Co-designing Digital Papyrology

Angelo Mario Del Grosso[1], Simone Zenzaro[1], Federico Boschetti[1], Graziano Ranocchia[2]
[1]Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), Pisa, Italy,
Email: {name}.{surename}@ilc.cnr.it
[2] Università di Pisa, Pisa, Italy,
Email: graziano.ranocchia@unipi.it

***Abstract*—Traditional papyrology has established effective methods for producing diplomatic editions of papyri and publishing critical editions of their textual content. Digital papyrology strives to bridge the gap with the digital age. Achieving the balance between familiarity for traditional scholars and the potential for computational analysis remains a challenge. This paper proposes an innovative co-design approach for developing digital papyrology tools, leveraging both Domain-Driven Design (DDD) and Domain-Specific Languages (DSLs). DDD emphasizes a collaborative understanding of the problem domain, while DSLs are formal languages tailored to specific domains like papyrology. The co-design process involves close collaboration with a team of papyrologists, philologists, linguists, and other humanities scholars. This ensures that the resulting tools are user-friendly and cater to the needs of traditional scholars. DSLs encode domain-specific knowledge, facilitating the creation of machine-actionable Digital Scholarly Editions (DSEs) that remain user-friendly. CophiEditor, a modular web environment designed within a micro-services architecture, implements the complete workflow for creating DSL-based DSEs. The co-design approach, the DSL definition, and the DDD paradigm guarantee that CophiEditor is familiar and produces interoperable and extensible data. The development of CophiEditor, within the ERC GreekSchools project, showcases the potential of this approach. It offers greater accessibility of digital tools for traditional philologists and opens doors for new possibilities in computational analysis of ancient texts.**

***Index Terms*—Computational Philology, Digital Humanities, Digital Papyrology, Digital Scholarly Editing, TEI/EpiDoc.**

I. Introduction

According to Monica Berti, providing digital editions of fragmentary texts means finding digital paradigms and solutions to express information about printed critical editions and their editorial and conventional features [1]. Berti also states that working on a digital edition means converting traditional tools and resources used by scholars into machine actionable contents [2], [4]. The aforementioned reasons highlight the critical need for a co-design and co-evolution process between text science and information science [5].

In our opinion, creating formal languages which adhere philological specific purposes is a possible path for reaching the aforementioned goal.

Currently, the digital representation of a textual resource is achieved through formal models of text, exploiting shared XML vocabularies that describe and define the structure and the semantics about the selected resource. Among these, the de-facto standard for scholarly editing is the XML schema maintained by the Text Encoding Initiative consortium (TEI) [6], or its specific profiles and extensions such as TEI/EpiDoc[1] - with regards to epigraphical resources - XML/MEI[2] - with regards to musical resources.

The TEI consortium also defines a set of “best practices” known as the TEI guidelines to encode “all of the phenomena” scholars encounter in texts and documents.[3]

From the TEI guidelines we read the following principles:

> *Because of its roots in the humanities research community, the TEI scheme is driven by its original goal of serving the needs of research, and is therefore committed to providing a maximum of comprehensibility, flexibility, and extensibility. More specific design goals of the TEI have been that these Guidelines should:*
>
> - *provide a standard format for data interchange*,
> - *provide guidance for the encoding of texts in this format*,
> - *support the encoding of all kinds of features of all kinds of texts studied by researchers*,
> - *be application independent*.

Textual scholars often express more frustration than enthusiasm when digitally encode their sources using declarative languages like XML/TEI, using whether the classical embedded approach [7] or the more flexible stand-off one [8]. Indeed, these declarative languages are perceived as [9]:

1) unfamiliar, 2) time-consuming, 3) inadequate to express complex textual phenomena, 4) inadequate to capture the traditional editing processes.

The paper, extended version of a contribution presented at CiSt’23 congress in Agadir, Morocco, was submitted for the review process on 15 march 2024. This work was supported by The GreekSchools project, “The Greek Philosophical School according to Europe’s earliest ‘history of philosophy’: Towards a new pioneering critical edition of Philodemus’ Arrangement of the Philosophers”. The project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020, Excellent Science (Grant agreement No. 885222, PI: G. Ranocchia).

Angelo Mario Del Grosso is Researcher with the Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), 56124, Pisa, PI, Italy (e-mail: angelo.delgrosso@ilc.cnr.it). Simone Zenzaro is Researcher with the Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), 56124, Pisa, PI, Italy (e-mail: simone.zenzaro@ilc.cnr.it). Federico Boschetti is Researcher with the Cnr-Istituto di Linguistica Computazionale “Antonio Zampolli” (CNR-ILC), 30123, Venezia, VE, Italy (e-mail: federico.boschetti@unive.it). Graziano Ranocchia is Full Professor of Papyrology with the Department of Philology, Literature and Linguistics, University of Pisa, 56124, Pisa, PI, Italy (e-mail: graziano.ranocchia@unipi.it).

[1] https://epidoc.stoa.org/gl/latest/index.html

[2] https://music-encoding.org/

[3] https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html

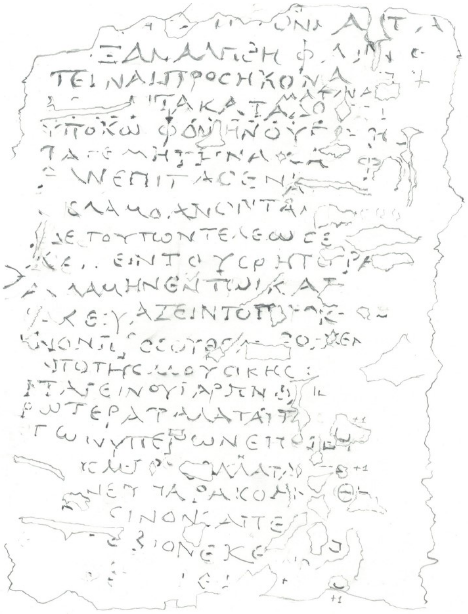


Figure 1. Image of a drawing of the col. n. 64 of the papyrus n. PHerc 1004

Consequently, valuable scholarly data and authoritative new critical editions remain locked outside the digital ecosystem (Fig. 2).

To overcome this issue, we are experimenting the adoption of domain specific-approaches[4] to formally describe the scholarly editing process, embracing the principles of Domain-Driven Design [10]. This strategy is based on the Euporia[5] annotation method [9] and has evolved into the DSL-based DSE methodology: Digital Scholarly Editions based on Domain Specific Languages [11]. This approach enables philologists to maintain continuity with traditional practices while enhancing the editorial process with digital, computational and collaborative capabilities, thereby enabling *machine actionability*.

Leiden+,[6] implemented within the papyri.info project, serves as a good example of a Domain Specific Language (DSL) in the domain of digital papyrology making traditional conventions machine actionable. Table I summarizes some prototypical examples for: 1) lost words supplied, 2) text inserted, 3) gap illegible, 4) omitted text Supplied, 5) surplus text, 6) deletion.

The paper is organized as follows: Section II provides an overview of existing research and initiatives relevant to the project; Section III introduces the GreekSchools ERC project, within which the research described in this paper is being developed; Section IV details the process we implemented to gather requirements for the project using the Domain-Driven Design framework. Section V outlines the overall workflow for designing and implementing the key components of the project: the design and the implementation of the core system for the DSL-based Digital Scholarly Edition (Subsection V-A); the CophiEditor scholarly digital environment developed for interacting with the DSE (Subsection V-B); Subsection V-C is a more technical overview of the implementation aspects of the project. Section VI illustrates some practical applications of the DSL-based DSE through real-world editing and querying examples. Two editing examples are detailed: Subsection VI-A demonstrates how the DSL facilitates editing paleographic apparatus and VI-B demonstrates how the DSL facilitates editing diplomatic transcriptions. Subsection VI-C provides an example of how to query the corpus using the DSL. Section VII describes current efforts to integrate character recognition capabilities into the scholarly platform. Finally, section VIII concludes the paper by summarizing the key findings and potentially outlining future directions for the project.

Table I
LEIDEN+ EXAMPLES

Editorial	Leiden+	XML encoding
Supplied	[ὁμο]λογῶ	<supplied reason="lost"> ὁμο</supplied>λογῶ
Inserted	\ὅλων/	<add place="above"> ὅλων</add>
Gap	.?	<gap reason="illegible" extent="unknown" unit="character"/>
Omitted	<ἀπεγραψάμην>	<supplied reason="omitted"> ἀπεγραψάμην</supplied>
Surplus	{ὀνόματος}	<surplus> ὀνόματος</surplus>
Deletion	⟦τοῖς κορασίοις⟧	<del rend="erasure"> τοῖς κορασίοις</del>

II. Related Initiatives

The creation of digital scholarly editions (DSEs), along with the development of computational tools devoted to accurately annotate scholarly phenomena of interest, has been the objective of numerous research initiatives over time.

If we were to mention only the most notable, we can include the following different groups of tools: (i) Textual Communities [12], SoSol [13], Digital Mappa [14], Proteus [15], Ediarum [16] for scholarly editing of textual resources; (ii) TEITOK [17], EVT [18], TEIPublisher [19], EFES [20] for publishing and browsing DSEs; (iii) Perseids [21], Annotation Studio [22], Recogito [23], Catma [24] for annotating scholarly works; and (iv) others noteworthy projects within the realm of DSEs [7].

The listed projects demonstrate the significant effort put into the development of environments for digital scholarly editing.

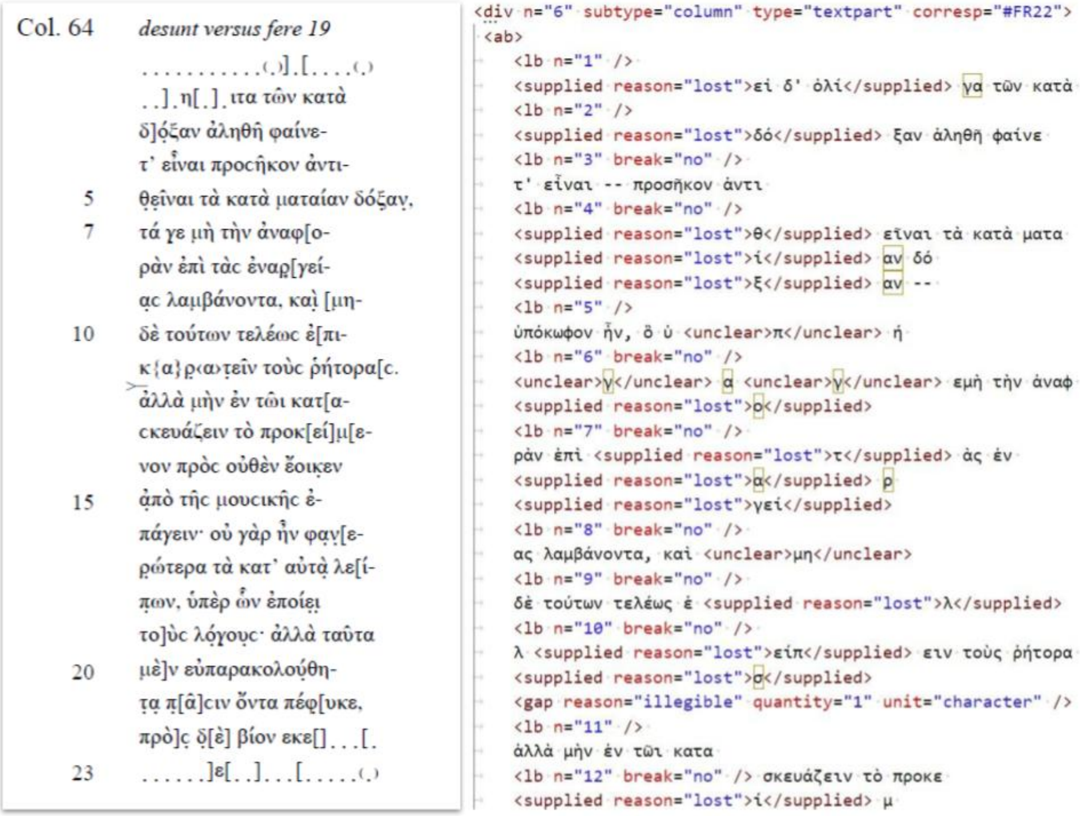
While the aforementioned tools are off-the-shelf applications, our ongoing initiative is currently customized for the ERC 885222-GreekSchools funded project (see section III).

[4] Domain-specific approaches clearly demonstrate a user-centered design philosophy, placing textual scholars, as the domain experts, at the core of the digital environment's development.

[5] Euporia is also a web-based annotation environment available on GitHub (https://github.com/CoPhi/euporia). It's built as an eXistDB Applet.

[6] https://papyri.info/docs/leiden_plus

[7] For a list of tools and initiatives within the digital scholarly editing field see [3] and [35] also the web sites of some research infrastructures such as www.clarin.eu for language resources and www.dariah.eu for the arts and humanities.



Col. 64 *desunt versus fere 19*
. . . . (.)].[. . . . (.)
. .].η[.].ιτα τῶν κατὰ
δ]ọ́ξαν ἀληθῆ φαίνε-
τ᾽ εἶναι προϲῆκον ἀντι-
5 θẹệναι τὰ κατὰ ματαίαν δόξαṿ,
7 τά γε μὴ τὴν ἀναφ[ο-
ρὰν ἐπὶ τὰϲ ἐναρ̣[γεί-
ạϲ λαμβάνοντα, καὶ̣ [μη-
10 δὲ τούτων τελέωϲ ἐ[πι-
κ{α}ρ̣‹α›ṭεῖν τοὺϲ ῥήτορα[ϲ.
ἀλλὰ μὴν ἐν τῶι κατ[α-
ϲκευάζειν τὸ προκ[εί]μ[ε-
νον πρὸϲ οὐθὲν ἔοιḳεν
15 ἀ̣πὸ τῆϲ μουϲικῆϲ ἐ-
πάγειν· οὐ γὰρ ἦν φαṿ[ε-
ρ̣ώτερα τὰ κατ᾽ αὐτὰ̣ λε[ί-
π̣ων, ὑπὲρ ὧν ἐποίẹι
το]ὺϲ λό̣γου̣ϲ· ἀλλὰ ταῦτα
20 μὲ]ν εὐπαρακολού̣θη-
ṭα̣ π̣[â]ϲιν ὄντα πέφ[υκε,
πρὸ]ϲ̣ δ̣[ὲ] βίον εκε[]. . .[.
23 . . . .]ε[. .]. . .[. . . . . (.)

```
<div n="6" subtype="column" type="textpart" corresp="#FR22">
 <ab>
    <lb n="1" />
    <supplied reason="lost">ει δ' ὀλί</supplied> γα τῶν κατὰ
    <lb n="2" />
    <supplied reason="lost">δό</supplied> ξαν ἀληθῆ φαίνε
    <lb n="3" break="no" />
    τ' εἶναι -- προσῆκον ἀντι
    <lb n="4" break="no" />
    <supplied reason="lost">θ</supplied> εἶναι τὰ κατὰ ματα
    <supplied reason="lost">ί</supplied> αν δό
    <supplied reason="lost">ξ</supplied> αν --
    <lb n="5" />
    ὑπόκωφον ἦν, ὃ ὑ <unclear>π</unclear> ή
    <lb n="6" break="no" />
    <unclear>γ</unclear> α <unclear>γ</unclear> εμὴ τὴν ἀναφ
    <supplied reason="lost">ο</supplied>
    <lb n="7" break="no" />
    ρὰν ἐπὶ <supplied reason="lost">τ</supplied> ὰς ἐν
    <supplied reason="lost">α</supplied> ρ
    <supplied reason="lost">γεί</supplied>
    <lb n="8" break="no" />
    ας λαμβάνοντα, καὶ <unclear>μη</unclear>
    <lb n="9" break="no" />
    δὲ τούτων τελέως ἐ <supplied reason="lost">λ</supplied>
    <lb n="10" break="no" />
    λ <supplied reason="lost">είπ</supplied> ειν τοὺς ῥήτορα
    <supplied reason="lost">σ</supplied>
    <gap reason="illegible" quantity="1" unit="character" />
    <lb n="11" />
    ἀλλὰ μὴν ἐν τῶι κατα
    <lb n="12" break="no" /> σκευάζειν τὸ προκε
    <supplied reason="lost">ί</supplied> μ
```

Figure 2. Differences between printed and digital edition - Excerpt of the papyrus n. PHerc 1004 Col. n. 64

Nevertheless, we designed the architecture of the platform to be as loosely coupled as possible, allowing us to accommodate new scholarly requirements in accordance with the principles of Domain-Driven Design and the DSL-based DSE methodology.

Among the wide range of similar works, Papyri.info[8] is one of the most closely aligned with our aims. Indeed, the Papyri.info environment supports collaborative editing, searching capabilities and browsing features for ancient documents. Moreover, the editing component of this tool also embraces the DSL approach by using the Leiden+ formalism.

As previously discussed, digital tools have long been used to study ancient texts. Nowadays, building on established computational methods, recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have the potential to radically transform the study of ancient texts. Indeed, these new approaches are empowering scholars to unlock insights of ancient texts that were unimaginable just a few months ago.

Projects like Ithaca [25] and Logion [26] demonstrate the potential of using Large Language Models (LLMs) for ancient languages - Greek in particular - as evidenced by initiatives which explore Ancient Greek with LLMs [27]–[29].

III. GreekSchools ERC Project

The context in which we are developing the DSL-based DSE methodology, together with the digital environment, is the GreekSchools project.[9]

GreekSchools is a multibeneficiary and multidisciplinary ERC Project (AdG 885222)[10] aiming to establish a new critical edition of Philodemus' *Arrangement of the Philosophers*. From this significant source, preserved by the Herculaneum papyri (Fig. 1 and Fig 3), we can derive a systematic account of the history of Greek philosophical schools, which is unparalleled in its kind.

Making a digital edition of such a textual object poses several challenges [30], [31], including: 1) virtually reading the text hidden on the verso of papyri, 2) detecting, classifying and replacing overlapping layers, 3) virtually reading the text concealed in the layers, 4) editing the text in a collaborative environment, 5) reviewing and commenting the editorial conjectures, 6) preserving conventional editorial practices.

CoPhi Editor is the digital environment that implements the DSL-based DSE methodology. Although initially developed for the digital papyrology domain, it is also applicable to the entire domain of digital textual scholarship.

The collaborative and cooperative nature of such an environment creates the opportunity to widen access to the text for

[8] https://papyri.info/

[9] https://greekschools.eu

[10] https://greekschools.eu/beneficiaries/co-beneficiaries/

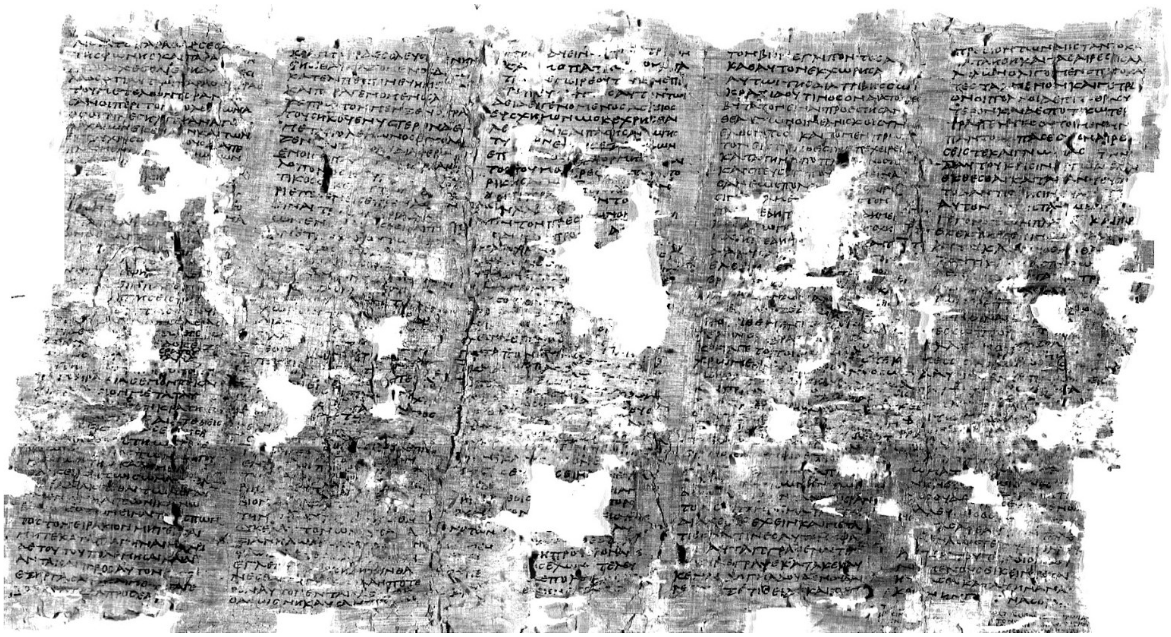


Figure 3. Image of the papyrus n. PHerc 1004

scholars, students, practitioners, and volunteers by lowering the entry barrier, and by allowing users to work remotely in a networked modality. As a consequence of editing the DSL-encoded text, the edition can be seen as an ongoing process that dynamically refines the text through collective effort [32].

IV. Scholar Requirements

Digital textual scholarship presents unique challenges from software engineering and from computational perspectives. Indeed, despite decades of research, philologists continue to struggle with a lack of effective tools and efficient procedures. Ideally, these would be organized as shared services readily adopted by scholars. Furthermore, there is no convergence on how to model software applications to meet philological requirements [33], [34]. Traditional philologists are dissatisfied with the current digital solutions as the digital models and currently available applications do not adequately reflect the real philological needs and methods in terms of their core concepts and domain procedures [35].

Not by chance, scholars have recently reported that philologists are still awaiting a digital environment that allows them to create a critical edition meeting the requirements for editing humanistic texts [30].

A. Silent Observation

Requirements gathering is the essential first step in designing valuable digital environments. By thoroughly understanding end-users' needs, we can build software applications that effectively meet those requirements. However, classical requirements elicitation methods, such as questionnaires and interviews, are often ineffective in the Digital Humanities. This is due to the complexity of the domain, the heterogeneous interactions among scholars belonging to different subdomains, and the lack of a common conceptual framework of reference among traditional humanists, computer scientists, and digital humanists. Moreover, common methodologies to gather requirements are inadequate for disciplines that are difficult to axiomatize, as is the case with the humanities. This frequently results in a significant dissatisfaction among traditional humanists who use digital tools in their daily work, as the tools fail to meet their expectations [34].

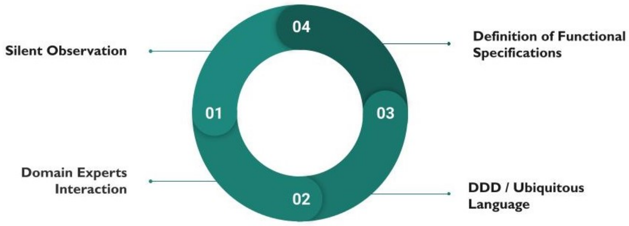
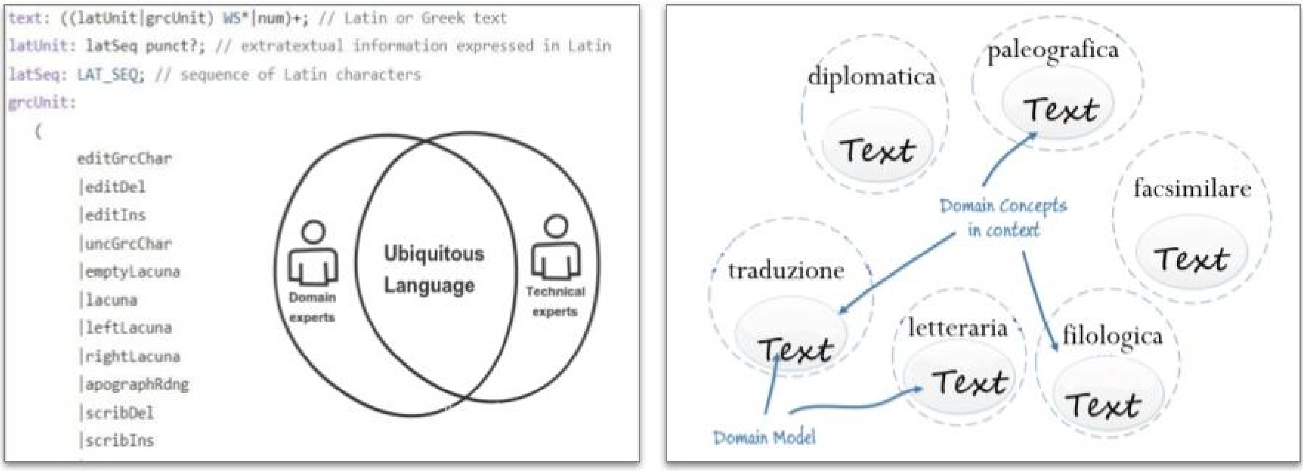

Silent Observation
01
Domain Experts Interaction
02
DDD / Ubiquitous Language
03
04
Definition of Functional Specifications

Figure 4. Domain-Specific Design Process encompasses different phases namely Silent Observation, Domain Expert Interaction, Ubiquitous and Specific Language Definition, Functional Requirements Definition

Instead, we can favor the co-evolution of multidisciplinary teams. A digital humanist serves as a mediator between software engineers and humanists through silent observation of groups of humanists with heterogeneous backgrounds that have already established a way to cooperate. In this reversed approach compared to questionnaires, the requirements are the result of careful induction from the observation and frequent dialogue with the domain experts. The observation must remain silent in the initial phase to avoid altering the environment that we want to model.

For example, papyrologists, paleographers, philologists, historians of the language, and historians of ancient philosophy



```
text: ((latUnit|grcUnit) WS*|num)+; // Latin or Greek text
latUnit: latSeq punct?; // extratextual information expressed in Latin
latSeq: LAT_SEQ; // sequence of Latin characters
grcUnit:
   (
        editGrcChar
        |editDel
        |editIns
        |uncGrcChar
        |emptyLacuna
        |lacuna
        |leftLacuna
        |rightLacuna
        |apographRdng
        |scribDel
        |scribIns
```

Figure 5. Domain-Driven Design and Domain-Specific Languages in implementing Digital Scholarly Editions for Greek Papyri

participating in the GreekSchools project regularly meet to discuss variant readings and conjectures from their disciplinary point of view. These meetings are a precious opportunity for us, digital philologists, to silently observe and take notes of their interactions, the analogical resources they rely on (such as lexica, printed critical editions, and scientific articles), and how and with whom they share supporting materials in the physical space. By observing them in silence, we can better capture the actual needs of a cross-disciplinary group, fostering Domain-Driven Design. This way, we aim to deeply understand the proven practices of such a multidisciplinary and cross-disciplinary domain.

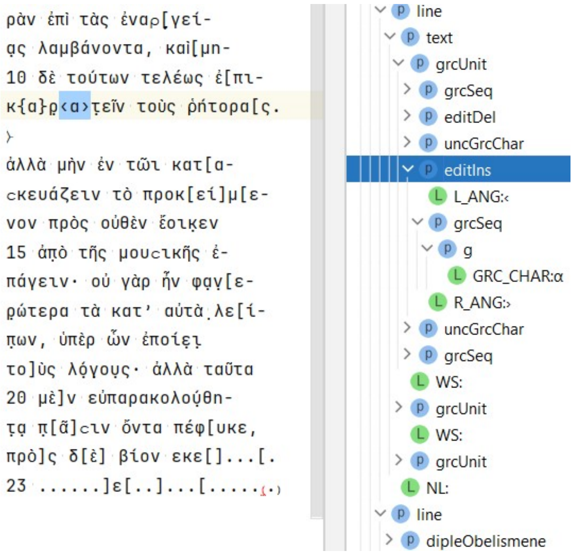
This task plays a crucial role in the effective development of formal models and useful computational tools for textual scholarship activities. Thanks to this approach in requirements gathering, we were able to distill core needs and implement useful features within the platform, as discussed in section V-C.

V. Method

The methodological framework we propose (Fig. 4) leverages Domain-Driven Design (DDD) [36], [37], and Domain Specific Languages (DSLs) [38]–[40]. The former is a software development approach that focuses on modeling complex domains by engaging domain experts in all phases of the development process, providing strategic principles and patterns to describe the problem space.

In particular, DDD's *bounded contexts* make the different data models that constitute the environment modular, such as the linguistic, the paleographic, and the textual context. DDD's *ubiquitous language* effectively describes the domain and the requirements by using a shared and unambiguous vocabulary.

The definition of the corresponding formal languages with a Context-Free Grammar (CFG), must be both familiar to the scholars and able to express the requirements of the domain in a machine-actionable manner (Fig. 5). DSLs embody data representation and processing operations that are directly defined by the domain experts with the mediation of digital humanists.



```
ρὰν ἐπὶ τὰς ἐναρ[γεί-
ạς λαμβάνοντα, καὶ[μη-
10 δὲ τούτων τελέως ἐ[πι-
κ{α}ρ̣‹α›ṭεῖν τοὺς ῥήτορα[ς.
›
ἀλλὰ μὴν ἐν τῶι κατ[α-
ϲκευάζειν τὸ προκ[εί]μ[ε-
νον πρὸς οὐθὲν ἔοιḳεν
15 ἀπ̣ὸ τῆς μουϲικῆς ἐ-
πάγειν· οὐ γὰρ ἦν φạγ̣[ε-
ρ̣ώτερα τὰ κατ’ αὐτὰ.λε[ί-
π̣ων, ὑπὲρ ὧν ἐποίẹι̣
το]ὺς λό̣γοụς· ἀλλὰ ταῦτα
20 μὲ]ν εὐπαρακολοῦ̣θη-
ṭạ π̣[ᾶ]ϲιν ὄντα πέφ[υκε,
πρὸ]ς δ[ὲ] βίον εκε[]...[.
23 ......]ε[..]...[.....(.)
```

```
v P line
  v P text
    v P grcUnit
      > P grcSeq
      > P editDel
      > P uncGrcChar
      v P editIns
          L L_ANG:‹
        v P grcSeq
          v P g
              L GRC_CHAR:α
          L R_ANG:›
      > P uncGrcChar
      > P grcSeq
      L WS:
    > P grcUnit
      L WS:
    > P grcUnit
      L NL:
v P line
  > P dipleObelismene
```

Figure 6. Digital Edition as Domain-Specific Language: Hierarchical View.

This methodology aims to provide scholars with a familiar and rich environment that empowers the editing process while retaining the long-standing and well established good practices of the philology domain (Fig. 6). To achieve this goal we define DSLs with the active participation of the domain experts (as explained in Sec. IV).

Consequently, the implementation of this method allows us to: 1) preserve the traditional expressiveness that textual scholarship practices have refined over time; 2) implement both generic tools and specific languages [41], making them usable and reusable.

The DSL-based methodology has also been applied outside the scholarly editing domain. In fact, one of the earliest attempts to use the DSL approach was in the scientific effort to describe natural languages within a formal framework. As an example, Figure 7 shows an application of parsing Greek text for natural language processing [29].

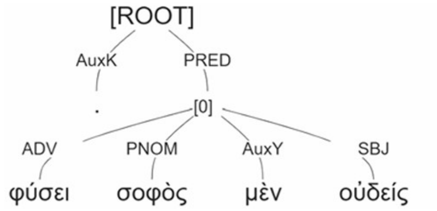


Figure 7. Example of linguistic parsing tree representation

A. DSL-based Digital Scholarly Editions

The scholarly editing methodology based on Domain Specific Languages (DSL-Based DSE) requires the definition of a formal language derived from the well-established ecdotic practices already adopted within the editorial workflow. This specification represents the text along with the corresponding philological interpretations in an analytical manner.

Figure 8 illustrates the DSL-based DSE concept with a ”Greek Textual Unit.” This figure depicts a graphical representation that includes all the editorial conventions used by philologists in textual criticism and text reconstruction.

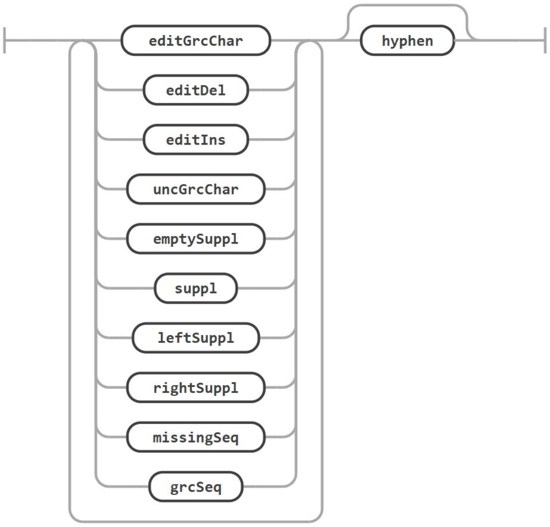


Figure 8. Representation of a production rule for a DSL-based DSE

Editorial conventions can be combined. For instance, the Greek text ῥ̣ήτορα[ς. records a right lacuna (indicated by the square brackets), a supplement (the letter sigma within the brackets), and a missing character (indicated by the dot). These editorial phenomena are:

1) Editorial Greek *Emendatio* (e.g. ξ_*) (editGrcChar)
2) Editorial Deletion (e.g., {α}) (editDel)
3) Editorial Insertion (e.g., <α>) (editIns)
4) Unclear Character (e.g., ρ̣) (uncGrcChar)
5) Empty Supplied (e.g., []) (emptySuppl)
6) Supplied Text (e.g., [εί]) (suppl)
7) Left Supplied Text (e.g., δ]) (leftSuppl)
8) Right Supplied Text (e.g., [ς) (rightSuppl)
9) Missing Sequence (e.g., . . .) (missingSeq)
10) Greek Sequence (e.g., τῆς) (grcSeq)
11) Hyphenation Greek Text (e.g., ἀντι-) (hyphen)

The textual practices used in scholarly editions can be considered as domain-specific languages, and the majority of these languages are formal enough to be derived from a Context-Free Grammar (CFG). A good example of a quasi-formal language is the critical apparatus (see Fig. 9 and Sec. VI).

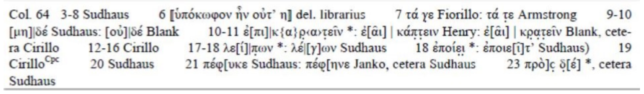

Col. 64 3-8 Sudhaus 6 [ὑπόκωφον ἦν οὐτ’ η] del. librarius 7 τά γε Fiorillo: τά ṭε Armstrong 9-10 [μη]|δέ Sudhaus: [οὐ]|δέ Blank 10-11 ἐ[πι]|κ{α}ρ̣‹α›ṭεῖν *: ἐ[ᾶι] | κάπ̣ṭειν Henry: ἐ[ᾶι] | κρ̣αṭεῖν Blank, cetera Cirillo 12-16 Cirillo 17-18 λε[ί]|π̣ων *: λέ|[γ]ων Sudhaus 18 ἐποίẹι̣ *: ἐποιε[ῖ]τ’ Sudhaus) 19 Cirillo^{Cpc} 20 Sudhaus 21 πέφ[υκε Sudhaus: πέφ[ηνε Janko, cetera Sudhaus 23 πρὸ]ς̣ δ̣[έ] *, cetera Sudhaus

Figure 9. Philological Apparatus in Printed Edition

Treating the text as machine actionable code, written in the formal language defined by a DSL, allows us to manage it as an Abstract Syntax Tree (AST) that represents the structure and the relationships of the textual phenomena (see Fig. 10). The CFG defines part of the domain model, and the AST representation is suitable for obtaining a digital representation of the text following available standard formats, such as the XML/TEI encoding schema. Since AST inherently represents the domain model and the functionalities to process it, we can actually define formal scholarly abstract data types. This feature automatically ensures actionability within the processing model of the data.

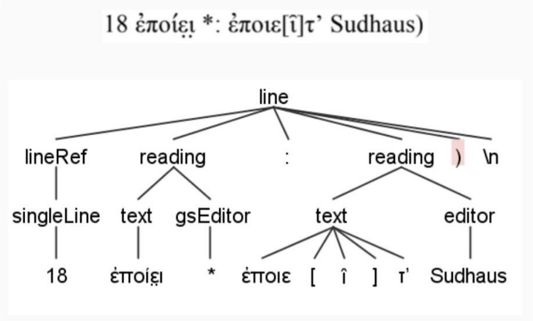


Figure 10. Philological apparatus and its Abstract Syntax Tree

B. CoPhi Editor Model

CoPhi Editor is a scholarly platform made of modular components following a micro-services architecture. This kind of design entails that the different parts of the model are reusable, in order to maximize the modularity of the whole system. The *data model* of the digital scholarly platform involves abstract entities and data types defined by leveraging the DDD ubiquitous language and implemented though the CFG of the DSL. For example, in Figure 10, the entities are *line*, *reading*, *text*, *editor*, etc.

For the same domain, we can define more than one language. Nevertheless, each one must be unambiguous within a *bounded context* identifying a specific semantic space within the domain. Moreover, the model derived from the DSL is

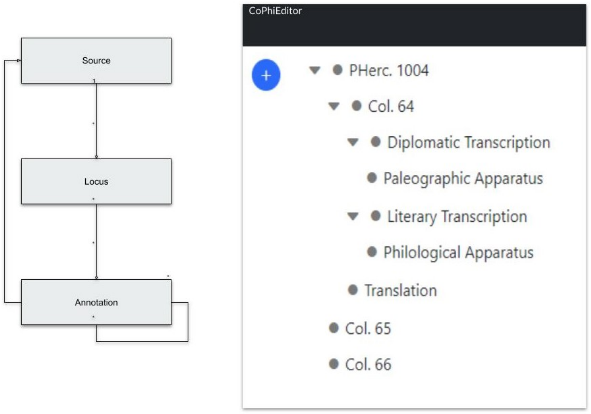


Figure 11. CoPhi Editor Data Model and workflow

combined with a general, agnostic and recursive model used for the collaborative work (see Fig.11).

This latter encompasses *Annotation*, *Locus* and *Source* entities which are based on a recursive data representation for the annotated text. Such a model allows scholars with an unlimited levels and an indefinite granularities of annotations. Additionally, the model can be implemented using the standard ontology of *Web Annotation Data Model* (WADM).[11] WADM "describes a structured model and format to enable annotations to be shared and reused".

C. Implementation

To support the *constitutio textus* of texts via CoPhi Editor, scholars work as they are accustomed to do (Fig. 12, Fig. 13):

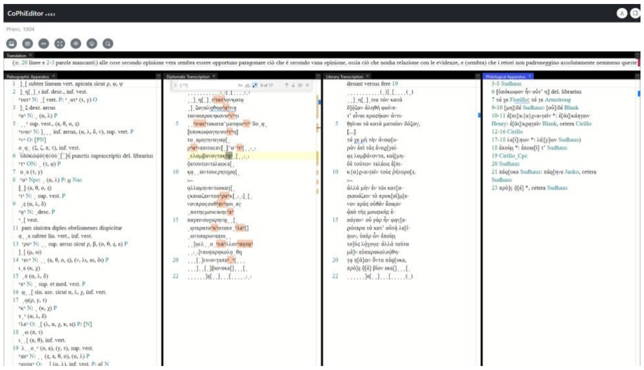
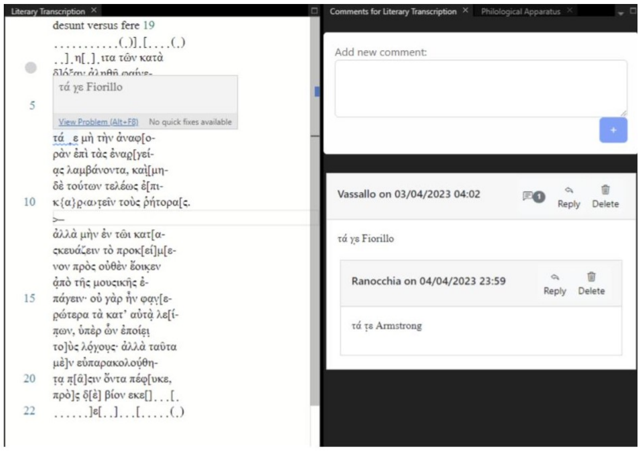


Figure 12. The CoPhi Editor Editing Platform

(i) transcribing the text from a primary source to create a diplomatic edition of the document (i.e., the selected column or the selected fragment); (ii) describing all sorts of assertions about character vestiges preserved on the support in a paleographic apparatus; (iii) producing a literary transcription (i.e., the *constitutio textus* conducted by the editor); (iv) describing in the philological apparatus the responsibility for each relevant reading alongside authoritative conjectures as well as other important variants that are compatible with the space and the preserved vestiges; (v) translating the reconstructed text (i.e., in Italian).



desunt versus fere 19
. . . . (.)].[. . . .(.)
. .].η[.].ιτα τῶν κατὰ
δ]όξαν· ἀληθῆ φαίνε-

τά γε Fiorillo

τά ̣ε μὴ τὴν ἀναφ[ο-
ρ̣ὰν ἐπὶ τὰς ἐναρ̣[γεί-
α̣ς λαμβάνοντα, καὶ̣[μη-
δὲ τούτων τελέως ἐ[πι-
κ{α}ρ̣‹α›ṭεῖν τοὺς ῥ̣ήτορα[ς.
>—
ἀλλὰ μὴν ἐν τῶι κατ[α-
ςκευάζειν τὸ προκ[εί]μ[ε-
νον πρὸς οὐθὲν ἔοικ̣εν
ἀ̣πὸ τῆς μουςικῆς ἐ-
πάγειν· οὐ γὰρ ἦ̣ν φα̣ν̣[ε-
ρ̣ώτερα τὰ κατ' αὐτὰ̣ λε[ί-
π̣ων, ὑπὲρ ὧ̣ν ἐποίε̣ι̣
το]ὺς λό̣γ̣ου̣ς· ἀλλὰ ταῦτα
μὲ]ν εὐπαρακολού̣θη-
ṭα̣ π̣[ᾶ]ςιν ὄντα πέφ[υκε,
πρὸ]ς δ̣[ὲ] βίον εκε[]. . .[.
. . . . . .]ε[. .]. . .[. . . . .(.)

Vassallo on 03/04/2023 04:02 1

τά γ̣ε Fiorillo

Ranocchia on 04/04/2023 23:59

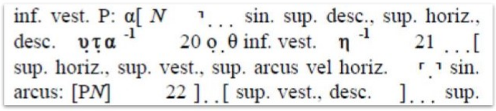
τά ṭε Armstrong

Figure 13. CoPhi Editor GUI editing and comments

Each step produces a special piece of text containing editorial conventions identified by special characters, whose meaning is typically shared by scholars within the same research domain or community of practice. Furthermore, CoPhi Editor provides scholars with cutting-edge features such as: 1) spell checking, 2) automated check for editorial conventions, 3) querying data and searching for metadata, 4) rich text editor, 5) cooperation and collaboration, 6) dynamic interface layout, 7) computer-assisted editing, 8) agnostic output format.

VI. DSL-Based DSE in practice

A. Paleographic Apparatus



inf. vest. P: α[*N* ⌝. . . sin. sup. desc., sup. horiz., desc. υ̣ṭα̣ $^{-1}$ 20 ο̣.̣θ inf. vest. η $^{-1}$ 21 . . .[sup. horiz., sup. vest., sup. arcus vel horiz. ⌜.⌝ sin. arcus: [P*N*] 22]. .[sup. vest., desc.]. . . sup.

Figure 14. Traditional Paleographic Apparatus

Practical examples can further illustrate the proposed workflow. Imagine a philologist writing text in the CophiEditor, adhering to their established conventions (as shown in Fig. 14, which depicts an excerpt of a printed palaeographical apparatus). For instance, when a philologist has to represent a fragment of text believed to be from a layer underlying the primary text on the physical support, he adds the corresponding entry, e.g., υ̣ṭα̣$^{-1}$, to the paleographic apparatus and makes no further actions (the superscript negative number refers to the first layer under the writing surface).

```
app: lem ( (witDetail wit*)
     | (witDetail? wit+ COLON rdg? witDetail?
       (wit+ | witsAbs) (COLON rdg? witDetail?
       (wit+ | witsAbs))*) );
lem: position? lectio;
witDetail: (layer | (velPalStat | palStat))
          (COMMA (velPalStat | palStat))*;
layer : sub | sup;
```

Listing 1. Contex Free Grammar in g4 format

[11] https://www.w3.org/TR/annotation-model/

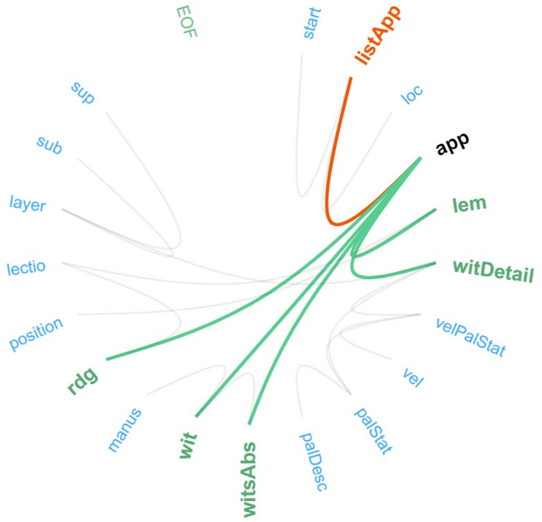


Figure 15. Graph Visualization for the DSL on Paleographic Apparatus

Thanks to the DSL toolkit, the digital scholarly environment recognizes the ecdotic conventions used in the apparatus and processes them as a formal language (Listing 1 shows a CFG excerpt, Fig. 15 also provides a visual representation of it), creating the AST of the parsed text (see Fig. 16 and Listing 2).

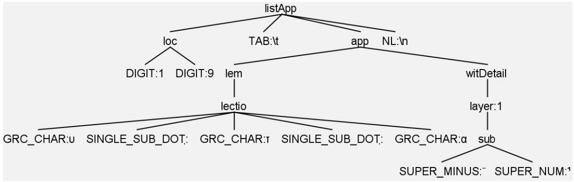


Figure 16. Abstract Syntax Tree on Paleographic Apparatus

Subsequently, XSL transformations (see Listing 4) enable the system to produce a standard representation in shared formats like TEI/EpiDoc[12] or PDF (see Listing 3).

```
<app><lem><lectio>υ̣τ̣α</lectio></lem>
 <witdetail>
  <layer><sub>−1</sub></layer>
 </witdetail></app>
```

Listing 2. XML serialization of the Abstract Syntax Tree

```
<app loc="19"><lem>υ̣τ̣α</lem>
 <note>underwritten text (−1 stratum)</note>
</app>
```

Listing 3. XML EpiDoc standard serialization of the the text excerpt

```
<xsl:template match="app">
 <xsl:variable name="loc"
   select="preceding-sibling::loc"/>
 <app loc={$loc}><xsl:apply-templates/></app>
</xsl:template>

<xsl:template match="lem">
 <xsl:element name="lem">
  <xsl:value-of select="current()/lectio" />
 </xsl:element></xsl:template>
```

```
<xsl:template match="witdetail">
 <xsl:choose>
   [...]
  <xsl:otherwise>
   <note><xsl:apply-templates /></note>
  </xsl:otherwise>
 </xsl:choose>
</xsl:template>

<xsl:template match="layer">
 <xsl:value-of
  select="if(./sub)
    then 'underwritten text'
     else 'overwritten text' " />
  (<xsl:value-of
   select="concat(sub|sup,' stratum')" />)
</xsl:template>
```

Listing 4. XSL Transformations of the AST derived from the text

B. Diplomatic transcription

Another illustrative example of our DSL-based DSE methodology is its application to the digital representation of papyri in diplomatic editions. Papyri editions may contain ecdotic representation for paleographical and scribal phenomena like apographs (variant readings), interlinear characters (written above the line), and unclear characters. Our DSL effectively recognizes and manages these elements, as shown in Figure 17. Consider the following Greek reconstruction sequence:

..⌈ιναι⌉ταϰατα\ματα̣ια⌈ν⌉/δο.α̣.

This sequence represents line 5 of PHerc 1004, col. 64, in diplomatic transcription.

Listing 5 showcases the XML serialization of the recognized text using dedicated diplomatic tags like <apographrdng>, <scribins>, <uncgrcchar>.

```
<line><text><grcunit>
 <u>.</u><u>.</u>
 <apographrdng>⌈
  <g>ι</g><g>ν</g><g>α</g><g>ι</g>
  ⌉</apographrdng>
  <g>τ</g><g>α</g><g>ϰ</g><g>α</g><g>τ</g><g>α</g>
  <scribins>\
    <grcunit><g>μ</g><g>α</g><g>τ</g>
     <uncgrcchar><g>α̣</g></uncgrcchar>
     <g>ι</g><g>α</g>
     <apographrdng>⌈<g>ν</g>⌉</apographrdng>
   </grcunit>/</scribins>
  <g>δ</g><g>ο</g><u>.</u>
  <uncgrcchar><g>α̣</g></uncgrcchar><u>.</u>
 </grcunit></text></line>
```

Listing 5. XML-serialization of the AST derived from the text

As described in the previous section, the DSL-XML serialization can be automatically converted into the interoperable TEI/EpiDoc format using XSL transformations (Listing 6).

```
<xsl:template match="grcunit/apographrdng">
    <xsl:element name="supplied" >
       [...]
    </xsl:element>
</xsl:template>

 <xsl:template match="grcunit/scribins">
    <xsl:element name="add" >
       [...]
    </xsl:element>
</xsl:template>
```

[12] https://epidoc.stoa.org/gl/latest/

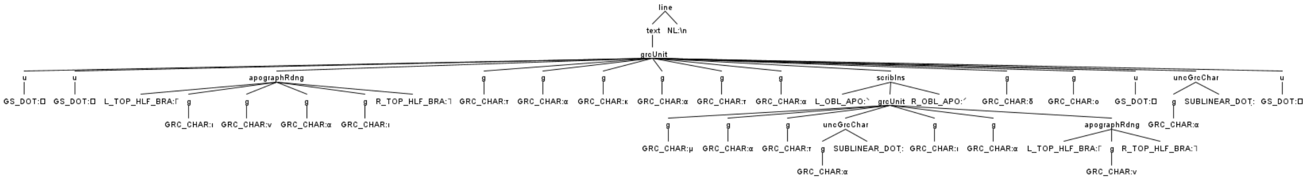
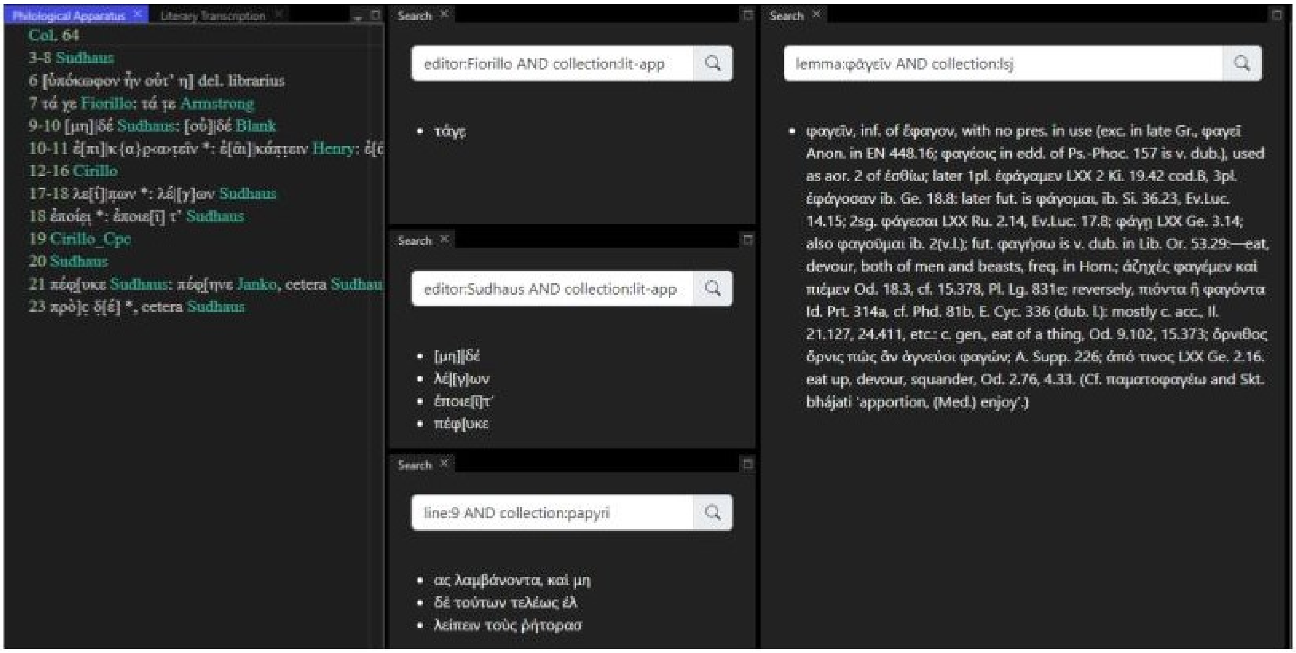


Figure 17. Example of the Abstract Syntax Tree for the Greek sequence ..⌈ιναι⌉τακατα\ματạια⌈ν⌉/δο.ạ. in a diplomatic transcription



Col. 64
3-8 Sudhaus
6 [ὑπόκωφον ἦν οὐτ' η] del. librarius
7 τά γε Fiorillo: τά γε Armstrong
9-10 [μη]|δέ Sudhaus: [οὐ]|δέ Blank
10-11 ἐ[πι]|κ{α}ρ‹α›τεῖν *: ἐ[ᾶι]|κάρτειν Henry: ἐ[ε
12-16 Cirillo
17-18 λε[ί]|πων *: λέ|[γ]ων Sudhaus
18 ἐποίει *: ἐποιε[ῖ] τ' Sudhaus
19 Cirillo_Cpc
20 Sudhaus
21 πέφ[υκε Sudhaus: πέφ[ηνε Janko, cetera Sudhau
23 πρὸ]ς δ[έ] *, cetera Sudhaus

editor:Fiorillo AND collection:lit-app

- τάγε

editor:Sudhaus AND collection:lit-app

- [μη]|δέ
- λέ|[γ]ων
- ἐποιε[ῖ]τ'
- πέφ[υκε

line:9 AND collection:papyri

- ας λαμβάνοντα, καὶ μη
- δέ τούτων τελέως ἐλ
- λείπειν τοὺς ῥήτορας

lemma:φαγεῖν AND collection:lsj

- φαγεῖν, inf. of ἔφαγον, with no pres. in use (exc. in late Gr., φαγεῖ Anon. in EN 448.16; φαγέοις in edd. of Ps.-Phoc. 157 is v. dub.), used as aor. 2 of ἐσθίω; later 1pl. ἐφάγαμεν LXX 2 Ki. 19.42 cod.B, 3pl. ἐφάγοσαν ib. Ge. 18.8: later fut. is φάγομαι, ib. Si. 36.23, Ev.Luc. 14.15; 2sg. φάγεσαι LXX Ru. 2.14, Ev.Luc. 17.8; φάγῃ LXX Ge. 3.14; also φαγοῦμαι ib. 2(v.l.); fut. φαγήσω is v. dub. in Lib. Or. 53.29:—eat, devour, both of men and beasts, freq. in Hom.; ἀζηχὲς φαγέμεν καὶ πιέμεν Od. 18.3, cf. 15.378, Pl. Lg. 831e; reversely, πιόντα ἢ φαγόντα Id. Prt. 314a, cf. Phd. 81b, E. Cyc. 336 (dub. l.): mostly c. acc., Il. 21.127, 24.411, etc.: c. gen., eat of a thing, Od. 9.102, 15.373; ὄρνιθος ὄρνις πῶς ἂν ἁγνεύοι φαγών; A. Supp. 226; ἀπό τινος LXX Ge. 2.16. eat up, devour, squander, Od. 2.76, 4.33. (Cf. παματοφαγέω and Skt. bhájati 'apportion, (Med.) enjoy'.)

Figure 18. Cophi Editor view for quering the corpus

```
  <xsl:template match="grcunit/uncgrcchar">
    <xsl:element name="damage" >
        <xsl:apply-templates />
    </xsl:element>
</xsl:template>

<xsl:template match="line/text/grcunit/u">
    <xsl:element name="gap">
        <xsl:attribute name="reason" />
        <xsl:attribute name="quantity" />
        <xsl:attribute name="unit" />
        <xsl:attribute name="ana" />
    </xsl:element>
</xsl:template>
```

Listing 6. XSLT transformation of XML-AST to XML-TEI/EPIDOC

Listing 7 displays the corresponding TEI/EpiDoc representation of the diplomatic text shown earlier. This includes:

1) the `<gap>` element, which encodes the lacunae (missing portions) in the text, 2) the `<supplied>` tag with the `@evidence` attribute equals to *parallel-apograph*, which indicates a supplied reading based on a *apograph* sources, 3) the `<add>` element with a nested `<supplied>` tag, which represents an addition and an editorial intervention made to the text, 4) the `<damage>` tag, which indicate that the character is incomplete.

```
<ab><lb n="1"/>
 <gap reason="illegible"
   quantity="1" unit="character" ana="#vestige"/>
 <gap reason="illegible"
   quantity="1" unit="character" ana="#vestige"/>
 <supplied evidence="apograph" reason="lost">
  ιναι</supplied>τακατα
 <add place="interlinear">ματ<damage>α</damage>ια
  <supplied evidence="apograph" reason="lost">
 ν</supplied></add>δο
 <gap reason="illegible"
   quantity="1" unit="character" ana="#vestige"/>
 <damage>α</damage>
 <gap reason="illegible"
   quantity="1" unit="character" ana="#vestige"/>
<ab>
```

Listing 7. XML TEI/EpiDoc standard for representing digital papyri

C. Querying the editions

The Cophi Editor platform, based on DSL-based Digital Scholarly Edition (DSE) technology, empowers scholars to improve their research workflows. This is achieved by enabling them to efficiently search and analyze all the encoded information. Scholars can leverage Cophi Editor's search functionalities to retrieve relevant passages, filter based on specific criteria, and explore connections between different elements within the text. Additionally, the platform facilitates deeper evaluation of the content by providing access to editorial annotations, historical context, and alternative readings.

The scholarly platform achieves this search capability through a real-time parsing process. The platform leverages predefined formal languages (see Fig. 18) to analyze and understand the structure and meaning of encoded textual resources.

Cophi Editor empowers scholars to conduct comprehensive research by enabling them to query a wealth of external resources directly within the platform. This includes dictionaries, lexica, parallel editions, and primary sources, among others.

Listing 8 shows the search capabilities mechanism. It utilizes the XQuery[13] language (a specialized search language) to efficiently retrieve all the alternative readings proposed by selected editors within the philological apparatus.

```
case "collection:lit-app" return
  let $editor := $param
  for $rdgs in $doc//app/rdg
  where $rdgs/@resp = $editor
  let $parent := $rdgs/parent::app
  return
   <li data-resp="{$rdgs/@resp}"
       data-loc="{$parent/@loc}">{data($rdgs)}
   </li>
```

Listing 8. XQuery fragment to further investigate the scholarly transcriptions

VII. Paleographic investigations

Critical editions of papyri aim to offer to the scholars the most likely reconstruction of the original text and the list of variant readings attested by different sources or conjectural supplements suggested by previous editors.The critical edition must be in agreement with the diplomatic edition of the same papyrus made by the same scholar or team of scholars. Indeed, the editor of the diplomatic edition describes possible readings of the document, and the editor of the critical edition decides which of these is the most likely.

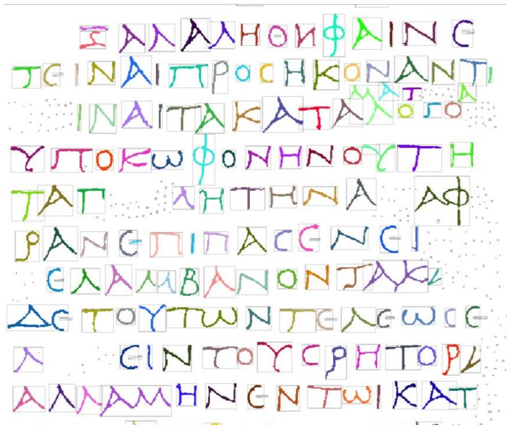
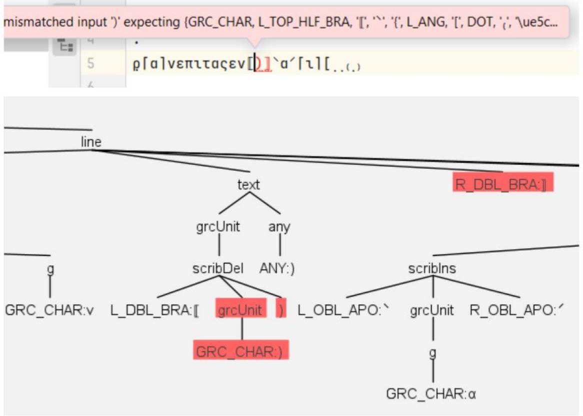


Figure 19. Example of connected component exploitation

For instance, the diplomatic edition registers that a couple of oblique signs may represent a lambda or part of a nearly illegible alpha and indicates the presence of any gap (lacuna), whereas the critical edition presents only the most probable choice between alpha or lambda and suggests how to fill the gap, according to linguistic and historico-philological contextual evidence. But the paleographic plausibility must be taken into account at any stage of the editorial process. First, recognition of single uncertain glyphs or parts of a glyph is based on comparison with other glyphs. Secondly, when filling textual gaps in a conjectural manner, it is crucial to ensure that the chosen glyphs are compatible with the average width of the specific letters involved (e.g., an M is larger than an I). The connected components technique [42] facilitates the identification of single glyphs and, consequently, the computation of the average width and height of them (see Fig. 19). Investigations in this direction and the related bibliography are discussed in [43] and [44].

VIII. Conclusion and future works

In this paper, we explored an innovative approach based on Domain-Driven Design (DDD) for co-designing and developing CophiEditor, a digital editing environment that caters to the needs of traditional papyrologists. This approach strikes a crucial balance between familiarity for traditional scholars and machine actionability for computational analysis and digital preservation. By leveraging Domain-Specific Languages (DSLs), CophiEditor empowers traditional scholars to create machine-actionable Digital Scholarly Editions (DSEs), facilitating both digital preservation and the development of new computational methods for scholarly editing (as shown in Fig. 20 with suggestions and error checking).



```
mismatched input ')' expecting {GRC_CHAR, L_TOP_HLF_BRA, '⟦', '`', '{', L_ANG, '[', DOT, ';', '\ue5c...
5    ρ[α]νεπιταςεν[)]`α´[ι][..(.)
```

Figure 20. DSL Error Checking and Suggestions capabilities

The context of digital scholarly editing has been enhanced through the implementation of a DSL-based DSE methodology. This approach has demonstrated its effectiveness and success in supporting textual scholarship within the GreekSchools ERC project.

While the computational scholarly environment offers a promising approach for digital papyrology, there is still room for improvement.

Further work will focus on conducting quantitative assessments of the platform's usability, along with collecting valuable statistics on textual encoding via DSL using the CoPhi Editor, in comparison with more traditional XML-based

[13] https://www.w3.org/TR/xquery-31/

encoding processes. This assessment will include factors such as time required for encoding and user feedback. Secondly, further improvements are needed for the DSL, including refining composition strategies and contextual syntax suggestions to further enhance user experience.

The integration of machine learning (ML) technologies and neural network architectures for natural language processing (NLP) presents exciting possibilities for future development. We plan to investigate the use of ML for various downstream tasks, such as: 1) filling in lost fragments of text; 2) detecting word boundaries more accurately; 3) checking the validity of previously proposed conjectures.

Finally, we plan to enhance characters recognition from facsimile images or papyrus drawings.

References

[1] M. Berti, Historical Fragmentary Texts in the Digital Age. Berlin, Boston: De Gruyter Saur, 2019, pp. 257–276.

[2] M. Berti, Digital Editions of Historical Fragmentary Texts, ser. Digital Classics Books. Heidelberg: Propylaeum, 2021, no. 5.

[3] N. Reggiani, Ed., Digital Papyrology I. Berlin, Boston: De Gruyter, 2017.

[4] N. Reggiani, Ed., Digital Papyrology II. Berlin: De Gruyter, 2018.

[5] M. K. Gold, Ed., Debates in the digital humanities. Minneapolis, MN: University of Minnesota Press, 2012.

[6] M. J. Driscoll and E. Pierazzo, Eds., Digital Scholarly Editing: Theories and Practices, ser. Digital Humanities Series. Open Book Publishers, 2016, vol. 4.

[7] D. Schmidt, “The inadequacy of embedded markup for cultural heritage texts,” Literary and Linguistic Computing, vol. 25, no. 3, 2010, pp. 337–356.

[8] P. Durusau, “Hypergraphs: Escaping the Surly Bonds of Syntax,” in Proceedings of Balisage: The Markup Conference 2023. Balisage Series on Markup Technologies, vol. 28, 2023.

[9] F. Boschetti et al., “Collaborative and Multidisciplinary Annotations of Ancient Texts: The Euporia System,” in The Ancient World Goes Digital, vol. 6, Leiden: Brill Academic Publishers, 2023, pp. 172–223.

[10] F. Boschetti and A. M. Del Grosso, “TeiCoPhiLib: A Library of Components for the Domain of Collaborative Philology,” Journal of the Text Encoding Initiative, no. 8, Jan. 2015.

[11] S. Zenzaro, A. M. Del Grosso, F. Boschetti, and G. Ranocchia, “Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d’uso GreekSchools,” in Culture Digitali. Intersezioni, Filosofia, Arti e Media. Università del Salento - AIUCD, 2022.

[12] P. Robinson, “Some principles for making collaborative scholarly editions in digital form,” Digital Humanities Quarterly, vol. 11, no. 2, 2017.

[13] R. Baumann, “The ’son of suda on-line’,” Bulletin of the Institute of Classical Studies. Supplement, no. 122, 2013, pp. 91–106.

[14] M. Foys, Virtual Mappa: Digital Editions of Early Medieval Maps of the World. Schoenberg Institute of Manuscript Studies (University of Pennsylvania), 2018.

[15] A. C. Williams, A. Santarsiero, C. Meccariello, G. Verhasselt, H. D. Carroll, J. F. Wallin, D. Obbink, and J. H. Brusuelas, “Proteus: A platform for born digital critical editions of literary and subliterary papyri,” in 2015 Digital Heritage, vol. 2, 2015, pp. 453–456.

[16] S. Dumont, N. Arndt, S. Grabsch, and L. Klappenbach, “ediarum.base.edit v2.0.0,” Dec. 2021.

[17] M. Janssen, “TEITOK: Text-faithful annotated corpora,” in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16). Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 4037–4043.

[18] R. Rosselli Del Turco, C. Martignano, C. Di Pietro, G. Cacioli, A. M. Del Grosso, and S. Zenzaro, “DSE Visualisation with EVT: Simplicity is Complex,” in Complexities, 2019.

[19] E. Morlock, “TEIPublisher for EpiDoc,” in Visible Words: Digital Epigraphy in a Global Perspective. Providence, United States: John Bodel, Brown University and Michèle Brunet, Université Lumière Lyon 2, HiSoMA, Oct. 2017.

[20] G. Bodard and P. Yordanova, “Publication, testing and visualization with EFES: A tool for all stages of the epidoc xml editing process,” Studia Universitatis Babes, -Bolyai Digitalia, vol. 65, no. 1, p. 17–35, Dec. 2020.

[21] B. Almas, “Perseids: Experimenting with infrastructure for creating and sharing research data in the digital humanities,” Data Science Journal, Apr 2017.

[22] K. E. Fendt, J. Paradis, and Massachusetts Institute of Technology, “Annotation Studio: Multimedia Annotation for Students,” 2016.

[23] G. del Rio Riande and V. Vitale, “Recogito-in-a-box: From annotation to digital edition,” Modern Languages Open, Aug 2020.

[24] J. Horstmann, “Undogmatic Literary Annotation with CATMA,” in Functions, Differentiation, Systematization, J. Nantke and F. Schlupkothen, Eds. Berlin, Boston: De Gruyter, 2020, pp. 157–176.

[25] Y. Assael et al., “Restoring and attributing ancient texts using deep neural networks,” Nature, vol. 603, no. 7900, pp. 280–283, Mar. 2022.

[26] C. Cowen-Breen, C. Brooks, J. Haubold, and B. Graziosi, “Logion: Machine Learning for Greek Philology.” 2023.

[27] F. Riemenschneider and A. Frank, “Exploring Large Language Models for Classical Philology,” in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 15181–15199.

[28] T. Sommerschield et al., “Machine Learning for Ancient Languages: A Survey,” Computational Linguistics, vol. 49, no. 3, Sep. 2023, pp. 703–747.

[29] A. Keersmaekers and T. Van Hal, “Creating a large-scale diachronic corpus resource: Automated parsing in the Greek papyri (and beyond),” Natural Language Engineering, 2023, pp. 1–30.

[30] J. H. Brusuelas, “Scholarly editing and AI: Machine predicted text and herculaneum papyri,” magazén, no. 1, jun 2021.

[31] F. P. Romano, E. Puglia, C. Caliri, D. P. Pavone, M. Alessandrelli, A. Busacca, C. G. Fatuzzo, K. J. Fleischer, C. Pernigotti, Z. Preisler, C. Vassallo, G. Verhasselt, C. Miliani, and G. Ranocchia, “Layout of ancient Greek papyri through lead-drawn ruling lines revealed by Macro X-Ray Fluorescence Imaging,” Scientific Reports, vol. 13, no. 1, p. 6582, Apr. 2023.

[32] Z. Bauman, Culture in a liquid modern world. John Wiley ’n Sons, 2013.

[33] A. M. Del Grosso, E. Giovannetti, and S. Marchi, “The Importance of Being... Object-Oriented: Old Means for New Perspectives in Digital Textual Scholarship,” in Advances in Digital Scholarly Editing, 2017, pp. 269–274.

[34] F. Boschetti, R. Del Gratta, and A. M. Del Grosso, “The role of digital scholarly editors in the design of components for cooperative philology,” in Advances in Digital Scholarly Editing, Sidestone Press, 2017, pp. 249–253.

[35] G. Mugelli, F. Boschetti, R. Del Gratta, A. M. Del Grosso, F. Khan, and A. Taddei, “A User-Centered Design to Annotate Ritual Facts in Ancient Greek Tragedies,” Bulletin of the Institute of Classical Studies, vol. 59, no. 2, 2016, pp. 103–120.

[36] E. Evans, Domain-Driven Design Reference: Definitions and Pattern Summaries. Dog Ear Publishing, 2014.

[37] S. Millett and T. Nick, Patterns, Principles and Practices of Domain-Driven Design. John Wiley ’n Sons, 2015.

[38] T. Parr, Language implementation patterns create your own domain-specific and general programming languages. Pragmatic Bookshelf, 2014.

[39] M. Fowler, Domain-Specific Languages, ser. Addison-Wesley Signature Series (Fowler). Pearson Education, 2010.

[40] A. Bucchiarone, A. Cicchetti, F. Ciccozzi, and A. Pierantonio, Domain-Specific Languages in Practice with JetBrains MPS. Springer, Jul. 2021.

[41] M. Voelter, Generic Tools, Specific Languages. Delft University of Technology, 2014.

[42] L. He, X. Ren, Q. Gao, X. Zhao, B. Yao, and Y. Chao, “The connected-component labeling problem: A review of state-of-the-art algorithms,” Pattern Recognition, vol. 70, 2017, pp. 25–43.

[43] F. Ponchio, M. Lamé, R. Scopigno, and B. Robertson, “Visualizing and transcribing complex writings through RTI,” in 2018 IEEE 5th International Congress on Information Science and Technology (CiSt). IEEE, 2018, pp. 227–231.

[44] M. Lamé, S. Rosmorduc, S. Polis, M. Dellepiane, G. Sarullo, A. Barmpoutis, E. Bozia, and F. Boschetti, “Technology and tradition: A synergic approach to deciphering, analyzing and annotating epigraphic writings,” Lexis, 2015, pp. 9–30.