Digital Scholarly Editions as Interfaces

edited by

Roman Bleier, Martina Bürgermeister, Helmut W. Klug,

Frederike Neuber, Gerlinde Schneider

2018

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 20. November 2018.

Encodings and Visualisations of Text Processes across Document Borders

Joshua Schäuble, with Hans Walter Gabler

Abstract

This essay exemplifies first inroads into interfacing a digitally born genetic edition (instanced by Virginia Woolf, "A Sketch of the Past" [1939–41; unfinished and fragmentary]). A basic requirement for digital genetic editing is to establish the records of text variation within individual documents and to concatenate these records in serial progression to trace and represent the genetic text processes across document borders. Coordinating the mark-up of revisional layering in individual documents within one encoding system establishes a synoptic record of the cross-document progression, stratified commonly into levels correlated to the carrier documents. Guidelines for genetic mark-up have been made available in the TEI P5v2 release. Their hands-on application to our sample material has resulted in some detailed critique and suggestions for modification. From capturing and marking-up the textual progression within each individual document of a series, we proceed to securing the continuity of the comprehensive digital record, and thus the permeability of the document border, by way of automatic collation and mark-up of the genetic text movements in the document interstice. The integrated digital record of a text progression across documents thereby effected carries in its mark-up the requisite information to visualise synoptically structured diachronic (genetic) text data under multiple perspectives. The interface visualisation as such must be realised through (sets of) visualisation software. From the construction-in-progress of one such set of modules, the essay demonstrates the design and describes the operation of one modular interface, a *Diachronic Slider*.

1 Introduction

Drafted writing in a single document or in a series of documents commonly goes through several stages of development. The document texts, in aggregate, bear witness to the textual genesis of the work. But a not inconsiderable amount of textual development happens *off the page*, as it were, in the rewriting of a document as clear, fair copy. The material evidence of the process of text development can be captured and represented – be it in the analogue of print, or through digital visualisation. Digital representation results from digital modelling of the genetic process as it is

materially evidenced and editorially and critically understood. This is the domain of scholarly editing – increasingly, of digital scholarly editing. When one emphasises textual development over the establishment of a critical text, this is digital genetic editing.

One requirement for digital genetic editing is to concatenate the record of text variation from a serial progression of documents. Each transition from one document to another marks a *border* in the materially traceable textual development. At each *document border*, the evidence for composition and revision is typically mixed: partly readable (and thus transcribable) from the materially extant documents, partly inferable only through critical assessment of the results of machine collation of document texts. Reading, transcribing, collating, and collation assessment need to be carried out in a constant interplay of digital automation and human (that is, critical) intervention. This constitutes the operative core of text-critical editing in the digital environment. The editorial assessments and decisions are modelled into the digital data through mark-up. From out of the genetically structured data, in turn, the visualisations to present the editorial perspectives are generated.

In this article, we describe the groundwork for unfolding the genetic development of textual progression *across* a document border. Our endeavour has been to fashion the digital record for Virginia Woolf's biography fragment "A Sketch of the Past" into the textual foundation for interface display and in-depth critical analysis. The diverse segments of this fragmentary writing exist as an incomplete set of draft manuscripts and a complete typescript. The typescript follows directly from the manuscript in the chain of textual transmission. The editorial practice described hereafter covers the computer-aided collation of these two documentary witnesses as well as the TEI encoding we developed to capture the textual genesis. Applying mark-up proposed by the TEI Guidelines hands-on to a specific task has naturally resulted in a many-faceted critique, with suggested modifications, of the guidelines. In a last subchapter, we discuss the benefits of encoding across the document border in order to derive dynamic visualisations.

2 Document collation with TUSTEP

Before processes of textual development across the document border can be encoded and visualised, they have to be critically assessed by a collation of the successive documents. "When there are two or more extant texts of a particular work or document, whether manuscript or printed, to collate them is to make a detailed comparison and to record systematically the differences between them" (Beal 79). For this process, the environment of the TUSTEP system of text-data processing was used.[1] At its

---

[1]   A print introduction into TUSTEP Version 2016 is given by Ott and Schäkle.

core, a robust collation module runs collations over pairs of files. The operation parameters for each collation run may be fine-tuned by user specification, as for instance (generally): 'collate natural-language content only; disregard (while keeping in place) all mark-up in angle brackets'; combined with (specifically): 'collate segmentally by (user-specified) match-points'. The collation outputs the variation in lists ordered by base-file reference. These lists can, in their turn, be modified to user specification.

To implement a *genetic collation* across a document border, the first text stage (the first text layer – *see below*) of a succeeding document text is collated against the last stage (=layer) of the textual development in the document immediately preceding it. All revisions visible on that preceding document are included, identified by suitable mark-up. All revisions on the succeeding document that were added in post-fair copy writing campaigns are marked up but, for the purposes of collation, are excluded (i.e. temporarily filtered out). This ensures that earlier revision campaigns on the preceding document and later revisions on the succeeding document do not falsely get assigned to the inter-documentary level of text progression.

The TUSTEP system for text-data processing is deployed for collating the states-of-text defined by their respective carrier documents and for merging them into a representation of the genetic text progression across the document border. TUSTEP is modularly organised, and its modules stand in a cog-wheel correlation to one another. This gives the system a comprehensive capacity to deploy the results from the operation of one processing module (the collation) to achieve automatically, by means of an updating module (separate, yet correlated to the preceding collation through re-use of its machine-generated results as update instructions), a desired (and again at will user-defined) modification of data format and/or content: that is, to achieve the integrated text marked up for visualising the process of the text development.

This means more specifically: the modification-and-updating functionalities of the TUSTEP system concatenate, for Virginia Woolf's "A Sketch of the Past", the text's manuscript-defined and typescript-defined levels so as to generate its genetic continuum in a fully marked-up XML-TEI P5v3.1.0 structuring. To this end, the list of the collation results is first rebuilt in terms of that data formatting. Text from the manuscript no longer present in the typescript gets automatically tagged as *deleted* in the transition from one document to the other. Reciprocally, text making its first appearance in the typescript gets tagged as *added* between the documents. Subsequently, the list of collation results so modified is used as the said instruction set to update a copy of the collation base file into a new composite file. In this composite file, text elements from the original base file and the original comparison file, already automatically tagged as they are with regard to their respective layer and level, are now also automatically merged. The merger represents the genetic manuscript-to-typescript text continuum of "A Sketch of the Past".

3 Levels and layers: encoding textual features with genetic TEI

With the implementation of "An Encoding Model for Genetic Editions" (Burnard et al.) in TEI P5v2 (TEI Consortium, TEI P5v2 § 11), the TEI Guidelines introduced elements to encode authorial revision campaigns. Following the tradition of genetic editing, this model enables the encoding of the material traces that an author leaves on a document. The model thus implies a doubly-focussed diplomatic encoding, encompassing the entire range of patternings of inscription on the document, on the one hand, and specifically all inscription recognised and readable as text, on the other hand. So comprehensively understood, the Guidelines do not limit the usage of the proposed elements to text encoding correlated to one individual document only. We argue that this diplomatic model can be used to encode the textual development over a spread of successive documents without contradicting the Guideline's current element descriptions (TEI Consortium, P5v3.1.0). The benefit of using the TEI model for genetic editions for inter-document encodings unfolds in its potential to derive dynamic visualisations and analyses of the textual genesis regardless of the physical documents. As opposed to an approach with a separate diplomatic encoding of each document, the critically assessed collation results are embedded in the mark-up and can be accessed and visualised without having to be constituted within the runtime of the user-interface (see subchapter "Visualisation Strategies").

To describe sets or single instances of authorial changes, TEI P5v2 introduced the elements `tei:listChange` and `tei:change` as part of the `tei:creation` statement. The creation statement is used within the `tei:teiHeader` to encode "all information relating to the genesis or production of a text" (TEI Consortium, P5v3.1.0 § 11.7). Within the `tei:creation`, one or more `tei:listChange` elements can be used to compile "a number of change descriptions associated with either the creation of a source text or the revision of an encoded text" (ibid.). Consequently, `tei:listChange` does not only assemble revision campaigns within individual source documents, but also within the more abstracted source text. In our understanding, and particularly in the case of the extant "A Sketch of the Past" material, the source text spans contiguous documents of transmission. Accordingly, the change descriptions that get grouped by a `tei:listChange` need not be materially evident on only one document but can relate to (contiguously) successive documents.

The elements grouped by `tei:listChange` are themselves either `tei:listChange` elements, or `tei:change` elements. As opposed to the `tei:listChange` element, the `tei:change` element clearly refers to one source document: "`tei:change` documents a change or set of changes made during the production of a source document [...]" (TEI Consortium, P5v3.1.0 <change> ref).[2] A strict interpretation of the Guidelines

---

[2]   Citations that are taken from a TEI-element's reference-page are given by (TEI Consortium, P5v3.1.0

might argue that a revision campaign or "set of changes" performed in the author's head in the act of transmission-turned-genesis between two documents cannot be described by using the `tei:change` element, because the grouped changes are not materially evidenced. But authorial changes, whether in material evidence on a source document or introduced silently in the act of text transfer from one document to another, always first arise in the author's mind and then manifest themselves on a document. Thus, a set of authorial changes, that by critical analysis are recognisable and can be grouped as a revision campaign, is always "made during the production of a source document".

The essential difference between sets of changes within a material document and across material documents is the way the individual changes themselves materialise and thus become identifiable. While revision campaigns within a document are visible on the document as interlinear additions, strike-through deletions, meta-mark transpositions and the like, the textual changes made in the act of text transmission across documents can only be traced by a critical analysis (or *Deutung*) of the textual collation (Zeller 55). Yet, this difference does not require separate TEI elements for revision campaigns within and across document borders, but rather calls for a consistent system of attributes to qualify the material evidence and the critical (i.e. editorial) certainty.

To categorise revision campaigns within documents and across document borders in terms of our project we rely on the distinction between textual *layers* and *levels*. Layers describe intra-document text stages: revisions made to a single document.[3] In contrast to intra-document *layers*, *levels* designate distinct stages of text in, as well as between, one document and another. The application of this classification according to our encoding guidelines can be seen in the following TEI example for the fifth text segment of "A Sketch of the Past":

`<elementname>` ref).

[3] Within the range of our experience, the term *layer* is based on the HyperNietzsche Markup Language (HNML), an SGML-based standard that provided an @lay attribute to describe the intra-documentary textual development in stages (Saller; Zapf). The 1984 edition of *Ulysses – A Critical and Synoptic Edition* (ed. Hans Walter Gabler) in its turn deployed a proprietary (pre-SGML) mark-up system to synoptically encode both the changes on various documents of transmission and the intra-document over*lays* on each individual document.

```
<creation xml:id="creationInfo">
  <listChange ordered="true">
      <change n="1" status="draft" xml:id="ms_stage_0" type="level">
        <label>MS 1</label>
        <desc>Manuscript: Currente Calamo Additions and Deletions</desc>
      </change>
      <change n="2" status="draft" xml:id="ms_stage_1" type="layer">
        <label>MS 2</label>
        <desc>Manuscript: Overlay Additions and Deletions</desc>
      </change>
      <change n="3" status="draft"
                          xml:id="ms_to_ts_transcription" type="level">
        <label>MS to TS</label>
        <desc>Typescript: Deletions, Additions and Substitutions that entered
                  in the transcription process</desc> </change>
      <change n="4" status="draft" xml:id="ts_stage_1" type="layer">
        <label>TS 1</label>
        <desc>Typescript: Overlay Additions and Deletions</desc>
      </change>
  </listChange>
</creation>
```

We identify four stages of textual development and encode each one as a `tei:change` element. These stages are chronologically ordered and listed within a `tei:listChange`. Each individual textual alteration identified on and between the two extant witnesses is assigned to one of these stages.

The first stage is classified as a level, since it marks a new (the first) document stage. According to the editorial precept of strictly referring every written word on the document to a genetic text stage, the first textual draft inscribed on the manuscript surface (every word that is not linked to any other stage) is assigned to this level. This level also includes authorial alterations materially evident as *currente calamo* additions and deletions. These are changes made directly in the act of first writing and not as later revision changes of the given document text: "Such features usually characterize writing produced by an author in the throes of composition, with corrections or revisions made immediately rather than later" (Beal 104). The level is thus defined as the constitution of a new version of the text on a new document. Hence, the definition covers both first inscription including visible *currente calamo* corrections on any one given document, on the one hand, and the materially fresh inscription by (authorial) text transfer from one document to its successor, on the other hand, including the non-materialised mental text alterations in the process of text transfer from one document to the next, identifiable only by comparison through collation of the successive document texts.

The layer, as said, is strictly intra-documentary. The second `tei:change` element in the given example is categorised as a layer. It contains all alterations made by Virginia Woolf when she worked over the initial draft in a later revision campaign. Each textual change that is linked to this layer is an alteration visible on the source document but not identifiable as a *currente calamo* inscription. On the manuscript, distinguishing

between changes made directly in the first composition process and changes made in the same ink in the course of a later intra-documentary revision presents to the genetic editor a challenge of combined document-analytical and text-critical discernment.

The core of this article centres on a more complex instance of *level* encoding: the third `tei:change` element in the given example. It describes the textual development between the last stage of the manuscript and the first manifestation of the text (altered) on a new document, the typescript. Distinguishing layers and levels of text in the proposed mark-up system permits eliciting in *virtual materiality* the author's decisions to alter the text in her acts of transcription. Each textual difference between the manuscript and the typescript is encoded *as if* it was physically highlighted on the typescript. Thus, the text as physically witnessed by the typescript can be correlated seamlessly to the last text stage of the manuscript. Collating the two text stages yields the textual differences between them as a textual extract in two layers. This provides the virtually material substratum to answering the critical question "How does the author transform the text in her mind?".

To ensure a consistent encoding of the document transmission, it is necessary to develop rules for encoding that clearly describe both – the identifiable types of inter-documentary changes and how instances of each type are to be encoded. To explain this problematic ambiguity, we give a first example: When transcribing her own manuscript into a typescript (which she did herself), Virginia Woolf changed the phrase "my mothers [*sic*] name" (Woolf MS.A.5.b,3) to "my mothers [*sic*] laughing nickname" (Woolf TS.A.5.a,54). The change can be detected with the aid of collation software, but it is for the editor to interpret it and provide it with an adequate TEI encoding. It could be encoded as a single substitution of the word "name" with "laughing nickname" or as an addition of the word "laughing" followed by a substitution of "name" with "nickname". If we tokenise on a finer level of granularity than the word, it could even be encoded as a single addition of the string "laughing nick" that builds a new compound with the following invariant string "name". All these solutions, when applied, produce the correct text of the typescript, yet the critical editor must decide for herself which encoding best models the writing act as text revision and ensure that whatever representation is decided for this instance is applied consistently to all changes of the same type. The level of granularity and the applicability of given encodings to individual alterations might be questioned, but not the method as such of digitally generating an accurate integrated transcript of the text from a preceding document and the text as witnessed on its successor document.

To capture the editor's certainty about her mark-up decisions, the guidelines provide the element `tei:certainty` with the attribute `@locus="name"`. With the given value, the attribute "indicates [...] whether the correct element or attribute name has been used" (TEI Consortium, P5v3.1.0 <certainty> ref). Due to the wide range of possible encodings for individual changes, such a certainty indication cannot be given for an

entire revision campaign. Instead, it can be encoded for types of alterations regardless of the layer or level they appear on, as well as for single instances of authorial alterations.

A similar problem arises with the encoding of the material evidence of a change. Per our differentiation, both *currente calamo* corrections that are materially evident on one document and changes that happen between the documents are comprised within *levels*. Therefore, the material evidence is not the same for all changes that are grouped under a level. Yet, both the *currente calamo* corrections and materially undocumented textual transformation in the author's head happen immediately in the act of inscription/typing and therefore rightly belong to the respective level of textual development. In other words, there is no predetermined correlation between a *level* alteration and its material evidence. In contrast, for changes that are classified as belonging to a *layer* by `tei:change` with the attribute `@type="layer"`, the encoding of the material evidence is redundant, since they are – by the definition of the *layer* – all materially evident on the respective source document.

4 Classification and certainty of inter-documentary alterations

The types of textual alteration that occur in the process of transmission from one document to another do not significantly differ from those that can be identified from intra-documentary revision campaigns. What differs is the degree of certainty about which type to choose for a given alteration. In the case of "A Sketch of the Past", the number of individual instances to be subsumed under a given type is significantly higher than on any other level or layer. In the act of typing her own manuscript, Woolf breaks away from the physical limits of the paper. In her head, she transforms the text more strikingly than it would be possible on the document. Nevertheless, she never completely loses the connection to the previous document. She carefully detaches text unit after text unit from her own manuscript, reworks the text in her mind and secures the result by typing it onto a new document. Revising such units, she omits words that she considers unsuitable, substitutes entire segments with alternative wordings and adds new phrases that enrich and further develop the previous version. Yet in no case does she completely discard the well-considered narrative of the earlier manuscript; her changes are embellishments and accretions rather than major departures.

Consequently, corresponding text units can be identified on both the typescript and the manuscript. By giving the starting points of these corresponding units on both documents, the TUSTEP collation algorithm can be manually realigned. This critical interaction with the collation software guarantees that an encoding of the textual development can always be achieved on a finer level of granularity than by way of

separate document transcripts. Even in the highly unlikely case that the collation algorithm cannot identify any further similarities or differences between two aligned text units, the corresponding text spans can still be encoded as single wide-ranging substitutions or variant readings. What this means is that a widely variant text unit will not confound collation of subsequent units. For the collation of the two extant "A Sketch of the Past" witnesses, this scenario of failing collation alignment does not occur. Instead, the collation across the document border succeeded in detecting more omissions, additions and substitutions in the text's genetic progression than at any other stage of the textual development. In other words, the act of fair copying is as creative and integral a phase of composition as further elaborating and revising that clean copy.

Although all inter-documentary alterations are verifiable by a comparison of the texts from the two consecutive witnesses, the moment of entry might be contested. One might dispute the immediate contiguity of the two extant documents by arguing for a missing intermediary document. Also, one might argue that in any case the two documents should not be considered in a relationship of textual development but should rather be seen as variant readings related collaterally. This perspective cannot be rejected out of hand. After critical assessment of the textual differences, however, there can be no doubt that the typescript text is a more advanced version of the text in the manuscript. The typescript text was produced in a linear workflow from the manuscript to the typescript and not as a collateral recasting of the text evidenced in the manuscript. With this critical insight, an encoding of every alteration as a collateral reading of equal validity, though theoretically possible, would give little insight into the textual development and the creative process taking place in the author's head. At the same time, however, it is also true that Virginia Woolf abandoned her project of "A Sketch of the Past" before arriving at a final version and approving it for publication. By reconstructing and visualising simply the genetic progress of her writing from manuscript to typescript, we make no claims for authorial preference nor do we imply a final authorial decision.

4.1 Additions

Together with the deletion, the addition is the simplest case of a textual alteration in the process of rewriting. Figure 1 gives an example of an inter-documentary addition between the manuscript and the typescript. The figure contrasts the manuscript and the typescript. Both image fragments show the first paragraph of the fifth text segment, in which Woolf sketches her memories of her half-sister Stella Duckworth, who died in 1897 at the age of 28 (McCracken 72–73).

In this, the first paragraph of the "Sketch", Woolf describes *the act of remembering* as a frame narrative embedding the actual memory. For the author, the act of
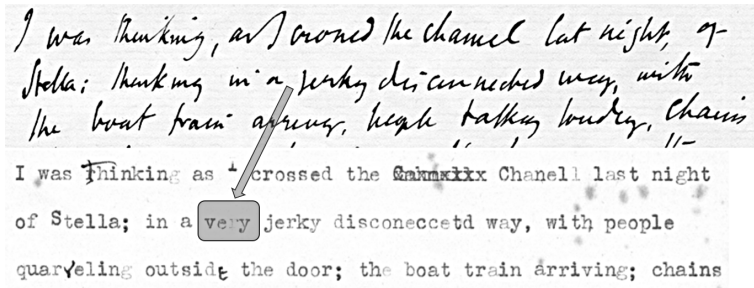
Figure 1: An addition in the text transmission. Top: MS.A.5.b,1 and bottom: TS.A.5.a,52.

remembering itself embodies a valuable and vivid memory that connects the past to the present. Neither the moment in time nor the circumstances are arbitrary as both the manuscript and the typescript are dated "June 20th 1939", although the latter might have been produced a couple of days later. The frame narrative is not replaced during the revision – which bears yet another act of remembering – but instead it is emphasised by textual alterations such as the simple addition of the word "very". The TEI we use for such instances looks as follows:

```
[...]in a
  <add change="#ms_to_ts_transcription" evidence="collation">
        very
  </add>
jerky[...]
```

A `tei:add` element is linked to the corresponding inter-documentary level (`tei:change`) with the attribute `xml:id="ms_to_ts_transcription"`, which is described in the `tei:creation` statement in the header. To distinguish between instant changes evident on the paper (like the instant correction of the still mistyped word "Chanell" in the first typescript line of figure 1) and those silently introduced *between the documents*, the attribute @evidence is used with one of the two values "collation" or "document". We thus extend the value options for the attribute rather than adopting the values "internal" / "external" as proposed by the guidelines. These, by implication, already support editorial decisions towards establishing critical texts. The values "collation" / "document", by contrast, merely record the attribute @evidence for the textual genesis without prejudicing subsequent critical editorial evaluation.

Further attributes, like the attribute class att.global.responsibility have been considered, which allow the editor to indicate "the agent responsible for some aspect of the text, the mark-up or something asserted by the mark-up, and the degree of certainty associated with it" (TEI Consortium, P5v3.1.0 att.global.responsibility). For

all inter-documentary changes with the attribute `@evidence="collation"`, this responsibility is shared between the author (Woolf, who introduced the addition) and the critical editor (Gabler, who elicited it with the aid of the collation tool). We thus consider the `@evidence="collation"` attribute to imply an indication of responsibility and do not encode *responsibility* explicitly for every single instance.

4.2 Deletions (omissions)

Deletions are encoded in analogy to additions. In the text segment displayed in figure 1, Woolf changed the phrase "thinking in a jerky disconnected way" (manuscript) to "in a very jerky disconeccetd [*sic*] way" (typescript). She not only adds the word "very" but also omits the word "thinking" at the beginning of the phrase. We encode this omission as follows:

```
[...]Stella;
  <del type="omission" change="#ms_to_ts_transcription" evidence="collation">
      thinking
  </del>
  in a[...]
```

While the encoding is self-explanatory, our decision to use the `tei:del` element for deletions between two documents may be questioned. The underlying problem is the definitional difference between a deletion and an omission. For deletions, the guidelines note:

> There is a clear distinction in the TEI between del [...] on the one hand and gap [...] on the other. del indicates a deletion present in the source being transcribed, which states the author's or a later scribe's intent to cancel or remove text. [...] gap or unclear, by contrast, signal an editor's or encoder's decision to omit something or their inability to read the source text. (TEI Consortium, P5v3.1.0 `<del>` ref)

This guideline demands evidence on the source document for the `tei:del` element. The authorial decision to omit text might therefore perhaps better be encoded with the `tei:gap` element. Yet, `tei:gap` is intended to encode illegible text or empty spaces (placeholders) within a document as "an editor's or encoder's decision" (ibid.). Consequently, the `tei:gap` element does not allow text content. From a document-centred perspective it is an oxymoron to encode the textual content of a gap or omission: text cannot be legible and omitted at the same time. Across the document border, however, this is perfectly possible. Neither `tei:del` nor `tei:gap` were initially intended to encode the textual development across a document border, but if one chooses to encode in this fashion, `tei:del` is the better fit. To us, a fall back on the generic `tei:mod` element or even a self-defined omission-element outside the TEI namespace does not seem necessary. The reason for the guidelines' "clear distinction" is to ensure
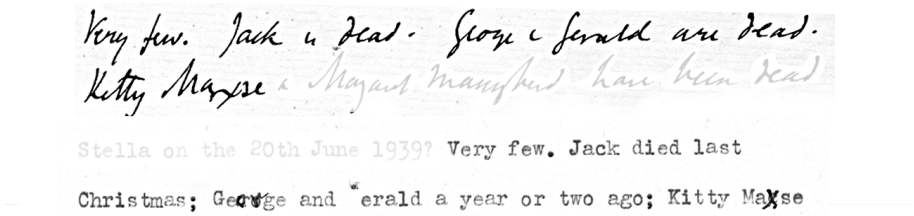
Figure 2: Substitutions across the document border. Top: MS.A.5.b,1 and bottom: TS.A.5.a,52.

the differentiation between intra-documentary phenomena. The possibility of an inter-documentary deletion was simply not considered.

4.3 Substitutions

Within a document – so on a layer or intra-document text stage – substitutions are visible as the cancellation of one phrase and its replacement by another. Both components, the cancelled phrase and its substitute, are materially evidenced on the same document, be it in the form of overwriting, an interlinear addition, a note in the margin or similar. Substitutions across the document border are, by contrast, highlighted on neither document. Only an alignment and comparison of variant and invariant textual units can render them visible.

Figure 2 compares the two witnesses for another fragment of the fifth text segment. Woolf is still reflecting on the circumstances that caused her to remember her half-sister Stella. She answers the rhetorical question "How many people are there still able to think about Stella on the 20th June 1939?" (typescript), by listing friends who, like Stella, are already dead. The manuscript reads "Jack is dead. George and Gerald are dead.", which on the typescript reads "Jack died last Christmas; George and Gerald a year or two ago;". Although the two versions are very similar, the typescript represents a slightly more detailed textual state. The motivation for the textual changes between the manuscript and the typescript is the same as for the addition in figure 1. In the act of transcription, Woolf enriches the text with further information. She distinctly substitutes two phrases to date the deaths of Jack and of her two half-brothers, which were before only listed in the manuscript.

To understand how substitutions are detected by a collation algorithm, an alignment of the variant and invariant text units of the manuscript and the typescript is helpful. Table 1 shows the substitutions as an alternation of invariant and variant text segments. The inter-documentary substitution is the combination of an inter-documentary omission and an addition, framed by invariant text. The surrounding invariant text (identical in both witnesses) defines the borders of the substituted text.

Invariant	Variant	Invariant	Variant	Invariant
MS: Very few. Jack	is dead.	George and Gerald	are dead.	Kitty [...]
TS: Very few. Jack	died last Christmas;	George and Gerald	a year or two ago;	Kitty [...]

Table 1: Inter-documentary substitutions as alternations between invariant and variant text.

A conservative TEI encoding would capture these instances as different readings (tei:rdg) of a textual variation within an apparatus structure. Since the emphasis here is not on the variation, but on the textual development, this was not an option. Instead, two inter-documentary substitutions are encoded. The phrase "is dead." is substituted by "died last Christmas;" and the phrase "are dead." is substituted by "a year or two ago;". In TEI this is encoded as follows:

```
Very few. Jack
  <subst change="#ms_to_ts_transcription" evidence="collation">
       <del>is dead. </del>
       <add>died last Christmas; </add>
  </subst>
  George and Gerald
  <subst change="#ms_to_ts_transcription" evidence="collation">
       <del>are dead. </del>
       <add>a year or two ago; </add>
  </subst>
       Kitty [...]
```

The level-reference has to be captured only for the tei:subst element. The tei:del and tei:add elements inherit this information, if no deviating revision campaign is assigned. If the added text gets modified again at a later text stage, this can be nested within the tei:add as encountered on the typescript. To differentiate between a substitution that is physically verifiable on the document and an inter-documentary substitution, the @evidence attribute is used again. If the attribute is set to "collation", the substitution is not visible on the witnesses and if it is set to "document", it is a *currente calamo* substitution that is visible on the typescript (the target document of a level definition). The latter case is usually qualified by further attributes such as @rendition="overwritten", to capture the type of documentary evidence.

To produce a transcript of a selected text stage, the processing software for each substitution iterates chronologically over all encoded text stages. For substitutions that happened before the selected text stage, the deleted text is hidden and the added text is shown, and for substitutions that happened on or after the selected text stage, the deleted text is shown and the added text is hidden. The @change attribute acts as a switch between the nested tei:del and tei:add. It controls which text is considered before and after the assigned revision campaign.
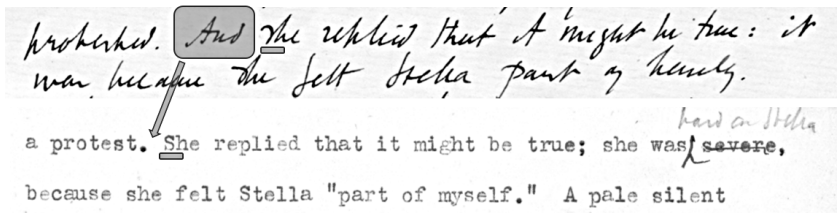
Figure 3: Omissions at the beginning of a sentence are encoded as substitutions.

	Variant	Invariant		Variant	Variant	Invariant
	Invariant					
MS:	. And she	replied that it might be true	:		it was	because [...]
TS:	. She	replied that it might be true	;		she was severe,	because [...]

Table 2: Collation alignment of a deletion at the beginning of a sentence.

For the collation algorithm, we tokenise and compare the textual transcripts at the word level. The comparison is case-sensitive. Inter-documentary additions and deletions at the beginning of a sentence often require changing the capitalisation of the succeeding word. Consequently, they are detected and encoded as substitutions. Figure 3 shows an instance of a deletion across the document border that affects the capitalisation of the succeeding word. With the omission of the capitalised "And" in the typescript, the succeeding "she" becomes the first word of the sentence and must be capitalised.

Again, a tabular collation alignment helps to comprehend the computational assessment of the textual development between the two documents (see Table 2). Since the TUSTEP collation algorithm is set to compare textual units at the word level and to distinguish capitalised from uncapitalised words, the words "she" and "She" are not assigned to the invariant succeeding phrase "replied that it might be true", but to the variant segment before. Therefore "And she" gets substituted by "She". By manual editorial intervention, such instances could alternatively be encoded with an inter-documentary deletion (here of the word "And") followed by a substitution of the following word with its capitalised variant (here "she" with "She"). We opted for the more straight-forward encoding of a single substitution, which does not demand manual intervention.

A vigilant reader of Table 2 might detect inconsistencies in the second and third variant column. First – and this is an automatable decision – the preceding variant punctuation was separated and encoded as an individual substitution. This is simply because it does not semantically belong to the following unit. The more interesting

intervention is in the third variant column. Here the aligned text fragments "it was" and "she was severe," both contain the invariant word "was" and should therefore be split into three columns. The string "it" would then be substituted by the string "she", the word "was" would be invariant in both document texts, and "severe" followed by a comma would become an inter-documentary addition. Here the table does not reflect the TUSTEP collation result, but a manual editorial intervention. Although the verb "was" is invariant, in the typescript it refers to a different grammatical subject. The authorial writing act was clearly aimed at replacing the whole clause as one cohesive alteration. This instance is a good example of the necessity of critically assessing the collation results. This assessment will lead to either affirmation of the collation output or manual editorial intervention.

4.4 Transpositions

Although transpositions are relatively common in authorial source documents, for a long time there was no applicable encoding model to represent them computationally. A transposition on a source document consists of one or more segments within the text that are marked by the author – whether with arrows, numbers, or any other *metamark* – to indicate that they should be reordered or moved to another position in the act of textual transfer. A model to encode transpositions must be able to represent the text before the transposition is applied, the text segments to be transposed, the metamarks that indicate the transposition and which segment they relate to and, finally, the target of each segment or *where in the text the segment is meant to move*.

With the integration of a model for genetic editing (Burnard et al.) into the P5v2 release of the guidelines, the Text Encoding Initiative provided elements that fulfilled the requirements for handling transpositions. To flag the text spans that are to be transposed, the guidelines provide the generic `tei:seg` element that can be qualified with the attribute `@function="transposition"`. To be recognisable, each segment gets assigned a unique identifier (`@xml:id`). The author's metamarks (arrows, numbers, lines) that indicate the transposition on the document can be encoded with the element `tei:metamark`. A metamark "contains or describes any kind of graphic or written signal within a document […] to determine how it should be read rather than forming part of the actual content of the document" (TEI Consortium, P5v3 `<metamark>` ref). These metamarks are linked to the `@xml:id` of the text segments they are meant to reorder. To describe how the flagged text segments are reordered, the guidelines provide two elements: `tei:listTranspose`, which "supplies a list of transpositions, each of which is indicated at some point in a document typically by means of metamarks" (TEI Consortium P5v3 § 11.3.4.5), and `tei:transpose`, which "describes a single textual transposition as an ordered list of at least two pointers specifying the order in which the elements indicated should be re-combined" (ibid.).
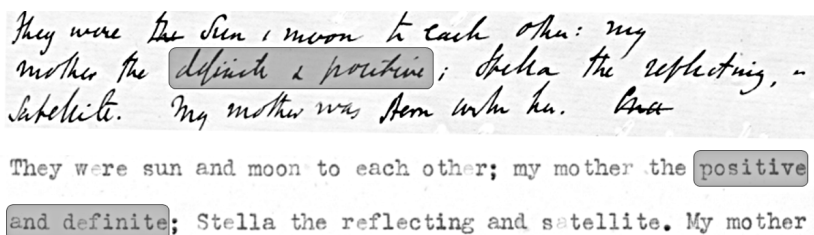
Figure 4: An inter-documentary transposition Top: MS.A.5.b,2 and bottom: TS.A.5.a,53.

These encoding guidelines as part of the model for genetic editing are clearly focused on diplomatic encoding. They answer to transpositions visibly indicated on the document, "typically by means of metamarks" (ibid.). Yet, the `tei:metamark` is not necessary to give an alternative order to addressable text segments within a `tei:transpose` element. If there are no metamarks to be encoded, e.g. because the transposition was not indicated but silently executed in the process of transcribing text from one document to another, the transposition can still be captured.

Figure 4 shows an example of a simple inter-documentary transposition. Here, Woolf describes the opposing characters of her mother and her half-sister Stella. Each one is characterised with two contrastive adjectives: "They were Sun and moon to each other: my mother the definite and positive; Stella the reflecting, & satellite." (MS.A.5.b,2). The grammatical structure of the sentence emphasises the celestial metaphor. The sun is aligned with her mother and the adjectives "definite and positive", the moon with her sister and the adjectives "reflecting, & satellite". When typing her own manuscript, Woolf realises that this clear alignment conflicts with the word order: the sentence is structured in parallelisms and yet the adjectives build a chiasmus. To correct this, she must change the order of one pair – only then "positive" opposes "reflecting" and "definite" opposes "satellite". She realises the mistake and without indicating it on the manuscript, rearranges the word order in the process of typing.

Simple movements of text can be detected by comparing additions and deletions across the document border. If a deletion and an addition, both occurring on the same level and within a reasonable textual distance from one another, contain the same text, they likely represent an instance of transposed text units – a textual shift. More complex transpositions can be detected by comparing substitutions. If two or more substitutions contain the same text, but reversed (i.e. the deleted text of one substitution is the added text of another substitution and vice versa), they are likely to be a single transposition of two or more text units – a rearrangement. The latter case can be observed when looking at a collation table for the previous example (see

Invariant	Variant	Invariant	Variant	Invariant
MS: [...]: my mother the	definite	and	positive	; Stella the reflecting and satellite [...]
TS: [...]: my mother the	positive	and	definite	; Stella the reflecting and satellite [...]

Table 3: Collation alignment of a transposition.

Table 3). Normally, we encode such alternating occurrences of invariant and variant text as a substitution: "definite" is substituted by "positive" and "positive" is substituted by "definite". Since the deletion within the first substitution is identical to the addition within the second, and vice versa, this is a strong indicator of a transposition.

We encode inter-documentary transpositions identically to the guidelines' proposed encoding for intra-document transpositions, with only the one difference that there are no metamarks to be encoded for, or referred to, the transposed text units. A TEI encoding for the instance of Table 3 looks as follows:

```
<seg xml:id="transp_sg5_3" function="transposition" n="1">definite</seg>
and
<seg xml:id="transp_sg5_4" function="transposition" n="2">positive</seg>
<listTranspose>
      <!-- previous transpositions... -->
      <transpose change="#ms_to_ts_transcription">
        <ptr target="#transp_sg5_4"/>
        <ptr target="#transp_sg5_3"/>
      </transpose>
      <!-- following transpositions... -->
</listTranspose>
```

The two transposed words are encoded as `tei:seg` elements with the specifying attribute `@function="transposition"`. In a `tei:listTranspose` element, the further specifications of all transpositions are listed for the complete text segment. Each transposition is here encoded with a `tei:transpose` element. The `@change` attribute on this element defines the level on which the transposition is executed. Nested within the `tei:transpose` elements are so-called pointers determining the order of the text segments after transposition. Here, the pointers can be read as follows: move the segment with the `xml:id="transp_sg5_4"` to the position of the first element of the current set. Then move the element with the `xml:id="transp_sg5_3"` to the position of the second element of the current set. The order of the `tei:ptr` elements specifies to which position of the previous order each element is transposed. By this proposed solution, the guidelines should successfully direct encoding related transpositions of multiple (more than two) elements.

This mode of encoding however involves a small technical difficulty. The TEI Guidelines never considered using this encoding to derive the text after transposition. Their only aim was to capture the indication of a transposition as it is visible on a

document: "By definition the result of a transposition is not present in the document, and should not therefore be encoded, if the intention is to represent the actual appearance of the document" (TEI Consortium, P5v3 § 11.3.4.5). If one nevertheless wants to design software to generate the transposition according to a new (revised) ordering given in the tei:transpose element, the difficulty met is that the tei:transpose element *only* contains information about the new order of the participating text segments, but not about the order witnessed by the document (*old order*). There is no explicit information given to tell the software which one of the participating segments is the first on the document. While a human reader might say: "it is simply the one that appears first in the text", software does not retrieve this information without analysing the text. What sounds complicated, is very easily solved, simply by ensuring that each segment of a transposition set accessed by the software knows its own relative position within the set. To achieve this, we have added the @n attribute to each tei:seg segment[4] (see encoding sample above).

4.5 Mixed Types

For "A Sketch of the Past", most instances of transposition across the manuscript-to-typescript border are not as neat as the previous example. In any case, they must be confirmed by a critical editor and in the majority of cases they are detected by comparative scrutiny of the material documents in the first place. This is because Woolf not only transposes units, but at the same time also revises parts of the text substituted in the transposition. Figure 5 shows an instance of such a mixed-type transposition:

This final example brings us back to the first paragraph of the fifth segment, which was already used to explain additions (see figure 1). All the inter-documentary changes in this paragraph are motivated by the same idea. Woolf enriches the frame narrative about the distracting circumstances that made her remember her half-sister Stella. When typing her own manuscript, she wants to emphasise her "very [addition!] jerky disconnected way" of thinking. To do so she substitutes the phrase "people talking loudly," by "people quar_eling [*sic*] outside the door;" (with the word "people" left invariant). *Quarrelling* is a stronger, more distracting action than *talking loudly*. It implies tension and a conflict and is therefore a more distinct irritant to induce her disconnected, conflicting memories than just loud talking. For Woolf, this substitution is not quite sufficient to express the close link between her disconnected thinking and the surrounding noise. She also expresses this closeness in the word order, transposing "the boat train arriving" by "people quar_eling outside the door". The latter is now

---

4   An alternative solution to locate each pointer's previous position is to parse over all pointers, look up the absolute position of the respective segment in the text – e.g. by using the XSLT position() function – and then to sort the results. Yet this solution seems needlessly complicated.
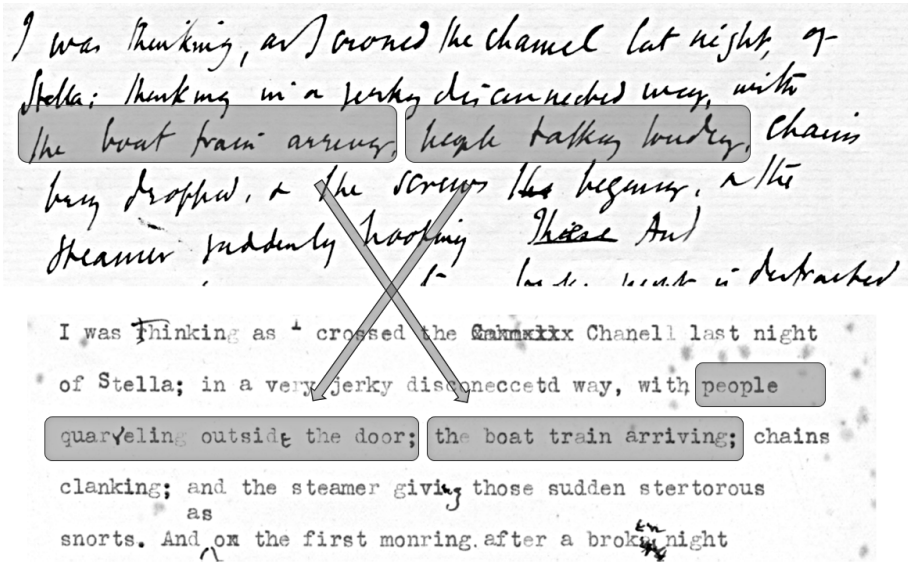
Figure 5: Mixed Types: A transposition of substituted text revised.

in direct succession to "thinking [...] in a very jerky disconnected way" and the first phrase connects better to the following description of chain noises, which aurally probably accompany "the boat train arriving".

We encode this as a transposed substitution. Effectively this encoding establishes regular inter-documentary substitutions as described above, nested within text segments that are part of a transposition. Note that both transposed text segments contain a substitution, because the punctuation was changed from commas to semicolons:[5]

```
<seg xml:id="transp_sg5_1" function="transposition" n="1">
     the boat train arriving
     <subst ...><del>, </del><add>; </add></subst>
</seg>
<seg xml:id="transp_sg5_2" function="transposition" n="2">
     people
     <subst change="#ms_to_ts_transcription" evidence="collation">
       <del>talking loudly, </del>
       <add>quarreling outside the door; </add>
     </subst>
</seg>
```

---

5   To keep the code example more legible, the level attributes of the punctuation substitution were omitted. Also, the handwritten addition of the letter "r" within the word "quar_eling" on the transcript is emended to simplify the example for the article.

```
<listTranspose>
  <transpose change="#ms_to_ts_transcription">
    <ptr target="#transp_sg5_2"/>
    <ptr target="#transp_sg5_1"/>
  </transpose>
</listTranspose>
```

Computationally, such instances are difficult to detect. Table 4 shows one possible result of a computational collation. A good algorithm detects the biggest invariant string, which is the phrase "the boat train arriving". The transposed "people quar_eling outside the door;" therefore has no corresponding text on the manuscript and is falsely encoded as an addition between the documents. The last part ", people talking loudly," aligns as variant text with a semicolon before the next invariant text ("chains", see figure 5).

	Addition	Invariant	Variant → Substitution
MS:		the boat train arriving	, people talking loudly,
TS:	people quar_eling outside the door;	the boat train arriving	;

Table 4: Collation alignment of a mixed type.

If we accepted this collation result, the instance would be encoded as an addition, followed by invariant text, followed by a substitution. Yet this by no means captures the author's assumed motivation for the set of text changes. The instance can so far only be detected as a transposition by critical scrutiny of the collation.

5 Visualisation strategies: an eXist-db based user interface

The main benefit of an inter-documentary encoding is its potential as seed-bed for dynamic visualisation. With this encoding, all the critically assessed collation results are present in a single TEI file and can be dynamically accessed for analysis and visualisation within a single user interface. It is not necessary to implement further collation algorithms at the backend to derive collation views of successive documents. At the same time, both a diplomatic and a textual transcript of every single text-stage can be dynamically reconstructed within the runtime of the user interface.

To provide a user interface for a digital edition of the encoded "A Sketch of the Past" material, we are developing a modularised eXist-db application. Figure 6 schematically visualises the backend design of this environment. Synchronous and asynchronous HTTP-requests are sent from the user to the eXist-db controller on the server. A sample request in natural language could be: "show me a textual (or diplomatic) visualisation of the first typescript layer of segment five". The XQuery-based controller
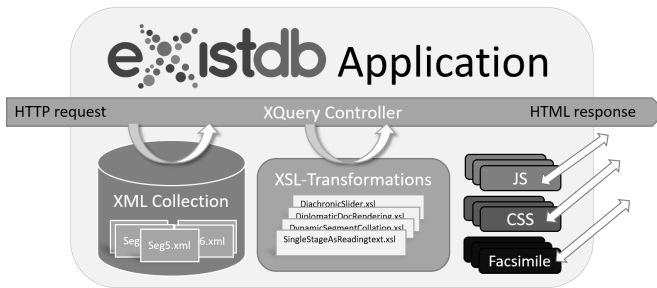
Figure 6: Simplified Controller Schema of the eXist-db Backend.

processes such requests. It first takes the requested TEI file (or parts of it) from the respective XML collection. Then an appropriate XSL-file is selected from a set of different possible scenarios, containing transformation templates to produce the requested visualisation. The integrated XSLT processor transforms the TEI contents according to the defined XSLT templates into an HTML output with references to CSS definitions and JavaScript libraries for dynamic browser functionalities and asynchronous AJAX requests. Finally, the resulting HTTP response (HTML including JavaScript, CSS and images, i.e. facsimiles) is sent to the client for display in a browser.

With this design, we approach a system-independent reusability of the individual XSLT transformations. Although all transformations could also be achieved using XQuery (and some simple asynchronous calls are indeed processed this way), XSLT allows the extraction of single modules from the eXist-db environment and their reusability in other environments.[6] On the other hand, existing XSLT stylesheets, such as those provided by the TEI (see Rahtz), can be plugged into a transformation chain. For example, in a first transformation, a more common TEI encoding (diplomatic or textual) confined to, and valid for, one single document (that is: the *intra-document* encoding for that given document) may be derived from the proposed document-embracing encoding. Such encodings specific to individual documents could then, in turn, be queued into an XSLT stylesheet provided by the TEI to produce a generic visualisation – though such a procedure would lose the text-genetic focus realisable through continuous mark-up of text progression across document borders which, as discussed above, may be generated from processing the results from collation of the textual states of contiguous documents.

While the application is still under development, some first modules have already been implemented. Two of these modules, the basic diachronic text viewer and what

---

[6] The basic prerequisite to reuse these XSLT stylesheets is of course a TEI encoding structured as described in this article.
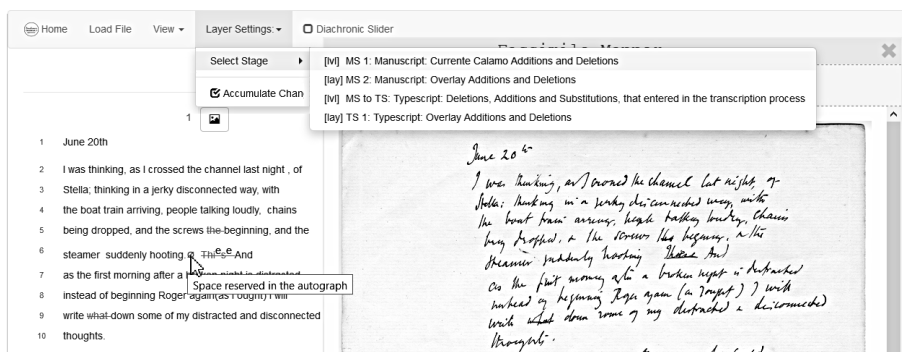
Figure 7: User Interface: Basic Text View.

we call the *Diachronic Slider*, will be introduced here to demonstrate the advantages of an inter-documentary TEI encoding.

Figure 7 shows a screenshot of the basic text viewer opened to the beginning of segment 5; it displays the same paragraph of Woolf's "A Sketch of the Past" we have used above to explain the different types of alterations discernible between documents. For each encoded page, the corresponding facsimile can be loaded into a scalable and draggable window by clicking on the facsimile symbol that precedes the page transcription. The image is loaded by an asynchronous call from the server.

In the current version, the user can select a text stage in the main menu and the changes of the corresponding *layer* or *level* are dynamically switched on or off in the transcript – without the text being reloaded from the server. Both the description and the order of the single text stages in this selection menu are taken from the `tei:listChange` element in the `tei:creation` statement (see the example code above). In the given example, the first text stage "MS 1: Manuscript: Currente Calamo Additions and Deletions" is selected.

The user can choose in the *Layer Settings*, whether she wants to display all layer- and-level changes accumulated on previous and the current document or only those of the document carrying the currently selected *layer*. At any given layer visualisation, changes that happened on previous *layers* and *levels* are shown in grey. Alterations at the selected stage are colour-coded with red plus strike-through for deletions, green for additions, superscripted plus green for substitutions and blue arrow-marks for transpositions. (This palette of colour indications is, of course, not reproduced in our black and white figures.) For various TEI elements, furthermore, mouse-over inform- ation is derived from the encoding; e.g., in line 6 a gap is encoded and commented with "Space reserved in the autograph" (see figure 7). Subsequent development will

Figure 8: User Interface: Diachronic Slider.

see synchronised highlighting of a selected line of text on both the transcript and the facsimile.

Figure 8 shows the *Diachronic Slider*, a tool that demonstrates the dynamic potential of a genetic encoding across the document border. It visualises the textual development of a user-selected text range by successively displaying each of the different stages.[7] When the user selects the button *Diachronic Slider* in the main menu, a new scalable overlay window opens. All modularised interface tools open in such windows. If the contents (text, metadata, images) a tool requires are not already included in the base text, the respective information is requested from the server and displayed in a tool sub-window. These requests are realised as asynchronous AJAX calls in the background. By this design contents are produced on demand on the server by the application of XSLT stylesheets on TEI documents. The results are seamlessly included into the frontend without interrupting the user's workflow by reloads of the complete page.

When the *Diachronic Slider* is first opened, the window is empty and the user is advised to click on the *Select* button in the slider's icon bar (figure 8). On clicking, the button turns grey and the mouse cursor changes into a paint brush to signal that the application is now in the marking mode. With the brush, the user selects and marks a text span in the main text area at will (avoiding only to set beginning or end into the middle of a textual alteration). Although additions on later stages and deletions on earlier stages are not visible in the display of the base text, the encoding of all

---

[7]   The tool development is influenced by Julie André's and Elena Pierazzo's interface-prototype for the genetic visualization of a Marcel Proust notebook (André and Pierazzo). Pierazzo emphasizes "how crucial is transiency (or time-based animation) in the case of genetic editions; this function could well represent the key factor in making digital editions worth pursuing for genetic editing" (Pierazzo 35).

alterations over all *layers* and *levels* is present in the linear succession of the text. Text segments invisible on the screen are not deleted at the back-end; their HTML representation in the browser is simply hidden by a dynamic JavaScript function. The marked text span which usually breaks the strictly tree-based *document object model* can be restored to build a correctly nested subtree which is passed to the *Diachronic Slider*, including all alterations at all text stages.

Immediately on release of the mouse click, the subtree that represents the selected text is copied once for each text stage. A visual textual representation of each stage is constructed: for each copy of the subtree, all alterations *until* the current text stage are applied. Previous deletions are struck through, future deletions are shown as normal text; previous additions are switched to *visible* and future additions are *hidden* (and so on). The results are mounted to the slider window. Figure 9 shows the same slider selection as figure 8 for two different states (top and bottom). All available text stages can be selected by clicking on the labels in the slider's icon bar. On *mouse over*, the labels are extended with little tooltips that contain the corresponding `tei:change` description from the metadata.

The top window in figure 9 shows the selected lines on the stage "MS1", the first manuscript version. When the user clicks on the label "MS to TS" in the icon menu, the text of the inter-documentary *level* slides in on the centre of the split-screen window from the right side until it fully covers the previous text (result: figure 9, bottom). The sliding direction signals the chronology. A slide from the right-side signals that a later text stage was selected and a slide from the left-side signals a preceding stage.[8] Directly underneath the central sliding part of the window is a static text version enabling eye-collation of the two parallel window parts. This static text is an identical copy of the initially selected base text. By changing the text stage in the main window before the slider is brought into the marking mode (main menu: Layer Settings →Select Stage), the static collation level can be set to the user's requirement. Note that figure 9 visualises the complex mixed-type transposition shown in figure 5. In the top state, the text is not transposed, while in the bottom state the two segments are swapped and marked as transposed with little arrows. A further highlighting of transpositions e.g. by an animation of the swap, may be considered at a later stage of the tool's development.

---

[8]   The dynamic appeal of this tool is difficult to describe and cannot be revealed by static screenshots on printed paper. The animation of the changing text stages can hardly be captured in this article. A live demo can be accessed via www.compositiongenetics.org.
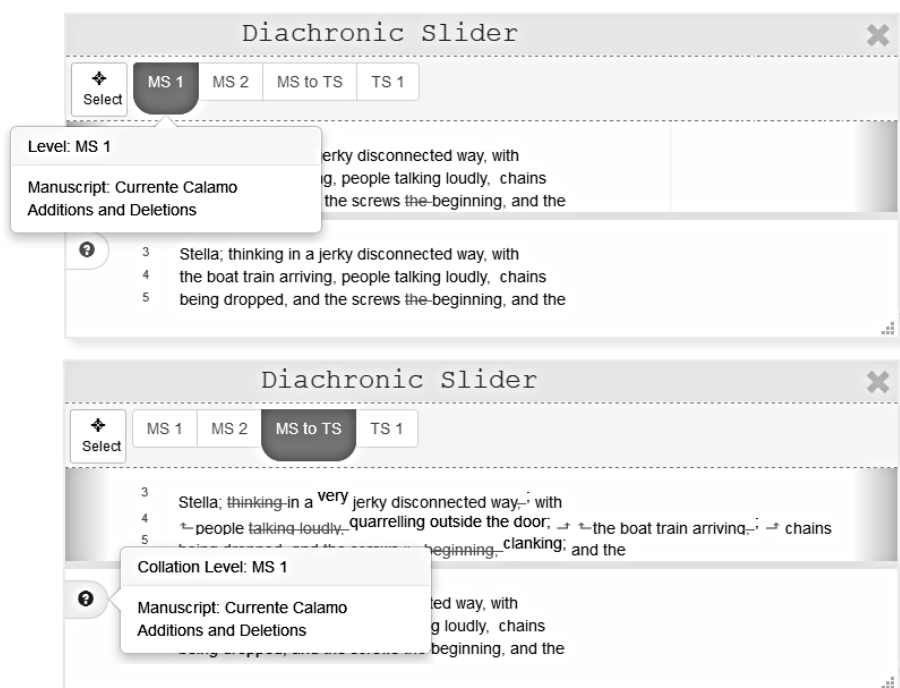
Figure 9: *Diachronic Slider* in two states. Top: first manuscript stage; Bottom: inter-document level.

6 Conclusion

Concluding a survey of work-in-progress means looking ahead. What we hope to have drawn attention to is that, while in the common situation of creative writing extending over sequences of documents the material evidence of the genesis of texts resides in individual documents, the text in the total development of its articulation and alteration in writing progresses across document borders. Capturing this fundamental textual condition digitally requires understanding, response and solutions beyond the attempts already made (e.g., by the TEI Consortium) to instigate conventions for analysing and marking-up many-layered writing in and on documents. Aids to standardisation have hitherto been focussed on use cases related to the individual document. Our proposition here has been how to deal with text genetics extending over sequences of documents: first, how to capture and pre-structure the material givens for intra-document and inter-document text progression; second, how, by means of automatic collation, to stratify the digital record of the material states-of-

text; third, how analytically and interpretatively to determine as layers and levels the stratifications elicited from the collation; fourth, to devise a differentiated mark-up to label the layers and levels and so render them retrievable according to their every analytical perspective encoded in the mark-up; and fifth, to devise dynamic modules for visualisation of the dynamic text progression so encoded. The analysis, structuring, and consequent mark-up amount to, and achieve, a 'synoptic' digital perception of a text progression across documents. Hence, visualisation operations linking into the synopsised record need not contend with, by preliminary coordinating, divergent document-specific encodings. Instead, all data from which to generate visualisations are already in one place in one comprehensively structured archival pool. How this can be utilised to advantage and ease, we have here demonstrated from the design and use of the *Diachronic Slider* tool. Developed in tandem with ongoing work on a Digital Critical and Synoptic Edition of James Joyce's *Ulysses*, and as only one of several modularised visualisation tools there in demand, the *Diachronic Slider* constitutes but one of several modularised interface designs under present development to visualise synoptically structured diachronic (genetic) text data under multiple perspectives. Interoperability is thus built into the development from the start. Our expectation is that user requests and use cases adduced will contribute to ramifying and strengthening our proposed approach to textual and critical investigation of multiple-text heritages.

Bibliography

André, Julie, and Elena Pierazzo. "Le codage en TEI des brouillons de Proust: vers l'édition numérique." *Genesis. Manuscrits–Recherche–Invention*, 36, 2013, pp. 155–161.

Beal, Peter. *A Dictionary of English Manuscript Terminology: 1450 – 2000.* Oxford University Press, 2011.

Burnard, Lou, et al. "An Encoding Model for Genetic Editions." *TEI - Text Encoding Initiative*, 2010, www.tei-c.org/Activities/Council/Working/tcw19.html. Accessed 11 August 2017.

Joyce, James. *Ulysses: A Critical and Synoptic Edition.* Prepared by Hans Walter Gabler with Wolfhard Steppe and Claus Melchior, Garland Publishing Inc., 1984, ²1986, New York. 3 vols.

McCracken, LuAnn. "The Synthesis of My Being": Autobiography and the Reproduction of Identity in Virginia Woolf." *Tulsa Studies in Women's Literature,* vol. 9, no. 1, 1990, pp. 59–78.

Ott, Wilhelm, and Kuno Schälkle. *TUSTEP 2016: Handbuch und Referenz in zwei Teilen.* pagina, 2015.

Pierazzo, Elena. *Digital scholarly Editing: Theories, models and methods.* Ashgate, 2015.

Rahtz, Sebastian. "XSL stylesheets for TEI XML." 2016, www.tei-c.org/release/doc/tei-xsl/. Accessed 11 Aug. 2017.

Saller, Harald. "HNML - HyperNietzsche Markup Language." *Jahrbuch für Computerphilologie*, 5, edited by Georg Braungart et al., Mentis, 2004, pp. 183–191, computerphilologie.uni-muenchen.de/jg03/saller.html. Accessed 11 Aug. 2017.

TEI Consortium. *TEI P5v2: Guidelines for Electronic Text Encoding and Interchange.* TEI Consortium, 2011, www.tei-c.org/Vault/P5/2.0.0/doc/tei-p5-doc/en/html/. Accessed 11 Aug. 2017.

—. *TEI P5v3.1.0: Guidelines for Electronic Text Encoding and Interchange.* TEI Consortium, 2016, www.tei-c.org/Vault/P5/3.1.0/doc/tei-p5-doc/en/html/. Accessed 11 Aug. 2017.

Woolf, Virginia. "A Sketch of the Past (TS.A.5.a, 1–69)." Monks House Papers, University of Sussex, 1939–1941.

—. "A Sketch of the Past (MS.A.5.b-A.5.d): Reminiscences of Family and Childhood by Virginia Woolf." Monks House Papers, University of Sussex, 1939–1941.

Zapf, Volker. *HNML: HyperNietzsche Markup Language.* University of Munich, 2006, www.hypernietzsche.org/events/sew/post/Slides%20and%20Texts_files/HNML.pdf. Accessed 11 August 2017.

Zeller, Hans. "Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition" *Texte und Varianten*, edited by Gunter Martens and Hans Zeller, 1971, München, pp. 45–90.