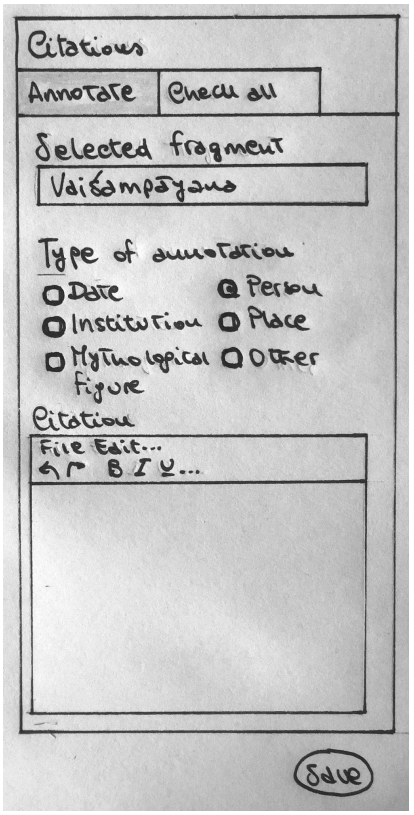


Fig. 39. Sketch of the form to associate a *citation* to the *textus* in the *Śivadharma Database*.

The form depicted in Fig. 38 allows for the formalisation of a citation of a specific entity within the *textus*. The form includes two essential details: (1) the type of cited entity, and (2) additional notes on the context of the citation.

5 ***Translation and notes***. In addition to the *complementary apparatuses*, editors have access to additional components such as *translation* and *notes* (Ch. 1.3.7). These are provided as two free fields that editors can fill in and format as required. The input translation/note should align automatically with the *textus* to make their consultation easier.

(a) **Translation**. Chapter 5.7.5 presents an illustrative example of a translation that can be prepared not only by the editor but also by other translators, including ancient ones. In the *Śivadharma Database*, it is possible to add a translated version of the text in any language and format it if required, along with relevant notes to formalise such characteristics (Fig. 40). The editor can decide on the translation unit, such as stanzas, chapters, and more.

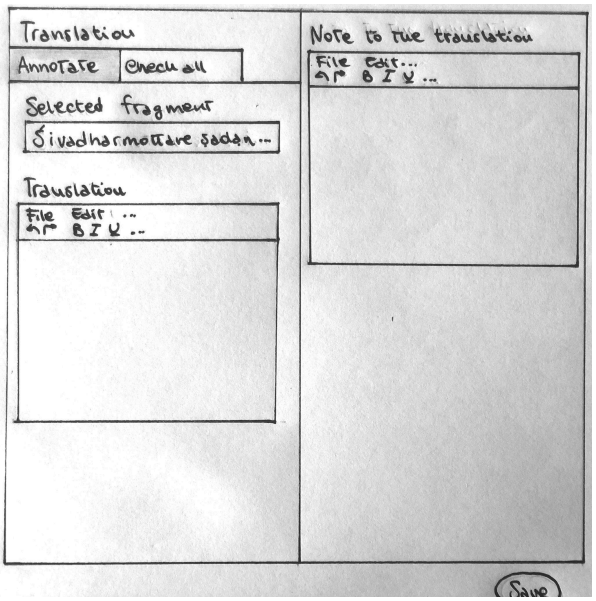


Fig. 40. Sketch of the form to associate a *translation* to the *textus* in the *Śivadharma Database*.

(b) **Notes**. The process of creating a *note* and linking it to the *textus* is straightforward. After selecting a textual fragment, it is required to insert its content into the textarea and format it appropriately, if needed (Fig. 41).

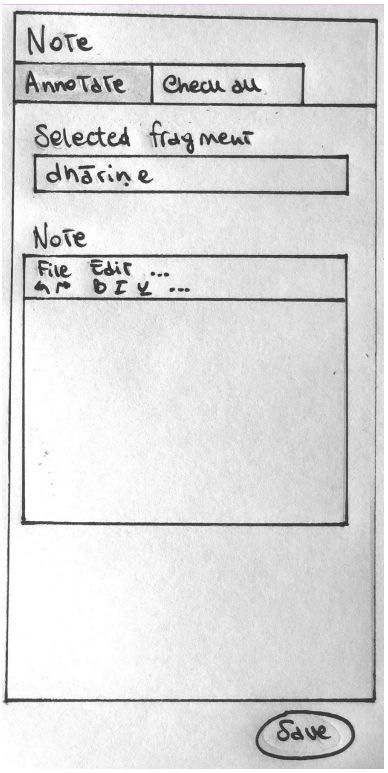


Fig. 41. Sketch of the form to associate a *note* to the *textus* in the *Śivadharma Database*.

5 ***Commentary***. Ch. 5.7.4 provides an example of the significance of the commentary component. The chapter highlights the relation between the *Bhikṣāṭanakāvya* and its associated commentary, the *Bhāvadīpikā*, shedding light on the importance of commentaries in providing interpretative material for texts. A form is made available in the *Śivadharma Database* to express such relations in an edition (Fig. 42).

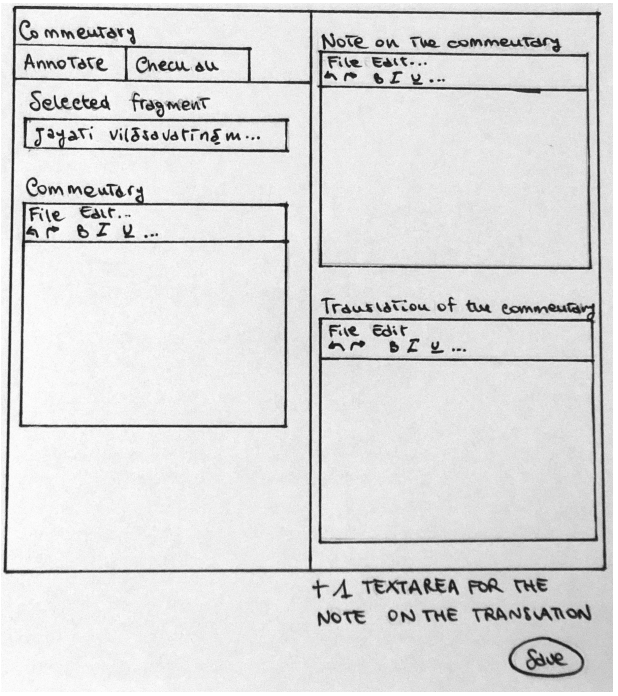


Fig. 42. Sketch of the form to associate a *commentary* to the *textus* in the *Śivadharma Database*.

6.5.5 Tools for encoding the edition: *highlighting* and *form-filling* strategy

The *Śivadharma Database* offers a user-friendly interface that allows for creating the various edition components identified in Ch. 6.5.4, including the *textus*, *apparatus*, notes, and more. However, as with any encoding method (Ch. 3.3), it is crucial to establish a formal connection between the *textus* and its complementary components, i.e., technically, the textual string object of annotation and its corresponding annotation.

Without going into technical details, which will be addressed in Ch. 7, it is sufficient to establish how *Śivadharma Database* allows editors to annotate a text at the interface level. We provide an example below. As exemplified in Fig. 43, the string highlighted in the text is identified as the *object of annotation*. At the same time, the related *apparatus* entry, which adheres to the scheme outlined in Ch. 1.3.4, is considered the corresponding annotation.

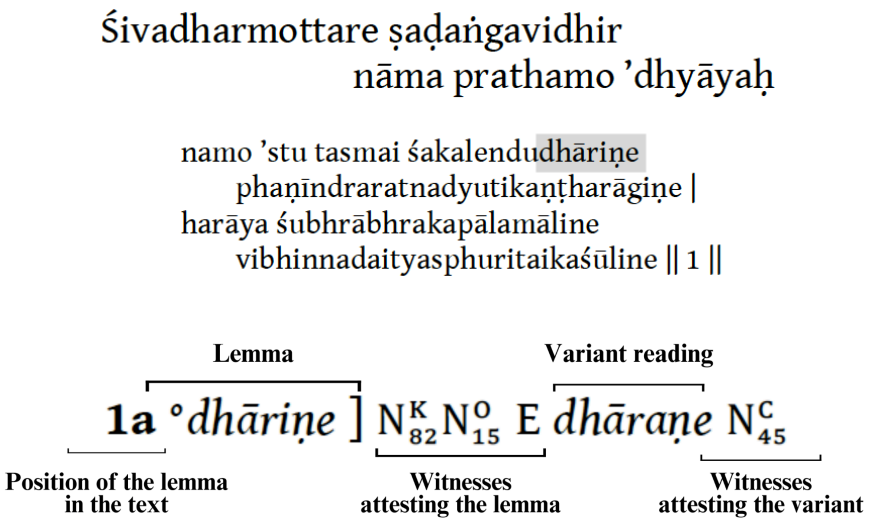


Fig. 43. Sample *apparatus entry* and corresponding reference string in the text.[136]

If a standard workflow were proposed in the *Śivadharma Database*, it would be necessary for the philologist to mark such constructs in XML/TEI (Ch. 3.3.1). A feasible *apparatus* entry formal representation in XML/TEI[137] can be referred to in Listing 8.

```
<l>
      namo &#8217;stu tasmai śakalendu
      <app xml:id="app-1-1-1a-1" resp="#flodes">
            <lem wit="#NK82 #NO15 #E" rend="circlefront">dhāriṇe</lem>
            <rdg wit="#NC45">dhāraṇe</rdg>
      </app>
</l>
```

Listing 8. Sample *apparatus entry* marked in XML/TEI.

The *apparatus* entry above has been encoded according to the TEI *parallel-segmentation* method.[138] Thus, it enables the integration of the contents of each entry within the text itself, thereby eliminating the necessity for any linking system (Griffiths and Janiak 2023, 67). By conjecturing the basic procedural measures adopted by an editor to establish this formal structure, five principal actions can be discerned:

[136] *Textus* and *apparatus* entry from *Śivadharmottara*, a Sanskrit work currently edited by Florinda De Simini (Ch. 5).

[137] The proposed encoding follows the DHARMA Encoding Guide for Critical Editions (Griffiths and Janiak 2023).

[138] TEI *parallel-segmentation* method: https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPPS.

1 **Identification of the preferred variant within the text**. The editor identifies a textual fragment on which there should be a variation to mark by an `<app>` element.

2 **Setting a relationship between the identified string and an *apparatus* entry**. The string identified earlier now serves as the *subject* of annotation. The editor encloses that string within an `<app>` element to establish a relationship between it and an *apparatus* entry.

3 ***Lemma* definition**. Consequently, the editor asserts the primary constituents of an *apparatus* entry. The string previously identified corresponds to the *lemma*, as it is the variant that has been selected for the *textus*. The editor encloses that string within a `<lem>` element to declare its *lemma* role.

4 **Assignment of witnesses attesting to the *lemma***. One or more witnesses attest to each variant. In order to declare the witness(es) attesting to the *lemma*, the editor assigns the attribute `@wit` to the `<lem>` element. The value(s) of this attribute should correspond to the witness identifier(s) attesting to the *lemma* as specified in the `<teiHeader>`.[139]

5 ***Variant readings* definition**. The editor creates an element `<rdg>` for each discarded variant.

6 **Assignment of witnesses attesting to the *variant readings***. In order to declare the witness(es) attesting to each *variant reading*, the editor assigns the attribute `@wit` to each `<rdg>` element. The value(s) of this attribute should correspond to the witness identifier(s) attesting to each variant as specified in the `<teiHeader>`.

The fundamental textual annotation workflow involves (1) *identifying a specific textual fragment* and (2) *adding information that describes specific aspects of the textual fragment*.
In order to make this process accessible to editors lacking coding skills, it is crucial to provide tools on the interface that enable them to carry out these operations without having to write code. Hence, we substitute the technical element of the procedure at every previously identified stage with a more user-friendly alternative in Table 15.

[139] The `<teiHeader>` TEI element provides a set of descriptive and declarative metadata associated with a digital resource or a collection of resources. This metadata enables an accurate description of the resource(s). It included the metadata related to the witnesses, corresponding to a catalogue description: https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiHeader.html.

Workflow to encode an *apparatus* entry in a critical edition on the *Śivadharma Database*				
		Methods		
		Standard XML/TEI method	>	User-friendly method
1	*Identification of the preferred variant within the text*			
		The editor visually identifies the textual fragment in the text.		
2	*Setting a relationship between the identified string and an apparatus entry*			
		The editor encloses the identified textual fragment in an `<app>` element.	>	The editor highlights the identified textual fragment and selects the annotation type corresponding to the edition component through a specific button.

		`<app>`
		`  fragment`
		`</app>`
			>	Text highlighting and selecting the annotation type.

3	*Lemma definition*			
		Inside `<app>`, the editor encloses the identified fragment in a `<lem>` tag.	>	A form opens next to the *textus* by filling in which the editor obtains the desired edition's component. A form contains an input for each information that must be present in the structure of the component. Inside the form to create an *apparatus* entry, it is possible to define the *lemma* by filling in the corresponding input.
		… `<lem>fragment</lem>` …	>	Filling in the *lemma* input in a form.

4	*Assignment of witnesses attesting to the lemma*			
		The editor assigns `@wit` as an attribute to the `<lem>` element and the witness identifier(s) as value(s).	>	Inside the form to create an *apparatus* entry, the editor defines the witness(es) attesting to the *lemma* by filling in the corresponding input. Prior to filling in the form, the editor is expected to have compiled a comprehensive list of all the witnesses to select from a predefined list of *sigla* when filling in the form.
		… `<lem wit="...">fragment</lem>` …	>	Compiling a list of all the witnesses and filling in the *witness(es) attesting to the lemma* input in the form.
5	*Variant readings definition*			
		The editor creates an element `<rdg>` for each discarded variant.	>	Inside the form to create an *apparatus* entry, the editor defines the variant reading(s) by filling in the corresponding input. A specific tool allows the editor to add more than one variant if required.
		… `<rdg>variant</rdg>` …	>	Filling in the *variant reading* input in the form.
6	*Assignment of witnesses attesting to the variant readings*			
		The editor assigns `@wit` as an attribute to the `<rdg>` element and the witness identifier(s) as value(s).	>	Inside the form to create an *apparatus* entry, the editor defines the witness(es) attesting to the *variant reading* by filling in the corresponding input.

				This input is automatically cloned when the editor adds a new variant.
		… <rdg wit="...">variant</rdg> …	>	Filling in the *witness(es) attesting to the variant reading* input in the form.

Table 15. Workflow to encode an *apparatus* entry in a critical edition on the *Śivadharma Database*. Shift from standard to *user-friendly* method.

This *highlighting* and *form-filling procedure*, which is summarised in Fig. 44, is the same for adding every other component and also for their updating. Each form contains specific inputs depending on the structure to obtain, as shown in the sketches in Ch. 6.5.4.

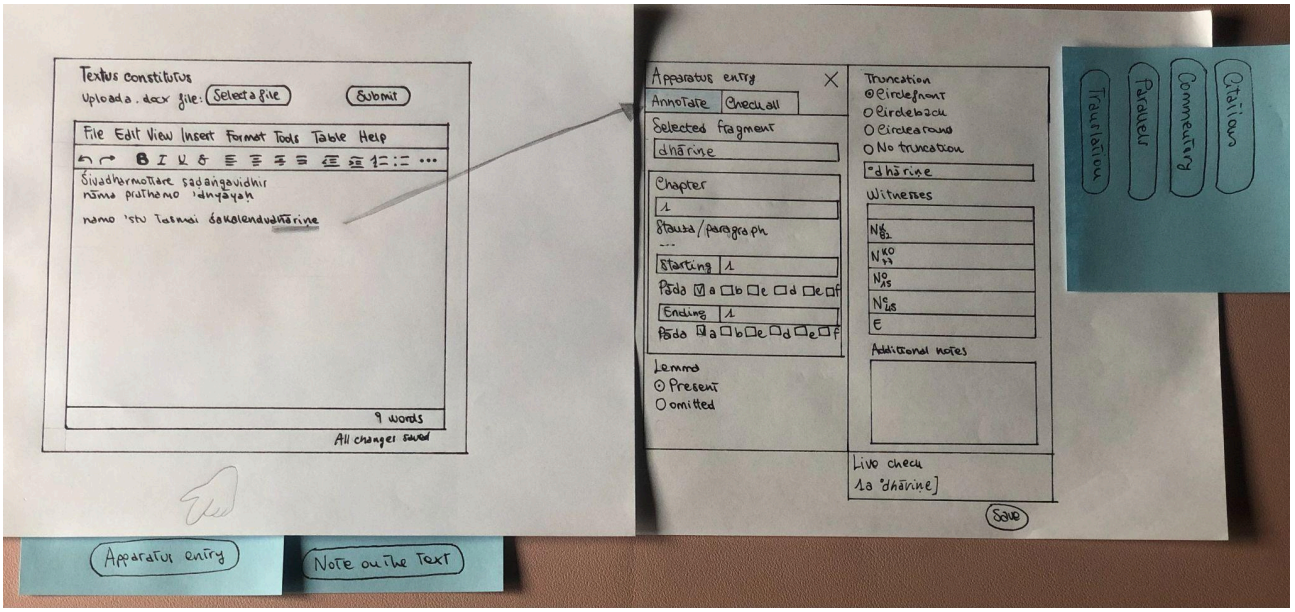


Fig. 44. Wireframing interaction flow to add an annotation, i.e., to create an *apparatus* entry, on the interface level.

It is noteworthy that the completion of all form fields is optional, except for those about the selected fragment location, which are essential for implementation purposes. Only the fields enabling the declaration of *relevant* information to the philologist may be filled in.

6.5.6 *Form-filling* implications

Deploying a form-based CMS for content creation and management is crucial in the *Śivadharma Database* project, as it greatly impacts the user-editor experience and the overall usability of the editing section. Forms serve as a *replacement for encoding*. By filling them out, editors can easily annotate texts without accessing or writing source code. The rationale behind using forms as widgets to complete these tasks is attributed to specific factors. Primarily, forms are *familiar* to users, as they come across daily, from logging into work accounts to signing in to credit card portals (Jarrett and Gaffney 2009, 2). Applying the abstract design principle of *familiarity* can significantly enhance the usability of an interactive system. *Familiarity* can be defined as the extent of correlation between a new user's pre-existing knowledge and the knowledge required for actual usage of the system. It is closely linked to the user's initial impression and is critical in determining how the user interacts with the system (Dix *et al.* 2004, 260–264). When users feel familiar with the system, it is more likely to be perceived as intuitive and user-friendly.

6.5.7 *Śivadharma Database* editing section final layout

During the process of wireframing interaction flows, a user-centric approach is adopted. By visualising the interface and the tasks from the users' point of view, we can gain insights into their intentions. This perspective enables us to identify the sets of actions that users are likely to repeat, thereby facilitating the development of more intuitive and user-friendly interfaces (Yang *et al.* 2016, 568). The functionalities of the editing section are summarised in the following wireframe (Fig. 45). The editing mode layout is designed to align with the layout of the edition published in the reading section, which is available under the *Editions* in the main navigation bar. The text occupies the left part of the screen, where a textarea is provided for writing or uploading the *textus*. Each edition component has a corresponding button in a different colour, which opens the compilation form on the right side of the *textus*. The button for each category of components is positioned in the same place where the respective component will appear in the edition. Saving the edition as a draft provides a more complete preview of the edition, available only to the editor and co-editors. The *Publish* button is available to make the edition available to all. In addition, the *Add Metadata* button allows the editor to describe the edition according to a specific scheme, including *title*, *editor*, *creation date*, etc. The editor can also add a philological note and declare the witnesses, as shown in Fig. 34, 36.

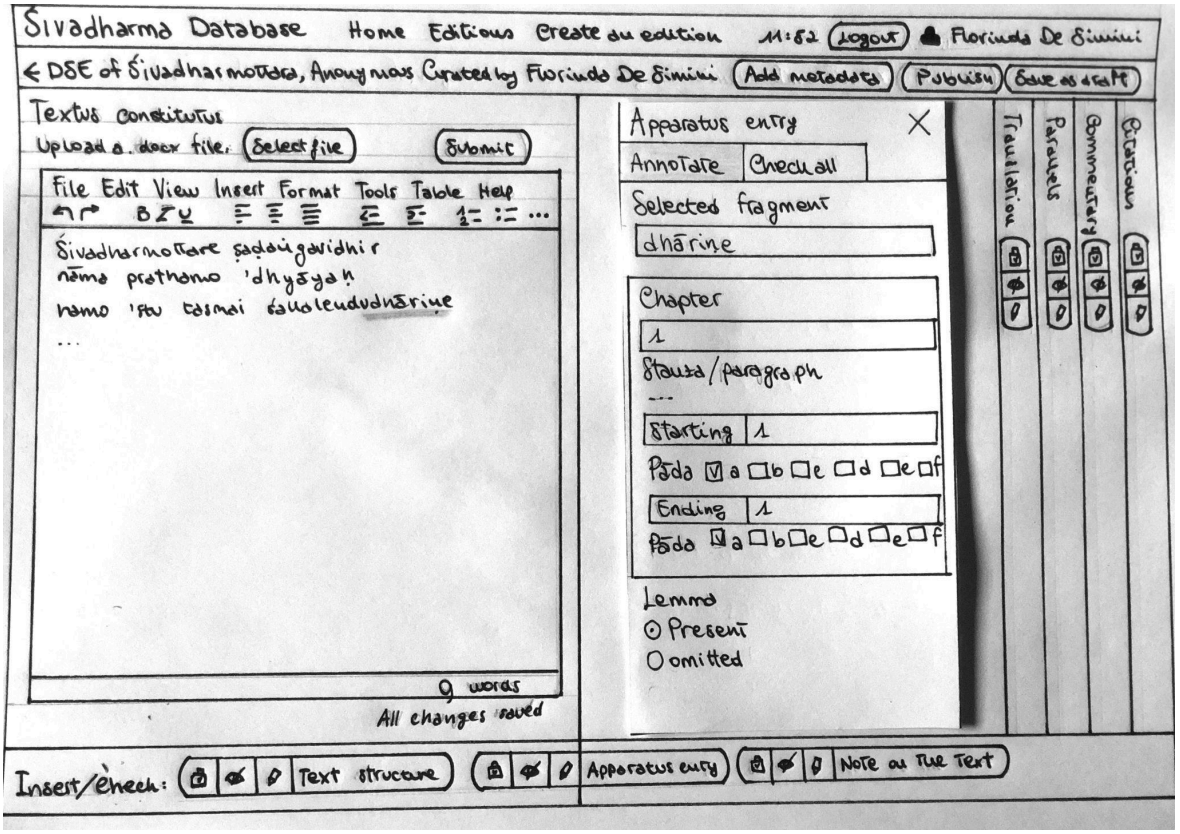


Fig. 45. *Śivadharma Database* editing section final layout.

6.6 Readers in the *Śivadharma Database*: tools to allow users to read the editions

As previously argued (Ch. 6.4), users can assume two distinct roles in the *Śivadharma Database* — that of an *editor* or a *reader*. For readers, there is no need for any specialised permissions. One can register and log in to gain access to the reading platform.

6.6.1 Tools for visualising the edition

The visualisation of the editions strictly depends on the edition model (Ch. 3.4.2). In the *Śivadharma Database*, only the content the user-editor has created is displayed to the readers. The components available in the editions are entirely at the editors' discretion, whether they are created or not. If the editors use all the features of the CMS, then the edition will be equipped with the components listed in Ch. 6.5.4:

1 ***Textus constitutus***.

2 **Note to the text**.

3 ***Apparatus***.

4 **Complementary apparatuses**, i.e., *parallels* and *citations*.

5 **Translation and notes**.

6 **Commentary**.

Moreover, editors hold the right to make changes to the components at any time, including adding new content, modifying existing ones, deleting components, or adding new ones, according to the *modularity* principle (Ch. 6.5.4).

6.6.2 *Śivadharma Database* reading section final layout

The functionalities of the *Śivadharma Database* reading platform are summarised in the prototype developed for the project (Fig. 46, 47).[140]

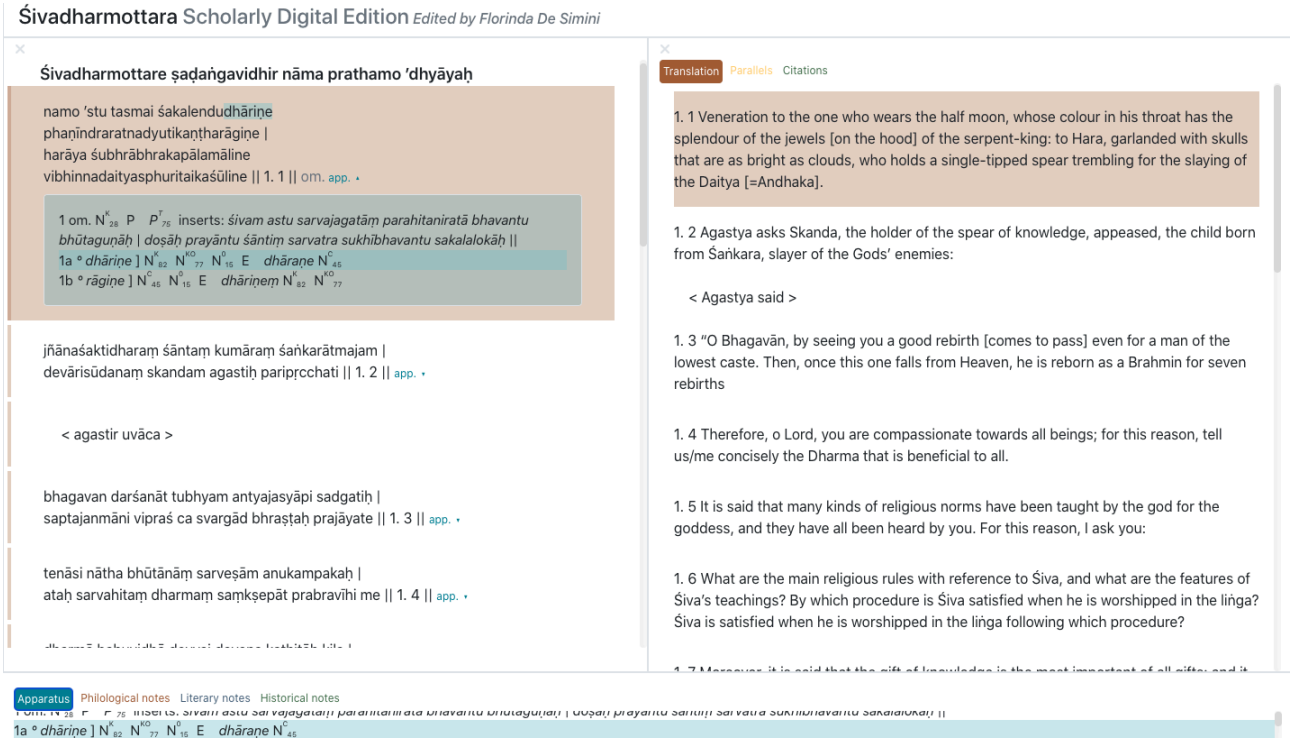


Fig. 46. *Śivadharma Database* reading platform. Focus on *apparatus* and *translation* components.

[140] Prototype of the *Śivadharma Database reading* platform: https://github.com/martinadellobuono/shivadharma-xslt-base-txt.

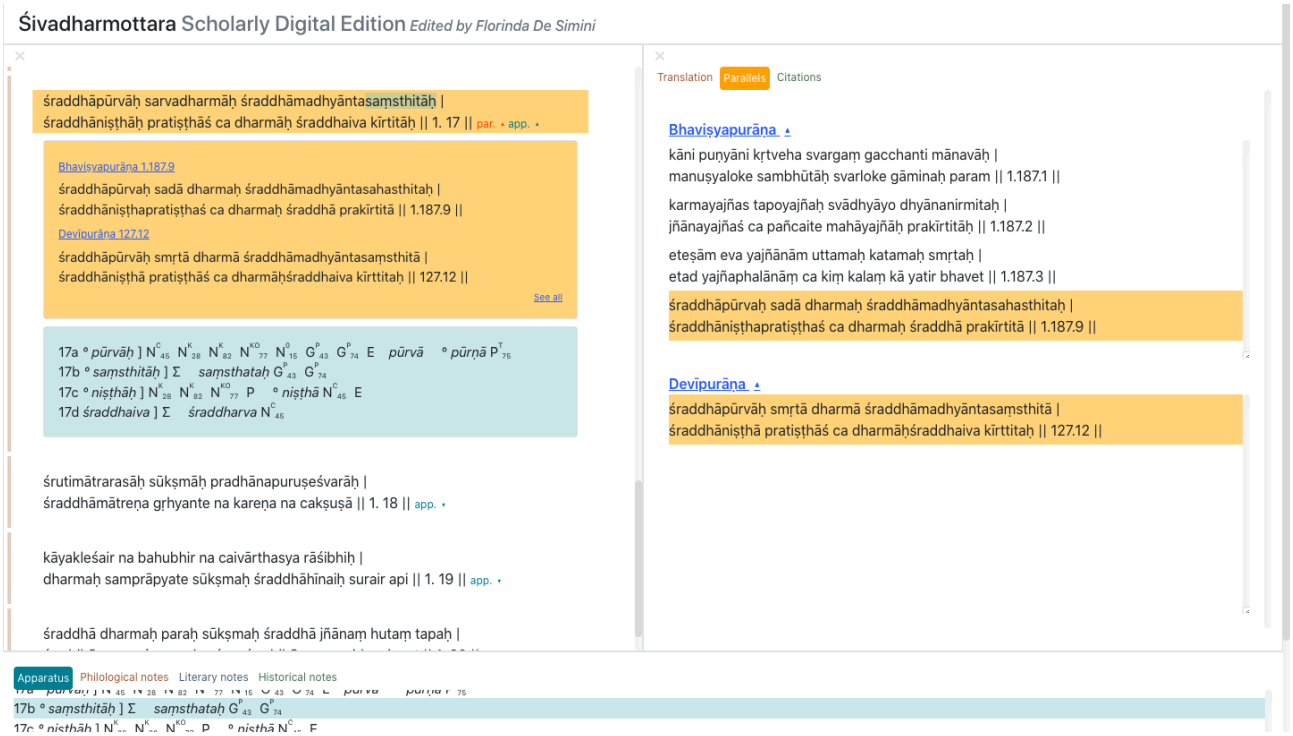


Fig. 47. *Śivadharma Database* reading platform. Focus on *apparatus* and *parallels* components.

The prototype illustrated in Fig. 45 and 46 depicts an edition published on the *Śivadharma Database*. A mirrored division of the page primarily distinguishes the layout into two halves. As in the editing section, the *textus* is positioned on the left side, while the right features supplementary components, i.e., *translations*, *parallels*, and *citations*. Above this division, the title and the editor's information are presented, while below it, the entire *apparatus* is accessible along with the *notes*.

The key to this design is not merely the *layout* but the *interactivity* that aims to augment the usage of a network of contents. While the alignment of the *textus*, *translation*, and *parallels* components intends to facilitate reading, the interactivity features play a crucial role in enhancing the overall user experience. Each component is linked to the text in a way that allows the user to access different kinds of information on the text easily. Specifically, it is possible to obtain information on (1) how the manuscripts report the text in the *apparatus*; (2) if and how other works report the same portions of text in the *parallels* and *citations*; (3) how the editor interprets the meaning of the text in the *translation*; and (4) how the editor interprets other details in the text in the *notes*.

6.7 Is there something missing in the *Śivadharma Database*? Checking the RIDE guidelines

The inventory of functional requirements for the *Śivadharma Database* is the outcome of an analysis of the issues, techniques, procedures, methodologies, and practices of philology at large, in conjunction with the specific case study in question and, lastly, with reference to the RIDE

Scholarly Digital Editions Guidelines (Sahle 2014). The aforementioned check of the RIDE Guidelines enables a thorough examination of the requirements outlined *vis-à-vis* those required during the revision stage of the edition. It facilitates the identification and fulfilment of any previously unidentified gaps. In the context of the *Śivadharma Database*, specific requirements pertinent to the validation process have not been accounted for thus far and, therefore, merit consideration:[141]

1 **Development of a search bar for browsing the editions**.
2 **Implementation of indexes for each edition**.
3 **Assignment of a DOI to each edition**.
4 **Conversion of editions into printable PDF format**.
5 **Selection of specific licences for each edition**.

6.8 *Śivadharma Database* final look and feel

As Agrebi and Boncori (2017, 624) argue, the *look and feel* is widely regarded as a reflection of Web Relational Proneness (WRP). One of the look and feel sub-dimensions is the *website design*. It is the first aspect that visitors see. Hence, it is crucial to enable users to appreciate the WRP and drive the *social presence* on the website. It is not just a matter of beauty but graphics and general content organisation. Screenshots of the *Śivadharma Database* CMS are shown below to provide an overall impression of the platform's user interface and design. Initially, the website greets visitors with sections for login, registration, and homepage (Ch. 6.8.1), which are followed by the editing (Ch. 6.8.2) and reading sections (Ch. 6.8.3). The website style is designed to achieve a certain uniformity in terms of shapes and colours, which distinguishes the functions of each section and widget.

6.8.1 Access to the *Śivadharma Database* look and feel

As assessed in Ch. 4.2, the reaction of scholars to technology in the field of digital philology is often of near exasperation. Therefore, the layout of the *Śivadharma Database*, characterised by vibrant hues and lively visuals, endeavours to counter this sentiment and present a welcoming and engaging browsing experience (Fig. 48, 49, 50)

[141] These requirements shall be implemented in a subsequent phase of the project.

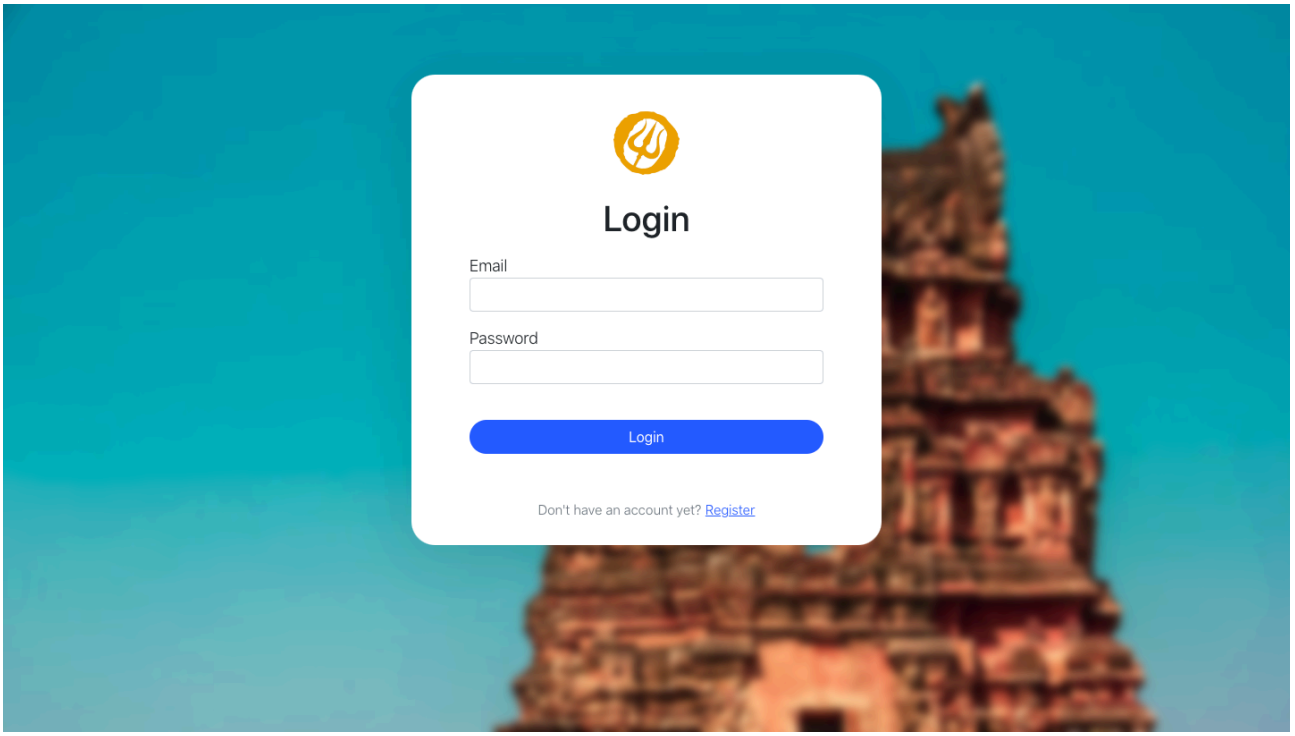


Fig. 48. *Śivadharma Database* login.

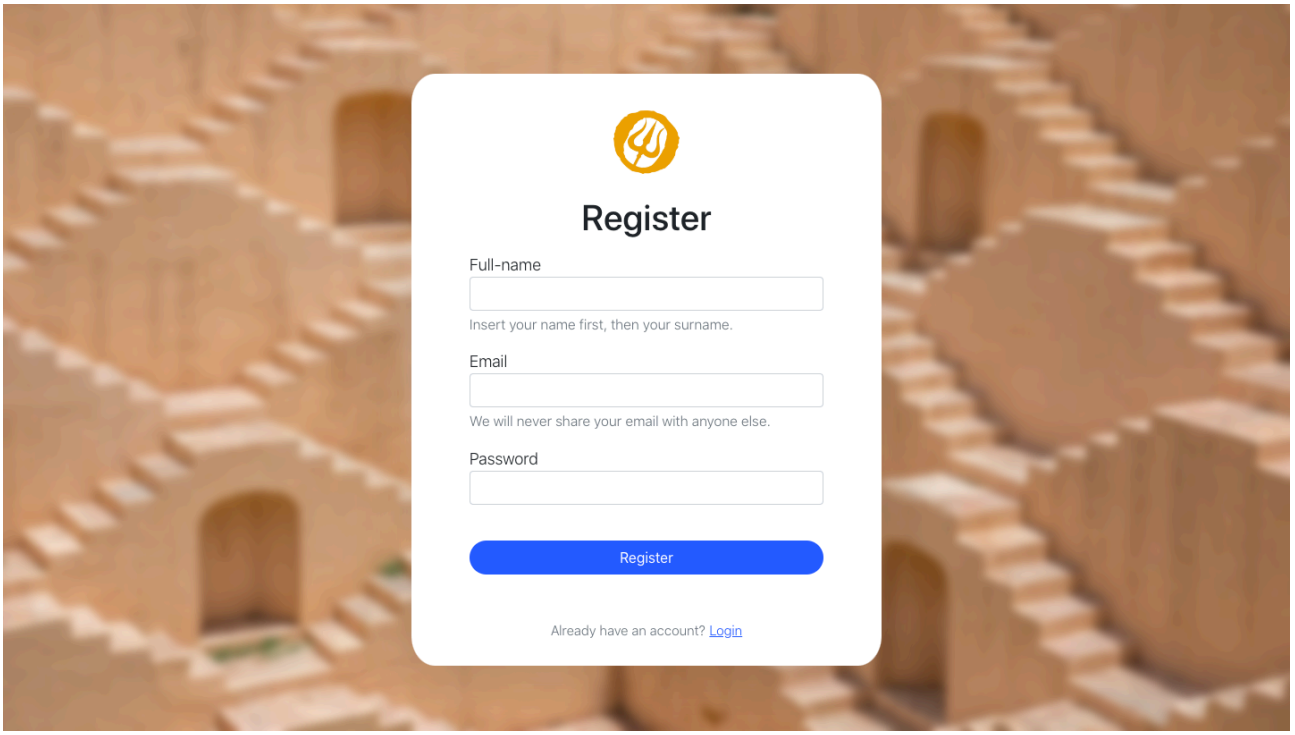


Fig. 49. *Śivadharma Database* register.

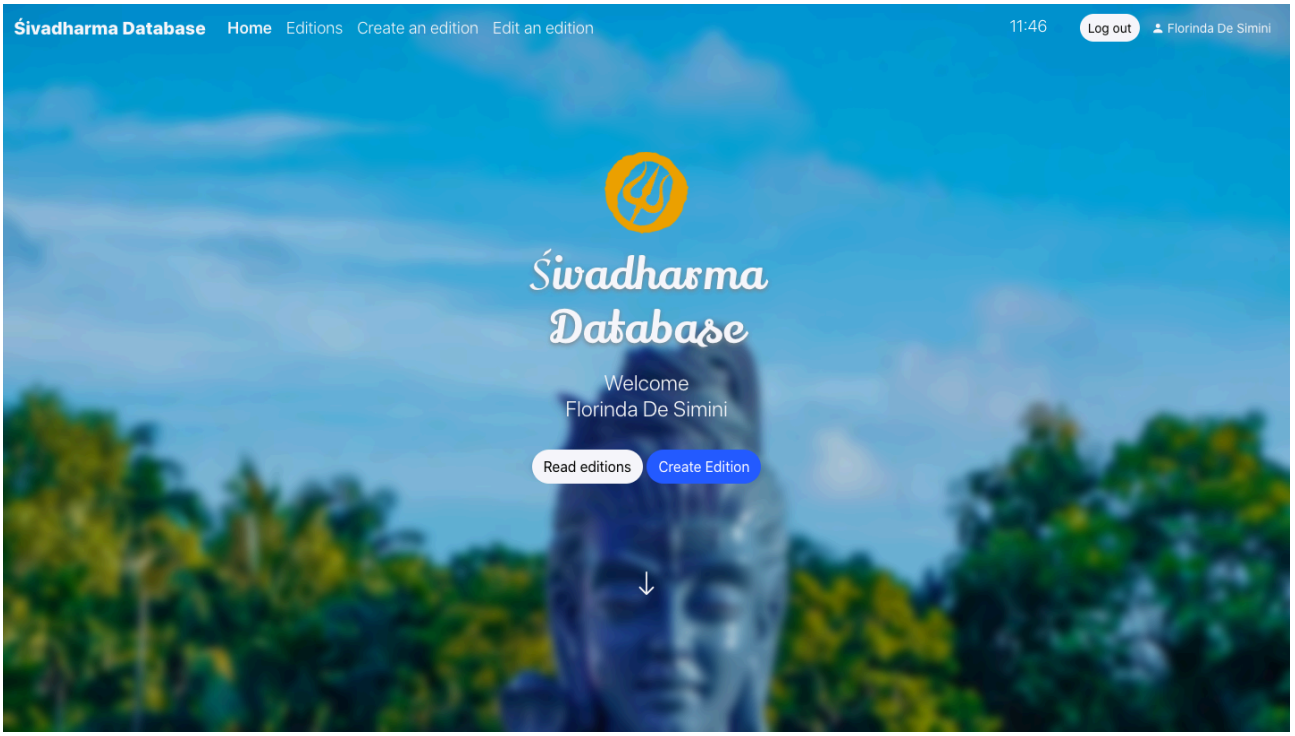


Fig. 50. *Śivadharma Database* homepage.

6.8.2 *Śivadharma Database* editing section look and feel

The design of the *Śivadharma Database* is a combination of two distinct styles. The welcome pages of the website (Ch. 6.8.1) and the editing section (Fig. 48, 49, 50) feature vibrant colours intended to be aesthetically pleasing and charming. In contrast, the sections dedicated to the actual work of the philologist, i.e., the editing section, use light colours to create a clean and orderly working environment, allowing editors to focus on their scholarly questions (Fig. 51, 52, 53, 54, 55, 56, 57, 58). The search for a cohesive stylistic approach with the rest of the platform is evident in the brightly coloured buttons used for inserting annotations into the text, which match the colour scheme of the rest of the site.

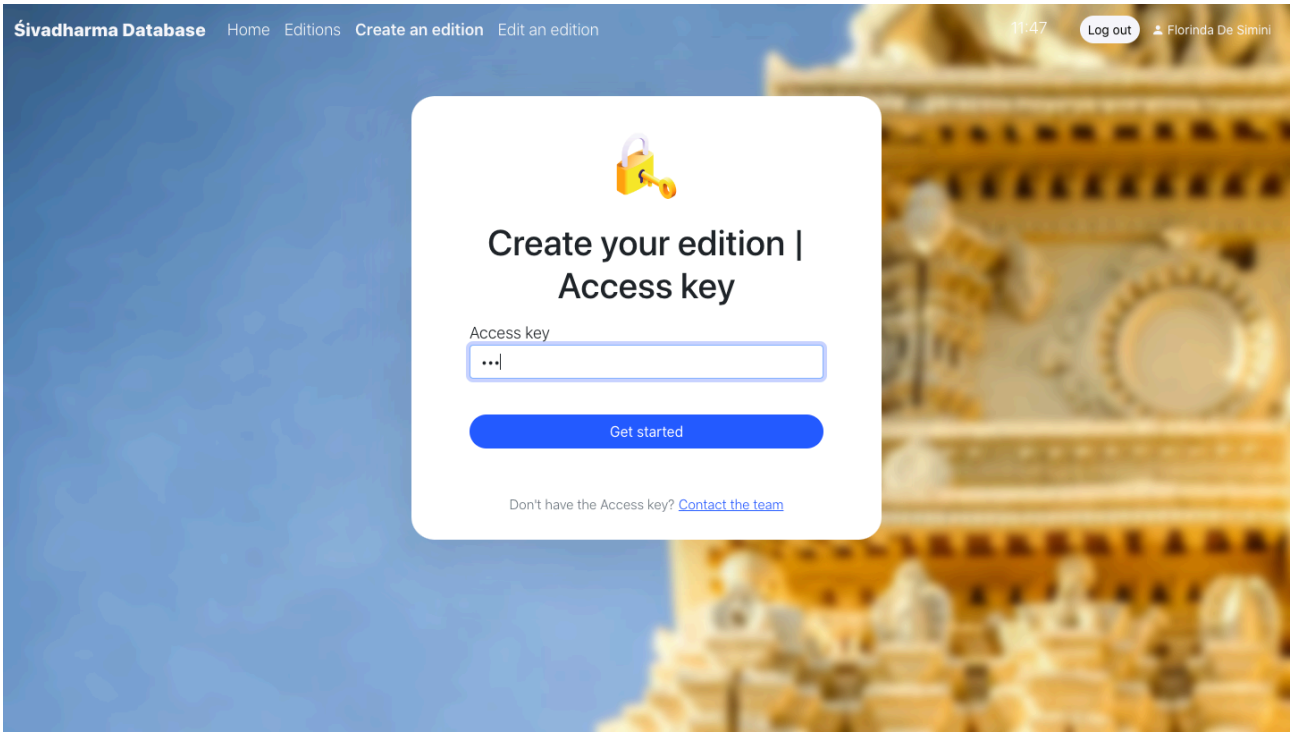


Fig. 51. *Śivadharma Database* access to the editing section.

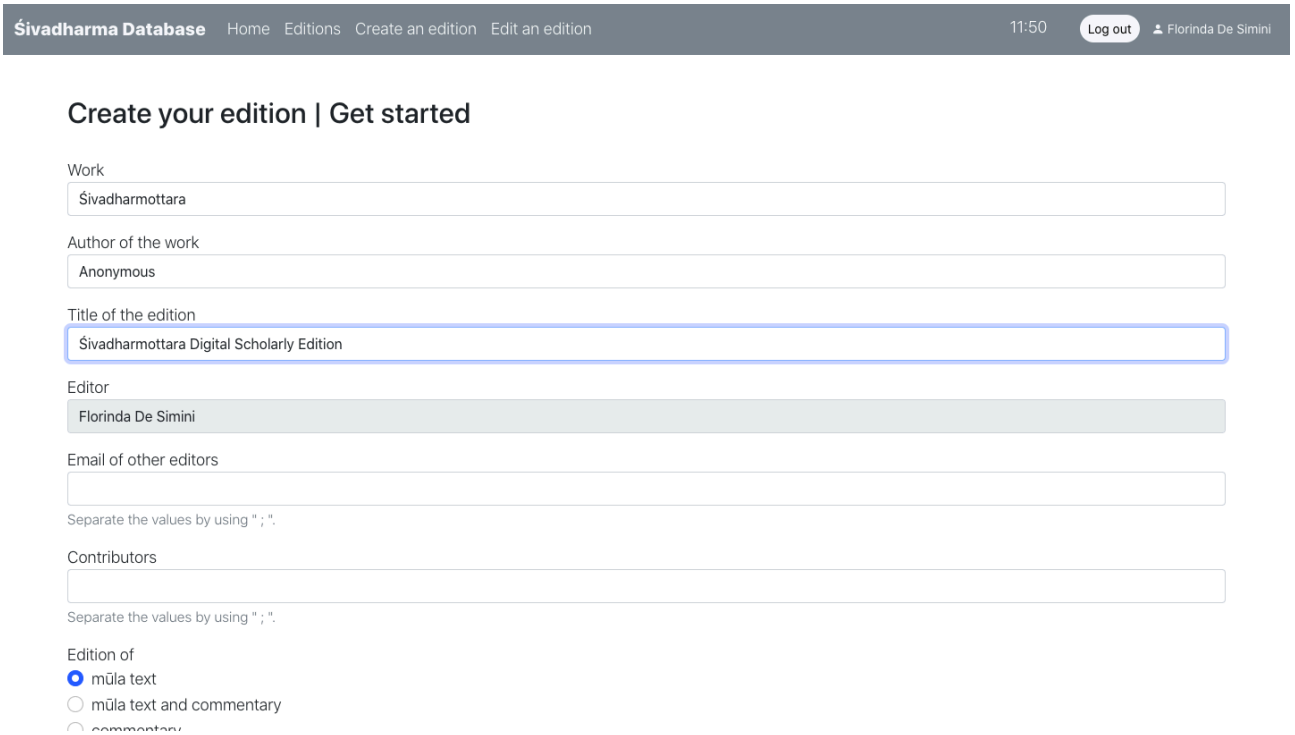


Fig. 52. *Śivadharma Database* editing section Get started.

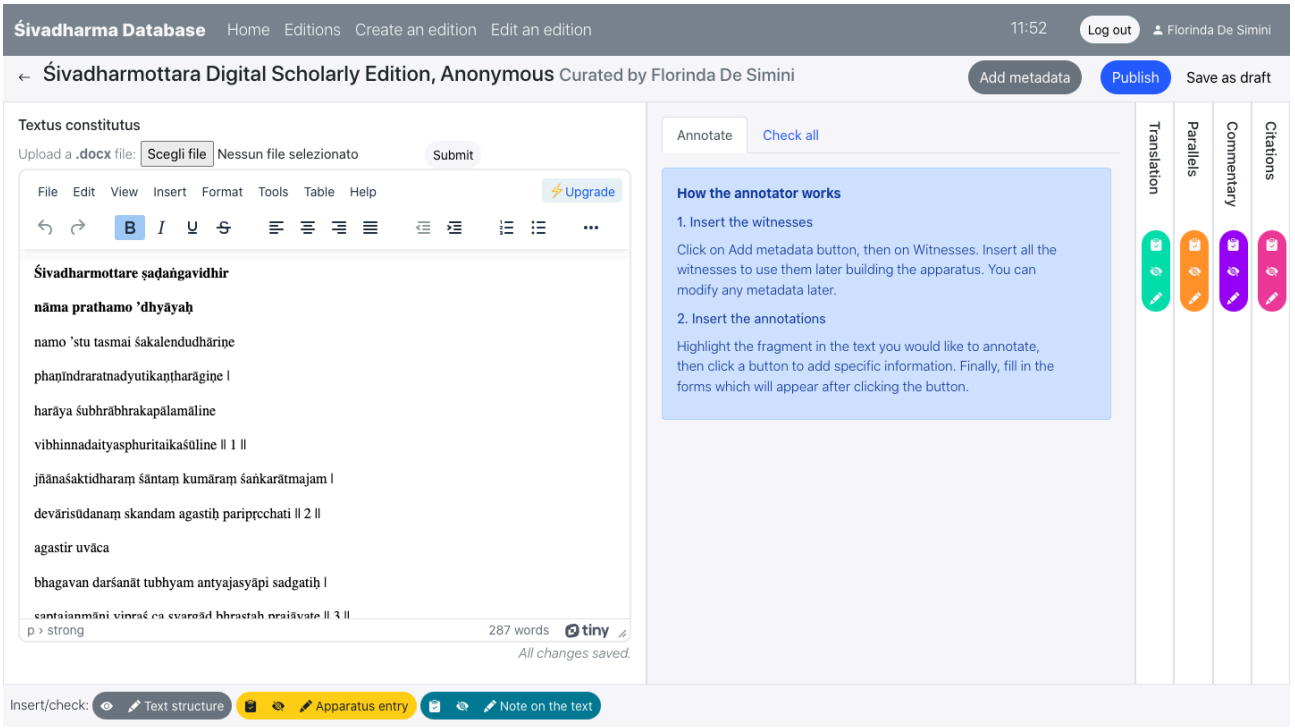


Fig. 53. *Śivadharma Database* editing section.

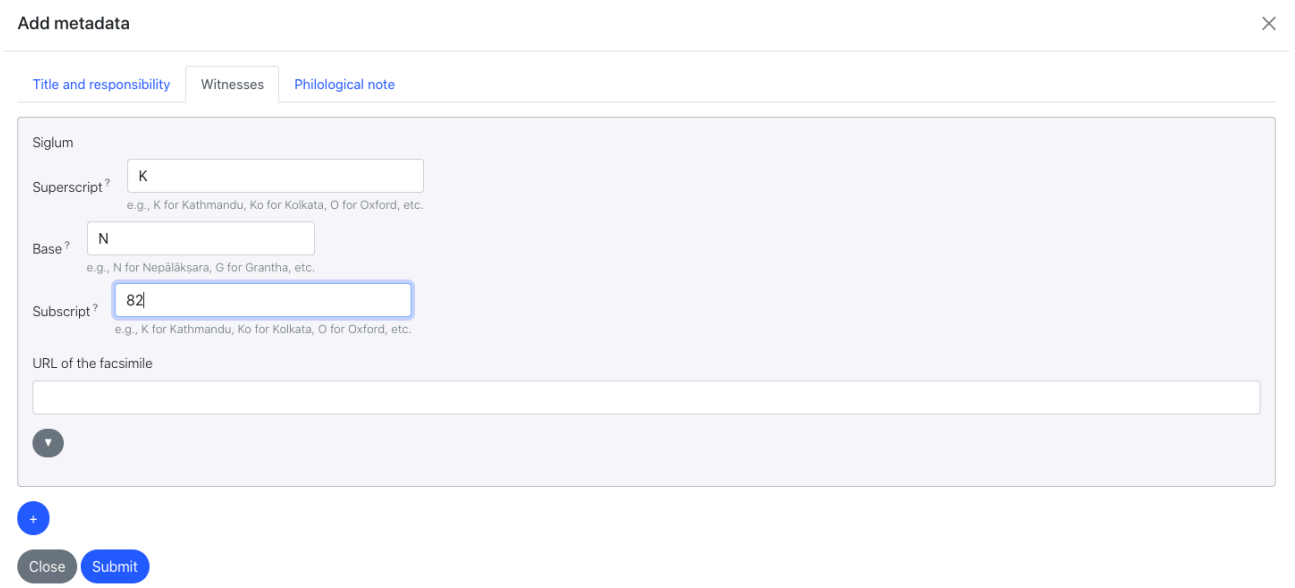


Fig. 54. *Śivadharma Database* platform for declaring the witnesses.

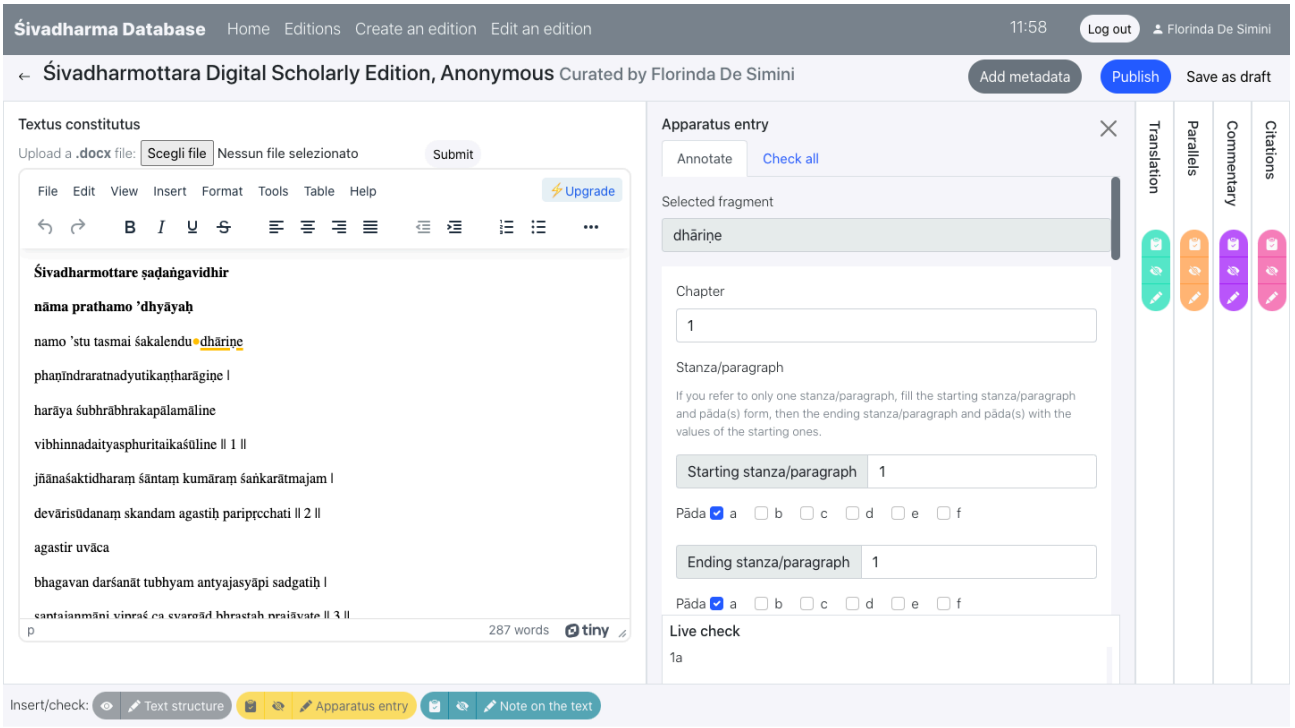


Fig. 55. *Śivadharma Database* editing section. Focus on the forms for creating the location of the *apparatus* entry.

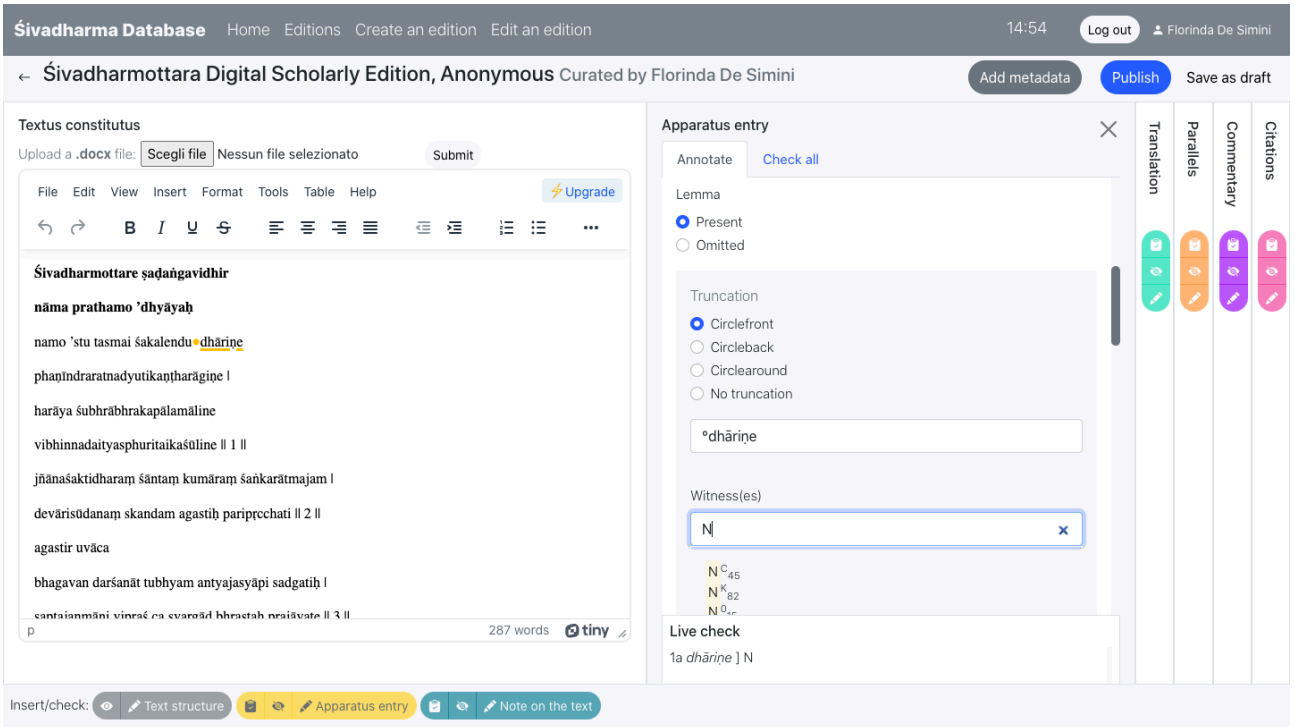


Fig. 56. *Śivadharma Database* editing section. Focus on the forms for creating the *lemma*.

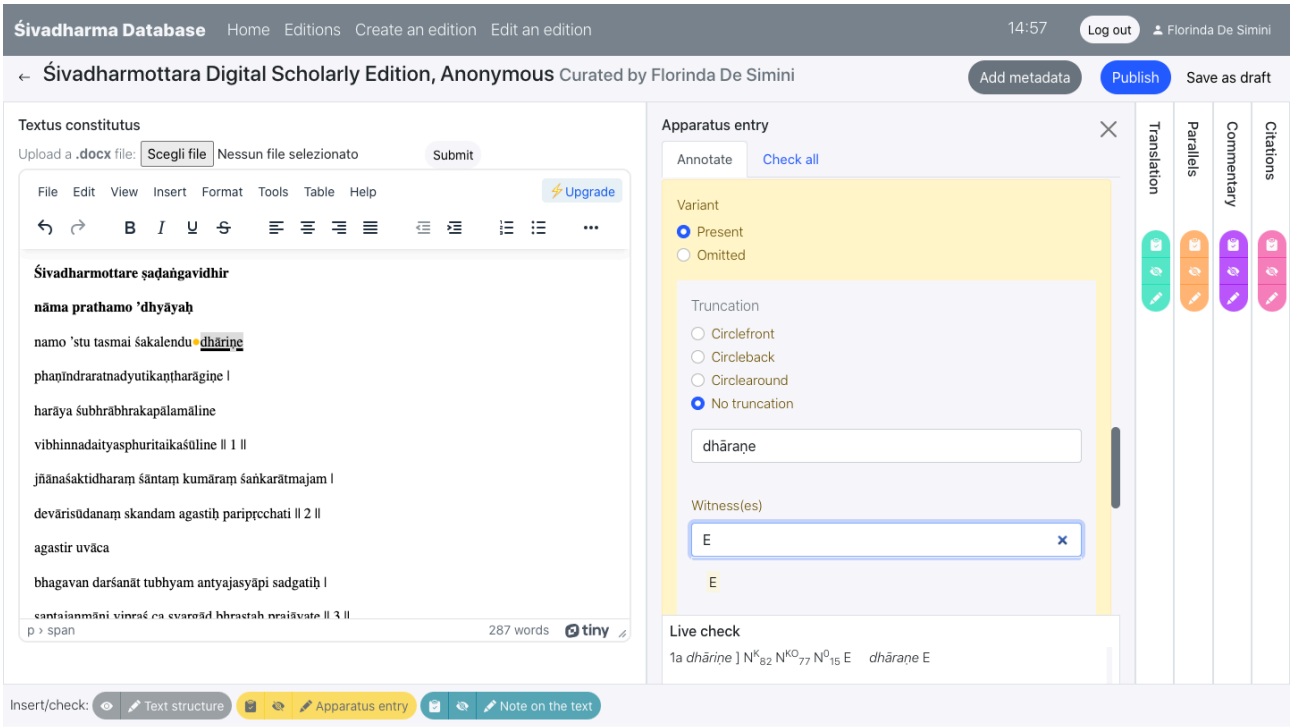


Fig. 57. *Śivadharma Database* editing section. Focus on the forms for creating a *variant*.

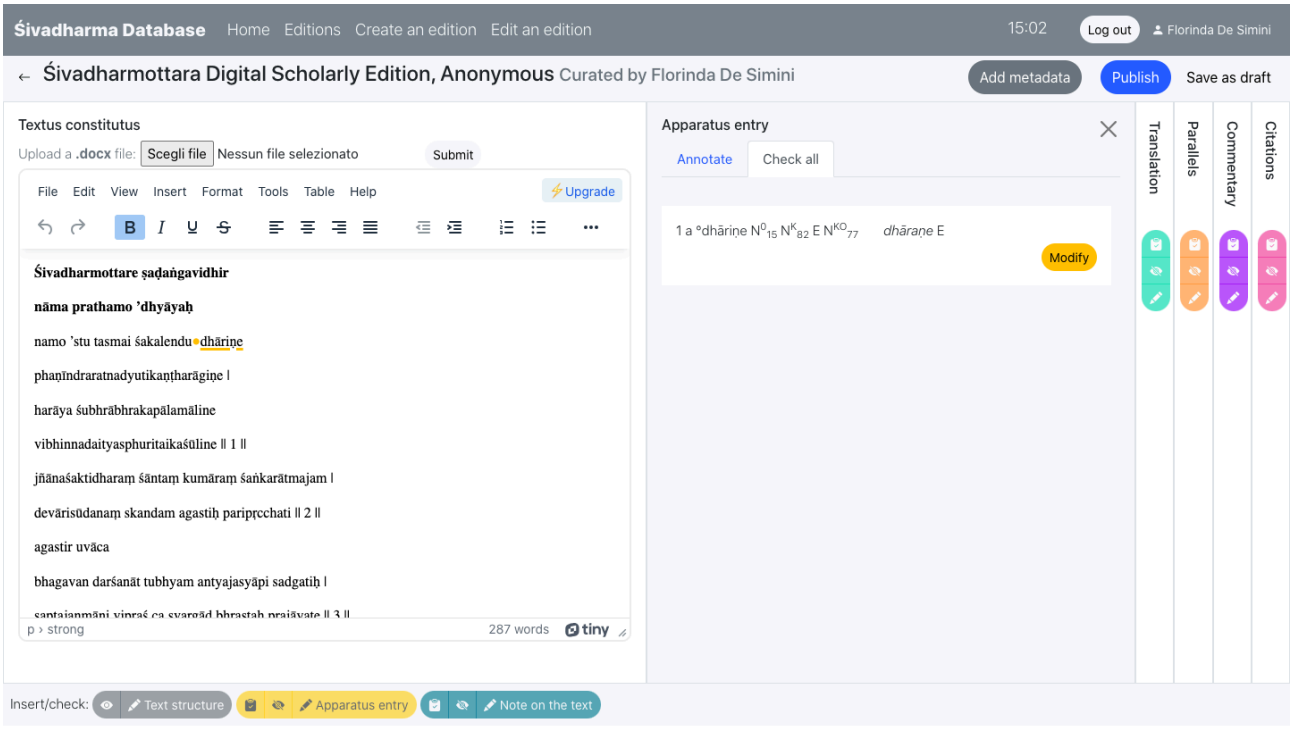


Fig. 58. *Śivadharma Database* editing section. Focus on the *Check all* function for modifying an *apparatus* entry.

6.8.3 *Śivadharma Database* reading section look and feel

In earlier discussions (Ch. 6.8.1, 6.8.2), we observed a distinct contrast in the colour scheme adopted by the welcome pages of the website and those of the editing section. The former employs vibrant colours to attract visitors to the site. At the same time, the latter adopts light colours to facilitate a clean and distraction-free environment for philologists to carry out their precision work that demands maximum concentration owing to its inherent complexity (Ch. 1).

The design scheme of the pages in the reading section balances the contrasting styles observed in the welcome pages and the editing section. While reading an edition demands significant effort, it is less intensive than the preparation phase. Therefore, we opted to blend the light colours of the editing section with a more pronounced utilisation of vibrant hues. Similar to the editing section, the colours distinguish the semantic differences of the various edition components. For instance, the button to access the *apparatus* is highlighted in yellow, consistent with the button in the editing section to create the *apparatus*.

Eventually, this section employs nested collapse elements (Fig. 60, 61, 62, 63, 64) extensively, which enable the display and concealment of contents upon the reader's request.

The list to access the editions is presented in Fig. 59, while Fig. 60, 61, 62, 63 and 64 show the reading section.

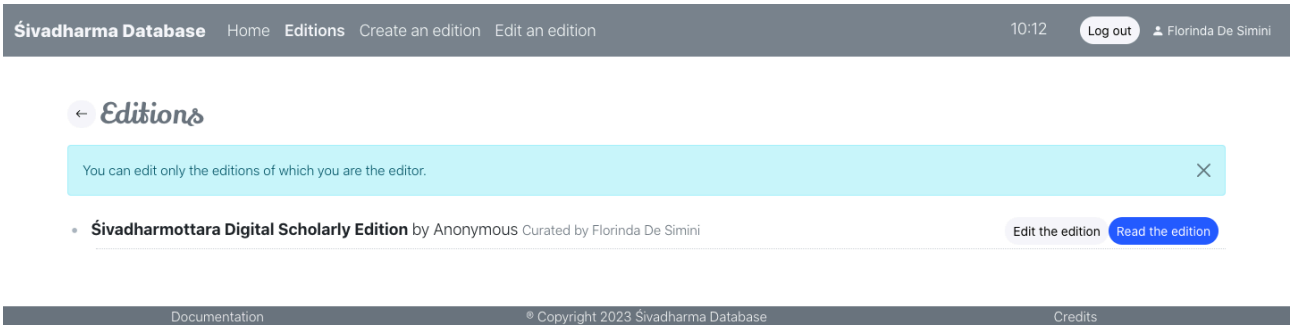


Fig. 59. *Śivadharma Database* reading platform. Focus on the *list of the published editions*.

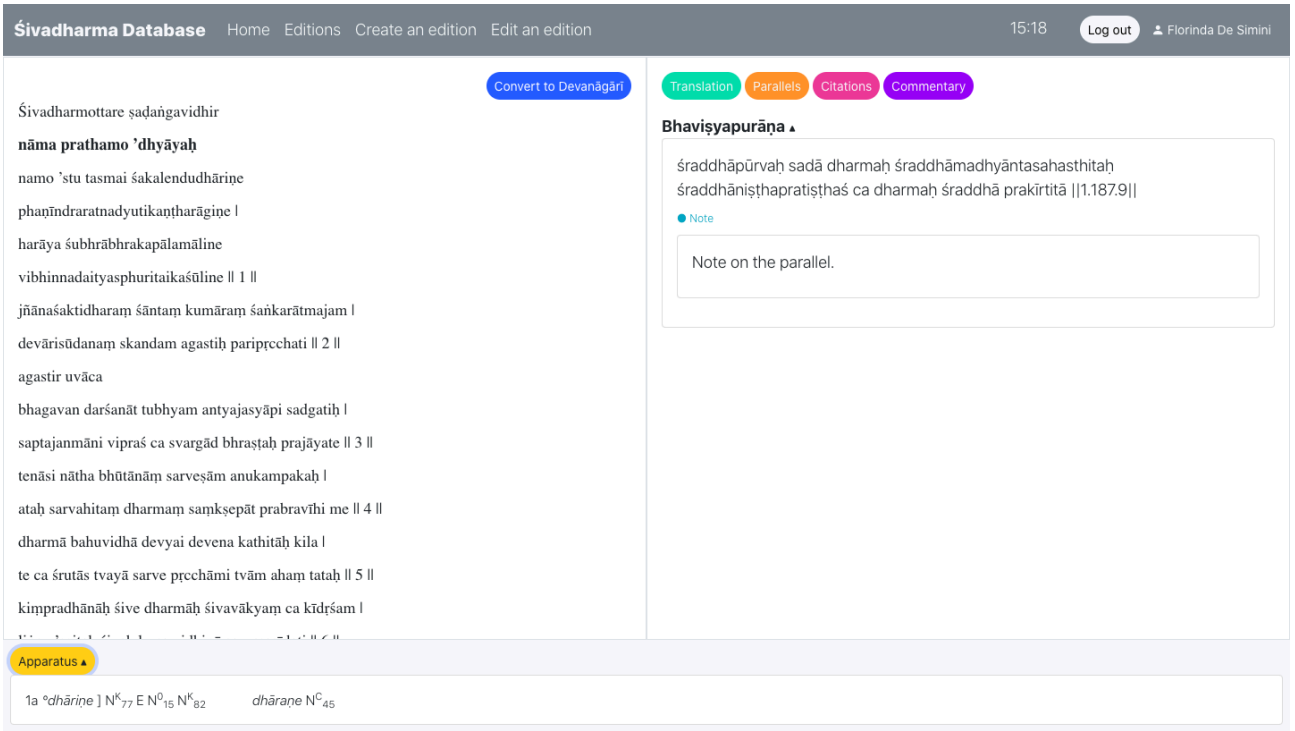


Fig. 60. *Śivadharma Database* reading platform. Focus on the *apparatus*.

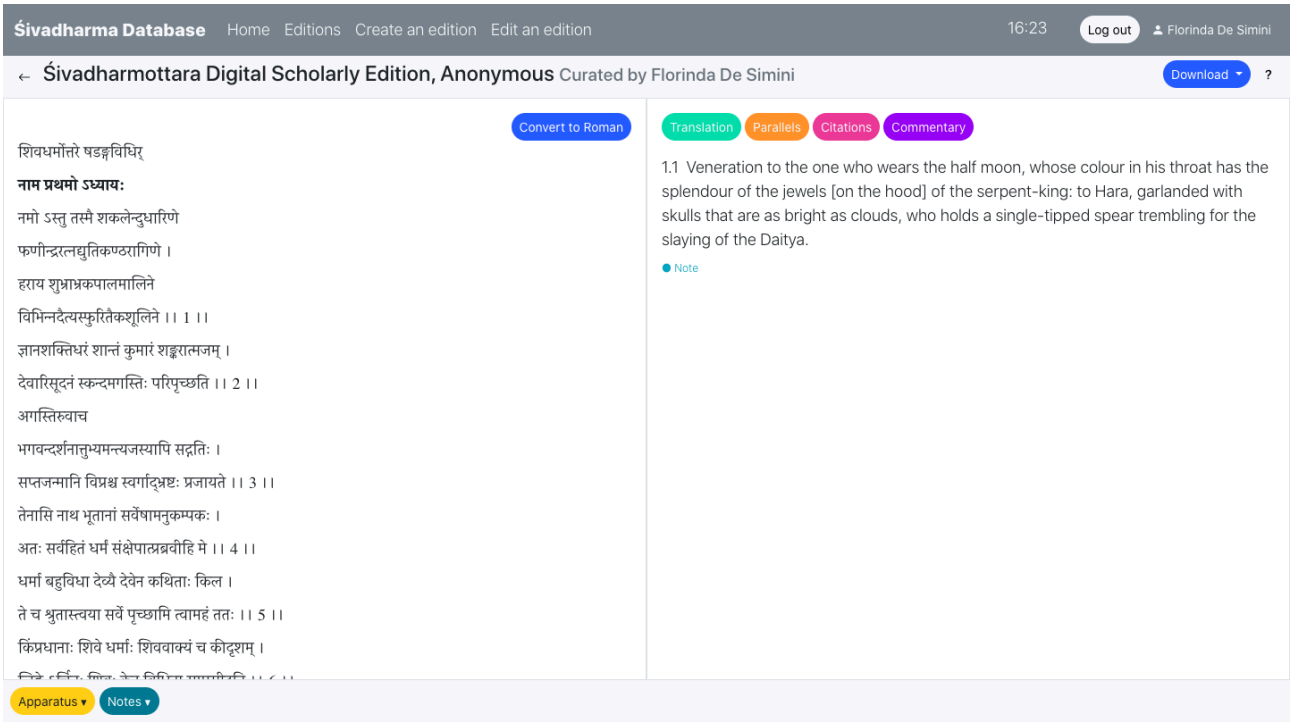


Fig. 61. *Śivadharma Database* reading platform. Focus on the conversion of the *textus* into *Devanāgāri*.

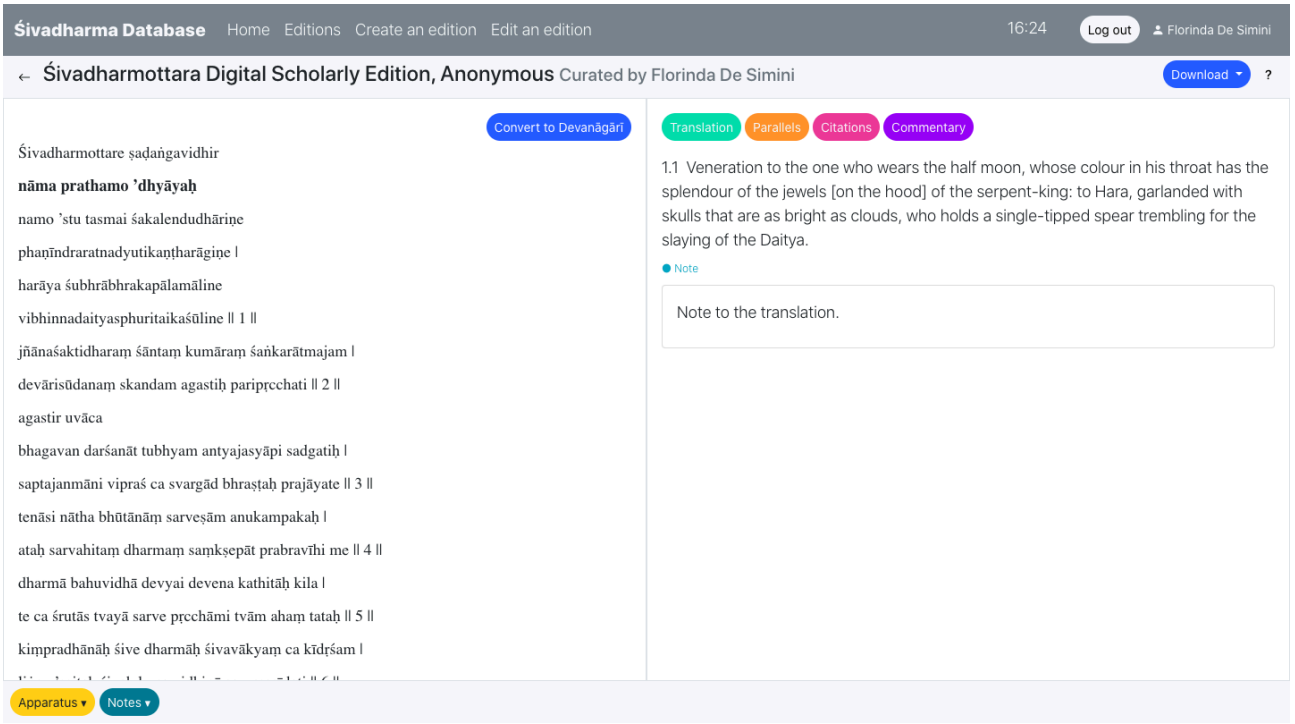


Fig. 62. *Śivadharma Database* reading platform. Focus on the *translation*.

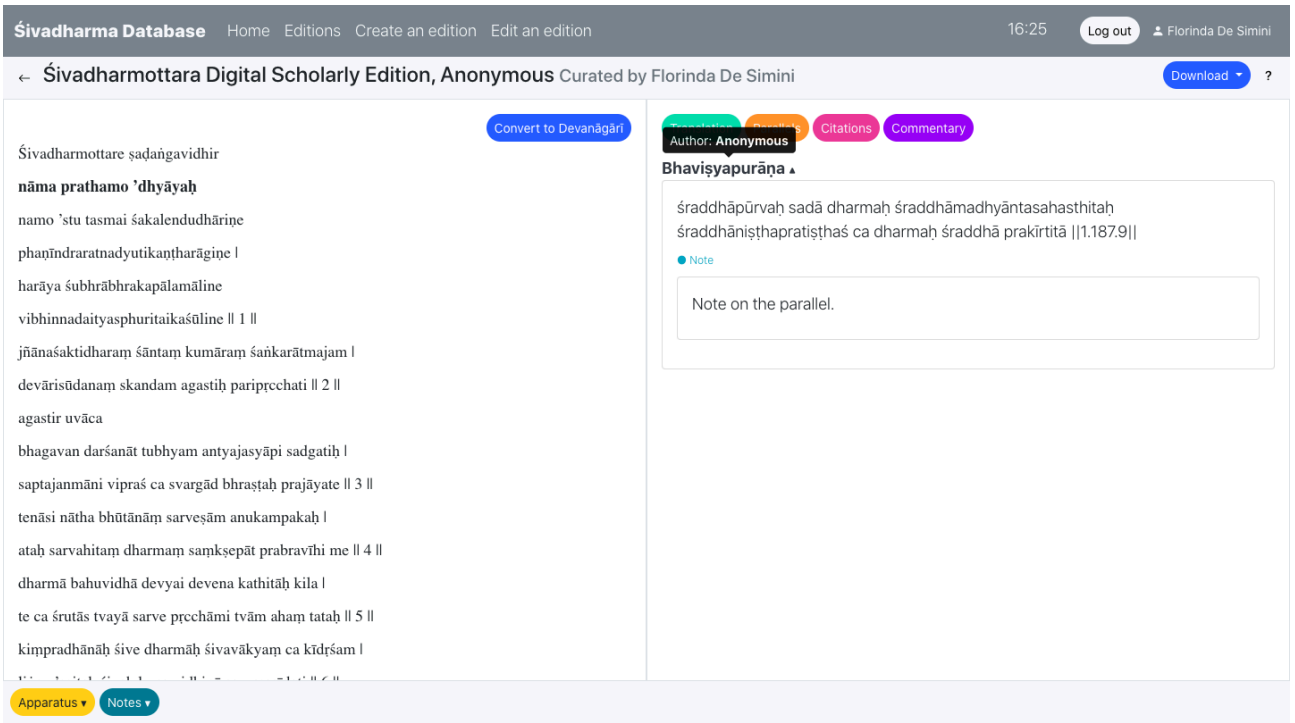


Fig. 63. *Śivadharma Database* reading platform. Focus on the *parallels*.

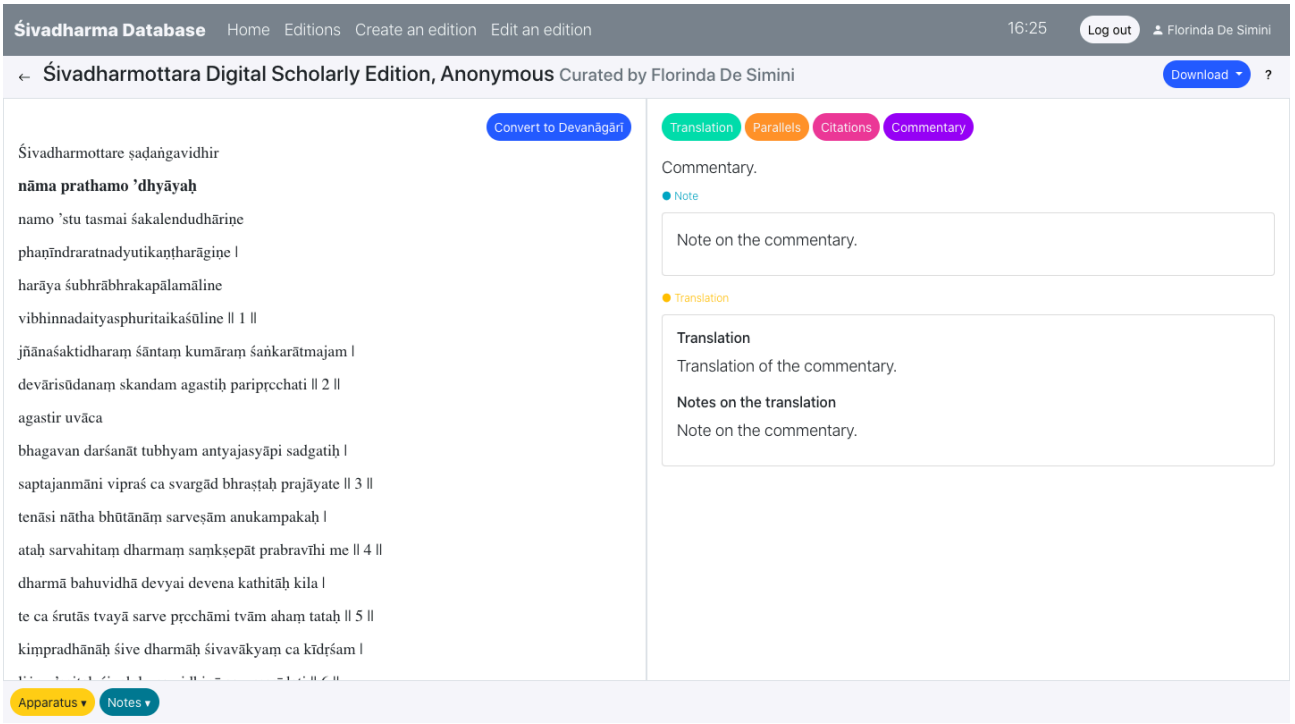


Fig. 64. *Śivadharma Database* reading platform. Focus on the *commentary*.

Chapter 7
Śivadharma Database: technological choices and development

One of the most challenging issues in Digital Scholarly Editing is determining the optimal *encoding system* for digital editions. Various approaches have been discussed in Ch. 3.3. The conventional approach is to use XML/TEI (Ch. 3.3.1), while a newer proposition is using RDF (Ch. 3.3.3) or relying on other *graph* systems (Ch. 3.3.5). Each has its advantages and disadvantages. XML/TEI is a popular and easy-to-learn solution. However, the vastness of its vocabulary limits the choice of descriptors by editors and reduces the interoperability of editions. Moreover, its limitations in representing complex textual phenomena, such as *overlapping hierarchies*, have been widely documented in the literature. On the other hand, RDF can overcome the inherent issues of an XML-based tree structure, but it requires extensive learning. Moreover, visualisation is outside the scope of RDF. Finally, an alternative approach could be to utilise stand-off annotations externally stored in a graph database, e.g., Neo4j. It effectively preserves the nonlinear structure of textual data, even if NoSQL databases are no standard options.

The *Śivadharma Database* proposes an annotation system based on HTML and a graph data structure relying on the Neo4j graph database to address these issues.

Starting with analysing the factors underlying the selection of the *graph* as the fundamental data structure in the *Śivadharma Database* (Ch. 7.1), we proceed to scrutinise the project data modelling decisions, with a focus on the textual phenomena observed in the *Śivadharma* case study expounded in Ch. 5 (Ch. 7.2). Subsequently, we thoroughly analyse the implementation choices regarding the *Śivadharma* annotation system (Ch. 7.3), which is primarily based on the Neo4j graph database. Finally, the *Śivadharma Database* CMS visualisation and comprehensive development (Ch. 7.4, 7.5) and its forthcoming undertakings (Ch. 7.6) are discussed.

7.1 Graph vs. Tree: *Graph* final choice

The examination conducted in Ch. 3 regarding the current techniques used for preparing digital editions is crucial in evaluating the implementation options that could be adopted for the *Śivadharma Database* project. By delving into the textual data modelling and encoding methodologies, we can deduce the most effective data structure that could serve as the foundation for the *Śivadharma* annotation system. Ch. 3.3 highlights that two structures are currently in use, namely the *tree* and the *graph*. Furthermore, it demonstrates that the *graph* is the most efficient data structure for formally representing textual data. Specifically, it excels in representing (1)

overlapping structures, as the graph is not strictly hierarchical (Ch. 3.3.3, 3.3.5), (2) relationships between elements, attributes, or attribute values, since it is possible to declare the meaning of each relation linking nodes in a graph, and (3) context where an element may be used with different meanings, as it is possible to set any property to nodes (Ch. 3.3.1).

It can be argued that this approach is likely viable only in the case of frequent occurrences of the reported phenomena. Otherwise, the adoption of the XML/TEI standard is a more cost-effective solution, although based on a tree structure. The analyses performed on the *Śivadharma* case study in Ch. 5.7 reveal that the text is an entirely non-linear entity, characterised by multifaceted, overlapping, and rich intra- and intertextual connections. The following case (Fig. 65) underscores the necessity for a more complex data structure than the tree-based one. Adopting a graph-based structure becomes a requirement, even for representing a basic textual phenomenon, such as the one shown in the example, i.e., the overlap between the *pāda* structure and the textual fragment annotated as *lemma*.

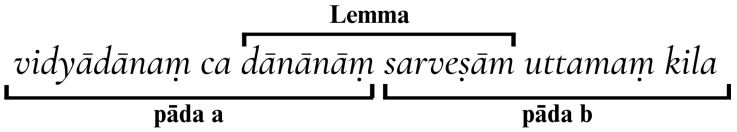


Fig. 65. *Overlapping* of *pāda* structures and *lemma*.

If the given overlapping phenomenon (Fig. 65) were encoded in XML/TEI following the *Parallel Segmentation* method, ensuring the correct nesting of tags would not be feasible. On the contrary, the TEI *Double End-Point Attachment* method[142] uses a syntactical trick to represent non-tree structures. Although this method enables circumvention of the issue, thus effectively providing a practical solution storing *stand-off* annotations, on the other hand, it conceals structural information about a secondary structure that later becomes difficult to retrieve through complex XSLT stylesheets (Di Iorio, Peroni and Vitali 2011, 1697). In contrast, by treating each textual fragment subject to annotation as a distinct node that multiple explicit relations can connect in a graph, we can elegantly and explicitly represent the overlap and make it easily searchable.

[142] The TEI Double End-Point Attachment method implies the use of `<milestone/>` and/or `<anchor/>` empty elements to indicate the boundaries of the content which is the *lemma* of an external `<app>` element. https://www.tei-c.org/release/doc/tei-p5-doc/it/html/TC.html#TCAPDE.

7.2 Data modelling in the *Śivadharma Database*

The data structure employed for text annotation in the *Śivadharma Database* is modelled graph-based. In the following sections, we follow the workflow to prepare an edition to delve into the modelling of the features we intend to support in the CMS (Ch. 6.5).[143] Many examples regarding textual features derive from authentic sources, as detailed in Ch. 5.7, which were instrumental in validating the presence of overlapping instances and determining the constituents to include in the editions. It is important to underline that this modelling was carried out on the basis of the visualisation designed (Ch. 6) for the reasons described in Ch. 3.4.1.

7.2.1 *Editor* entity

The first step to creating an edition in the *Śivadharma Database* is the user's registration on the website. Once registered, the user is formalised as an editor entity, which entails possessing the following properties: (1) `<id>` for unique identification, (2) `email`, (3) `name`, and (4) `password`. At present, there are no relationships established in the graph (Fig. 66).

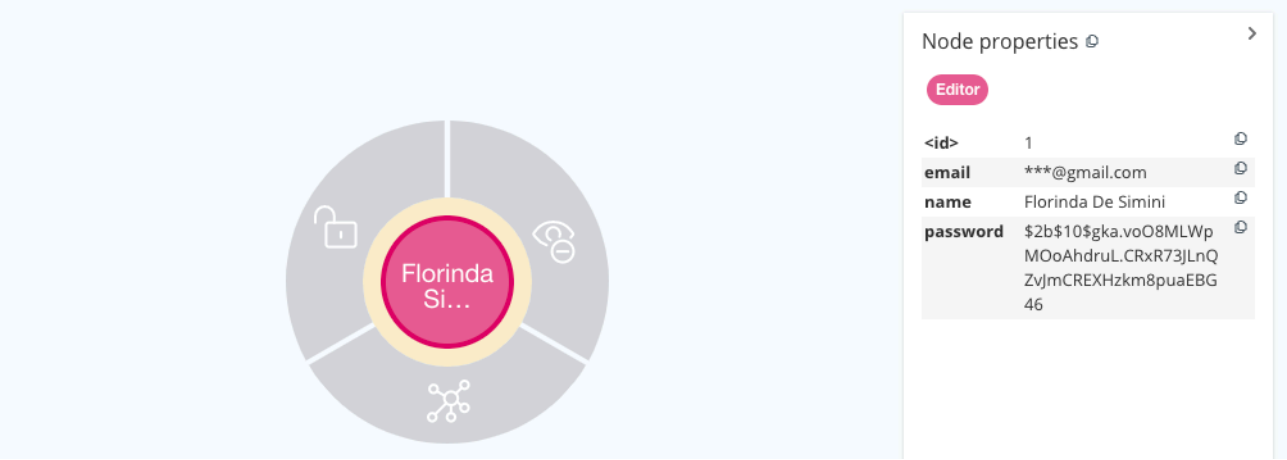


Fig. 66. Modelling of the *editor* entity as a node in a graph.

In the example in Fig. 65, *Florinda De Simini* is an *editor* node.

[143] The query language employed is Cypher, which is a syntax supported by Neo4j. The graphs have been implemented using Neo4j.

7.2.2 *Work*, *edition*, and *author* entities

After login, the editor can initiate the process of edition creation. From a modelling perspective, a *work* node is initially created, as a single work can have multiple editions that distinct editors may edit. It has the following properties: (1) `<id>`, and (2) `title`. In addition, it is linked to another node entity, i.e., *edition*, through the `HAS_MANIFESTATION` relation.[144] This entity represents the specific edition that the editor is going to prepare and it is described by the following properties: (1) `<id>`, (2) `authorCommentary`, (2) `editionOf`, (3) `publishType`, and (4) `title` . In addition, the *work* node is linked to the *author* node via the `WRITTEN_BY` relation. The *author* has these properties: (1) `<id>`, and (2) `name`. Finally, the `IS_EDITOR_OF` relation connects the node created in the previous step (Ch. 7.2.1), i.e., *editor*, to the *edition* (Fig. 67).

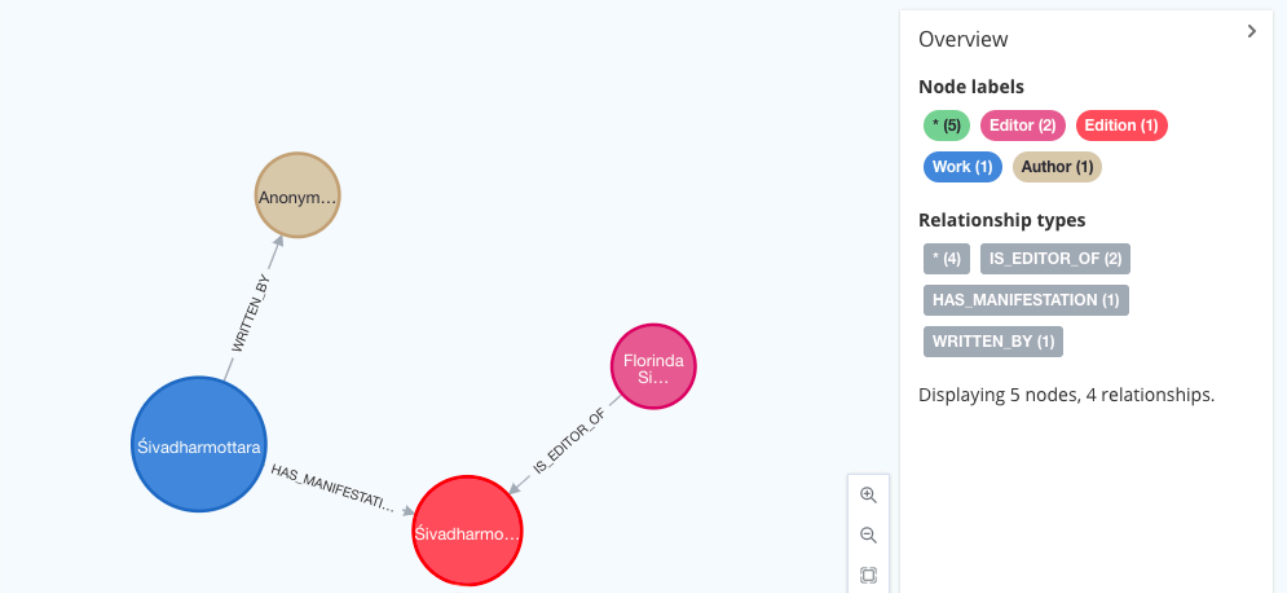


Fig. 67. Modelling of the *work*, *edition*, and *author* entity as nodes and related relations in a graph.

In the example in Fig. 66, the *Śivadharmottara* work, written by an anonymous author, has the edition curated by *Florinda De Simini* as manifestation.

7.2.3 *Witness* entity

The subsequent action an editor will likely perform involves the witness statement on which the edition will be based. We model these entities as follows. For each witness, a *witness* node connects

[144] The reference model is FRBR (Peroni 2015): http://www.sparontologies.net/ontologies/frbr.

to the *edition* node through the `USED_IN` relation. The *witness* node has the following properties to catalogue the witness: (1) `<id>`, (2) `antigraph`, (3) `author`, (3) `authorRecord`,[145] (3) `bibliography`, (4) `binding`, (5) `classmark`, (6) `colophon`, (7) `condition`, (8) `date`, (9) `dimensionH`, (10) `dimensionW`, (11) `editions`, (12) `explicit`, (13) `extent`, (14) `finalRubric`, (15) `foliation`, (16) `format`, (17) `incipit`, (18) `initialRubric`, (19) `language`, (20) `location`, (21) `marginalia`, (22) `material`, (23) `notes`, (24) `people`,[146] (25) `provenance`, (26) `repository`, (27) `scripts`, (28) `secondaryLiterature`, (29) `siglum`, (30) `siglumTex`, (31) `state`, (32) `structuralTypology`, (33) `title`, and (34) `urlFacsimile`. By extracting these data from the database, generating and displaying a comprehensive cataloguing description for each witness will be possible (Fig. 68)

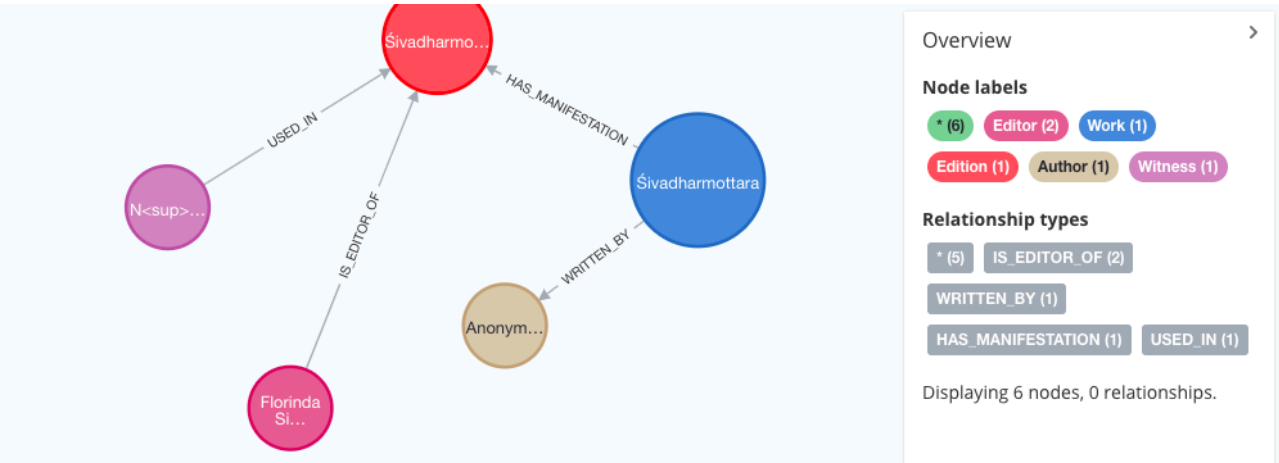


Fig. 68. Modelling of the *witness* entity as node and related relations in a graph.

In the example in Fig. 68, a node of a witness used in the creation of the edition of the work *Śivadharmottara* edited by Florinda De Simini is added.

7.2.4 *File* entity

After the witnesses have been inserted, the editor proceeds with the actual reconstruction of the text by typing it into the designated textarea or uploading it (Ch. 6.5.3). The *textus* is a *file* entity containing a sequence of characters. This entity is a node characterised by its (1) `<id>` and (2) `name` properties. It connects to other nodes through two different relationships: (1) with the `IS_ITEM_OF`

[145] The `authorRecord` property refers to the author of the catalogue record.

[146] The `people` property refers to the people related to the witness, e.g., a scribe.

relationship to the *edition* node, and (2) with the `PRODUCED_BY` relationship to the *editor* node (Fig. 69)

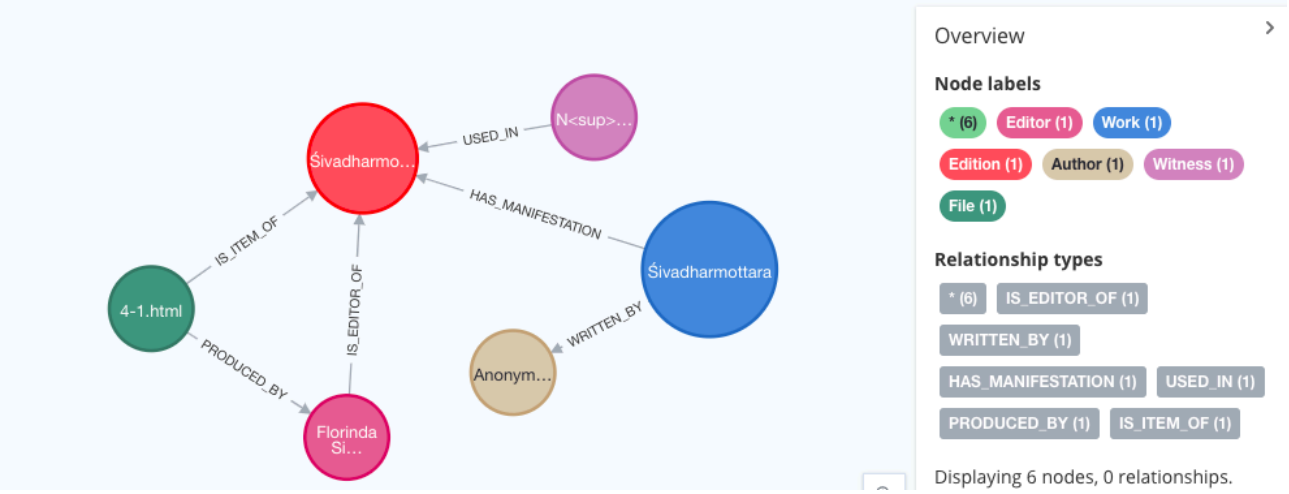


Fig. 69 Modelling of the *file* entity as node and related relations in a graph.

In the example in Fig. 69, a node of the file containing the edition of the work *Śivadharmottara* edited by Florinda De Simini is an item of the edited work *Śivadharmottara*.

7.2.5 Entities related to an *apparatus entry*

After the witnesses have been inserted, the editor proceeds with the actual reconstruction of the text, documenting their interpretation according to the scheme provided in the *apparatus entry* module in Fig. An *apparatus* entry is modelled as follows (Fig. 70)

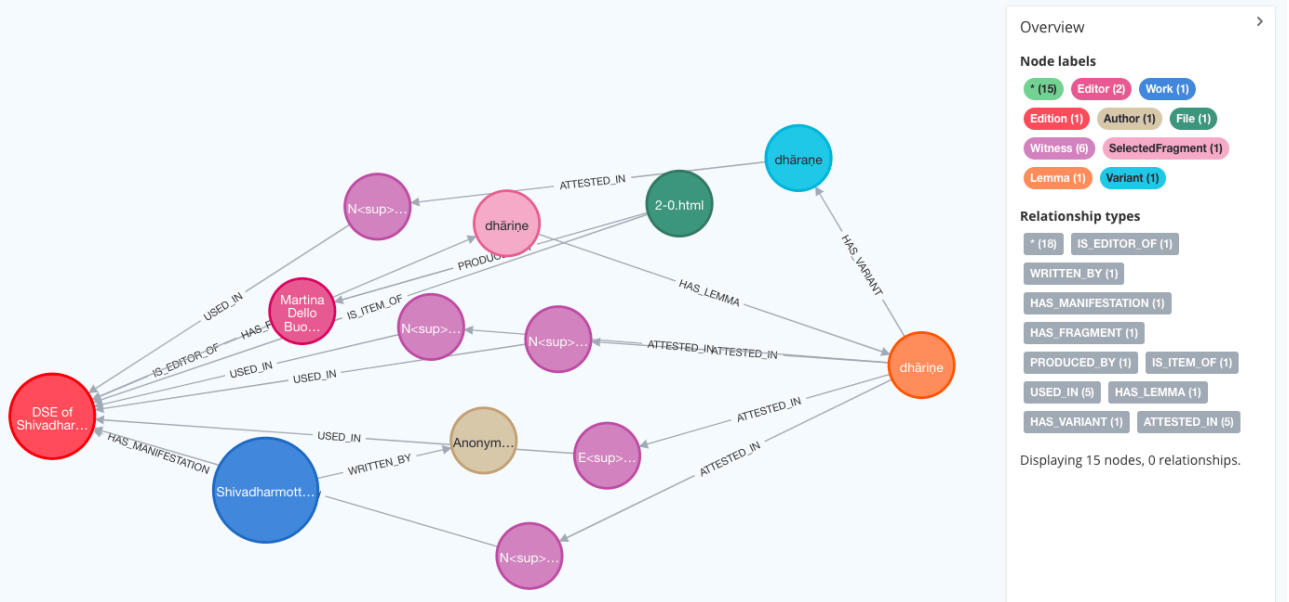


Fig. 70. Modelling of an *apparatus entry* in a graph.

As shown in Fig., the selected fragment `:selectedFragment` is the subject of the relation `:HAS_LEMMA`. The object `:lemma` is, in turn, the subject of the relation `:HAS_VARIANT`. Then both `:lemma` and `:variant` are subjects of the relation `:ATTESTED_IN` to link them to their witnesses.

The properties of a *lemma* are: (1) `<id>`, (2) `idLemma`, (3) `notes`, (4) `truncation`, and (5) `value`. Instead, the properties of a *variant* are: (1) `<id>`, (2) `idVariant`, (3) `notes`, (4) `truncation`, and (5) `value`.

The result of this data structuring corresponds to the device entry shown in Fig. 3.

7.2.6 Entities related to a *parallel*

The concept of a *parallel*, which could be added to an edition as the following presented edition components (Ch. 7.2.7, 7.2.8, 7.2.9, 7.2.10), is formally expressed as follows. The `:selectedFragment` node is connected to a `:parallel` via the `:HAS_PARALLEL` relation. At the same time, the source work of the parallel and its author are made explicit (Fig. 71).

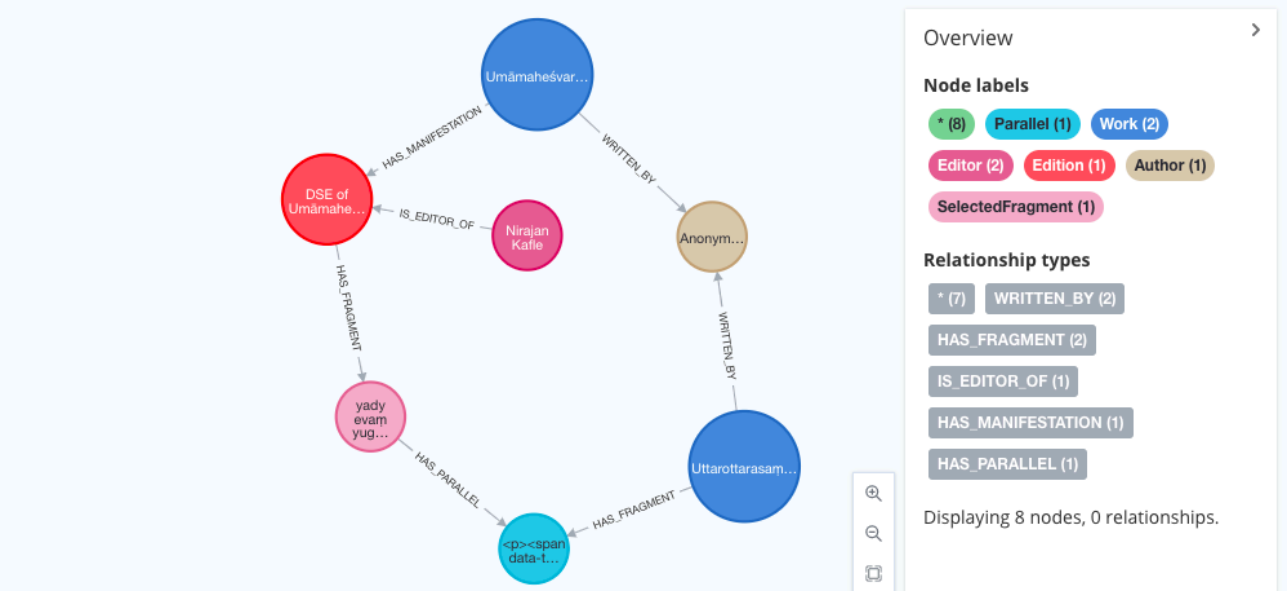


Fig. 71. Modelling of a *parallel* in a graph.

The example above refers to the parallel between the works the *Umāmaheśvarasaṃvāda* and the *Uttarattaramahāsaṃvāda* identified in Ch. 5.7.2.

The properties related to the *parallel* are: (1) `<id>`, (2) `bookChapter`, (3) `bookStanza`, (4) `idAnnotation`, (5) `note`, (6) `padaStart`, (7) `stanzaStart`, (8) `stanzaEnd`, and (8) `value`.

7.2.7 Entities related to a *citation*

The instance employed to exemplify citation modelling is extracted from Ch. 5.7.1. In the literary work *Vṛṣasārasaṃgraha*, a distinct character named *Vaiśampāyana* is referenced. To formally convey this association, we utilise the quote entity *vis-à-vis* the `:selectedFragment` entity, that is, the name of the character, employing the relationship `:IS_A_CITATION_OF` (Fig).

The citation entity has the following properties: (1) `<id>`, (2) `idAnnotation`, (3) `stanzaStart`, (4) `stanzaEnd`, (5) `typeOfCitation`, and (6) `value` (Fig. 72).

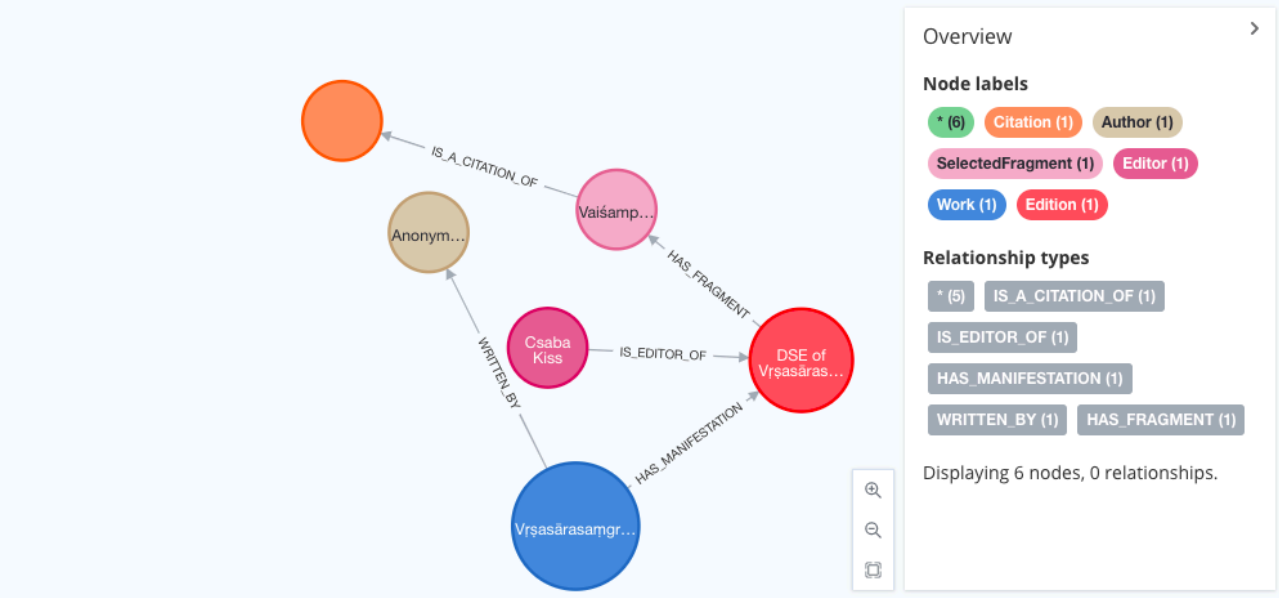


Fig. 72. Modelling of a *citation* in a graph.

7.2.8 Entities related to a *translation*

Translation is an integral component of numerous editions. For instance, in Chapter 5.7.5, the correlation between *textus* and translation is demonstrated through the translation of *Śivadharmottara* into Tamil. This relation is formalised as follows. The `:selectedFragment`, i.e., the *textus*, is in relation with a node :translation via the `:HAS_TRANSLATION` relation (Fig.).

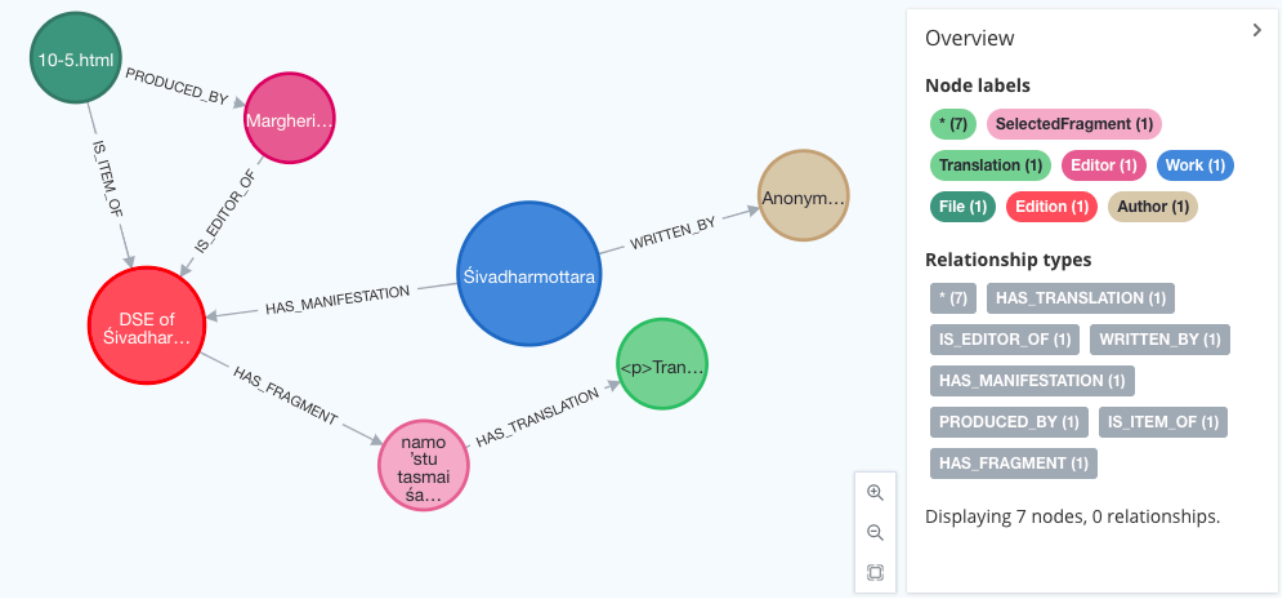


Fig. 73. Modelling of a *translation* in a graph.

The *translation* entity properties are (1) `<id>`, (2) `idAnnotation`, (3) `note`, and (3) `value`.

7.2.9 Entities related to a *note*

The fundamental association between a textual string and its corresponding explanatory note is established through an arc in the graph. Specifically, the `:selectedFragment` is connected to a `:note` node via the `:IS_DESCRIBED_IN` relationship (Fig. 74).

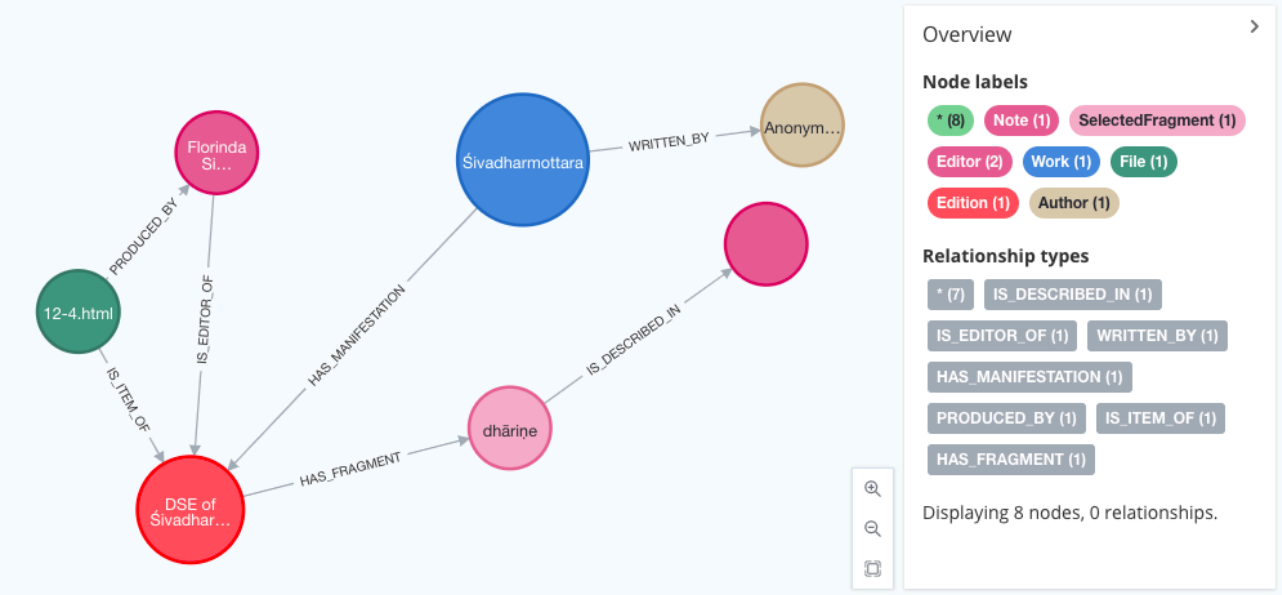


Fig. 74. Modelling of a *note* in a graph.

The *note* entity properties are (1) `<id>`, (2) `idAnnotation`, and (3) `value`.

7.2.10 Entities related to a *commentary*

On the example of *textus*-commentary relationship reported in Ch. 5.7.4, the formalisation of this relationship in the *Śivadharma Database* is reported. The `:selectedFragment` subject to a comment is linked to a node :commentary via the `:IS_COMMENTED_IN` relation (Fig. 75).

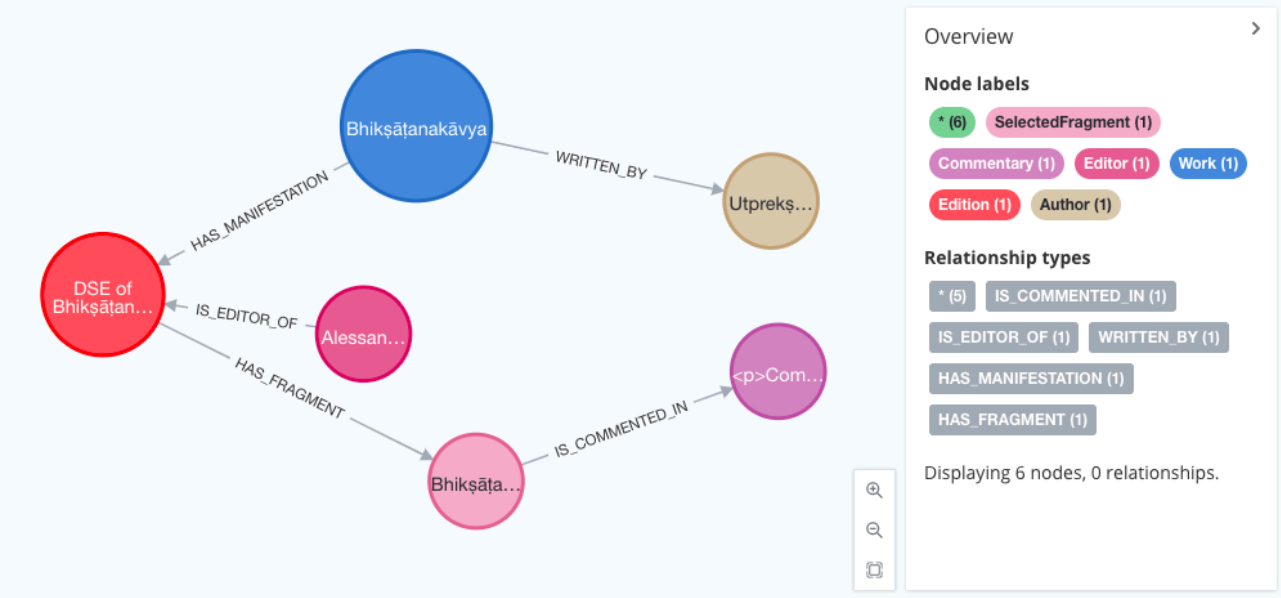


Fig. 75. Modelling of a *commentary* in a graph.

The *commentary* entity properties are (1) `<id>`, (2) `idAnnotation`, (3) `note`, (4) `translation`, (5) `translationNote`, and (6) `value`.

7.2.11 *Philological note* entity

Upon completion of the work, a philologist may document the practices employed in preparing the edition. This feature is incorporated within the *Śivadharma Database*, whereby the `:edition` entity is linked to the `:philologicalNote` node via the `:IS_DOCUMENTED_IN` relationship. Similar to the edition file (Ch. 7.3.3), only the filename containing the philological note is stored in the database to conserve storage (Fig. 76).

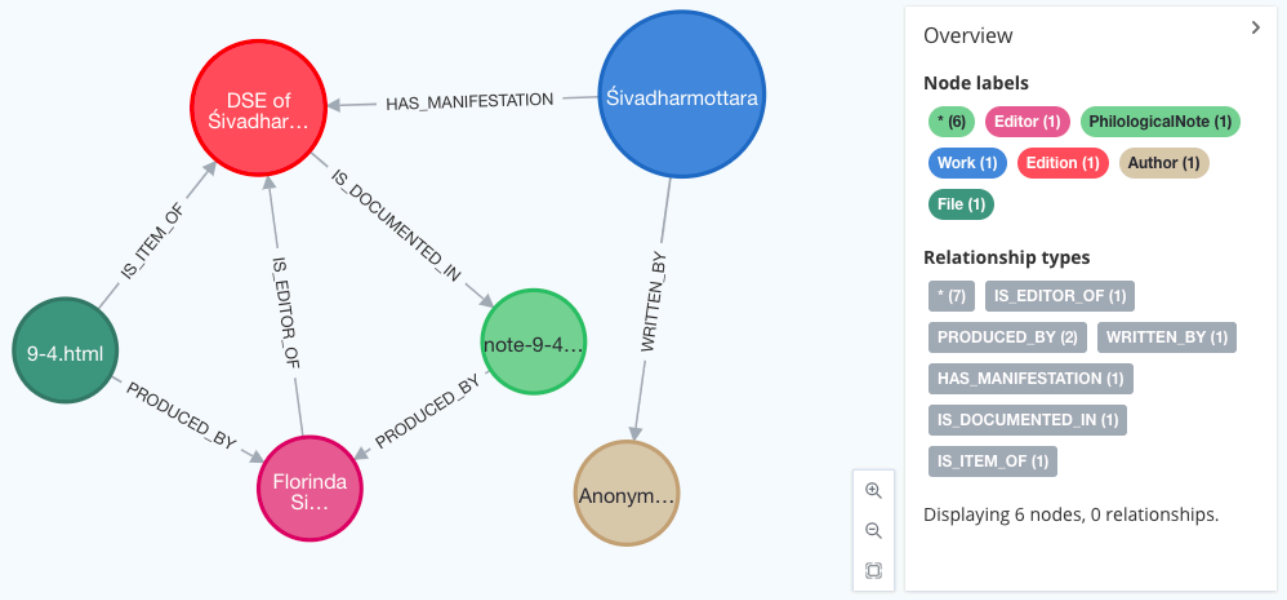


Fig. 76. Modelling of a *philological note* in a graph.

7.3 Annotation system development

The *Śivadharma* annotation system design puts forth the use of HTML and graph data structure as the foundation for encoding digital editions (Ch. 7.3.1). This approach is practical for multiple reasons. Firstly, HTML is an ideal choice for visualisation purposes on the web, as it does not require specific conversions. Particularly when the digital edition is a short-term project, HTML encoding can effectively fulfil the basic requirement of displaying the edition.[147] Additionally, it can serve as the basis for conversion into other languages. On the other hand, the graph data structure allows the formal representation of complex and overlapping textual phenomena. This solution contrasts with the traditional choice of encoding editions in XML/TEI, where formally complex and overlapping structures are represented through workarounds and therefore not adequately valued (Ch. 7.1).

7.3.1 Algorithm to create textual annotations

HTML is the starting point for modelling *Śivadharma* edition data as a graph. More precisely, an HTML-identified textual fragment is employed as an atomic unit for multiple relations within an external graph database. Specifically, the CMS system is built upon a Neo4j graph database. This approach creates a *stand-off* layer for editors' interpretation, thereby enhancing text readability and improving data interoperability.

[147] The visualisation aspect should not be underestimated, as argued by Turska, Cummings, and Rahtz (2016, 2), as it is what scholars, funding bodies, students, and the general public actually see.

Retracing the workflow of preparing an edition in the *Śivadharma Database* (Ch. 6.5) from a technical standpoint, the editor inputs the *textus* into the CMS system through the designated textarea. Upon input, the system automatically saves the *textus* as a file in the .html format. Only the filename is stored in the database and linked to other entities previously entered into the database through input, as shown in Ch. 7.2. The filename is unique and generated automatically. It includes the ID assigned to the *editor* entity (Ch. 7.2.1), followed by a hyphen, and the ID assigned to the *edition* entity (as discussed in Chapter 7.2.2). Storing solely the filename in the database results in a space-saving outcome.

At this stage, the editor initiates the actual annotation process, highlighting a relevant textual fragment subject of annotation. This fragment is automatically included within a `<span>` element, corresponding to an HTML milestone identified by a unique ID assigned to the milestone `@data-annotation` attribute. This ID serves as a link between the HTML text and the database, as the annotation node in the database related to that specific fragment refers to the fragment ID via the `idAnnotation` property. Indeed, the editor utilises the form to assign specific information to the fragment, creating the *annotation*. The annotated fragment is sent to the database. It constitutes the fundamental unit to which other nodes, corresponding to the information sent via forms as modelled in Ch. 7.2, are connected. Fig. 77 summarises the algorithm working.

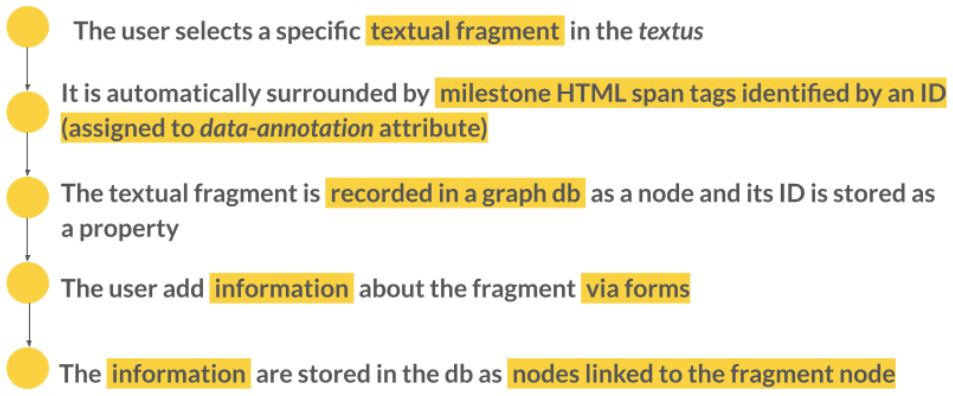


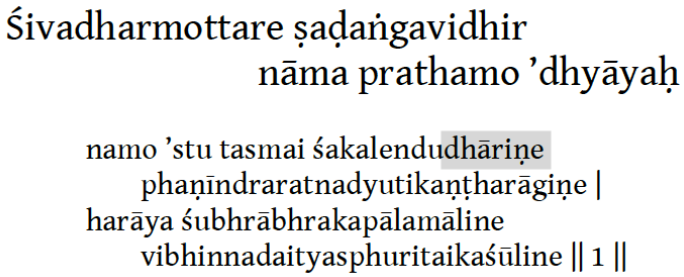
Fig. 77. *Śivadharma* annotation algorithm.[148]

[148] Excerpt from the AIUCD 2023 presentation *Śivadharma Database CMS. HTML and graph as a starting point for digital editions* by Martina Dello Buono and Francesca Tomasi: https://doi.org/10.6084/m9.figshare.24433486.v1.

7.3.2 Algorithm to create textual annotations: an example

The *Śivadharma Database* annotation system allows the creation of complex edition components such as *apparatus*, notes, parallels, citations, translations, and more, by leveraging the *graph* data structure. By treating *textual fragments* as atomic units to associate with multiple relations, building and visualising the edition components becomes accessible. A possible technical workflow to create an *apparatus entry* is presented below.

1 **Identification and selection of textual fragments to annotate**. The user identifies and selects a textual fragment in the *textus*, as shown in Fig. 78.


Śivadharmottare ṣaḍaṅgavidhir
nāma prathamo 'dhyāyaḥ

namo 'stu tasmai śakalendudhāriṇe
phaṇīndraratnadyutikaṇṭharāgiṇe |
harāya śubhrābhrakapālamāline
vibhinnadaityasphuritaikaśūline || 1 ||

Fig. 78. Highlighted textual fragment by the editor.[149]

The textual string *dhārine* is automatically annotated in the HTML file as follows (Listing 9).

```
<p>namo &rsquo;stu tasmai śakalendu<span data-type="milestone"
data-subtype="apparatus" data-start="start"
data-annotation="#apparatus3d2fe13898cbe1685947509772"></span>dhāriṇe<span
data-type="milestone" data-subtype="apparatus" data-end="end"
data-annotation="#apparatus3d2fe13898cbe1685947509772"></span></p>
```

Listing 9. Sample *apparatus entry* and corresponding reference string in the text.[150]

The desired output that the editor aims to achieve is the following *apparatus* entry in Fig. 79.

[149] Sample *textus* of *Śivadharmottara*, edited by Florinda De Simini (Ch. 5).

[150] *Ibid.*

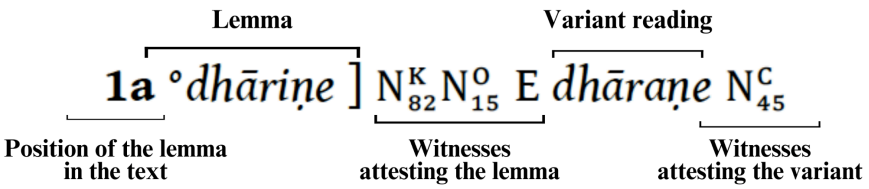


Fig. 79. Highlighted textual fragment by the editor.[151]

2 **Filling in the form for annotating the selected textual fragment**. The editor fills in the form shown in Fig. 37 by incorporating *dhārine* as *lemma* and N_{82}^{K}, N_{15}^{O}, and E as witnesses attesting to it. Then, *dhāraṇe* is added as *variant* with N_{45}^{C} as *witness* attesting to it. The selected fragment is sent and stored in the database. It is identified by assigning to its `idAnnotation` property the same ID assigned to the `@data-annotation` in the HTML file (Fig) to connect it to the *textus* file.

3 **Saving the data**. Upon the editor submitting the information through the *Save* button, the data is automatically organised into a graph data structure within the database as follows (Fig. 80).

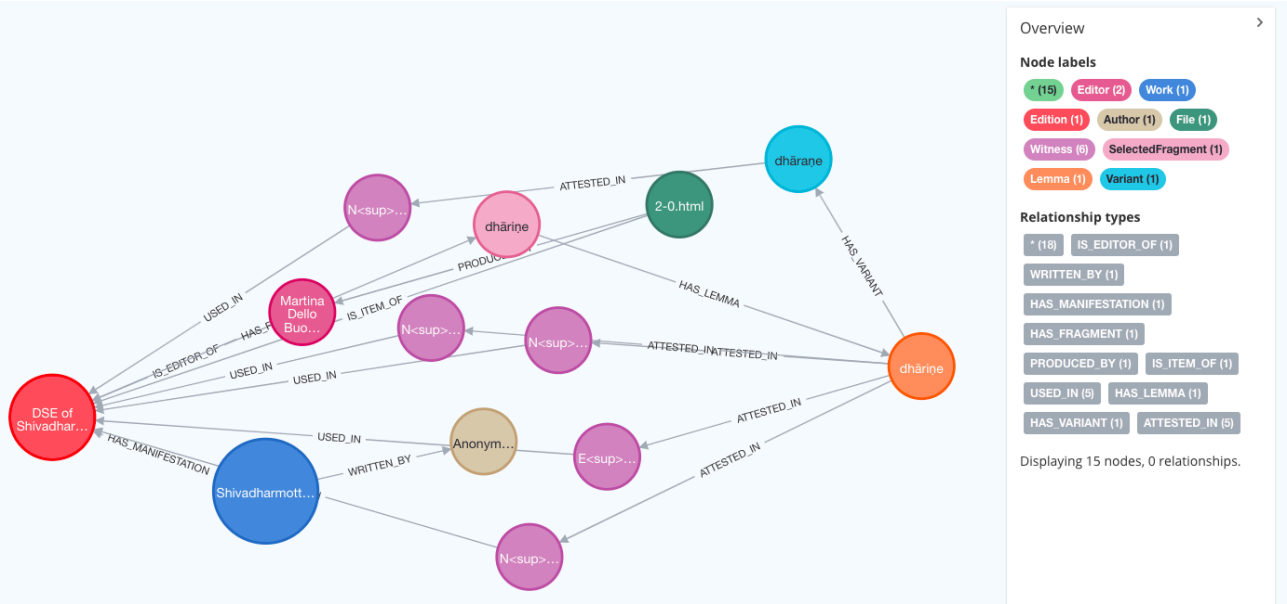


Fig. 80. Resulting graph from the editor's input of an *apparatus* entry.

Hence, the selected fragment *dhārine* is formalised as a `:selectedFragment` node. It is the subject of the relation `:HAS_LEMMA`. The object `:lemma`, i.e., *dhārine*, is, in turn, the subject of the relation `:HAS_VARIANT`, i.e., *dhāraṇe*. Then, both `:lemma` and `:variant` are subjects of the relation `:ATTESTED_IN` to link them to their witnesses.

[151] Sample *apparatus* entry included in the *Śivadharmottara*, edited by Florinda De Simini (Ch. 5).

7.3.3 Implementation of the algorithm to create annotations

According to the implementation strategy, the editor works on an HTML file for annotating and saving the *textus*, as outlined in Ch. 7.3.2. The file is stored on the server, and its name is recorded in a graph database, i.e., Neo4j on which the *Śivadharma Database* is built. The *textus* file is pivotal as all the annotations link to it in the database via ID properties assigned to the selected fragments in the database. We opted for a graph database over semantic technologies due to its simplified maintenance from a developmental perspective (Sippl, Burghardt, Wolff 2021, 191). This approach is intended to promote reusability and customisation of the CMS for future projects.

Since the *Śivadharma Database* relies on a database-centric approach, the information submitted via forms is sent to the Neo4j database through a query in Cypher,[152] Neo4j's query language. The complexity of the queries is contingent upon the degree of complexity of the desired structure. A highly elaborate query is presented here, enabling information organisation within an *apparatus* entry (Listing 10).

```
`
MATCH (edition:Edition)<-[:IS_EDITOR_OF]-(editor:Editor)
WHERE id(edition) = ${idEdition} AND id(editor) = ${idEditor}

MERGE (selectedFragment:SelectedFragment {idAnnotation: "${req.body.idAnnotation}"})
ON CREATE
SET selectedFragment.value = "${req.body.selectedFragment}", selectedFragment.chapter =
"${req.body.chapter}", selectedFragment.stanzaStart = "${req.body.stanzaStart}",
selectedFragment.padaStart = "${req.body.padaStart}", selectedFragment.stanzaEnd =
"${req.body.stanzaEnd}", selectedFragment.padaEnd = "${req.body.padaEnd}"
ON MATCH
SET selectedFragment.value = "${req.body.selectedFragment}", selectedFragment.chapter =
"${req.body.chapter}", selectedFragment.stanzaStart = "${req.body.stanzaStart}",
selectedFragment.padaStart = "${req.body.padaStart}", selectedFragment.stanzaEnd =
"${req.body.stanzaEnd}", selectedFragment.padaEnd = "${req.body.padaEnd}"
MERGE (edition)-[:HAS_FRAGMENT]->(selectedFragment)

MERGE (lemma:Lemma {idLemma: "${req.body.idLemma}"})
ON CREATE
SET lemma.value = "${lemmaReq}", lemma.truncation = "${req.body.lemmaTruncation}",
lemma.notes = "${req.body.lemmaNotes}"
ON MATCH
```

[152] Cypher: https://neo4j.com/docs/getting-started/cypher-intro/.

```
SET lemma.value = "${lemmaReq}", lemma.truncation = "${req.body.lemmaTruncation}",
lemma.notes = "${req.body.lemmaNotes}"
MERGE (selectedFragment)-[:HAS_LEMMA]->(lemma)

FOREACH (wit IN split("${lemmaWits}", ",") |
MERGE (witness:Witness {siglum: wit})
MERGE (lemma)-[:ATTESTED_IN]->(witness)
)

MERGE (variant:Variant {idVariant: "${req.body[idVariant]}"})
ON CREATE
SET variant.value = "${variantReq}", variant.number = "${i}", variant.notes =
"${req.body[notesVariant]}"
ON MATCH
SET variant.value = "${variantReq}", variant.number = "${i}", variant.notes =
"${req.body[notesVariant]}"
MERGE (lemma)-[:HAS_VARIANT]->(variant)
FOREACH (wit IN split("${req.body[manuscriptVariant]}", " ; ") |
MERGE (witness:Witness {siglum: wit})
MERGE (variant)-[:ATTESTED_IN]->(witness)
)

RETURN *
`
```

Listing 10. Cypher query to store an *apparatus* entry in the Neo4j database.[153]

In Listing 10, we initially match the edition on which the editor is working through the editor and edition ID check.[154] Subsequently, the fragment selected by the user is linked to the *edition* node via the following syntax: `(edition)-[:HAS_FRAGMENT]->(selectedFragment)`. The same structure is also applied for the *lemma* that connects directly to the fragment, which is denoted as `(selectedFragment)-[:HAS_LEMMA]->(lemma)`, and for the witnesses connected to the *lemma*, which is represented as `(lemma)-[:ATTESTED_IN] ->(witness)`. According to the same structure, variants are connected to the *lemma*, i.e., `(lemma)-[:HAS_VARIANT]->(variant)`, and witnesses to the variants, i.e., `(variant)-[:ATTESTED_IN]->(witness)`.

Data and their relationships in databases are formally represented in JSON, thus easily convertible in any other format.[155]

[153] The `${}` allows importing a Javascript (JS) variable into a Cypher query.

[154] We do not rely only on the edition title, as there may be multiple editions with the same title.

[155] JSON: https://www.json.org/json-it.html.

7.3.4 Updating and deleting annotations

Database data updating and deletion operations are performed through Cypher queries following a specific syntax. For instance, as illustrated in the Listing 10, the `ON MATCH SET` syntax permits the assignment of new values to the properties of pre-existing nodes. The editor sends the new values via the appropriate forms for this purpose.

7.3.5 What if the annotated text changes?

The *Śivadharma* annotation system offers comprehensive control over annotation by allowing modifications to the annotated text while ensuring its atomic unit of annotation, i.e., fragments, remains intact. The editor selects the fragment directly from the text, assigns it a unique identifier (ID), and sends it to the database via a form. If the editor modifies a pre-existing fragment, the system automatically detects the change by verifying its ID and promptly sends it back to the database. This process, summarised in Fig. 81, ensures that the assigned annotation remains linked to the fragment and unchanged.

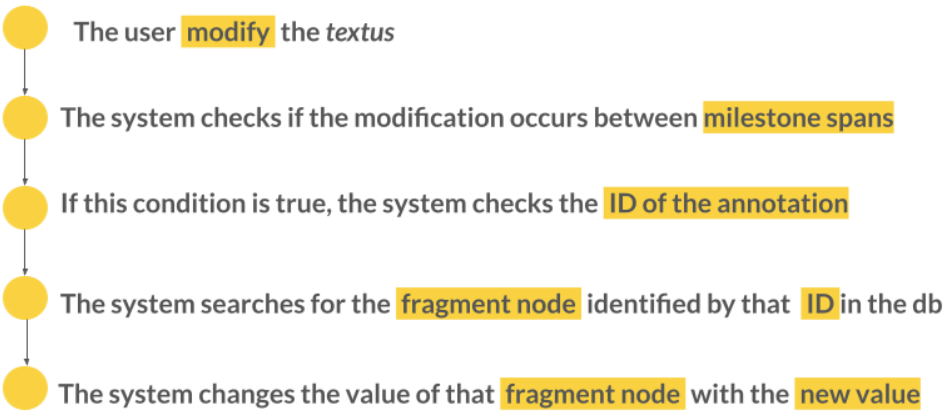


Fig. 81. *Śivadharma* algorithm for updating annotated textual fragment values.[156]

7.4 Visualisation system development

In terms of visualising editions, the *Śivadharma* system uses two distinct mechanisms. The *textus* is an HTML file, so we can *visualise* it directly without needing additional processing. However, other components, such as the *apparatus*, require intermediate data processing. It involves extracting the data required for that component from the database through specific backend Cypher queries,

[156] Excerpt from the AIUCD 2023 presentation *Śivadharma Database CMS. HTML and graph as a starting point for digital editions* by Martina Dello Buono and Francesca Tomasi: https://doi.org/10.6084/m9.figshare.24433486.v1.

sorting it according to preset criteria, and finally displaying it to the user. As an instance, in order to display an *apparatus* entry, it is mandatory to procure two distinct sets of data: (1) the *lemma* along with its corresponding *witnesses* and (2) the *variants* along with their corresponding *witnesses*. Thus, the data extraction query below (Listing 11) comes into play.

```
`
MATCH
(author:Author)<-[:WRITTEN_BY]-(work:Work)-[:HAS_MANIFESTATION]->(edition:Edition)<-[:IS_
EDITOR_OF]-(editor:Editor)
WHERE id(edition) = ${idEdition} AND id(editor) = ${idEditor}
…
OPTIONAL MATCH (edition)<-[:USED_IN]-(witness:Witness)
OPTIONAL                    MATCH                    lemmaWitness                    =
(selectedFragment)-[:HAS_LEMMA]->(lemma:Lemma)-[:ATTESTED_IN]->(lw:Witness)
OPTIONAL                  MATCH                  lemmaVariantWitness                =
(lemma)-[:HAS_VARIANT]->(variant:Variant)-[:ATTESTED_IN]->(vw:Witness)
RETURN *
`
```

Listing 11. Cypher query to extract an *apparatus* entry in the Neo4j database.

The data extraction query involves a systematic approach, commencing with matching the specific editor and edition ID. Once the identification is complete, the *apparatus* components are extracted through the relationships of selected fragments, *lemma*, *variants*, and *witnesses*. The query is executed by implementing the following logical sequence: `(selectedFragment)-[:HAS_LEMMA]->(lemma:Lemma)-[:ATTESTED_IN]->(lw :Witness)` and `(lemma)-[:HAS_VARIANT]->(variant:Variant)-[:ATTESTED_IN]->(vw:Witness)`.

7.5 CMS development

Śivadharma Database is a CMS since it allows users to create, publish, and update contents via an interface. It has been developed from scratch in Node.js,[157] Express,[158] Vanilla JS,[159] EJS,[160] CSS Sass (SCSS),[161] and Neo4j.

[157] Node.js: https://nodejs.org/en.

[158] Express: https://expressjs.com/.

[159] Vanilla JS (pure JS): http://vanilla-js.com/.

[160] EJS: https://ejs.co/.

[161] Sass: https://sass-lang.com/.

7.5.1 Languages choice: JavaScript everywhere

The languages used in the *Śivadharma Database* primarily revolved around a single programming language, namely JavaScript. The backend infrastructure is entirely developed in JavaScript, leveraging the following technologies: (1) Node.js, an open-source and cross-platform JavaScript runtime environment; (2) JavaScript libraries to implement various backend functions, including Express, a web framework for Node.js that is particularly useful for managing *routing*,[162] and (3) Neo4j, a graph database written in Java that structures data in JSON, i.e., a format for data exchange based on the JavaScript programming language. Adopting a single programming language ensures code consistency and facilitates smooth data exchange between different components of the project. Also the frontend leverages JS. Specifically, EJS, chosen for building the *views*, allows embedding JS lines of code in HTML (7.5.5).

7.5.2 *Śivadharma Database* routing

The *Śivadharma* CMS is built on the Node.js framework, and employs the Express library for routing management. Each CMS page, developed in EJS, corresponds to an HTML file that includes embedded JS lines of code in practice, and is fetched via the Express HTTP protocol, i.e., `app.get()`, in response to client requests. For instance, the homepage corresponds to the index.ejs file (Fig. 82)

[162] *Routing* pertains to the manner in which an application's endpoints, also known as URIs, respond to requests made by clients: https://expressjs.com/en/guide/routing.html.

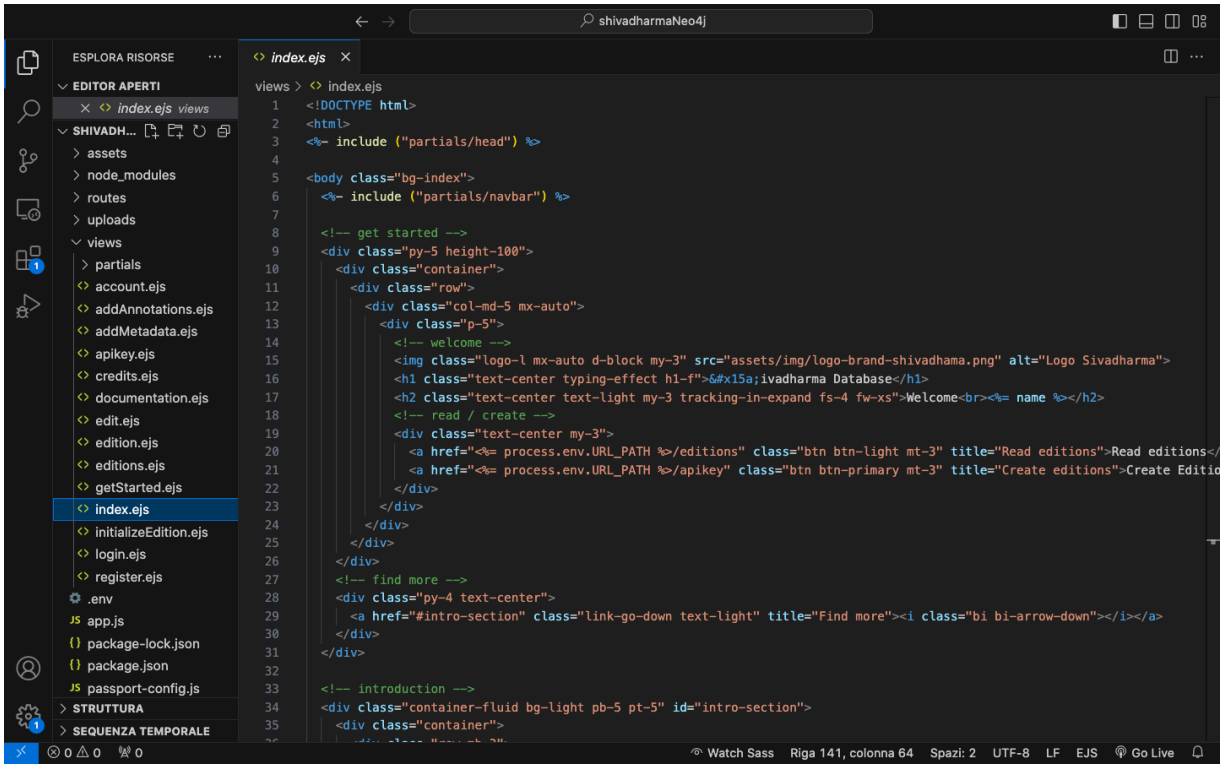


Fig. 82. index.ejs file corresponding to the CMS homepage.

The page *index*, as all the other pages, is requested (Fig. 83) and returned to the client (Fig. 84) via Express.

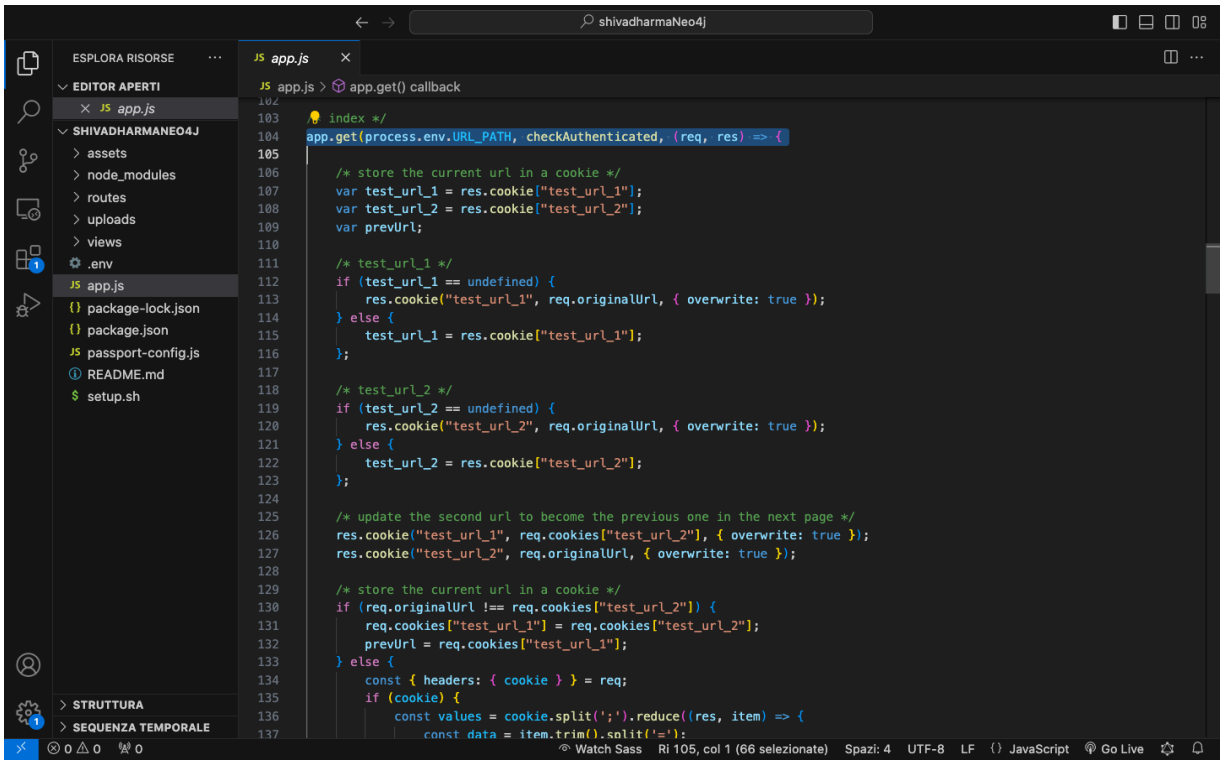


Fig. 83. Example of *routing* in Express.[163]

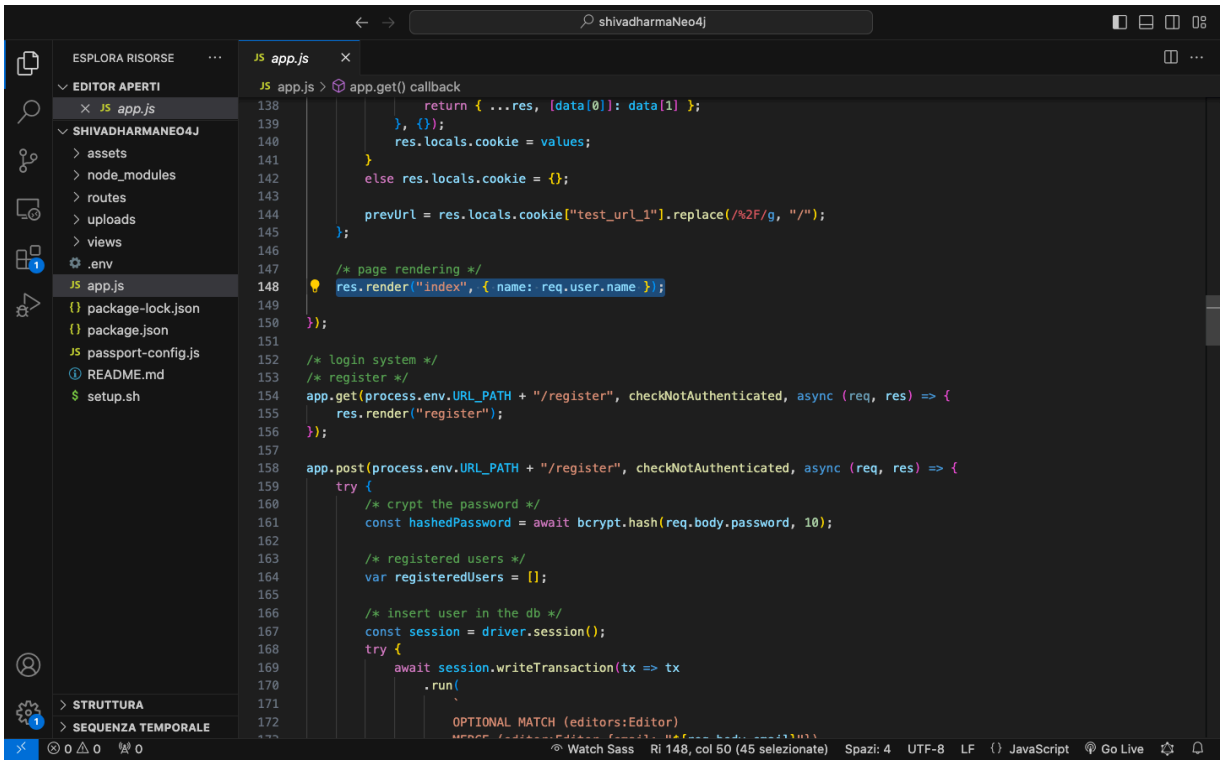


Fig. 84. Page rendering in Express.

7.5.3 GET request method to obtain data from the database

Requests from the client are of two types: GET and POST. GET is used as in Fig. 85 to request data from the server. This method is used not only for routing the *Śivadharma* pages, but also to pass the data relating to the editions from the backend to the frontend. Taking an *apparatus* entry for example, we can observe that a Cypher query requests the data related to the *apparatus* from the database (Ch. 7.4). This query is included in a GET request that responds to the client's request for a page (Fig. 85). The query results obtained are formalised in JSON and therefore passed to the frontend as shown in Fig. 86.

[163] The functions displayed in the screens have been intentionally condensed to highlight the most significant lines of code. However, the complete functions are accessible in the official GitHub repository of the project: https://github.com/martinadellobuono/shivadharma-database.

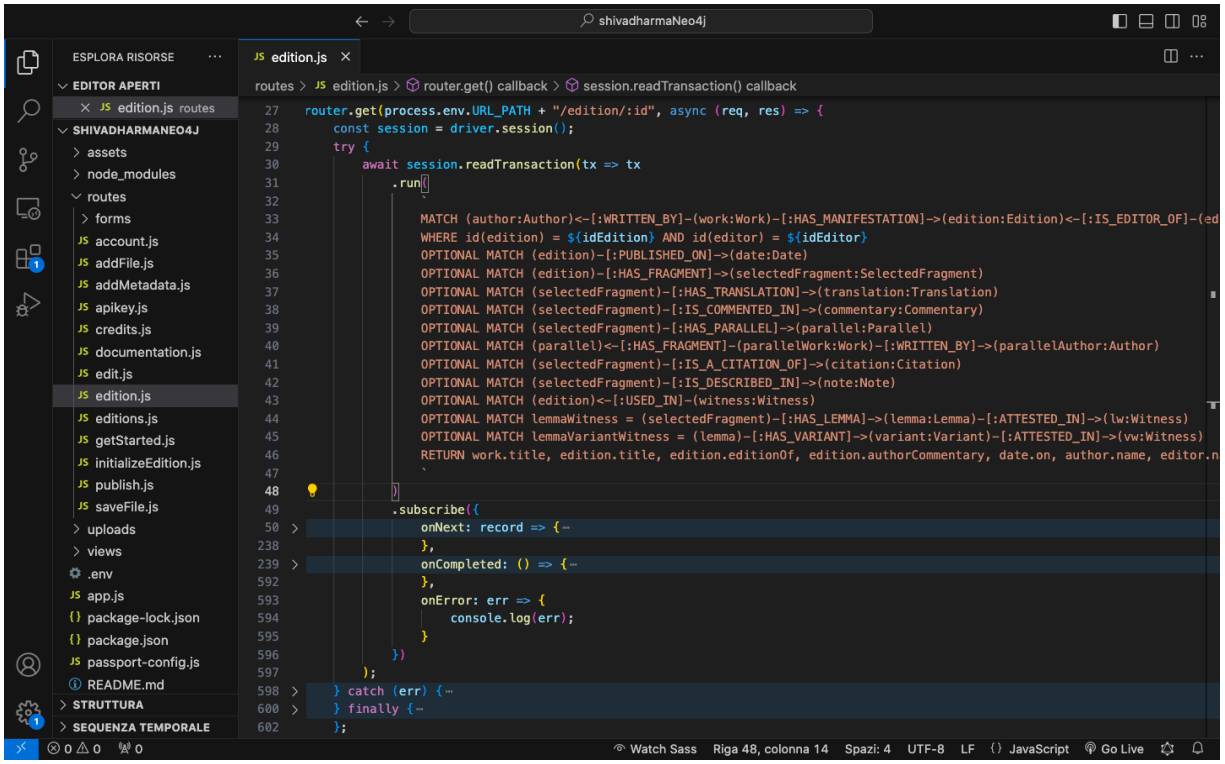


Fig. 85. Cypher query requesting the data related to the *apparatus* to the database.

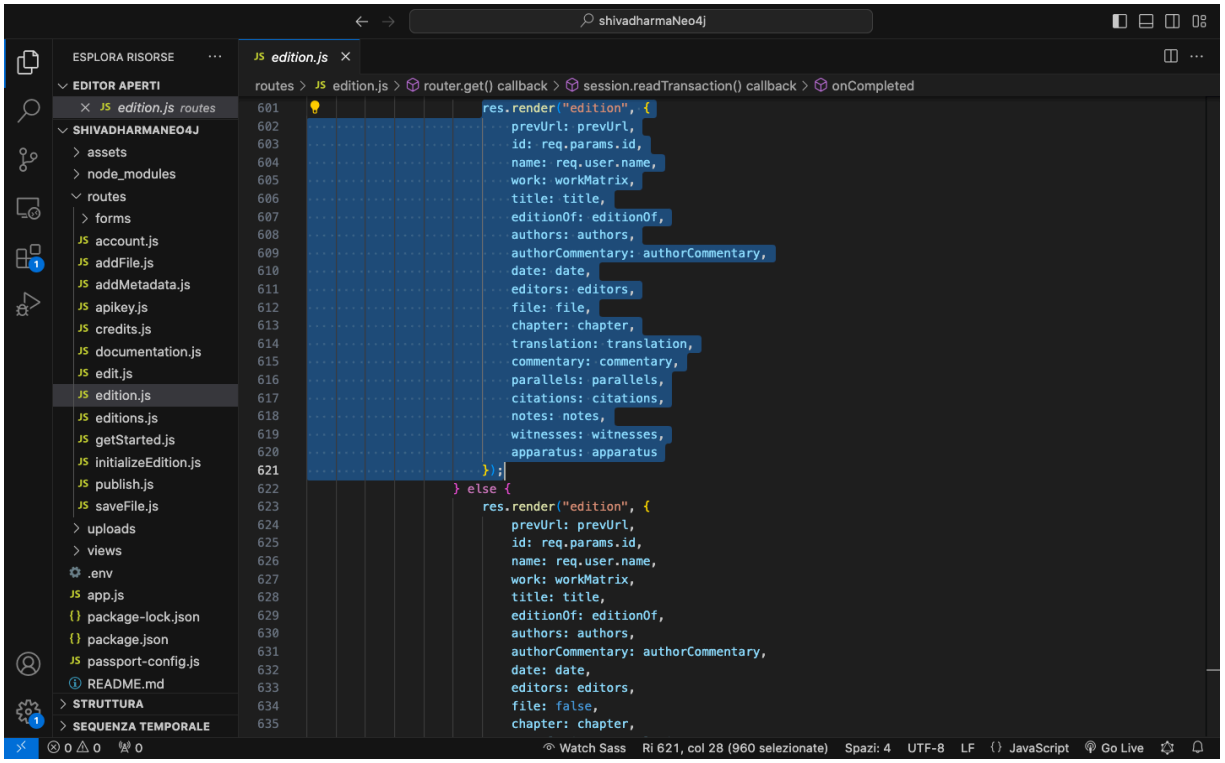


Fig. 86. Results of the query to the database passed to the frontend.

7.5.4 POST request method to send data to the database

As data production is feasible through the *Śivadharma* CMS, the transmission of data from the frontend to the database is carried out by POST requests. Like in GET requests (Ch. 7.5.3), the Cypher queries (Ch. 7.3.3) are encompassed in POST requests. Fig. 87 shows a POST request sending data related to an *apparatus* entry to the database.

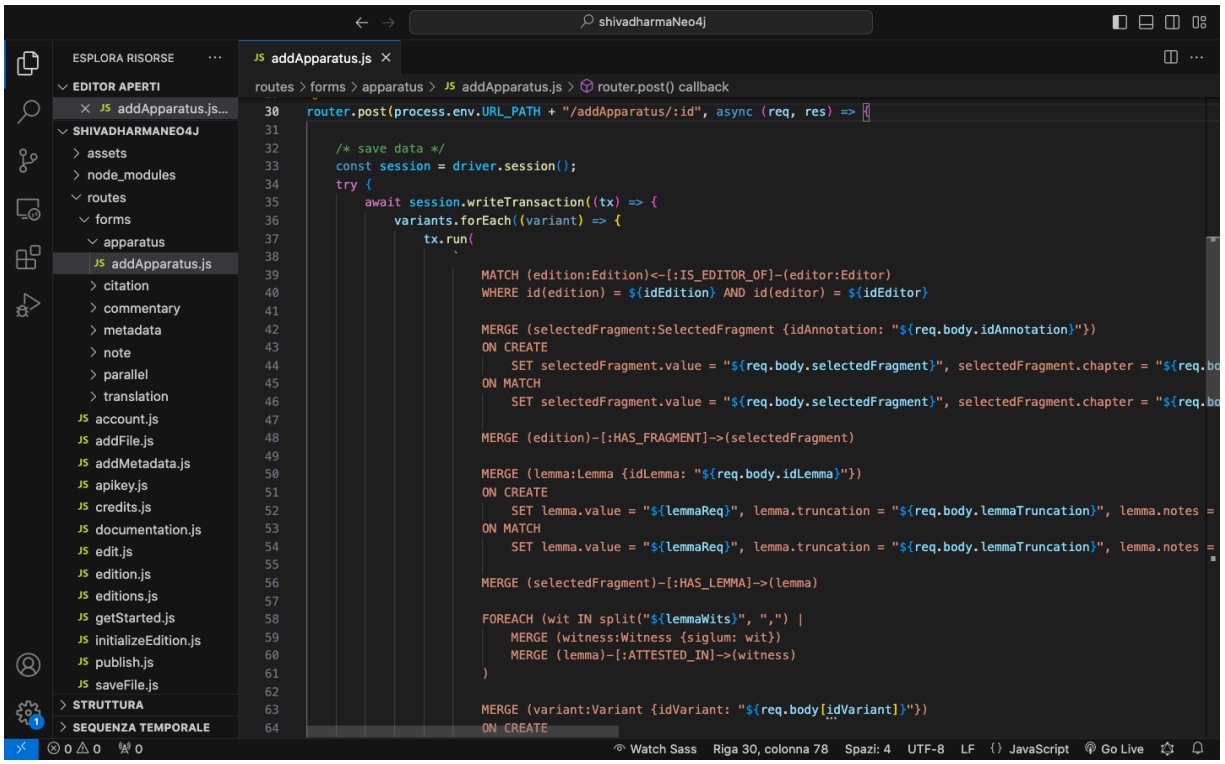


Fig. 87. Cypher query posting the data related to the *apparatus* to the database.

As can be observed in Fig. 87, for each edition component there is a file containing the POST requests.

7.5.5 Data visualisation via EJS

Specific lines of JS code embedded in HTML accomplishes the task of displaying the data once returned by the server (Ch. 7.5.3). Fig. 88 shows an example related to the rendering of the *apparatus* entries. Specifically, it shows that a specific JSON is sent upon request of the page. The JS embedded in HTML displays the data contained in that specific JSON.

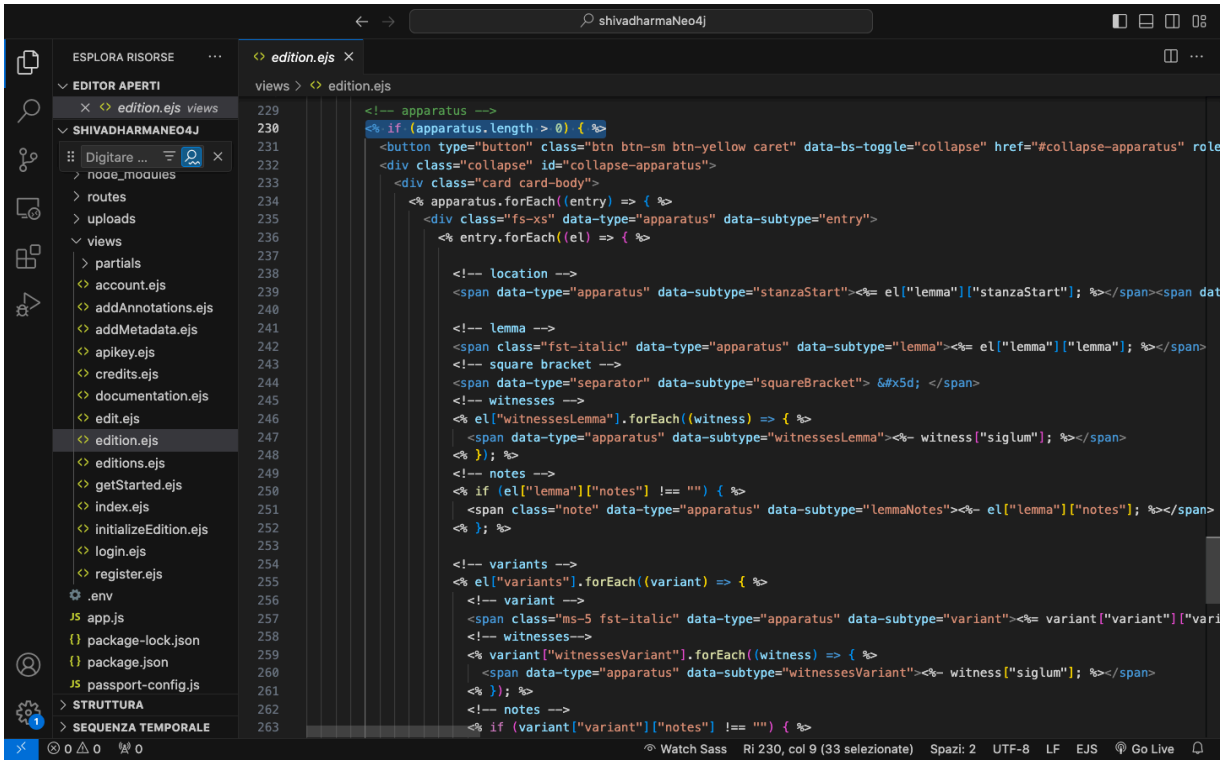


Fig. 88. *Apparatus* entry rendering in EJS.

7.6 Future work

The basic functionalities of the *Śivadharma Database* are released in open source on the project repository.[164] The future work entails the completion and testing of the platform. Specifically, our objective is to meet the following requirements.

1 **Flexibility**. The *Śivadharma* CMS has been tailored to cater specifically to the domain of Sanskrit, with an attempt during the design phase to incorporate the requirements of digital philology in a more general sense (Ch. 1, 2). Nevertheless, it is plausible that the features of the CMS may not encompass all the requisites needed by philologists of other domains. Hence, testing sessions are necessary to identify shortcomings and enhance the CMS's flexibility, making it applicable to preparing editions in other domains.

2 **Reusability**. Given that one of the primary goals of this project is to ensure the reusability of the CMS in various domains, we intend to extend the project documentation to meet these specific reusability requirements.

[164] *Śivadharma Database* GitHub repository: https://github.com/martinadellobuono/shivadharma-database.

3 **Interoperability**. The Shivadharma Database stores data related to the *textus* in an external database using stand-off methods. They are in JSON format and thus easily convertible to any other format. In particular, there is an intention to implement a feature for converting editions into XML/TEI, as it is the standard in the domain of Scholarly Digital Editing. Specifically, the text files could be converted from HTML to XML/TEI, while the annotations extracted directly from the database would be reported in an XML/TEI file external to the *textus* using the TEI *Double End-Point Attachment* method. Such a solution is easily implementable since each *textus* file already contains references to the IDs of annotations saved in the database (Ch. 7.3).

4 **New functionalities**. In order to address the first requirement and to make the editions created on *Shivadharma Database* more sophisticated, we aim to develop advanced features and release them on the platform. These features may include automated reconstruction of witnesses, tools for data visualisation, and edition filtering. Our objective is to enhance the platform's capabilities and provide a more comprehensive *digital* — and not *digitised* — philology experience to the users.
Finally, a commit-based system will be implemented to enable editors to track the earlier versions of their editions with each update. Each commit will comprise of a *title*, an *agent* corresponding to the actual author of the update, a *date*, and an optional *description* of the operations performed. This approach enables the tracing of every modification to editions and their provenance, allowing them to be sorted chronologically.

Conclusion

The *Śivadharma* research reflects on various topics, from the technical need to discover a data structure that accommodates complex and overlapping textual phenomena to the more humanistic aspects of modelling, design, and philological analysis. Drawing upon a thorough investigation of Sanskrit and general philologists' work, the *Śivadharma Database* led to extensive discussions and reflections on the underlying implications of the case, particularly the relationship between philology and technology, where Digital Humanities naturally find their place. Through the *Śivadharma* experience, we found that technology is not merely a *means* to an end, a technical exercise solely to create a new visualisation of interactive content. Technology carries its paradigm and must be approached as such. Modelling textual data necessitates a comprehensive understanding of the characteristics of a text and reflects the philologist's interpretive perspective. The selection of technological methods for text encoding is paramount as it influences not only the expressiveness of annotations but also the effort required to learn and apply these technologies. The literature presents two trends in Digital Scholarly Editing. On the one hand, the encoding process in the traditional XML/TEI is relatively straightforward from the purely practical point of view of writing code. However, selecting the descriptors and handling complex textual structures are demanding tasks. The less consolidated RDF proposes a data structure that can handle innate textual complexity but requires extended practical learning. Moreover, visualisation does not fall within its objectives. The present research contributes to this ongoing debate by proposing a solution based on HTML and graph data structure. This solution presents both methodological and practical advantages. Firstly, HTML encoding immediately satisfies the fundamental requirement of Digital Scholarly Editing, i.e., displaying the editions. It also lends itself to multiple conversions into other languages, such as XML/TEI, for reuse and preservation. The graph data structure, instead, ensures a good level of expressiveness of the annotations, with the textual fragment as the pivot. The *Śivadharma Database* implements this solution by providing scholars with user-friendly tools to prepare, publish, and update Scholarly Digital Editions from scratch via an annotation interface. Providing an interface reduces the scholars' required effort on purely technological issues and guarantees complete control over the annotations, even when updating the text.

Future work includes the completion and testing of the platform, together with the release of additional functionalities, such as the automatic reconstruction of witnesses and tools for data visualisation and edition filtering, and the extension of its documentation to meet reusability needs. Since the tools available are not strictly related to the Sanskrit field, this application is valuable

regardless of the application domain. Even the replicability of the CSM itself is ensured by the online availability of its open-source code and instructions for its running.

Undoubtedly, the task of designing for philology is arduous and intricate. The multifarious nature of texts, coupled with the complexities of philological practices, make it challenging to incorporate their intricacies into computable frameworks. Thus, despite the initial progress towards developing the *Śivadharma Database*, it cannot be deemed a complete solution. Instead, it represents the beginning of a collaborative effort between philology and DH to find an answer to the Digital Scholarly Editing issues.

References

Agrebi, Meriem, and Anne Laure Boncori. 'What Makes a Website Relational? The Experts' Viewpoint'. European Management Journal 35, no. 5 (1 October 2017): 617–31. https://doi.org/10.1016/j.emj.2016.11.001.

André, Jacques. *Règles et recommandations pour les éditions critiques: série latine*. Vol. 1. Paris: Les Belles Lettres, 1972.

Andrews, Tara. 'Transcription and collation'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 160–175. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384-004.

Aralikatte, Rahul, Neelamadhav Gantayat, Naveen Panwar, Anush Sankaran, and Senthil Mani. 'Sanskrit Sandhi splitting using seq2 (seq)^2. *arXiv:1801.00428* (15 July 2019).

Balogh, Dániel, and Arlo Griffiths. 'DHARMA Encoding Guide for Diplomatic Editions. [Technical Report]'. EFEO; Humboldt-Universität (Berlin); CEAIS - Centre d'Études de l'Inde et de l'Asie du Sud, 2020. https://shs.hal.science/halshs-02888186.

Barabucci, Gioele, Elena Spadini, and Magdalena Turska. 'Data vs. Presentation. What Is the Core of a Scholarly Digital Edition?' In *Advances in Digital Scholarly Editing*, 37–46. Leiden: Sidestone Press, 2017. https://serval.unil.ch/notice/serval:BIB_09C6C598108A.

Barzaghi, Sebastian. *La modellazione dei dati nell'Edizione Nazionale delle Opere di Aldo Moro. Tech. rep.* Bologna: University of Bologna, 2021. https://doi.org/10.5281/zenodo.5144961.

Battistini, Alessandro. 'A first look at an unpublished commentary on the *Bhikṣāṭanakāvya*'. In *Śivadharmāmṛta: Essays on the Śivadharma and its Network*, edited by Florinda De Simini and Csaba Kiss, 233–253. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-228-1.

Bausi, Alessandro. '«Philology» as Textual Criticism. «Normalization» of Ethiopian Studies'. *Ethiopian Philology* 1, no. 1 (November 2008): 13–46.

Bernardo, Emilia Di, Marilena Maniaci, Nina Sietis, and Nicola Tangari. 'MeMo – Memory of Montecassino: Un sistema digitale integrato per la conservazione, la fruizione e la valorizzazione del patrimonio manoscritto dell'Abbazia di Montecassino'. *DigItalia* 18, no. 1 (4 August 2023): 113–28. https://doi.org/10.36181/digitalia-00063.

Besta, Maciej, Robert Gerstenberger, Emanuel Peter, Marc Fischer, Michał Podstawski, Claude Barthels, Gustavo Alonso, and Torsten Hoefler. 'Demystifying Graph Databases: Analysis and Taxonomy of Data Organization, System Designs, and Graph Queries'. *ACM Computing Surveys* 56, no. 2 (15 September 2023): 31:1–31:40. https://doi.org/10.1145/3604932.

Biardeau, Madeleine. *Le Mahābhārata*. Vol. 2. Paris: Seuil, 1986.

Bidez, Joseph, and Anders B. Drachmann. *Emploi des signes critiques, disposition de l'apparat dans les éditions savantes des textes grecs et latins. Conseils et recommandations*. Edited by Armand Dellatte and Albert Severyns. Bruxelles: Union Académique Internationale and Paris: Les Belles Lettres, 1938.

Bisschop, Peter C. 'A Parallel in Bhaviṣyapurāṇa 1.171'. In *A Śaiva Utopia: The Śivadharma's Revision of Brahmanical Varṇāśramadharma*, edited by Florinda De Simini. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-215-1.

—— *Universal Śaivism: The Appeasement of All Gods and Powers in the Śāntyadhyāya of the Śivadharmaśāstra*. Leiden, Boston: Brill, 2018. https://doi.org/10.1163/9789004384361.

Bisschop, Peter C., Nirajan Kafle, and Timothy Lubin. *A Śaiva Utopia: The Śivadharma's Revision of Brahmanical Varṇāśramadharma*. Edited by Florinda De Simini. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-215-1.

Boccali, Giuliano, Stefano Piano, and Saverio Sani. *Le letterature dell'India. La civiltà letteraria indiana dai Veda a oggi: principi, metodologie, storia*. Torino: UTET libreria, 2000.

Bolioli, Andrea, Riccardo Tasso, and Roberto Rosselli Del Turco. 'ASED – Annotazione Semantica per Edizioni Digitali'. In *AIUCD2016 - Book of Abstracts*, edited by Federico Boschetti, 167–168. Firenze: Associazione per l'Informatica Umanistica e la Cultura Digitale, 2017.

Bonsi, Claudia, Angelo Di Iorio, Paola Italia, and Fabio Vitali. 'Manzoni's electronic interpretations'. *SEMICERCHIO*, LIII, no. 2 (2015): 91–99.

Bonsi, Claudia, Angelo Di Iorio, Paola Italia, Francesca Tomasi, Fabio Vitali, and Ersilia Russo. 'PhiloEditor®: simplified HTML markup for interpretative pathways over literary collections'. In *La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica. Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)*, edited by Cristina Marras, Marco Passarotti, Greta Franzini, and Eleonora Litta, 47–54. 2020.

Bright, William. ‘The Devanagari Script’. In *The World’s Writing Systems*, edited by Peter T. Daniels and William Bright, 384–390. New York, Oxford: Oxford University Press, 1996.

Buzzetti, Dino. ‘Ambiguità diacritica e markup. Note sull’edizione critica digitale’. In *Soluzioni informatiche e telematiche per la filologia, Atti del Seminario di studi (Pavia, 30-31 marzo 2000)*, edited by Simone Albonico. Pavia: Università degli Studi di Pavia, 2000.

—— ‘Informatica umanistica : Codifica del testo e intelligenza artificiale’. *Schede umanistiche : rivista semestrale dell’Archivio Umanistico Rinascimentale Bolognese* (ARUB), no. 1 (2003): 1000–1027. https://doi.org/10.1400/12676.

—— ‘Rappresentazione digitale e modello del testo’. In *Il ruolo del modello nella scienza e nel sapere, Atti del Convegno (Roma, Accademia Nazionale dei Lincei, 27-28 ottobre 1998)*, 127–161. Roma: 1999.

Buzzoni, Marina. ‘A Protocol for Scholarly Digital Editions? The Italian Point of View’ In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Digital Humanities Series, vol. 4, 59–82. Cambridge: Open Book Publishers, 2016. https://doi.org/10.11647/OBP.0095.04.

—— ‘Text-critical analysis’. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 380–405. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Carbé, Emmanuela. *Digitale d’autore: macchine, archivi e letterature*. Firenze: Firenze University Press; Siena: USiena Press, 2023. https://doi.org/10.36253/979-12-215-0023-3.

Carro, Massimo. ‘NoSQL Databases’. *arXiv preprint arXiv*:1401.2101 (2014).

Caviglia, Giorgio, Paolo Ciuccarelli, and Nicole Coleman. ‘Communication design and the digital humanities’. In *Proceedings of the 4th International Forum of Design as a Process*. 2012.

Cinquilli, Mattia, Dave Evans, Stephen C. Foulkes, Dirk Hufnagel, Marco Mascheroni, Marc Norman, Zdenek Maxa, et al. ‘The CMS Workload Management System’. *Journal of Physics: Conference Series* 396, no. 3 (December 2012): 1–11. https://doi.org/10.1088/1742-6596/396/3/032113.

Ciotti, Fabio. ‘A Formal Ontology for the Text Encoding Initiative’. *Umanistica Digitale*, no. 3 (18 November 2018). https://doi.org/10.6092/issn.2532-8816/8174.

Ciotti, Fabio, Marilena Daquino, and Francesca Tomasi. 'Text Encoding Initiative Semantic Modeling. A Conceptual Workflow Proposal'. In *Digital Libraries on the Move*, edited by Diego Calvanese, Dario De Nart, and Carlo Tasso, 48–60. Communications in Computer and Information Science. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-41938-1_5.

Ciotti, Fabio, and Francesca Tomasi. 'Formal Ontologies, Linked Data, and TEI Semantics'. *Journal of the Text Encoding Initiative*, no. Issue 9 (24 September 2016). https://doi.org/10.4000/jtei.1480.

Ciotti, F., Silvio, P., Francesca, T., Fabio, V. (2016). 'An OWL 2 Formal Ontology for the Text Encoding Initiative'. In *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, pp. 151–153.

Cohoner, Véronique, and Juliette Hueber. 'L'utilisation des technologies IIIF dans les publications scientifiques à forts contenus visuels : les expérimentations de la pépinière de revues DeVisu, membre du réseau Repères'. *Culture et recherche*, no. 144 (2023): 101–103.

Conti, Aidan. 'A typology of variation and error'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 242–254. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384-005.

Contini, Gianfranco. *Esercizî di lettura sopra autori contemporanei, con un'appendice su testi non contemporanei. Edizione aumentata di «Un anno di letteratura»*. Torino: Einaudi, 1974.

Contini, Gianfranco, and Ludovica R. Di Meana. *Diligenza e voluttà*. Milano: Garzanti, 1989.

Cummings, Lance. 'Content Management System 3.0: Emerging Digital Writing Workspaces'. In *Digital Writing Technologies in Higher Education : Theory, Research, and Practice*, edited by Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, 261–75. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-36033-6_17.

D'Avella, Victor B. *Notes on the Sanskrit «Śloka»*. 2019.

Dahlström, Mats. 'Critical Editing And Critical Digitisation'. In *Text Comparison and Digital Creativity*, 77–97. Leiden, Boston: Brill, 2011.

Daquino, Marilena, Francesca Giovannetti, and Francesca Tomasi. 'Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini'. *Umanistica Digitale*, no. 7 (18 December 2019). https://doi.org/10.6092/issn.2532-8816/9091.

Del Gratta, Riccardo, Federico Boschetti, Angelo Mario Del Grosso, Luigi Bambaci, e Simone Zenzaro. ‘La Filologia come sistema dinamico’. *Umanistica Digitale*, no. 13 (21 October 2022): 1–20. https://doi.org/10.6092/issn.2532-8816/13684.

Dello Buono, Martina. ‘Paolo Bufalini's Notebook: A Semantic Scholarly Digital Edition’. figshare, September 27, 2023. https://doi.org/10.6084/m9.figshare.24204507.v1.

De Simini, Florinda. ‘Śivadharma Manuscripts from Nepal and the Making of a Śaiva Corpus’. In *One-Volume Libraries: Composite and Multiple-Text Manuscripts*, edited by Michael Friedrich and Cosima Schwarke, 233–86. Berlin, Boston: De Gruyter, 2016a. https://doi.org/10.1515/9783110496956-009.

—— *Of Gods and Books: Ritual and Knowledge Transmission in the Manuscript Cultures of Premodern India*. Berlin, Boston: De Gruyter, 2016b. https://doi.org/10.1515/9783110478815.

—— ‘When Lachmann’s Method Meets the Dharma of Śiva. Common Errors, Scribal Interventions, and the Transmission of the Śivadharma Corpus’ In *Indic Manuscript Cultures through the Ages: Material, Textual, and Historical Investigations*, edited by Vincenzo Vergiani, Daniele Cuneo and Camillo Alessio Formigatti, 505–548. Berlin, Boston: De Gruyter, 2017. https://doi.org/10.1515/9783110543100-017.

De Simini, Florinda, and Nina Mirnig. ‘Umā and Śiva’s Playful Talks in Detail (*Lalitavistara*): On the Production of Śaiva Works and their Manuscripts in Medieval Nepal’ In *Indic Manuscript Cultures through the Ages: Material, Textual, and Historical Investigations*, edited by Vincenzo Vergiani, Daniele Cuneo and Camillo Alessio Formigatti, 587–654. Berlin, Boston: De Gruyter, 2017. https://doi.org/10.1515/9783110543100-019.

Di Donato, Francesca, and Susanne Müller. ‘Burckhardtsource.Org: The Unpublished Correspondence to Jacob Burckhardt’. *Lexicon Philosophicum: International Journal for the History of Texts and Ideas*, no. 2 (14 March 2014). https://doi.org/10.19283/lph-20142.409.

Di Iorio, Angelo, Silvio Peroni, and Fabio Vitali. ‘A Semantic Web approach to everyday overlapping markup’. *Journal of the American Society for Information Science and Technology* 62, no. 9 (2011): 1696–1716. https://doi.org/10.1002/asi.21591.

Dix, Alan, Janet Finlay, Gregory D. Abowd, Russel Beale. *Human-computer interaction*. Harlow: Pearson Education, 2004.

Fagin Davis, Lisa. 'Old Books, New Technologies: Medieval Manuscript Fragments and IIIF (International Image Interoperability Framework)'. *Digital Initiatives Symposium*, 18 April 2023.

Fiormonte, Domenico, Teresa Numerico, Francesca Tomasi, Desmond Schmidt, and Christopher Ferguson. *The Digital Humanist: A Critical Inquiry*. Goleta: Punctum Books, 2015.

Fischer, Franz. 'Digital Classical Philology and the Critical Apparatus' In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, edited by Monica Berti, 203–220. Berlin, Boston: De Gruyter Saur, 2019. https://doi.org/10.1515/9783110599572-012.

—— 'Representing the critical text'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 405–427. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Formisano, Luciano. 'Gaston Paris e i «nouveaux philologues»'. *Ecdotica*, no. 2/2005 (2005): 5–22. https://doi.org/10.7385/98685.

Frank, István. 'The Art of Editing Lyric Texts'. In *Medieval Manuscripts and Textual Criticism*, edited by Christopher Kleinhenz, 123–138. Chapel Hill: University of North Carolina Press, 1976. [trans. of Frank 1955].

Franzini, Greta, Melissa Terras, and Simon Mahony. 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities'. *Journal on Computing and Cultural Heritage* 12, no. 1 (13 February 2019): 1:1–1:23. https://doi.org/10.1145/3230671.

Franzini, Greta, Simon Mahoni, and Melissa Terras. 'A Catalogue of Digital Editions' In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Digital Humanities Series, vol. 4, 161–182. Cambridge: Open Book Publishers, 2016. https://doi.org/10.11647/OBP.0095.09.

Garrett, Jesse J. *The Elements of User Experience: User-Centered Design for the Web and Beyond*. 2nd ed. Berkeley: New Riders, 2011.

Giannouli, Antonia. 'Critical Editions and the Complementary Apparatuses to a Critical Apparatus'. Edited by Bausi, Alessandro, Buzi, Paola, Maniaci, Marilena, Melissakis, Zisis, Parodi, Laura E., and Sokolinski, Eugenia. *Comparative Oriental Manuscript Studies Bulletin* 1, no. 1 (2015): 21–28.

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

Gumbert, Peter Johan. 'Codicological units. Towards a terminology for the stratigraphy of the non-homogeneous codex'. Edited by Edoardo Crisci and Oronzo Pecere. *Segno e Testo, International Journal of Manuscripts and Text Transmission*, 2 (2004): 17–42.

Hacker, Paul. *Kleine Schriften*. Wiesbaden: Harrassowitz Verlag, 1978.

Hagel, Stefan. 'The Classical Text Editor. An attempt to provide for both printed and digital editions'. In *Digital philology and medieval texts*, edited by Francesco Stella and Arianna Ciula, 77–84. Pisa: Pacini Editore, 2007.

Hahn, Michael. *A brief introduction into the Indian metrical system (for the use of students)*. https://www.academia.edu/6353023/Michael Hahn.

Haynes, Ronald. 'Evolving Standards in Digital Cultural Heritage – Developing a IIIF 3D Technical Specification'. In *3D Research Challenges in Cultural Heritage III: Complexity and Quality in Digitisation*, edited by Marinos Ioannides and Petros Patias, 50–64. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-35593-6_3.

Haugen, Odd E. 'The Making of an Edition: Three Crucial Dimensions.' In *Digital Critical Editions*, edited by Daniel Apollon, Claire Bélisle, and Philippe Régnier, 203–245. Urbana, Chicago, Springfield: University of Illinois Press, 2014.

—— 'Types of editions'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 359–379. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Havet, L. *Règles et recommandations générales pour l'établissement des éditions Guillaume Budé*. Paris: Durand, 1925.

Hellwig, Oliver. 'Using Recurrent Neural Networks for Joint Compound Splitting and Sandhi Resolution in Sanskrit'. In *4th Biennial Workshop on Less-Resourced Languages*, 2015. https://www.semanticscholar.org/paper/Using-Recurrent-Neural-Networks-for-joint-compound-Hellwig/9e54eda9a0d1889010d405ac8b2b7156fde65b91.

Hinrichs, Uta, Florian Windhager, Mennatallah El-Assady, Eric Alexander, Adam Bradley, and Mark-JanBludau. 'From Sketching to Coding: Visualization as a Thinking Process'. In *Digital Humanities 2023: Book of Abstracts*, edited by Anne Baillot, Walter Scholger, Toma Tasovac, and Georg Vogeler, 43–44. Graz: Zentrum für Informationsmodellierung - Austrian

Centre for Digital Humanities, University of Graz, 2023. https://doi.org/10.5281/zenodo.7961822.

Holzner, Steven. *Inside XML*. Indianapolis: New Riders, 2001.

Hopkins, Edward W. *The great epic of India. Its character and origin*. New York: Charles Scribner's Sons, 1901.

Hunter, Michael. 'The Apparatus'. In *Editing Early Modern Texts: An Introduction to Principles and Practice*, edited by Michael Hunter, 92–101. London: Palgrave Macmillan UK, 2009. https://doi.org/10.1057/9780230228788_8.

Italia, Paola. *Editing Duemila. Per una filologia dei testi digitali*. Roma: Salerno Editrice, 2020.

—— 'Il lettore Google'. *Prassi Ecdotiche della Modernità Letteraria* 1 (2016): 13–26. https://doi.org/10.13130/2499-6637/6971.

—— 'Le mano (e la mente) dell'autore: Storia e prospettive della Filologia d'autore'. *Anuario Lope de Vega Texto literatura cultura* 29 (31 January 2023): 324–50. https://doi.org/10.5565/rev/anuariolopedevega.488.

—— 'Methods'. In *What is Authorial Philology?*, edited by Paola Italia and Giulia Raboni, 29–69. Cambridge: Open Book Publishers, 2021. http://books.openedition.org/obp/18888.

Irigoin, Jean. *Règles et recommandations pour les éditions critiques: série grecque*. Vol. 2. Paris: Les Belles Lettres, 1972.

Jagamogan, Reevan Seelen, Saiful Adli Ismail, Noor Hafizah, and Hassan Hafiza Abas. 'A Review: Penetration Testing Approaches on Content Management System (CMS)'. In *2021 7th International Conference on Research and Innovation in Information Systems (ICRIIS)*, 1–6, 2021. https://doi.org/10.1109/ICRIIS53035.2021.9617087.

Jänicke, S., G. Franzini, M. F. Cheema, and G. Scheuermann. 'Visual Text Analysis in Digital Humanities'. *Computer Graphics Forum* 36, no. 6 (2017): 226–50. https://doi.org/10.1111/cgf.12873.

Jarrett, Caroline, and Gerry Gaffney. *Forms that Work: Designing Web Forms for Usability*. Burlington: Morgan Kaufmann, 2009.

Kafle, Nirajan. 'The Umāmaheśvarasaṃvāda of the Śivadharma and its network'. In *Śivadharmāmṛta: Essays on the Śivadharma and its Network*, edited by Florinda De Simini and Csaba Kiss, 233–253. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-228-1.

Katajamäki, Sakari, and Veijo Pulkkinen. ‘Introduction: The Widening Circles of Genetic Criticism’. In *Genetic Criticism in Motion: New Perspectives on Manuscript Studies*, edited by Sakari Katajamäki, Veijo Pulkkinen, and Tommi Dunderlin, 13:7–14. Helsinki: Finnish Literature Society, 2023. https://www.jstor.org/stable/jj.10067050.3.

Kaur, Uttamjit, and Gagandeep Singh. ‘A Review on Software Maintenance Issues and How to Reduce Maintenance Efforts’. *International Journal of Computer Applications* 118, no. 1 (20 May 2015): 6–11.

Kessler, Brett. ‘Sandhi and syllables in Classical Sanskrit’. Edited by. In *The proceedings of the Twelfth West Coast Conference on Formal Linguistics*, edited by Erin Duncan, Donka Farkas, and Philip Spaelti, 35–50. Stanford: Center for the Study of Language and Information, 1994. http://spell.psychology.wustl.edu/sandhi-WCCFL/WCCFL-sandhi.html.en.utf8.

Kiss, Csaba. ‘«...not satisfied with the Mahābhārata…» (śrutvā bhāratasaṃhitām atṛptaḥ): the function of the Vṛṣasārasaṃgraha in the Śivadharma corpus’. In *Śivadharmāmṛta: Essays on the Śivadharma and its Network*, edited by Florinda De Simini and Csaba Kiss, 183–201. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-228-1.

Kiss, Dániel. ‘Catullus Online: A Digital Critical Edition of the Poems of Catullus with a Repertory of Conjectures’. *DIGITALE ALTERTUMS-WISSENSCHAFTEN* (2020): 99.

Lausberg, Heinrich. *Linguistica romanza, vol. 1: Fonetica*. Milano: Feltrinelli, 1971.

Leonardi, Lino. ‘Filologia digitale del Medioevo italiano’. *Griseldaonline* 20, no. 2 (15 December 2021): 77–89. https://doi.org/10.6092/issn.1721-4777/12817.

Li, Charles. ‘Critical Diplomatic Editing: Applying Text-Critical Principles as Algorithms’. In *Advances in Digital Scholarly Editing*, edited by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Aodhán Kelly, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk van Hulle, 305–310. Leiden: Sidestone Press, 2017. http://dx.doi.org/10.17613/M6NV92.

—— ‘helayo: Reconstructing Sanskrit Texts from Manuscript Witnesses’. *Journal of Open Source Software* 7, no. 71 (22 March 2022): 1–3. https://doi.org/10.21105/joss.04022.

—— ‘Philology and Digital Humanities’. In *Routledge Handbook of Yoga and Meditation Studies*, edited by Suzanne Newcombe and Karen O’Brien-Kop, 383–392. London: Routledge, 2020. http://dx.doi.org/10.17613/vzyw-sj55.

Maas, Philip A. ‘On What to Do with a Stemma – Towards a Critical Edition of *Carakasaṃhitā Vimānasthāna* 8. 29’. In *Medical Texts and Manuscripts in Indian Cultural History*, edited by Dominik Wujastyk, Anthony Cerulli, and Karin Preisendanz, 29–61. New Delhi: Manohar, 2013.

Manovich, Lev. *Software Culture*. Milano: Edizioni Olivares, 2010.

Marinelli, Paolo, Fabio Vitali, and Stefano Zacchiroli. ‘Towards the Unification of Formats for Overlapping Markup’. *New Review of Hypermedia and Multimedia* 14, no. 1 (January 2008): 57–94. https://doi.org/10.1080/13614560802316145.

Masud, Luca, Francesca Valsecchi, Paolo Ciuccarelli, Donato Ricci, Giorgio Caviglia. ‘From data to knowledge – Visualizations as transformation processes within the data-information-knowledge continuum’. In *14th International Conference on Information Visualisation*, edited by Ebad Banissi, Stefan Bertschi, Remo Burkhard, John Counsell, Mohammad Dastbaz, Martin Eppler, Camilla Forsell, Georges Grinstein, Jimmy Johansson, Mikael Jern, Farzad Khosrowshahi, Francis T. Marchese, Carsten Maple, Richard Laing, Urska Cvek, Marjan Trutschl, Muhammad Sarfraz, Liz Stuart, Anna Ursyn, and Theodor G Wyeld, IV, 445–449. Los Alamitos: IEEE Computer Society, 2010. https://doi.org/10.1109/IV.2010.68.

McAllister, Patrick. ‘Quotes, Paraphrases, and Allusions: Text Reuse in Sanskrit Commentaries and How to Encode It’. *Journal of the Text Encoding Initiative* 13 (2021): 1–32. https://doi.org/10.4000/jtei.3324.

McKeever, Susan. ‘Understanding Web content management systems: evolution lifecycle and market’, *Industrial Management & Data Systems* 103, no. 9 (1 December 2003): 686–692. https://doi.org/10.1108/02635570310506106.

Memon, Jamshed, Maira Sami, Rizwan Ahmed Khan, and Mueen Uddin. ‘Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)’. *IEEE Access* 8 (2020): 142642–142668. https://doi.org/10.1109/ACCESS.2020.3012542.

Michailidou, Ioanna, Charlotte Haid, and Udo Lindemann. ‘How to Evaluate Use Scenarios and Stories’. In *Design, User Experience, and Usability: Design Discourse*, edited by Aaron

Marcus, 609–620. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015. https://doi.org/10.1007/978-3-319-20886-2_57.

Monella, Paolo. 'Why are there no comprehensively digital scholarly editions of classical texts?'. In *Digital Philology: New Thoughts on Old Questions*, edited by Adele Cipolla, 141–159. Padova: libreriauniversitaria.it edizioni, 2018. https://hdl.handle.net/10447/294132.

Moro, Aldo, *Edizione Nazionale delle Opere di Aldo Moro*, voll., Bologna, University of Bologna, 2021. https://doi.org/10.6092/unibo/aldomoro.

Most, Glenn W. 'What is a critical edition?' In *Ars Edendi Lecture Series*, edited by Barbara Crostini, Gunilla Iversen, and Brian M. Jensen, vol. IV, 181–207. Stockholm: Stockholm University Press, 2016. https://doi.org/10.16993/baj.g.

Moureau, Sébastien. 'The Apparatus Criticus' In *Comparative Oriental Manuscript Studies: An Introduction*, edited by Alessandro Bausi, Pier Giorgio Borbone, Françoise Briquel-Chatonnet, Paola Buzi, Jost Gippert, Caroline Macé, Marilena Maniaci, Zisis Melissakis, Laura E. Parodi, and Witold Witakowski, 348–352. Hamburg: Tredition, 2015. http://hdl.handle.net/2078.1/155652.

Neill, Iian, and Andreas Kuczera. 'The Codex – an Atlas of Relations'. *Zeitschrift für digitale Geisteswissenschaften / Sonderbände* 4 (8 May 2019). https://doi.org/10.17175/sb004_008.

Neill, Iian, and Desmond Schmidt. 'SPEEDy. A Practical Editor for Texts Annotated with Standoff Properties'. In *Graph Data Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler, 15: 45–54. Norderstedt: BoD, 2021. http://kups.ub.uni-koeln.de/id/eprint/55224.

Nguyen, Thi Tuyet Hai, Adam Jatowt, Mickael Coustaty, and Antoine Doucet. 'Survey of Post-OCR Processing Approaches'. ACM Computing Surveys 54, no. 6 (13 July 2021): 124:1–124:37. https://doi.org/10.1145/3453476.

Oren, Eyal, Knud Hinnerk Möller, Simon Scerri, Siegfried Handschuh, and Michael Sintek. 'What are Semantic Annotations?'. *DERI Galway* 9(62), no. 14 (2006).

O'Sullivan, Sinéad. 'Transmission of text'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 15–23. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Palumbo, Giovanni. 'Criticism and controversy'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 88–108. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Paris, Gaston, and Léopold Pannier. *La Vie de Saint Alexis, poème du XIe siècle et renouvellements des XIIe, XIIIe et XIVe siècles*. Paris: Franck. https://gallica.bnf.fr/ark:/12148/bpt6k33044x/f3.image.texteImage.

Pasquali, Giorgio. *Storia della tradizione e critica del testo*. Florence: Le Monnier, 1934.

—— *Storia della tradizione e critica del testo. Reprint with a new introduction and three appendices*. Florence: Casa Editrice Le Lettere, 1952.

Peroni, Silvio. *Example of use of FRBR DL #1*. Figshare. http://dx.doi.org/10.6084/m9.figshare.1540739

Peroni, Silvio, and Fabio Vitali. 'Annotations with EARMARK for Arbitrary, Overlapping and out-of Order Markup'. In *Proceedings of the 9th ACM Symposium on Document Engineering*, 171–180. DocEng '09. New York, NY, USA: Association for Computing Machinery, 2009. https://doi.org/10.1145/1600193.1600232.

Pierazzo, Elena. 'Dalla codifica all'edizione'. In *Che cos'è un'edizione scientifica digitale*, 73–87. Roma: Carocci, 2020.

—— 'Metodologia delle edizioni scientifiche digitali'. In *Che cos'è un'edizione scientifica digitale*, 47–71. Roma: Carocci, 2020.

—— 'What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter'. *International Journal of Digital Humanities* 1 (2019): 209–220. https://doi.org/10.1007/s42803-019-00019-3.

Pierazzo, Elena, and Élise Leclerc. 'L'Edizione Scientifica al Tempo Dell'Editoria Digitale'. *Ecdotica*, no. 1/2015 (2015). https://doi.org/10.7385/99232.

Renear, Allen, David Dubin, and C. M. Sperberg-McQueen. 'Towards a Semantics for XML Markup'. In *Proceedings of the 2002 ACM Symposium on Document Engineering*, 119–126. DocEng '02. New York, NY, USA: Association for Computing Machinery, 2002. https://doi.org/10.1145/585058.585081.

Rizzo, Silvia. *Il Lessico Filologico Degli Umanisti*. Roma: Edizioni di Storia e Letteratura, 1973.

Robinson, Peter. 'Current Issues in Making Digital Editions of Medieval Texts—or, Do Electronic Scholarly Editions Have a Future?'. *Digital Medievalist* 1 (20 April 2005). https://doi.org/10.16995/dm.8.

—— 'The Digital Revolution in Scholarly Editing' In *Ars Edendi Lecture Series*, edited by Barbara Crostini, Gunilla Iversen, and Brian M. Jensen, vol. IV, 181–207. Stockholm: Stockholm University Press, 2016. http://dx.doi.org/10.16993/baj.h.

—— 'Where we are with electronic scholarly editions, and where we want to be'. 2003.

Roelli, Philipp. 'Introduction'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 1–8. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Roelli, Philip, Caroline Macé, Marko K. Halonen, Marina Buzzoni, Odd E. Haugen, Aidan Conti, Tuomas M. Heikkilä, and Teemu T. Roos. *Parvum Lexicon Stemmatologicum*. 2015. https://wiki.hiit.fi/display/stemmatology/Parvum+lexicon+stemmatologicum.

Rosselli Del Turco, Roberto. 'Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology)'. *Textual Cultures* 12, no. 2 (29 August 2019): 91–111. https://doi.org/10.14434/textual.v12i2.27690.

—— 'The Battle We Forgot to Fight: Should We Make a Case for Digital Editions?' In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Digital Humanities Series, vol. 4, 219–238. Cambridge: Open Book Publishers, 2016. https://doi.org/10.11647/OBP.0095.12.

Rosselli Del Turco, Roberto, Giancarlo Buomprisco, Chiara Di Pietro, Julia Kenny, Raffaele Masotti, and Jacopo Pugliese. 'Edition Visualization Technology: A Simple Tool to Visualize TEI-Based Digital Editions'. *Journal of the Text Encoding Initiative*, no. Issue 8 (28 December 2014): 1–29. https://doi.org/10.4000/jtei.1077.

Ruppel, Antonia M. *The Cambridge Introduction to Sanskrit*. New York: Cambridge University Press, 2017.

Sahle, Patrick. 'A catalog of Digital Scholarly Editions'. Institut für Dokumentologie und Editorik. 2020. https://v3.digitale-edition.de/index.html.

—— ‘What Is a Scholarly Digital Edition?’ In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Digital Humanities Series, vol. 4, 19–39. Cambridge: Open Book Publishers, 2016. https://doi.org/10.11647/OBP.0095.02.

Sahle, Patrick in collaboration with Georg Vogeler and the members of the IDE. ‘Criteria for Reviewing Scholarly Digital Editions, Version 1.1’. 2014. http://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/.

Sanderson, Alexis. ‘Śaivism and the Tantric Traditions’ In *The World’s Religions*, edited by Peter Clarke, Friedhelm Hardy, Leslie Houlden, Stewart Sutherland, 660–704. London: Routledge, 1988.

Sanderson, Robert, Paolo Ciccarese, and Herbert Van de Sompel. ‘Designing the W3C Open Annotation Data Model’. In *Proceedings of the 5th Annual ACM Web Science Conference*, 366–375. WebSci ’13. New York, NY, USA: Association for Computing Machinery, 2013. https://doi.org/10.1145/2464464.2464474.

Schmidt, Desmond. ‘The Current State of the Digital Scholarly Editions and Three Challenges’. (2018). https://www.academia.edu/37585331/The_Current_State_of_the_Digital_Scholarly_Edition_and_Three_Challenges.

—— ‘The Inadequacy of Embedded Markup for Cultural Heritage Texts’. *Literary and Linguistic Computing* 25, no. 3 (1 September 2010): 337–356. [https://doi.org/10.1093/llc/fqq007]

—— ‘Towards an Interoperable Digital Scholarly Edition’. *Journal of the Text Encoding Initiative*, no. Issue 7 (12 November 2014): 1–20. https://doi.org/10.4000/jtei.979.

—— ‘True Interoperability for Digital Scholarly Editions’. *Umanistica Digitale*, no. 10 (9 September 2021): 1–23. https://doi.org/10.6092/issn.2532-8816/12604.

Segers, Hannelore. ‘Edition Visualization Technology Project (Review)’. *Digital Philology: A Journal of Medieval Cultures* 10, no. 2 (2021): 340–42. https://doi.org/10.1353/dph.2021.0020.

Sippl, Colin, Manuel Burghardt, and Christian Wolff. ‘Modelling Cross-Document Interdependencies in Medieval Charters of the St. Katharinenspital in Regensburg’. In *Graph Data Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler, 15: 181–203. Norderstedt: BoD, 2021.

https://kups.ub.uni-koeln.de/55227/.

Spadini, Elena, and Francesca Tomasi. 'Introduction'. In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler, 15:1–6. Norderstedt: BoD, 2021.

Sperberg-McQueen, C. M., and Claus Huitfeldt. 'GODDAG: A Data Structure for Overlapping Hierarchies'. In *Digital Documents: Systems and Principles*, edited by Peter King and Ethan V. Munson, 139–160. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004. https://doi.org/10.1007/978-3-540-39916-2_12.

Stokes, Peter A., Benjamin Kiessling, Daniel Stökl Ben Ezra, Robin Tissot, and El Hassane Gargem. 'The EScriptorium VRE for Manuscript Cultures'. *Classics@ Journal, Ancient Manuscripts and Virtual Research Environments*, 18 (2021). https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/.

Stoppelli, Pasquale. *Filologia della letteratura italiana*. Roma: Carocci Editore, 2019.

Sutton, Nicholas. *Religious Doctrines in the Mahābhārata*. Delhi: Motilal Banarsidass Publishers Private Limited, 2000.

TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.6.0. Last updated on 4th April 2023. TEI Consortium. http://www.tei-c.org/Guidelines/P5/.

Tomasi, Francesca. 'L'Edizione Digitale e La Rappresentazione Della Conoscenza. Un Esempio: Vespasiano Da Bisticci e Le Sue Lettere'. *Ecdotica*, no. 1/2012 (2012): 264–286. https://doi.org/10.7385/99154.

—— *Organizzare la conoscenza: Digital Humanities e Web semantico*. Milano: Editrice Bibliografica, 2022.

—— 'Vespasiano Da Bisticci, Lettere. Knowledge Site 2020. Versione 3.0'. /DH.arc, Alma Mater Studiorum, University of Bologna. CRR-MM, 2020. https://doi.org/10.6092/unibo/vespasianodabisticciletters.

Tomasi, Francesca, Fabio Ciotti, Maurizio Lana, Fabio Vitali, Silvio Peroni, and Diego Magro. 'Dialogue and linking between TEI and other semantic models'. In *The linked TEI: Text Encoding in the Web*, edited by Arianna Ciula and Fabio Ciotti, 145–158. Rome: DIGILAB Sapienza University & TEI Consortium, 2013. https://hdl.handle.net/11585/185113.

Tomasi, Francesca, Marilena Daquino, Francesca Giovannetti, Martina Dello Buono. *Paolo Bufalini. Quaderno di appunti (1981-1991)* [Semantic Scholarly Digital Edition]. Bologna: Digital Humanities Advanced Research Centre (/DH.arc), 2019. https://doi.org/10.6092/UNIBO/AMSACTA/6415.

Tomasin, Lorenzo. 'Nove tesi e mezza per la filologia nell'era della liquidità digitale'. *Storie e linguaggi* 5, no. 1 (2019): 19–33.

Trento, Margherita. 'Translating the Dharma of Śiva in sixteenth-century Chidambaram: Maṟaiñāṉa Campantar's Civatarumōttaram. With a preliminary list of the surviving manuscripts'. In *Śivadharmāmṛta: Essays on the Śivadharma and its Network*, edited by Florinda De Simini and Csaba Kiss, 233–253. Naples: UniorPress, 2021. https://doi.org/10.6093/978-88-6719-228-1.

Trovato, Paolo. 'Neo-Lachmannism'. In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli, 109–138. Berlin, Boston: De Gruyter, 2020. https://doi.org/10.1515/9783110684384.

Turska, Magdalena, James Cummings, and Sebastian Rahtz. 'Challenging the Myth of Presentation in Digital Editions'. *Journal of the Text Encoding Initiative*, no. Issue 9 (24 September 2016). https://doi.org/10.4000/jtei.1453.

Van Buitenen, Johannes A. B., and James L. Fitzgerald. *The Mahabharata, Volume 1: Book 1: The Book of the Beginning*. Chicago, London: University of Chicago Press, 1973.

Van Hulle, Dirk. 'Modelling a Digital Scholarly Edition for Genetic Criticism: A Rapprochement'. *Variants. The Journal of the European Society for Textual Scholarship*, no. 12–13 (31 December 2016): 34–56. https://doi.org/10.4000/variants.293.

Van Zundert, Joris J. 'On Not Writing a Review about Mirador: Mirador, IIIF, and the Epistemological Gains of Distributed Digital Scholarly Resources'. *Digital Medievalist* 11, no. 1 (3 August 2018): 1–48. https://doi.org/10.16995/dm.78.

—— 'The Case of the Bold Button: Social Shaping of Technology and the Digital Scholarly Edition'. *Digital Scholarship in the Humanities* 31, no. 4 (1 December 2016): 898–910. https://doi.org/10.1093/llc/fqw012.

Vogeler, Georg. '«Standing-off Trees and Graphs»: On the Affordance of Technologies for Assertive Edition'. In *Graph Data Models and Semantic Web Technologies in Scholarly*

Digital Editing, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler, 15: 73–94. Norderstedt: BoD, 2021. https://kups.ub.uni-koeln.de/55227/.

Yang, Qian, John Zimmerman, Aaron Steinfeld, and Anthony Tomasic. 'Planning Adaptive Mobile Experiences When Wireframing'. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 565–76. DIS '16. New York, NY, USA: Association for Computing Machinery, 2016. https://doi.org/10.1145/2901790.2901858.

Zaccarello, Michelangelo. 'Filologia materiale e nuove tecnologie: il testo digitale fra edizione e archivio'. *Medioevo Letterario d'Italia* 14 (2017): 117–25. https://dx.doi.org/10.19272/201705601009.

Zenzero, Simone, Angelo M. Del Grosso, Federico Boschetti, and Graziano Ranocchia. 'Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso GreekSchools'. In *AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Preceedings della 11a conferenza nazionale, Lecce 2022*, edited by Fabio Ciracì, Giulia Miglietta, and Carola Gatto: 20–25. 2022.