Umanistica Digitale – ISSN: 2532-8816 – n. 23, 2026: Special Issue
F. Lazzerini, M. Di Franco – *Creating a digital critical edition of a classical text with XML/TEI*
DOI: http://doi.org/10.60923/issn.2532-8816/23486

Creating a digital critical edition of a classical text with XML/TEI

Federica Lazzerini

University of Turin, Italy
federica.lazzerini@unito.it

Matteo Di Franco

University of Naples "Federico II", Italy
matteo.difranco@unina.it

Abstract

Il contributo analizza la possibilità di realizzare una edizione critica digitale di un testo classico in XML/TEI, evidenziandone limiti e potenzialità. Dopo aver richiamato la complessità della tradizione testuale, si discutono i due metodi TEI per collegare testo e apparato: *Parallel Segmentation* e *Double End-Point Attachment*. Nel progetto RETI è stata sviluppata una variante del metodo DEPA in *stand-off mode*, integrata in EVT3, che consente di rappresentare famiglie di manoscritti, mani correttive e gerarchie di testimoni. Metadati specifici permettono di classificare le varianti per rilevanza stemmatica, rendendo visibile e interrogabile il processo della *constitutio textus*.

Parole chiave: Edizione critica digitale; XML/TEI; DEPA; Varianti testuali; Stand-off Markup

The paper analyses the feasibility of a digital critical edition of a classical text using XML/TEI, highlighting its limits and potential. After outlining the complexity of textual transmission, it discusses the two TEI methods for linking text and apparatus: Parallel Segmentation *and* Double End-Point Attachment. *Within the RETI project, a variant of the DEPA method in stand-off mode integrated in EVT3 enables the representation of manuscript families, corrective hands, and witness hierarchies. Specific metadata classify variants by stemmatic relevance, making the* constitutio textus *process visible.*

Keywords: Digital Critical Edition; XML/TEI; DEPA; Textual Variants; Stand-off Markup

Copyright © 2026 The Author(s)
The text in this work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0

1. State of the Art[1]

The objective announced in the title of this paper is better framed as a question: Is it possible to create a digital critical edition of a classical text with XML/TEI? The answer must take into account all the elements embedded in this question. While critical editions—many of them of classical texts—have been produced since the 19th century, and various kinds of digital scholarly editions are available online (including some critical editions of mediaeval and modern texts), the feasibility of a *digital critical* edition of a *classical* text has long been the object of contention.

Greek and Roman texts belong to the category of literature transmitted through a complex process that is plural both "vertically" (spanning many centuries) and "horizontally" (through the existence of many roughly coeval manuscripts). The result is a tradition with multiple witnesses, no two of which are identical; consequently there exists no single version of "the text", but almost as many versions as there are witnesses. It is therefore the editor's task to reconstruct the version of the text (*constitutio textus*) that they consider most likely to approximate the original. This is achieved by gathering and comparing as many textual variants as possible (a process known as "collation") and evaluating them on the basis of the reliability of each witness, the plausibility of each reading, and other factors.

Whereas the task of collation allows for a certain degree of objectivity, the critical evaluation of the variants — and the resulting choice of a preferred reading — is the product of the editor's intellect and expertise. The editor makes part of their reasoning visible to the reader by reporting a selection of discarded textual variants which they judge to be significant or interesting, and by attributing certain variants only to some of the witnesses that transmit them. This selection—which has scholarly value in its own right—is made based on various criteria, as established by Karl Lachmann in the 19th century: for example, one tends not to report purely orthographical variants that distinguish between witnesses of minor importance; it is also common to omit references to witnesses that merely copy a certain reading from an older witness (a principle known as eliminatio codicum descriptorum). The selected variants are reported in the critical apparatus, which, in traditional printed editions, runs beneath the established text.

The annotation in the apparatus is often complex, as a result of the substantial and layered data produced by a complex tradition. For example, the editor of a classical text often finds it important to distinguish between different hands in a manuscript, which may offer textual variants: they might, for example, have to report in the apparatus that the second hand of a manuscript gives the preferred reading, but does so as a correction to the first hand, which gave a variant. Furthermore, variants do not always affect the same stretch of text: one string of words may have variation in one or two words in one witness, or three or four in another, and these differences might not overlap (one witness may lack some words, another may have all the same words but in a different order, and so on).

This brief overview, though possibly redundant for many, is necessary to lay the groundwork and clarify the challenges posed by editing a classical text. Such texts come loaded with substantive data, which are both numerous and diverse, and require sophisticated, multi-layered annotation to represent this diversity. Hence the question with which we began: can this complexity be represented digitally at all? If so, what digital infrastructure and tools are best

[1] Section 1 is authored by F. Lazzerini, Section 2 by M. Di Franco.

suited to the task? And, if it is possible to create an edition that is truly *critical* (with all that that entails) in digital form, can it present any added value compared to traditional critical edition?

A survey of the state of the art reveals that the digital landscape of classical texts is vast and rich, yet it offers few examples of truly digital *critical* editions.[2] The resources that are currently available mostly fall into two categories. On the one hand, there are digital corpora and archives (such as *Perseus Digital Library*,[3] *Latin Library*,[4] *Musisque deoque*,[5] *DigilibLT*,[6] and the library of the Packard Humanities Institute (PHI) [7]), which offer remarkable completeness and accessibility, but function essentially as large-scale databases rather than critical editions.[8] On the other hand, there are digital libraries provided by academic publishers (such as Loeb,[9] Oxford University Press,[10] or Cambridge Core)[11]; while valuable, these are mostly digital reproductions (or retro-digitizations) of traditional printed editions. They are not the product of genuinely digital scholarship (born digital and designed using digital tools and logic) but are simply made digital *a posteriori*.[12]

The *de facto* standard markup language for scholarship in the humanities is XML, specifically in accordance with the TEI conventions: most of these digital editions are encoded with XML,[13] and the majority adopt TEI or TEI-derived schemas.[14] Because of its versatility and its

[2] The criteria for the definition of a digital critical edition, and (or contrasted with) a digital scholarly edition, have themselves been made the object of scholarly discussion: see [6], critiqued by [9] and [10]: 150-151. In what follows, I follow many of the points established by [11].

[3] https://www.perseus.tufts.edu/hopper.

[4] http://www.m.thelatinlibrary.com/.

[5] https://www.mqdq.it.

[6] https://digiliblt.uniupo.it.

[7] https://latin.packhum.org.

[8] Musisque Deoque is (partially) an exception, providing, in some cases, an essential critical apparatus.

[9] https://www.loebclassics.com.

[10] https://www.oxfordscholarlyeditions.com.

[11] Cambridge Classical Texts and Commentaries: https://www.cambridge.org/core/series/cambridge-classical-texts-and-commentaries/51252139188DDBF618222EC72D9ED59F.

[12] This is one of the conclusions stressed by [11]: «while printed editions normally give exactly one version of a text, the deeply marked up textual code of the digital edition theoretically covers several views of the text and may lead to various presentations generated by specific algorithms. This fundamental difference in paradigm and its consequences for the reality of editions in our digital media landscape lead to the following important conclusion: A digitised edition is not a digital edition.». This fact carries another implication: «A digital edition cannot be given in print without significant loss of content and functionality.».

[13] Some exceptions are Latin Library, encoded in plain HTML, and the PHI, which employs «(some derivation of) the original format used on the PHI CD-ROM» [4].

[14] For example, Oxford Scholarly Editions Online uses an XML schema which borrows elements from TEI [3]. Projects that employ entirely different infrastructures are rare: for example, The

pervasiveness in digital humanities, it makes sense to turn primarily to XML/TEI for the purpose of creating a digital critical edition. The TEI core module already includes various elements that are useful for textual criticism (tags to encode quotations, glosses, lists, notes, bibliographic references, and others), and many additional ones are available through the critical module (manuscript description, primary source transcription, and critical apparatus). However, there are also shortcomings and limitations in TEI, which will be discussed in this section, making it a problematic—though still viable—solution to the challenge of encoding a digital critical edition.

1.1 Limitations to TEI encoding

While TEI offers full representability, in a sense, and somewhat paradoxically, it struggles to represent some of the very phenomena that define classical textual traditions; one may gain the impression that the TEI conventions for critical editions were not designed for texts with such complex traditions. To reiterate a keen assessment made by Francesco Stella now over a decade ago, TEI treats manuscripts as autonomous artefacts rather than witnesses to a reconstructed text.[15] This sub-section discusses some of the consequences that this outlook entails: namely, the issues of encoding witnesses and complex annotations in the critical apparatus.

*1.1.1 Hierarchies of witnesses (***stemma codicum***)*

The main feature required by a critical edition is the possibility of reporting variants and attributing them to the witnesses that transmit them or to the scribe who introduced them. This requires, firstly, that the witnesses be encoded in the <teiHeader> as this specific type of item and—like other items such as persons, places, and objects—assigned an @xml:id, which can be referred to throughout the TEI document.

The TEI offers a <witness> element which can be used to encode manuscripts and editions (anything that qualifies as a "witness" transmitting textual variants), and which can include a number of useful elements for describing that witness in great detail. In <msDesc> ("manuscript description") one can give details about the manuscript's name and where it is stored (<msIdentifier>) and its content (<msContents>); one can describe the manuscript as a physical object (<physDesc>, <objectDesc>) and distinguish different hands (<handDesc>); and one can include detailed information on the manuscript's history and transmission (<history>, <provenance>), also including bibliographic references. This structure can condense information that would otherwise span many pages in the preface of a traditional edition. This set of elements to describe a witness serves very well the purposes of digital editions of individual manuscripts.[16]

Homer Multitext project (https://www.homermultitext.org/) does not use markup language at all, but organizes its content through CITE/CTS architecture [2].

[15] «L'universo TEI considera il manoscritto come un veicolo di se stesso, da trattare nella sua unicità, non come testimone di un oggetto di comunicazione assai più complesso, la cui forma originale debba essere ricostruita o restaurata anche grazie a tecniche filologiche.» ([12]: 350).

[16] For example, see the editions (created for EVT, see below) of some manuscripts of the Biblioteca Nazionale Universitaria di Torino (https://bnuto.cultura.gov.it/biblioteca-digitale/manoscritti-in-evt), such as J.I.22-23, containing Pliny the Elder's Naturalis Historia, or D V 33, containing

For classical texts (which are often transmitted by dozens of witnesses) one can of course encode as many witnesses as needed, one after the other in the <teiHeader>. However, in multi-witness traditions witnesses are not all related to one another equally: rather, it is usually possible to determine which manuscripts were copied from others, or which were produced in the same *scriptorium*; based on their histories and on the number of shared readings, manuscripts are also often grouped into "families", denoted by *sigla* (usually letters of the Greek alphabet). These relationships are represented as a family tree—a *stemma codicum*. In fact, any critical edition of a classical text (when based on a so-called closed *recensio*) relies on a *stemma codicum* (either an already established one or one revised by the author of the new edition) to provide the reader with a *conspectus codicum*, which is necessary to decipher the critical apparatus.

Currently, however, the TEI guidelines do not provide a straightforward way to encode familial relationships between witnesses in the <teiHeader>. The only way to represent a hierarchy of witnesses is to exploit one of the defining features of XML (namely, its tree structure) and nest witnesses within witnesses, using the <listWit> element, which can be made to correspond to a family. This can be done potentially *ad infinitum*; however, it is not technically how <listWit> is intended to be used, and the result can become convoluted and redundant (see below, section 2.1).

Furthermore, the usefulness of grouping manuscripts into families lies in the fact that, if a certain reading is common to all the members of that family, in the critical apparatus one can attribute that reading to all of them simply by writing the corresponding *siglum*. However, if one uses <listWit> to represent families, this solution clashes with this established practice, because, according to the TEI conventions, a reading must be attributed to an item encoded as either a <witness> or a <hand>: it cannot be attributed to a <listWit>.

The only alternative, at present, is to give up encoding the *stemma codicum* as a structure, and either leave it out of the TEI document or, if at all, to encode it as a picture; in that case, the witnesses would be encoded one after the other, all within a single <listWit>.

1.1.2 Linking the apparatus to the text

In a printed edition, the critical apparatus runs beneath the text; in most editions, line numbering makes it easier to match the variants in the apparatus to the correct section of the text. Ideally, the visual output of a digital critical edition should achieve the same effect, allowing the text and the apparatus to run synchronously in the user interface. In any case, it is necessary to encode digitally the link between a specific textual segment and its apparatus notes.

The TEI envisages three methods to do so.[17] The first, "Location-Referenced", is only used to encode the digital version of a printed edition, and therefore falls outside the scope of this paper. The second, "Double End-Point Attachment", and third, "Parallel Segmentation", are viable for encoding a digital-born edition. However, while the Parallel Segmentation (PS) method is intuitive and straightforward, it cannot represent the complexity of the annotations required by a multi-witness text; as for the Double End-Point Attachment (DEPA) method, although it

Leonardo Bruni's Latin translation of Aristotle's Nicomachean Ethics (see under Project Info > File description).

[17] https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPLK.

allows complex annotation, it is quite cumbersome to implement. This constitutes a further limitation of TEI when it comes to digital critical editions.

In the PS method, the critical apparatus is encoded in-line within the text itself, breaking its flow whenever a segment has variation.[18] At first glance, it is easy to implement and effective; however, these advantages are predicated on its application to a limited amount of data. For this reason, it works well when texts have few witnesses, or at least few variants, which can be easily handled through in-line encoding.

In principle, these conditions could also be met in the edition of a classical text, provided that the editor applies principles that reduce the number of reported variants substantially (as is most often done in printed edition, for the reasons outlined above). The question, then, becomes what mission statement underlies the digital editor's endeavour—or, in other words, how the editor intends to make use of the digital medium and its affordances. Arguably, a digital edition that, despite the inherent advantage of an expandable space, offers no more data than what could be included in a printed edition does not represent a meaningful scientific advance; rather, it would be desirable for a digital edition to provide the reader with a broader view of the multilayered tradition of a text than can be conveyed in a few printed lines of critical apparatus. As stated above, there is unquestionable scholarly value in the editor's critical evaluation of variants; however, whereas in printed editions this work becomes visible mainly through the selection of variants to report or discard, in a digital format one can envisage alternative ways of making this process visible.

Therefore, if the edited text derives from a sizeable number of witnesses and variants, the authors of this paper argue that the digital edition should account for as many as possible. However, this course of action renders PS an impractical method.

Furthermore, the PS method entails a limitation that persists even in simpler textual traditions: it offers no way of handling overlapping variants. This expression describes a situation in which variants exist for different parts of the same segment of text, and often the same word appears in multiple readings of varying extension.

As an example, consider a sentence from Cicero's *Lucullus*, § 8: *Hoc autem liberiores et solutiores sumus, quod integra nobis est iudicandi potestas nec ut omnia quae praescripta et quasi imperata sint defendamus necessitate ulla cogimur*.[19] Let us take one segment from this sentence: *quod integra nobis est iudicandi potestas*. Even in such a short string, multiple words have variants in readings that cannot be superimposed:

- *nobis est* appears as *est nobis* in one variant (transposition in the witness Ambr1) and in another it is replaced with *sit* (transmitted by Harl);
- *iudicandi* appears in the variant *iudicandi est* (transposition in Gud), in *iudicandi potestas est* (a different transposition of a more extensive section of the text, in Harl3), and in *potestas iudicandi* (yet another transposition in Vat6); it is also

[18] Of course, the break in the encoded text would not necessarily affect the user interface.

[19] "And in this respect we have more freedom and autonomy: that out ability to form judgements is intact and we are not in any way forced to uphold all the views that have been prescribed—as it were, compelled—to us." Translation my own.

replaced by other single words in various witnesses (*disputandi* in Magl and Par4; *indicandi* in Matr; *iudicanda* in Pan)

In such a case, an <app> note attached to *est*, *iudicandi*, or *potestas* would end up being incomplete; the only way to have complete and consistent <app> notes would be to report extended strings even when the variant is only on one word.[20] While this problem can arise with any text, it goes without saying that the more witnesses and variants exist (the more complex and layered the tradition), the more likely such cases are to present themselves.

Conversely, these and similar cases are handled effectively through the DEPA method, which entails encoding the apparatus in a separate section of the TEI document (in stand-off mode)[21] and linking each <app> note to a segment of the text, marked off precisely by means of markup elements that function as cornerstones; in this way, each word can be included in multiple <app> notes. The TEI suggests using <anchor> elements (each provided with an @xml:id) which, placed before and after words of the text as required, can be used to encode as the beginning and end of each segment on which there is variation. So, for example, by placing <anchor> tags appropriately in that segment from the *Lucullus* (Fig. 1):

```
integra <anchor xml:id="example1"/> nobis <anchor xml:id="example2"/> est <anchor xml:id="example3"/>
iudicandi <anchor xml:id="example4"/> potestas <anchor xml:id="example5"/>
```

Fig. 1 - Example of use of <anchor>

one can encode as many <app> notes as needed and link them to their exact match in the text. Let us encode the readings within those <app> notes, using the *siglum* of each witness as their @xml:id (Fig. 2)

[20] Thus, for our example: *est nobis* Ambr1 : *sit nobis* Harl. And: *est iudicandi potestas* Gud : *iudicandi potestas est* Harl3 : *est potestas iudicandi* Vat6 : *est disputandi potestas* Magl4 : *est indicandi potestas* Matr : *est iudicanda potestas* Pan.

[21] The TEI Guidelines state that «both the location-referenced and the double end-point methods may be used with either in-line or external apparatus.» (https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html). The disadvantages of a in-line DEPA are comparable to those of the PS method.

```
<app from="#example1" to="#example3">
 <rdg wit="#Ambr1">est nobis</rdg>
</app>
<app from="#example2" to="#example3">
 <rdg wit="#Harl">sit</rdg>
</app>
<app from="#example3" to="#example4">
 <rdg wit="#Magl #Par4">disputandi</rdg>
 <rdg wit="#Matr">indicandi</rdg>
 <rdg wit="#Pan">iudicanda</rdg>
</app>
<app from="#example2" to="#example4">
 <rdg wit="#Gud">iudicandi est</rdg>
</app>
<app from="#example2" to="#example5">
 <rdg wit="#Harl3">iudicandi potestas est</rdg>
</app>
```

Fig. 2 - Example of use of link in <app>

In this way, the DEPA method solves both problems encountered with PS: it can handle a large number of variants without interrupting the flow of the edited text, and it allows complex annotation in the <app> notes.

There are two main problems with implementing the DEPA method in an edition project. Firstly, it is cumbersome to implement, as is already evident from the example above. Because it is a stand-off mode, it entails keeping track of two different sections of the TEI document (the text and the apparatus) at the same time, going continually back and forth to match each note to the precise segment marked off by the <anchor> tags; furthermore, those tags must be placed manually into the text. The other problem is that the stand-off apparatus requires a specific parser to render it into a legible form in the user interface, whatever this may be. These two issues are likely the main reason why, to our knowledge, the DEPA method has not yet been successfully applied to an edition project of a multi-witness text.

Yet, unlike in the case of PS, the problems with DEPA are not structural or inherent to the method itself, but arise from external constraints. If the encoding workflow could be streamlined to mitigate its cumbersomeness, and if dedicated tools for parsing and visualizing the stand-off apparatus were developed, there would be no reason why DEPA could not serve as a viable solution for producing a digital critical edition encoded in XML/TEI.

2. The RETI Methodology for Digital Critical Editions with EVT3

The RETI project (REndering Texts and Images) is a PRIN 2022 research project carried out by the Universities of Pisa, Turin, Molise, and Naples "Federico II".[22] It aims to provide a methodology for producing digital scholarly editions using the software EVT3 (*Edition Visualization Technology*).[23] Within the project, four case studies were developed:

1. Cicero, *Lucullus* (45 BCE);
2. Aelius Aristides, *Panegyric to Cyzicus* (166 CE);
3. Correspondence between the marquises Ludovico Gonzaga and Barbara of Brandenburg and the diplomat Bartolomeo Bonatti (1461);
4. *Cronaca* by Alessandro Streghi (1470);

The latter two are texts transmitted by a *codex unicus*, where the focus is on the transcription and visualization of the manuscript images. The other two, classical texts transmitted by multiple witnesses (the *Lucullus* by 74, the *Panegyric* by 60), were developed as digital critical editions.

RETI pursued two main goals:

- To design and test a unified methodology for creating digital editions suited to different textual situations (from single to multiple witnesses);
- To contribute to improving EVT3, extending its capabilities to meet the specific needs of critical editions based on the DEPA method in stand-off mode.

To address the main obstacles of the DEPA method outlined in the previous section, the RETI methodology operates on two fronts. First, it proposes a specific encoding strategy that partially automates and streamlines the annotation process, reducing the burden of manual tagging, as discussed below (section 0). Second, it relies on the newly extended capabilities of EVT3 to parse and visualize the resulting stand-off apparatus. The following sections detail how this combined approach handles the encoding of witnesses, hands, and variant readings and shows how this encoding is currently visualized in EVT.

2.1 Encoding Witnesses and Hands, and Representing Families and Sub-Families

In the case of the two critical editions, the text reconstructed by the editors is encoded within the <body> element, while the critical apparatus is encoded in stand-off mode within <back>. Each manuscript and previous edition is encoded as a <witness> element with its own @xml:id, and each variant reading (<rdg>) is attributed to the relevant witnesses through the @wit attribute. On EVT, the text and the apparatus are displayed on parallel windows and can be consulted synchronously; the parts of the text on which there is variation are marked by an apex and correspond to individual apparatus notes (Fig. 3).

[22] See also [1].

[23] http://evt.labcd.unipi.it/. See [7], [8].

Fig. 3 - Visualization of the first paragraph of the *Lucullus* (text and apparatus)

As discussed above, the TEI provides tools to describe manuscripts and scribal hands (<handNote> within <handDesc>). In RETI's model, hands are encoded both as <handNote> elements (to which scribal interventions such as deletions and additions can be attributed) and as independent <witness> elements (Fig. 4).

```
<witness xml:id="F" corresp="#F.hand.1">F
    <msDesc>
        <msIdentifier>...</msIdentifier>
        <msContents>... </msContents>
        <physDesc>
            <objectDesc>... </objectDesc>
            <handDesc hands="4">
                <handNote xml:id="F.hand.1">first hand before corrections</handNote>
                <handNote xml:id="F.hand.c">first hand correcting itself</handNote>
                <handNote xml:id="F.hand.2">second hand</handNote>
                <handNote xml:id="F.hand.3">later hand(s)</handNote>
            </handDesc>
        </physDesc>
        <history>...</history>
    </msDesc>
    <listWit>
        <witness xml:id="F.c" corresp="#F.hand.c"/>
        <witness xml:id="F.2" corresp="#F.hand.2"/>
        <witness xml:id="F.3" corresp="#F.hand.3"/>
    </listWit>
</witness>

<rdg type="lectio.sing" wit="#F"><subst hand="#F.hand.2">
    <del>omnes</del>
    <add place="superscript">omnis</add>
</subst>
<note resp="#FLa">emendav. F²</note></rdg>

<rdg wit="#B #F.3 #M #Laur2 #Par6 #Reg #Urb">percontanda</rdg>
```

Fig. 4 - Example of encoding of hands for the manuscript F (witness of the *Lucullus*) and their use as <handNotes> or <elements> in the critical apparatus

In EVT, scribal corrections are displayed as different stages of the text of the same witness, but hands can also figure as witnesses of variant readings (Fig. 5).

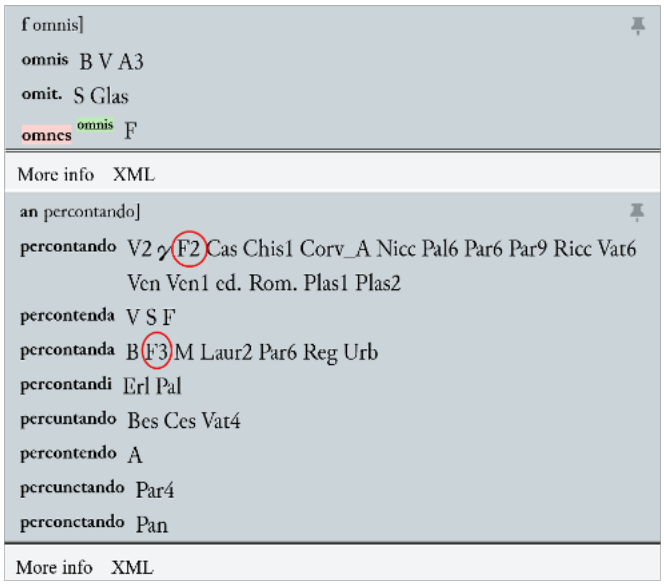

Fig. 5 - Visualization of scribal corrections (colour-coded) and hands as witnesses

The limitations of representing stemmatic relationships in TEI encoding, discussed in section 1.1.1, required the development of a syntax allowing variants to be attributed to entire families. RETI's approach was to encode each family as a <witness> containing a <listWit> of its members (Fig. 6).

```
<witness xml:id="ϕ"> <desc>Agreement between F's apographi: K, M, Bes, and the manuscripts of <ref target="#π"/>
</desc>
    <listWit>
        <witness xml:id="K"/>
        <witness xml:id="M"/>
        <witness xml:id="Bes"/>
        <witness xml:id="π"> <desc>Agreement between Vat4 and its apographi: Ambr1, Ces</desc>
            <listWit>
                <witness xml:id="Vat4"/>
                <witness xml:id="Ambr1"/>
                <witness xml:id="Ces"/>
            </listWit>
        </witness>
    </listWit>
</witness>
```

Fig. 6 - Example of nested <listWit> to represent families and sub-families of the *stemma* of the *Lucullus*

2.2 Encoding Variant Readings

The RETI approach replaces the use of <anchor/> elements (see section 1.1.2) with the wrapping of each word of the text in a <w> element, each assigned its own @xml:id (Fig. 7).

```
<text>
  <body>
    <div type="book" n="2">
      <p n="1-3">
        <seg n="1" xml:id="Luc-001">
          <w xml:id="Luc-001-1">Magnum</w>
          <w xml:id="Luc-001-2">ingenium</w>
          <w xml:id="Luc-001-3">L.</w>
          <w xml:id="Luc-001-4">Luculli</w>
          <w xml:id="Luc-001-5">magnumque</w>
          <w xml:id="Luc-001-6">optimarum</w>
          <w xml:id="Luc-001-7">artium</w>
          <w xml:id="Luc-001-8">studium</w><pc>,</pc>
```

Fig. 7 - Encoding of each word of the beginning of the text of the *Lucullus*

This methodological shift offers three advantages:

- whereas <anchor> tags must be placed manually, encoding each word in this way and assigning it an @xml:id with a progressively numbered value can be automated through scripts, significantly streamlining the editor's workflow;
- <app> notes can be linked to single or multiple words;
- the text remains suitable for further processing (e.g. lemma-based searches).

The connection is declared in the apparatus by means of the @from attribute, which identifies the first word involved, and, when multiple words are affected, optionally by the @to attribute, thus overcoming the problem of overlapping variants described above.

Moreover, each <rdg> is described by two further attributes, @type and @cause, which classify the variant according to, respectively, its stemmatic relevance and the nature or cause of the change. The TEI provides general values for these attributes, but they are not fully adequate for classical texts with multiple witnesses; therefore, RETI introduced customized classifications. @type defines as a typology that conveys the relevance of each variant for the purposes of the *constitutio textus* and in relation to the *stemma codicum.* As for @cause, we adopted the classification of causes from Reynolds and Wilson [5].

The values for @type are:

- archetypal — erroneous but widespread readings, probably deriving from the archetype;
- sub-archetypal — readings distinguishing the main branches;
- family-distinctive — readings shared within a family;
- associative — readings shared by manuscripts frequently agreeing;
- polygenetic — independently recurring readings;

- lectio.sing — individual readings;
- editorial — conjectures by ancient or modern editors.

For @cause:

- palaeographicConfusion — mistakes due to letter confusion, *scriptio continua*, or abbreviations;
- diction — orthographic or pronunciation-based mistakes;
- omission — haplography or omissions by *saut du même au même*;
- addition — dittography or interpolations;
- transposition — inversion of letters, words, or lines;
- contextInduced — changes influenced by the surrounding context;
- intentional — deliberate corrections or interpolations;
- multiple — multiple possible causes.

```
<w xml:id="Luc-011-19">Heraclitus</w>
<w xml:id="Luc-011-20">Tyrius</w><pc>,</pc>

 <app from="#Luc-011-20">
   <lem wit="#A #V #S #F #Nicc">tyrius</lem>
   <rdg type="sub-archetypal" cause="diction" wit="#B">tirius</rdg>
   <rdg type="polygenetic" cause="diction" wit="#Mon3 #Pat">thirius</rdg>
   <rdg type="lectio.sing" cause="multiple" wit="#T">tirus</rdg>
   <rdg type="lectio.sing" cause="multiple" wit="#Par4">turius</rdg>
 </app>
```

Fig. 8 - Encoding of variants for *Tyrius* in *Luc*. 11

The introduction of the @type and @cause metadata makes the editor's critical evaluation of variants visible without sacrificing comprehensiveness; furthermore, it is designed to support alternative visualizations of variant selection (with the attributes functioning as filters), giving the user control over what they want to see of the apparatus (Fig. 8).

The implementation of an interactive filtering system based on these specific attributes represents the next major milestone for the software. By embedding this deep semantic layer into the XML/TEI markup, this methodology ultimately realizes one of the key advantages that a digital critical edition should have over a traditional printed one: the ability to explore and represent the editorial reasoning in a dynamic, customizable way.

References

[1] Barbero, Chiara, Matteo Di Franco, Federica Lazzerini, and Annamaria Persia. forthcoming. "Il Progetto RETI (REndering Texts and Images): Metodologia e primi risultati". In *Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica Nell'Era Dell'Intelligenza Artificiale. Proceedings Del XIV Convegno Annuale AIUCD, Verona 11-13 Giugno 2025*.

[2] Blackwell, Christopher William, and Neel Smith. 2016. "Modeling Citable Textual Analyses for the Homer Multitext". *Data Science Journal* 15: 17. https://doi.org/10.5334/dsj-2016-017.

[3] Dunn, Mark. 2015. "Devising Data Structures for Scholarly Works". *OUPblog*. https://blog.oup.com/2015/07/scholarly-editions-data-model/.

[4] Kozák, Dániel. 2018. "Review of *Classical Latin Texts*. A Resource Prepared by The Packard Humanities Institute (PHI)". *RIDE* 8. https://doi.org/10.18716/ride.a.8.2.

[5] Reynolds, Leighton D., and Nigel G. Wilson. 2013. *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*. Oxford University Press.

[6] Robinson, Peter. 2002. "What Is a Critical Digital Edition?". *Variants: The Journal of the European Society for Textual Scholarship* 1: 43–62.

[7] Rosselli Del Turco, Roberto. 2019. "Designing an Advanced Software Tool for Digital Scholarly Editions: The Inception and Development of EVT (Edition Visualization Technology)". *Textual Cultures* 12 (2): 91–111. https://doi.org/10.14434/textual.v12i2.27690.

[8] Rosselli Del Turco, Roberto, and Chiara Di Pietro. 2019. "La Visualizzazione Di Edizioni Digitali Con EVT: Una Soluzione per Edizioni Diplomatiche e Critiche". *Ecdotica* 16: 148–73.

[9] Sahle, Patrick. 2007. "Digitales Archiv und Digitale Edition: Anmerkungen zur Begriffsklärung". In *Literatur und Literaturwissenschaft auf dem Weg zu Den Neuen Medien*, edited by Michael Stolz, Lucas Marco Gisi and Jan Loop, 64–84. Germanistik.ch.

[10] Sahle, Patrick. 2013. *Digitale Editionsformen: Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. BoD.

[11] Sahle, Patrick. 2016. "What Is a Scholarly Digital Edition?". In *Digital Scholarly Editing. Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–39. Open book publisher. https://books.openedition.org/obp/3397.

[12] Stella, Francesco. 2015. "Il Problema della codifica nelle edizioni filologiche digitali". In *Nel segno del testo. Edizioni, materiali e studi per O. Pecere*, edited by

Lucio Del Corso, Francesco De Vivo, and Antonio Stramaglia, 347–357. Edizioni Gonnelli.