Ongoing reflections on ongoing work: a discussant response to Lombardo's "Modelling"

[Reflexões em curso sobre um trabalho em curso: uma resposta crítica a "Modelling", de Lombardo**]**

Elli Bleeker*

Keywords

TEI XML encoding, Genetic digital editing, Manuscript fragmentariness, Editorial interpretation, Digital scholarly editions.

Abstract

This discussant response examines the TEI encoding methodology presented by Elena Lombardo for *The Duke of Parma* digital archive. It explores three interconnected themes: (1) the challenges, choices, and consequences of encoding complex genetic material; (2) the tension between editorial interpretation and documentary neutrality; and (3) the implications of publication platforms for long-term sustainability and data reuse. Drawing on Paul Eggert's call for "reports from the editorial trenches," it engages Lombardo's transparent account of encoding decisions as a productive site for methodological reflection. While acknowledging the project's strengths, the paper argues for simplifying the encoding model, embracing editorial interpretation as advantageous, and addressing long-term maintenance and reusability in digital scholarly editions.

Palavras-chave

Codificação TEI XML, Edição genética digital, Fragmentariedade manuscrita, Interpretação editorial, Edições digitais académicas.

Resumo

Esta resposta crítica analisa a metodologia de codificação TEI apresentada por Elena Lombardo para o arquivo digital de *The Duke of Parma*. O texto explora três eixos interligados: (1) os desafios, escolhas e consequências da codificação de material genético complexo; (2) a tensão entre interpretação editorial e neutralidade documental; e (3) as implicações das plataformas de publicação para a sustentabilidade a longo prazo e a reutilização dos dados. Retomando o apelo de Paul Eggert a "relatos da linha da frente editorial", o texto aborda a descrição transparente das decisões de codificação como um espaço produtivo de reflexão metodológica. Reconhecendo os pontos fortes do projeto, defende-se a simplificação do modelo de codificação, a valorização da interpretação editorial e a atenção à manutenção e reutilização a longo prazo nas edições digitais académicas.

* DHLab / Huygens Institute.

Elena Lombardo's paper "Modelling the unfinished and fragmentary: the digital archive of the tragedy *The Duke of Parma*" offers a methodologically rich and unusually transparent account of the encoding decisions made in producing the digital archive of one Fernando Pessoa's unfinished dramatic works. The project addresses the considerable challenge of encoding 183 manuscript leaves that are fragmentary, unfinished, and resistant to conventional editorial arrangement. Lombardo's explicit commitment to reducing inference and conjecture to a minimum, and her detailed documentation of encoding practices make this work a great case study in the practical challenges of genetic digital editing.

In his 2009 monograph *Securing the Past*, Paul Eggert argued that "what we need now, I believe, are the practical responses: the reports from the editorial trenches […] Theory is far too important to be left to the theorists: the practitioners should have their say too" (EGGERT, 2009: 90). Lombardo's paper answers this call. Her work documents not only the solutions adopted by *The Duke of Parma* (DoP) project but also the unresolved tensions and compromises that characterize any attempt to model complex textual phenomena in TEI XML. My response as a discussant aims to engage with these tensions by considering them productive sites for scholarly reflection.

My comments are organized around three themes: encoding challenges, choices, and consequences (Part 1); the relationship between encoding and interpretation (Part 2); and the implications of publication platforms for data reuse and long-term sustainability (Part 3). Throughout, I aim to open dialogue rather than to close it, recognizing that the DoP project as well as its TEI XML encoding model represent ongoing work, and that many of the issues I raise are shared challenges across the field of genetic digital editing.

1. Encoding challenges, choices, and consequences

The first section of Lombardo's paper establishes the project's core principles: documentary fidelity, minimal editorial inference, and organizational transparency. These principles shape specific encoding decisions, three of which I examine here for their broader methodological implications.

1.1. Encoding relationships: the *dramatis personae* network

The paper describes the encoding of character relationships using <particDesc> and <listRelation> elements, with relationships specified through @type and @name attributes (e.g., @type="personal" name="parent"). This approach is technically sound and follows TEI conventions for encoding prosopographical data. What struck me in examining the resulting network visualization, however, was the high degree of isolation among characters. Many figures in the *dramatis personae* list exist as nodes

without connections, suggesting either that Pessoa had not yet developed their relationships within the dramatic world, or that the fragmentary state of the manuscripts prevents us from recovering those relationships.

This raises a question about the intended functionality of relationship encoding in a fragmentary archive. If the goal is to enable future scholars to identify patterns or hypothesize about the play's development, the encoding has succeeded in making explicit what can be known. But if the goal is to provide readers with a sense of the dramatic world Pessoa was constructing, the network's sparseness may reinforce the archive's opacity rather than illuminate it. I am curious whether the project team considered alternative visualization strategies that might foreground the dramatic fragments that do exist, rather than emphasizing the gaps in our knowledge.

1.2. Meta-text and the <rs> element

The most theoretically intriguing and potentially problematic encoding decision concerns the use of the <rs> (referencing string) element to encode what the paper calls "meta-text": authorial self-commentary embedded in the writing process. This category includes structural labels, evaluative comments, procedural notes, and indicators of choice or hesitation. The paper argues that these instances share a functional characteristic: they stand for operations rather than contributing to the "fictional world" and thus warrant treatment under a unified element.

The difficulty is that the TEI defines <rs> as containing "a general purpose name or referring string". Specifically, a string that points to a named entity such as a person, place, or organization (TEI Consortium, 2026). Meta-textual commentary does not refer to external entities; it enacts authorial judgment, planning, or evaluation. The semantic mismatch between the TEI's definition and the DoP project's usage creates both conceptual and practical problems.

Conceptually, the use of <rs> for meta-text conflates two distinct textual functions: referring (pointing to something) and commenting (expressing an attitude or performing an action). Practically, this conflation makes queries less precise rather than more so. An XPath query for <rs> elements will now return both character references (legitimate named entity uses) and process commentary, requiring constant @type filtering to distinguish between them. Lombardo argues that encoding meta-text with <rs> allows to introduce "an analytical and querying perspective," but this benefit derives from the fact of encoding itself, not necessarily from the particular choice for the <rs> element. Any semantically appropriate element with a controlled @type attribute would enable the same analytical possibilities and, potentially, more cleanly.

In fact, the TEI provides several elements that may be better suited to the project's needs:

- <note type="authorial"> for evaluations, disclaimers, and procedural notes
- <label> or <head> for structural markers like “Act I” or “Scene 2”
- <stage type="authorial"> for meta-theatrical planning notes

These alternatives would preserve conceptual distinctions between types of authorial intervention while maintaining queryability. I recognize that the desire for a “single access point” has practical appeal, but this approach risks what could be called “tag abuse”: using a general-purpose element to avoid the interpretive work of distinguishing between different kinds of meta-textual phenomena.

A related issue concerns the separation of meta-textual marks from their targets. Lombardo notes that hesitation marks encoded as <rs> are kept “separate from the segment to which [they refer].” But this separation loses the explicit link between mark and target. The TEI’s @target attribute is designed for this purpose, and its use would preserve the connection while maintaining the encoding’s analytical clarity.

1.4. Meta-text vs. meta-marks: a blurred boundary

The encoding model distinguishes between “meta-text” (self-communication through language) and “meta-marks” (directive graphic signs), encoding the former with <rs> and the latter with <metamark>. This conceptual distinction is potentially useful, but its application in practice appears inconsistent. For instance, hesitation strokes in some examples are classified as meta-text and encoded with <rs>, while the hesitation line under “bear” in example 25 is encoded as <metamark>. Both are non-linguistic graphic signs with the same instructional function: marking uncertainty or reconsideration. Similarly, the word “Backward” (example 22) is classified as a meta-mark, while “good!” (example 18) is classified as meta-text, despite both being linguistic and directive.

The boundary between these categories seems to hinge on whether the function can be determined “with sufficient certainty,” but this is a matter of degree rather than a categorical difference. I would be interested to discuss the rationale for maintaining this distinction and whether they considered alternative approaches that might handle this continuum of meta-textual phenomena more flexibly.

2. Encoding and interpretation

The second major theme of Lombardo’s paper concerns the relationship between encoding and interpretation. The project articulates a principle of reducing “inferences and conjectures” to a minimum, positioning documentary fidelity as the primary editorial value. This commitment is both admirable and problematic. Admirable because it resists the temptation to over-interpret fragmentary evidence, problematic

because interpretation is both unavoidable and constitutive of any act of scholarly editing. So, should it be avoided?

2.1. The inevitability of interpretation

It can be argued that any encoding decision documented in Lombardo's paper involves interpretation, even those framed as purely descriptive. To name a few examples: assigning @subtype="approval" versus @subtype="dissatisfaction" to authorial commentary requires the editor to interpret Pessoa's affective stance. Mapping the abbreviation "M" to "Minister" (even with a documenting <note> acknowledging uncertainty) involves inferential reasoning about character identity. Distinguishing between a "textual alternative" encoded as <seg type="alt"> and a "meta-textual choice marker" encoded as <rs subtype="choice"> requires a judgment about textual function.

After Table 3, the paper introduces a category of marks whose "function cannot be determined with sufficient certainty." These are encoded with the generic <hi> element, effectively throwing up editorial hands. But this move exposes the tension at the heart of the "minimal inference" principle: if the goal is to avoid interpretation, why not encode all marks as <hi>? The fact that the project does not do so reveals that interpretation is not something to minimize but something to acknowledge, document, and defend.

The rhetoric of "minimal inference" may inadvertently mask the substantial editorial work the project is doing. Every interpretive decision in the DoP encoding is grounded in scholarly expertise, informed by close engagement with the manuscripts, and documented transparently. Rather than positioning interpretation as a problem to be avoided, the project might more productively frame it as a necessary dimension of scholarly editing: one that enables rather than compromises the edition's value.

2.2. Documentary Encoding and User Experience

The commitment to documentary neutrality has significant implications for the archive's accessibility. As Lombardo notes, the project refrains from asserting "relationships" between documents, treating each manuscript leaf as an autonomous entity. This approach is principled: it avoids imposing a false narrative coherence on materials that Pessoa left in a fragmentary state. But it also creates what I, as a lay person, experienced as slight disorientation when browsing the archive. Am I reading the play in its intended order? In archival order? In some hypothetical reconstruction?

The question is not whether the project's restraint is justified—it clearly is—but rather who can meaningfully use the archive in its current form. For scholars

already familiar with Pessoa's manuscripts and the scholarly debates surrounding their organization, the documentary approach provides a valuable starting point that may be considered neutral (or as neutral as possible, see above). But for students, general readers, or scholars encountering the material for the first time, the archive's neutrality may register as opacity.

This tension is not unique to the DoP project. It echoes the challenges faced by earlier ambitious digital archives such as HyperNietzsche, which struggled to balance archival completeness with navigational affordances (D'IORIO *et al.*, 2001). More recent projects have explored ways to provide editorial guidance without compromising documentary principles. The LdoD Archive, which Lombardo cites as a model for future development, offers diplomatic transcriptions alongside reading texts and enables users to construct "virtual editions" based on fragments in the archive (PORTELA & SILVA, 2014). Similarly, the Melville Electronic Library's presentation of *Typee* provides both a documentary transcription and a reading text enriched with what John Bryant calls a "revision narrative": an editorial construct that uses scholarly knowledge of the text's genesis to present revisions in a compelling narrative form (BRYANT, 2002).

Bryant argues that "exposing readers to a fluid text […] through the physical complexities of the change itself—requires a high degree of editorial guidance" (BRYANT, 2002: 138). A revision narrative is necessarily subjective and interpretive, but it serves a pedagogical function: it disentangles layered corrections and presents them in a hypothetical chronological order that makes the revision process legible to readers who do not have the editor's expertise.

Lombardo writes of "projections regarding future developments towards more interpretative digital editions". I am genuinely curious what forms these developments might take. Could scholarly hypotheses about document ordering be encoded as alternative sequences, such as toggleable layers that users could engage or disengage? Could the project provide guided pathways through the fragments without asserting a definitive order? These questions can be read as invitations to imagine how the archive might serve both expert and non-expert users without compromising its documentary foundations.

3. Publication platforms and the longevity of digital editions

The final theme concerns the relationship between scholarly encoding and publication platforms. Lombardo acknowledges that several encoding decisions were shaped by the requirements and limitations of EVT (Edition Visualization Technology) version 2, the beta software chosen for visualizing the archive. These include the use of the <front> element to hold metadata (because EVT cannot retrieve the <teiHeader> of individually included files), the workaround for empty <space> elements (filling them with <desc>□</desc> to satisfy EVT's processing requirements),

and possibly the flattening of meta-textual categories under <rs> for simpler processing.

3.1. How platforms shape models

These accommodations raise a broader methodological question: to what extent has the choice of a beta visualization platform shaped the scholarly encoding model itself? The <front> workaround, while computationally redundant and not fully compliant with TEI guidelines, solves a practical problem. But it also illustrates how platform affordances can constrain intellectual work. If the project were encoding for maximum TEI semantic precision – or for a different platform entirely – would some choices differ?

This is not a criticism of the DoP project, but an invitation to reflect on a challenge facing all digital scholarly editions. We often speak of how digital methods open new possibilities for textual scholarship, but we speak less often about how they foreclose others. Every platform embeds assumptions about what matters in a text, how users will interact with it, and what counts as a meaningful scholarly intervention. These assumptions are rarely neutral.

3.2. Data silos and long-term sustainability

Present-day textual scholars can no longer afford to create "data silos": bespoke datasets that require constant maintenance and generally do not outlast the project's funding cycle. The choice of EVT 2 beta, a platform that the project acknowledges "lacks some functionalities" and that was superseded by version 3 during the project's development, raises questions about long-term maintenance and data reuse.

To what extent will the choice of EVT 2 influence the stable availability and potential reuse of the project's data? This is not a hypothetical concern. Many digital editions from the early 2000s are now inaccessible or difficult to use because they were built on platforms that are no longer maintained (HOLMES & TAKEDA, 2023). The TEI encoding itself may be stable, but if it has been shaped by the constraints of a specific visualization tool, migrating to a different platform may require substantial re-encoding.

3.3. Toward interoperability? The VIDIT working group

These concerns point to the broader challenge of interoperability in digital textual scholarship. In this respect, I won't let the opportunity slide by to point to the VIDIT (Visualising and Investigating Differences in Texts) working group, an international community of scholars and developers studying textual variation. VIDIT consists of two subgroups focused on developing the best practices for text variation tools and

establishing a shared vocabulary for textual variation. The goal is not to impose a single encoding standard—scholars are understandably attached to their carefully developed models—but to identify points of convergence that might enable data exchange and comparison across projects.[1]

The DoP project's transparent documentation of its encoding decisions makes it an ideal candidate for this kind of comparative work. By articulating not only the solutions adopted but also the problems those solutions address, Lombardo's paper is more than a report from the tranches: it makes a significant and important contribution to the collective methodological knowledge of the field.

Conclusion

Elena Lombardo's paper offers a methodologically rich account of the challenges involved in modeling fragmentary, unfinished textual material in TEI. *The Duke of Parma* project's commitment to the documentary editing tradition, providing transparent documentation, and iterative refinement of the encoding guidelines represents a valuable contribution to genetic digital editing. As said in the outset of this paper, my remarks should therefore be read as a means to open dialogue about the productive tensions that characterize this work.

Three themes have structured this response. First, I have examined specific encoding decisions: the use of <rs> for meta-text and the boundary between meta-text and meta-marks or their broader implications regarding semantic precision, queryability, and interoperability. Second, I have engaged with the tension between the project's stated principle of minimal inference and the interpretive work necessarily involved in any act of scholarly encoding. Third, I have reflected on how the constraints of publication platforms shape encoding models and raised questions about long-term data sustainability and reuse.

Throughout, I have suggested that the project might benefit from three shifts in emphasis: simplifying the encoding model where possible to enhance semantic clarity and interoperability; accepting editorial interpretation as advantageous rather than problematic (since it is both unavoidable and valuable); and attending proactively to questions of long-term maintenance and data reusability.

These suggestions emerge from deep engagement with work that I find genuinely thought-provoking. The DoP project represents ongoing textual scholarship, and Lombardo's transparency about unresolved questions invites exactly the kind of collaborative methodological reflection that advances the field. I look forward to seeing how the project develops and to continuing this conversation.

[1] For more information, see the VIDIT website: https://wg-vidit.github.io/

Bibliography

BRYANT, John (2002). *The Fluid Text: A Theory of Revision and Editing for Book and Screen*. Ann Arbor: University of Michigan Press.

DEPPMAN, Jed; FERRER, Daniel; GRODEN, Michael (2004) (eds.). *Genetic Criticism: Texts and Avant*-Textes. Philadelphia: University of Pennsylvania Press.

DILLEN, Wout (2015). *Digital Scholarly Editing for the Genetic Orientation: The Making of a Genetic Edition of Samuel Beckett's Works*. Ph.D. Thesis. Antwerp: University of Antwerp.

D'IORIO, Paolo, *et al.* (2001) (eds.). *HyperNietzsche*. Paris: Association HyperNietzsche, École Normale Supérieure.

EGGERT, Paul (2009). *Securing the Past: Conservation in Art, Architecture and Literature*. Cambridge: Cambridge University Press.

HOLMES, Martin; TAKEDA, Joey (2023). "From Tamagotchis to Pet Rocks: On Learning to Love Simplicity through the Endings Principles." *DHQ: Digital Humanities Quarterly*, vol. 17, issue 1.

LOMBARDO, Elena (2026). "Modelling the unfinished and fragmentary: the digital archive of the tragedy *The Duke of Parma*."International conference *Unveiling The Duke of Parma*, February 23-24, Università di Parma.

PORTELA, Manuel; RITO SILVA, António (2014). "A Model for a Virtual LdoD." *Literary and Linguistic Computing*, vol. 29, issue 3, pp. 354-370.

TEI Consortium (2026). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.11.0. Last updated on 18th February 2026, revision 358d2e48e. https://tei-c.org/release/doc/tei-p5-doc/en/html/

VANHOUTTE, Edward (2007). "Traditional Editorial Standards and the Digital Edition." *Learned Love: Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet*. Els Stronks and Peter Boot (eds.). The Hague: Huygens ING, pp. 157-174.

VIDIT Working Group (2024). "Visualising and Investigating Differences in Texts." https://wg-vidit.github.io/

ELLI BLEEKER is a researcher at the Huygens Institute for the History of the Netherlands. As a Research Fellow in the Marie Skłodowska-Curie–funded DiXiT network (2013–2017), she received advanced training in manuscript studies, text modelling, and XML technologies. She completed her PhD at the Centre for Manuscript Genetics at the University of Antwerp (2017), focusing on the role of the scholarly editor in the digital environment. She specializes in digital scholarly editing, with a focus on modern manuscripts, genetic criticism, and semi-automated collation. She currently works as a researcher at the Huygens Institute and the DHLab of the KNAW Humanities Cluster. She is Treasurer of the European Society for Textual Scholarship, coordinator of the VIDIT working group, and a member of the Dutch Society for Textual Scholarship.

ELLI BLEEKER é investigadora no Huygens Institute for the History of the Netherlands. Como bolseira de investigação na rede DiXiT (2013–2017), financiada pelas ações Marie Skłodowska-Curie, recebeu formação avançada em estudos de manuscritos, modelação textual e tecnologias XML. Concluiu o doutoramento no Centre for Manuscript Genetics da Universidade de Antuérpia (2017), com uma tese sobre o papel do editor académico no ambiente digital. Especializou-se em edição académica digital, com enfoque em manuscritos modernos, crítica genética e colação semiautomatizada. Atualmente, é investigadora no Huygens Institute e no DHLab do KNAW Humanities Cluster. É tesoureira da European Society for Textual Scholarship, coordenadora do grupo de trabalho VIDIT e membro da Dutch Society for Textual Scholarship.