

Un-mixing the re-mix. Publishing the complete manuscripts of Anne Frank

Peter de Bruijn¹, Elli Bleeker², Marielle Scherer³, Peter Boot¹,
Karina van Dalen-Oskam¹, Gijsjan Brouwer⁴, Ronald Haentjens
Dekker², and Bas Doppen²

¹Huygens Institute for the History and Culture of the
Netherlands

²Humanities Cluster (KNAW)

³National Archives (The Netherlands), formerly Huygens
Institute for the History and Culture of the Netherlands

⁴Independent software developer, formerly Humanities Cluster
(KNAW)

We present a new and complete edition of the manuscripts of Anne Frank. While rewriting her diary into a novel, Anne ‘re-mixed’ her own work; when after the war her diary was published, the texts underwent another remix. The digital edition aims to present the manuscripts in their authentic form, with an integral transcription including deletions and additions. It provides historical context to the described events and highlights the themes discussed in the manuscripts. The possibilities of the digital medium help clarify the textual relationships between the manuscripts. The edition represents an ‘un-mixing’ of Anne Frank’s work that makes it possible, for the first time, to study Anne’s personal growth and her development as a writer in the years she spent in hiding.

Keywords: scholarly editing, diary, text encoding, text variation, Anne Frank

1 Introduction

In this article we discuss the methodology the Anne Frank Manuscripts project employed in reconstructing, encoding, and visualising Anne Frank’s writing process in a digital scholarly edition of her work. On the face of it, the diary of Anne Frank does not seem to require any introduction. Since its publication in 1947 the work has been read and celebrated worldwide. It may be the best sold originally Dutch work of all time.¹ However, it is less generally known that the diary as it is most commonly read

¹ https://en.wikipedia.org/wiki/List_of_best-selling_books. All urls checked 10 June 2023.

is in fact a compilation of two different works by Anne Frank: her original diary and an unfinished novel based on her diary. The paper discusses how the digital edition “un-mixes” Anne Frank’s diary and foregrounds her writing and composition process. In September 2021, a digital scholarly edition of her complete works was launched by the Belgian Vereniging voor Onderzoek en Ontsluiting van Historische Teksten [Society for Research and Publication of Historical Texts] (Frank, 2021). It is available online at <https://annefrankmanuscripts.org> in all countries where copyright laws permit it (including Belgium and Luxembourg, but not the Netherlands). The edition is the result of more than a decade of research by the Anne Frank Manuscripts project, a joint effort of the Huygens Institute for History and Culture of the Netherlands and the Anne Frank House in Amsterdam (Huygens Institute, 2021). We will discuss the research that resulted in the digital edition, focusing on the ways in which we have examined and encoded Frank’s writing process. We will also illustrate how Frank’s texts have been mixed and remixed over the years, and how the digital “un-mixing” of her work sheds new light on her creative working methods and her literary talent.

2 Related work

As we will see, Anne Frank’s manuscripts have a double character: they are historical documents, providing knowledge about events in the past, but they can also be considered as literary manuscripts. There are many digital editions of diaries; Sahle’s catalog lists 47 entries (Sahle, 2023). Many of these provide similar functionality to our edition in terms of historical and contextual annotation (see for example the travel diaries of Hans Posse² or William Godwin’s diary)³ or a distinction between reading text and transcription (see e.g. the diaries of Michael Faulhaber⁴ or those of Christian II. von Anhalt-Bernburg⁵). The edition of the diaries of Karl Wiesinger⁶ is indexed by theme, somewhat comparable to what we did for Frank’s manuscripts. Wiesinger was a literary author, and his diaries contain many personal reflections. This brings us in a more personal and literary domain than for instance the mostly factual notes about the events of the day that we find in diaries such as Godwin’s. Well-known digital editions of authors’ diaries are those of Robert Graves⁷ and Arthur Schnitzler.⁸ Graves’ diary is of the factual type; Schnitzler is more personal and discusses e.g. dreams and personal relations. But neither of these diaries have been reworked into a novel in diary form.

A comparison of various text versions is sometimes used for historical documents (see e.g. the editions of the *Holinshed Chronicle*⁹ or Foxe’s *Acts and Monuments*)¹⁰ but is more common in the study of the genesis of literary works. A number of tools has been developed for parallel display of multiple text versions (Bleeker et al., 2019). Examples of editions of literary texts which feature parallel display of multiple text

² <https://editionhansposse.gnm.de/>.

³ <http://godwindiary.bodleian.ox.ac.uk/index2.html>.

⁴ <https://www.faulhaber-edition.de/index.html>.

⁵ <http://diglib.hab.de/edoc/ed000228/start.htm>.

⁶ <https://gams.uni-graz.at/context:wiesinger>.

⁷ <https://graves.uvic.ca/index.html>.

⁸ <https://schnitzler-tagebuch.acdh.oeaw.ac.at/>.

⁹ http://english.nsms.ox.ac.uk/holinshed/texts.php?text1=1587_1262&text2=1577_1662#p4638.

¹⁰ <https://www.dhi.ac.uk/foxe/>.

versions are Lodewijk van Deyssel's prose-poem *Menschen en Bergen*,¹¹ Coleridge's *Wanderings of Cain*¹² or the archive of Pessoa's *Livro do Desassossego*.¹³ The variants between text versions can be classified in multiple ways. Spadini (2019) mentions four ways of doing so: the category of the change (addition, deletion, substitution and transposition), the linguistic aspect (orthography, morphology, syntax, lexis), the direction, and the material aspect of the variant. In some editions the interface makes it possible for the users to select the categories of variation they want to see. An example of such an edition is the *Bucolica* by Calpurnius Siculus.¹⁴

We should admit, however, that we did not perform an extensive analysis of similar editions before starting work on the edition of the diary. We were well aware of what was going on in the world of the digital edition and the Text Encoding Initiative and wrote a functional design based on that awareness and our ambitious requirements.

3 Works of Anne Frank

Anne Frank began keeping her diary a few months before going into hiding. The diary, written (mostly) in the form of letters to her imaginary friend Kitty, served her as a friend and *confidante* during this difficult period, but it also offered her an environment to develop and refine her writing skills. Frank enjoyed writing: in addition to her diary, she wrote several short stories and fables (later published as the collection *Tales from the Secret Annex*) in which she sometimes “mixed” elements from her diary. On March 28 1944, Frank heard a radio broadcast from Dutch minister Gerrit Bolkestein, in which he announced the government's intention to collect war diaries and letters from Dutch people once the war was over, because of their importance for future historiography.¹⁵ Inspired by that broadcast, in May 1944 Frank began to rework her diary into a novel she entitled *Het Achterhuis* [literally: *The Back House*]. In the span of three months, Frank rewrote large parts of her diary and transformed it into a literary work, complete with narrative arcs and compelling characters based on her fellow hidiers.

There is certainly no one-to-one relation between the entries in the diary and in *The Back House*. In her rewriting, Frank could merge texts from letters of multiple dates to create a new text. Some of the events that she wrote about in her diary would return in her novel and then be fictionalised into one of her tales.

After the raid on 4 August 1944, in which Frank's family and the other people in hiding were arrested, Miep Gies, one of the helpers, found some of Frank's manuscripts and kept them safe. After the war she handed them to Frank's father, Otto, who was the only one of the family to survive the war. When it comes to the diary, three notebooks are left; one or more have disappeared. *The Back House* is written on loose sheets of paper. Figure 1 presents an overview of Frank's writing activity and the remaining physical evidence, including the story *Cady's Life* and *The Book of Beautiful Sentences*, in which Frank recorded striking fragments she came across in her reading.

¹¹ <http://menschenenberg.nl/huygens.knaw.nl>.

¹² <https://romantic-circles.org/editions/cain/parallelcanto2.html>.

¹³ https://ldod.uc.pt/fragments/fragment/Fr286/inter/Fr286_WIT_ED_CRIT_SC.

¹⁴ https://ldlt.digitallatin.org/editions/calpurnius_siculus/.

¹⁵ A transcript of the broadcast can be found in the NIOD collection and on annefrank.org.

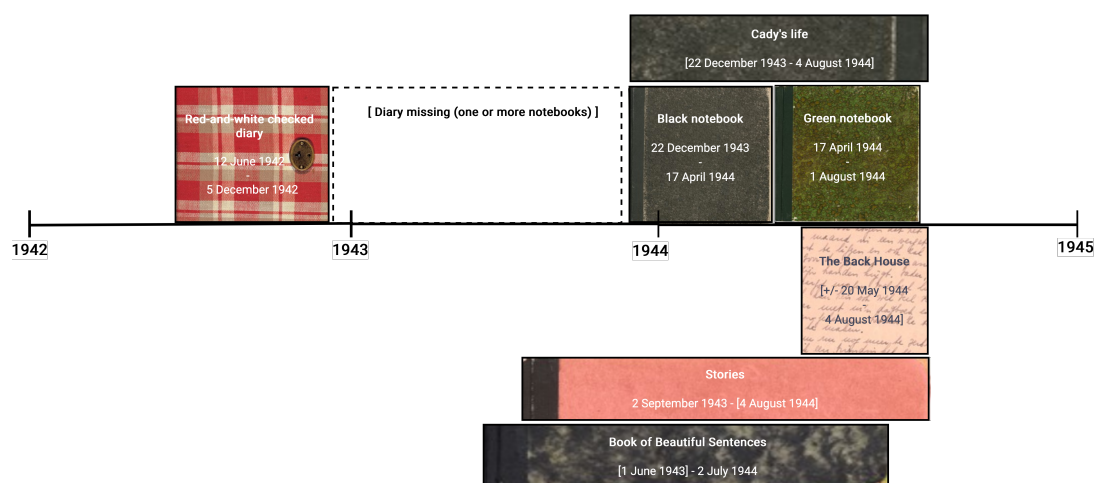


Figure 1: Frank's remaining manuscripts and their dating.

4 Remixing

The diary of Anne Frank as the world knows it has been compiled by Otto Frank. This compilation – in a sense, a “remix” of Anne Frank’s work – is a composition of fragments from the diary, *The Back House*, and Anne’s stories. Otto Frank also shortened the text and excluded certain passages that he considered too sensitive. The result was first published in 1947 as *Het Achterhuis* [known in English as *The Diary of a Young Girl* or as *The Secret Annex*] (Frank, 1947). In 1991 an extended edition was published, edited by Miriam Pressler, that included previously omitted passages (Frank, 1991). In the United States, this edition was marketed as the ‘definitive edition’ (Frank, 1995).

In a first attempt to un-mix this composite work, researchers of the Netherlands Institute for War Documentation (NIOD) published a comprehensive edition of Frank’s diary in 1986 (Frank, 1986).¹⁶ The NIOD edition distinguishes between the texts of the diary, Anne’s novel, and Otto Frank’s 1947 edition. Each page shows (a) a fragment of a diary entry; (b) the corresponding passage in the novel, and (c) the corresponding passage in the Otto Frank edition (Figure 2). By reading the NIOD edition “vertically”, readers can compare the three versions and work out the changes made by Anne Frank and her father, respectively. However, not all of Frank’s texts lend themselves to a vertical juxtaposition: as we’ll see later, her (re)writing and mixing processes are too complex for a straightforward comparison. The in-depth, cross-referenced textual comparison that would help to understand the relationships between Frank’s works and to illustrate her writing and composition process is impeded by the limited affordances of the print medium.

5 Digitally Unmixing

To this end, the Huygens Institute and the Anne Frank House set out to study Anne Frank’s manuscripts and work methods, and to present the findings in a digital environment that offers users the opportunity to examine the texts for themselves (Frank,

¹⁶ Henceforth referred to as “the NIOD edition”

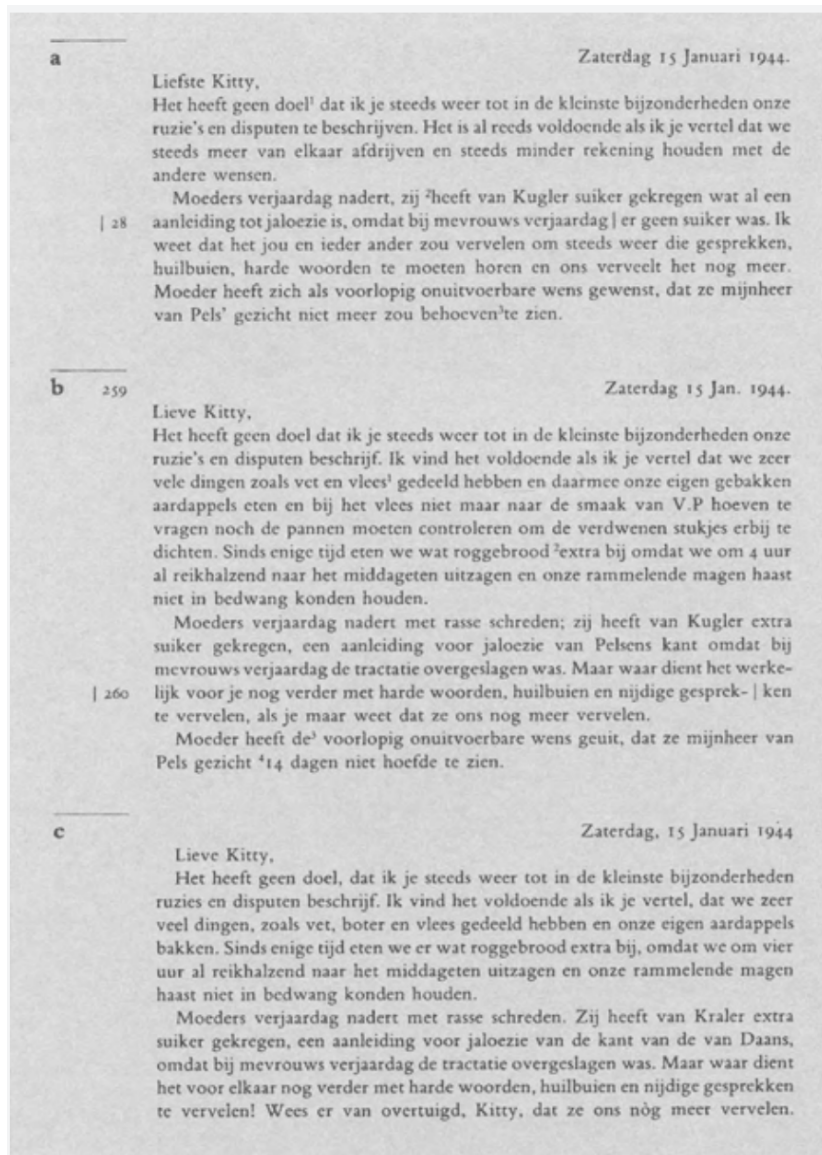


Figure 2: Example of the NIOD edition's presentation of a fragment from (a) Anne Frank's diary, (b) its corresponding fragments in the text of *The Back House*, and (c) the Otto Frank edition.

2021). Historical research done at the Anne Frank House resulted in a knowledge base. Transcription, editing and enrichment was done at the editorial group at the Huygens Institute. The group used the knowledge base as well as its own research in annotating the edition.

To facilitate further research, we formulated three main scholarly objectives for the edition:

1. to unite the manuscripts in a digital environment where they can be accessed, read, compared, and studied from different points of view (e.g., chronological, material, or thematic);
2. to annotate and contextualise Frank's manuscripts and the persons and places she mentions, and the historical circumstances in which the manuscripts were authored;
3. to provide the means to study the textual relationships between Frank's texts in great detail so as to gain insight into her writing process.

The website should stay as close as possible to the original material (the manuscripts) and make the additional content (the context) accessible in a user-friendly and unobtrusive way.

The manuscripts have been given prominence in the design by displaying the zoomable facsimile manuscript pages with the transcription, as well as by the facsimile-only display of whole manuscripts. We also chose to include thumbnail images in the table of contents and in the navigation bar on the manuscript display pages. For each of the manuscripts, an image of the manuscript cover is used as an icon to clarify references to that manuscript throughout the site.

At the time of writing, the edition is available only in Dutch. Users can browse the manuscripts of every surviving work by Frank: i.e., her original diary, *The Back House*, the *Tales*, the short novel *Cady's Life*, and *The Book of Beautiful Sentences*. An extension with an English translation of all text is under construction and its publication is foreseen for the end of 2023. Various other enhancements in the edition's usability as a research tool are being developed and will be included with this English version. In our discussion of the edition we will sometimes refer to this future publication.

6 First objective: access to the manuscripts

The edition foregrounds the materiality of the manuscripts: it shows the pictures, postcards, and notes that Frank pasted into her diary, and it displays physical traces like deletions or additions of Frank's writing process. All manuscripts have been described and transcribed in TEI/XML (TEI Consortium, 2020) and the transcription includes a full account of the writing process of the texts, as far as it was possible to reconstruct that. The text can be displayed in an original, diplomatic version (which includes a representation of material and textual details, see Figure 3) or in a reading version. Both versions include a facsimile display, as well as access to extensive annotation and a bar of thumbnails providing the manuscript context for the displayed entry. In the upcoming release the manuscript can also be displayed in a facsimile-only view, where the relative absence of editorial intervention will provide perhaps the most immediate experience of the manuscript. In that release, the user will have the option of displaying an English reading text with annotations. A diplomatic version of the

text will not be available in English, as in translation most of the textual details lose their relevance. However, the user will have the opportunity to view a side-by-side rendering of the Dutch and English texts.



Figure 3: A screenshot of the diary entry of 6 January 1944 (future release). The left panel shows a digital facsimile of the manuscript page. The middle panel shows the original version of the manuscript text; a “p” indicates a place in the text where the text was modified. Clicking the “p” shows the steps of that process in detail. The right panel provides access to various forms of annotation. Hovering on the thumbnail bar brings up a summary of the letter(s) on that page and clicking the thumbnail brings the user to that page.

The manuscripts, as all other material in the edition, are also accessible through a search interface. The search engine facilitates searches by key word and/or facet, such as date, persons, themes or document.

One of the fundamental aims of the edition was to reconstruct, encode, and visualise Frank’s writing process. Especially in the novel, the revisions reveal how Anne Frank experimented with different wordings, rephrased certain fragments in a more literary way, and contextualised events in view of a future readership. In short, the revisions on the manuscripts show a writer who is honing her literary skills. Anne Frank’s revisions are encoded with the TEI-XML elements `` and `<add>`, and the `@seq` attribute to express our interpretation of the chronology of the revisions (see Figure 4 and Figure 5 for a small example).¹⁷

Apart from the additions and deletions in the text, the transcription also faithfully records other material aspects of Frank’s writing, such as retraced characters and the writing utensils she used, as well as the placing of text on the manuscript page. A remarkable characteristic of Frank’s writing is that she used both a running and a block script. In the original text version, the edition reproduces that distinction. The edition also notes where Frank made errors against Dutch grammar or spelling. A very conservative approach was taken with respect to correction.

While in general we tried to stick to the elements provided by the TEI, in some cases

¹⁷ TEI Guidelines P5, chapter 3.5.3. “Additions, Deletions, and Omissions” and chapter 11.3.1.4. “Additions and Deletions”, see <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

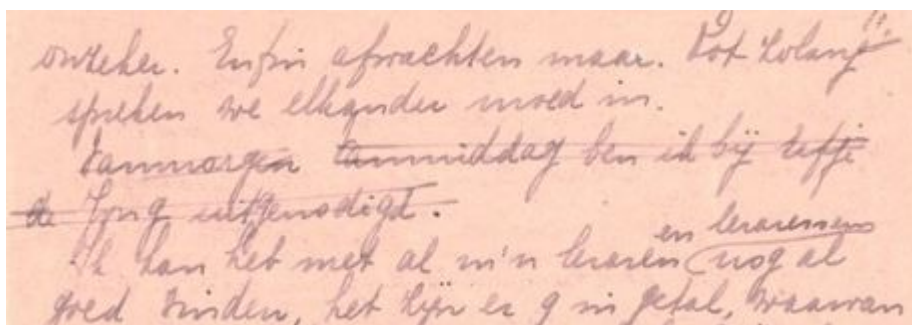


Figure 4: Fragment of the manuscript of Het Achterhuis showing a strategic deletion.

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI>
  <text>
    <body>
      <div>
        <p>
          <seg>
            <del seq="1"><del seq="0" rend="direct">Vanmorgen</del><add seq="0" place="inline">Vanmiddag</add>
              ben ik bij Eefje</del><del seq="1"><lb/>de Jong uitgenodigd.</del>
            </seg>
          </p>
        </div>
      </body>
    </text>
  </TEI>
```

Figure 5: TEI-XML transcription of the text of the fragment, simplified for presentation purposes.

we decided to extend the TEI model and to create elements in our own namespace. For one thing, we felt the note element, already used for various types of editorial annotation, would be semantically overloaded by also using it for those cases where Frank added notes to her writing. We defined an `af:orignote` element to transcribe those notes. We also created an `af:sepLine` element to be able to describe the many different separating lines Frank drew between the sections of her *Book of Beautiful Sentences*. We added a milestone version of the `rs` (referring string) element (`af:rsMilestone`) in order to have referring strings that crossed hierarchical boundaries.

Besides defining new elements for our encoding, we also made some other extensions. For example, as in most of the Huygens Institute's other edition projects, we relaxed the content model of `div` (textual division) to allow (among other things) paragraphs that follow subdivisions within a division. We also added some date-related attributes to `div`, such as `when`, `notBefore` and `sortDate`, in order to structurally describe the date of writing of various units of textual content. Because of the need to describe, at almost any level of the text, aspects of the handwriting, we made `hand` into a global attribute. For a similar reason, we added the `place` attribute to a number of elements.

An interesting physical characteristic of especially the first diary notebook is that Frank pasted in various documents (letters and postcards). Near the end of the notebook, when she was afraid of lacking space to continue her diary, she also glued in sheets of paper extending the amount of space in the notebook. We devised an encoding scheme in which what we called `af:patchLocations` indicated the locations where sheets or other objects were pasted into the diary.¹⁸ The texts written on the

¹⁸ We felt the `add` element does not cover this situation, as it refers to textual additions, not

additional sheets use the `on` attribute, which we defined for that purpose, to refer to the location that they were pasted in. At those locations in the diary, the text structure may become rather complicated, as Frank sometimes continued regular entries on the extra sheets, but could also use successive extra sheets for parallel entries and brief interjected texts. In the website, the texts on sheets that were pasted in are shown in a frame. At the location where the sheets were pasted in, the edition provides references to the text on those sheets.

When looking back on the encoding process, one aspect that we may have expected too much from is the precision of the XML encoding when representing the location of each individual word on the page. Especially in the openings and closings of the letters, the dates and salutations may be placed in varying locations that can be hard to assign to specific manuscript lines or horizontal positions. A date, for instance, might be split over two lines, the second part coming after the salutation. It wasn't easy to automatically turn this encoding into a consistent text presentation. Because of this, at a rather late stage we decided where necessary to add a second dateline to the encoding (with `resp=ed(itor)`) to normalize the presentation of the letters' openings in the reading view. Given the presence of facsimile images in the edition, one can wonder how much detail is really necessary in the transcription.

7 Second objective: contextualize Frank's works

While Frank was working on a novel, and therefore a work targeted at the public, her writing is also highly private and based on her own world. This world was much smaller than it should have been, due to the extraordinary circumstances of her life. Users for whom the Second World War is more than 75 years away need clarification of the historical circumstances, the persons that surrounded Anne (from the hiders and helpers in her immediate environment to the politicians involved in the war effort), the books that she read and the places that she mentioned. The edition provides three levels of annotation: first, explanatory notes attached to specific positions in the text, explaining manuscript details or historical circumstances; second, contextual information that can be provided thanks to detailed encoding of occurrences of persons, locations, themes, media and places; and third, essays that provide in-depth discussions of the manuscripts, Anne's personal development and the phenomenon of going to into hiding in the wartime Netherlands.

The edition provides exhaustive information for each letter, accessible from the information menu on the right (Figure 6). Apart from the explanatory notes, editorial details and a summary, the menu also gives access to the structural information. Figure 6 also shows how, for example, the corresponding text fragment is highlighted when the 'Themes' option is chosen and the 'Writing' theme is selected. The user can search for occurrences of that theme or read a further discussion of the theme's relevance to Anne and her manuscripts.

Similar functionality is available for the other options, such as persons and places. For all persons that Frank refers to, a description or short biography is available (Figure 7). There are also descriptions of all rooms in the front and back house and places referred to in the manuscript (Figure 8). The wireframe of the house and the locations on the map are hyperlinked to further descriptions, which are also accessible from the entries.

physical extensions of the writing space.

Contents

Contents

Editorial introduction

Explanatory notes (11)

Persons (25)

Locations (6)

Media (0)

Themes (5)

Images (4)

Donderdag 6- Januari 1944.

Lieve Kitty,

Vandaag moet ik je drie dingen bekennen, die heel wat tijd in beslag zullen nemen, maar die ik aan iemand moet vertellen en dat kan ik dan toch maar het best aan jou doen, omdat ik stellig weet dat je altijd en onder alle omstandigheden zult zwijgen. Het eerste is over moeder. Je weet, dat ik vaak geklaagd heb over moeder en dan toch steeds weer moeite deed aardig tegen haar te zijn. Plotseling is het me nu duidelijk geworden wat er aan haar mankeert. Moeder heeft ons zelf verteld, dat ze ons meer als haar vriendinnen dan als haar dochters beschouwd; dat is nu natuurlijk wel heel mooi, maar toch kan een vriendin de plaats van een moeder niet vervangen. Ik heb er behoefte aan, mijn moeder als een voorbeeld te nemen en eerbied voor haar te hebben en mijn moeder is in de meeste dingen wel een voorbeeld voor mij, maar juist een voorbeeld zoals ik het niet moet doen. Ik heb het gevoel dat Margot in al deze dingen zo anders denkt en wat ik je nu verteld heb, nooit zou begrijpen. En vader ontwijkt alle

Thema's (5)

Buiten het Achterhuis

Voor de onderduik / flashbacks

Oorlogsomstandigheden

Dagelijks leven

Ontspanning

Verzorging en bevoorrading van de onderduikers

Lichaamsverzorging en gezondheid

Lezen, leren en praten in het Achterhuis

Godsdienst

Lezen

Schrijven

Persoonlijke omstandigheden en ontwikkeling

Figure 6: On the left, the information menu for the letter of 6 January 1944 (Diary). On the right, the result when selecting the Themes option.

Filter

People in hiding

Helpers

Family and friends

Acquaintances, people from the neighbourhood and classmates

Business relations

Public persons


Historical, mythical and biblical names

Writers


Fictitious persons

Anonymous and other


People in hiding (8)




Otto Frank




Edith Frank




Margot Frank




Anne Frank




Hermann van Pels



Auguste van Pels



Peter van Pels



Fritz Pfeffer

Figure 7: The entrance page to the person descriptions, here opened on the page referring to the people in hiding.

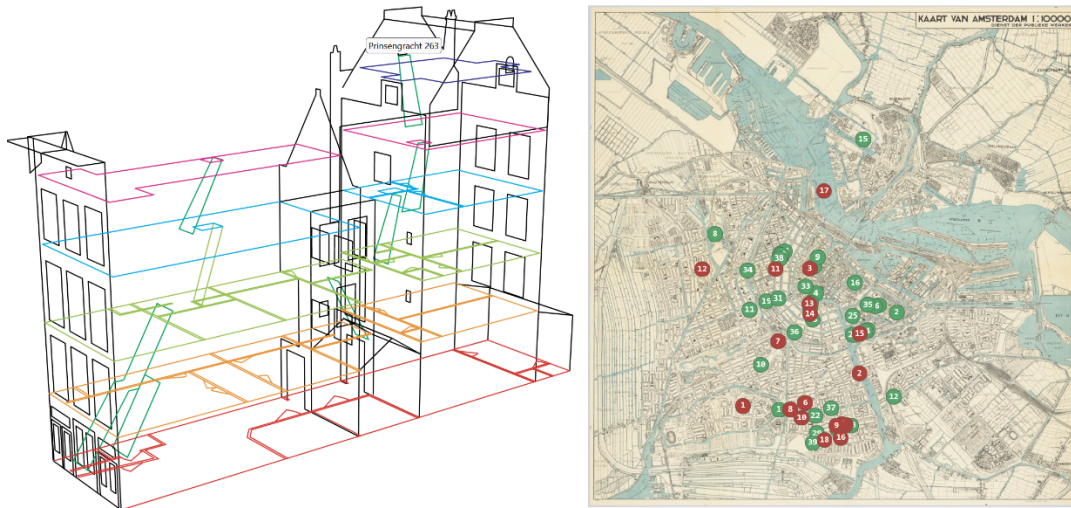


Figure 8: Places referred to in the manuscripts. On the left, a wireframe of the front and back house. On the right, a contemporary map of Amsterdam.

8 Third objective: display of textual relationships

As mentioned above, Frank reworked the experiences described in her diary into a novel and into tales. It is natural that the texts of her productions should show similarities as well as differences, as her viewpoint changed and her skills matured. That is why it is important for the edition to facilitate in-depth study of how Frank described the same events in her various texts.

Frank's reworking of her diary texts has many aspects of 'bricolage': in creating the letters in her novel, she could reuse diary entries from up to six different days. The diary and the novel are not essentially a single text with variants, but two or more texts inspired by the same sequence of events. An ongoing variant apparatus would be inappropriate. We chose to create a selective apparatus for passages with agreements in the various manuscripts, independent of the manuscript transcriptions, also because of the level of encoding already present in the manuscript transcriptions. The apparatus was created manually. We experimented with CollateX (Haentjens Dekker and Birnbaum, 2018), but it turned out that the agreements between the diary entries were too fragmentary for CollateX to produce a useful result.

Figure 9 shows an example from the variant apparatus. The sections in the apparatus (*ab* elements) are anchored to ranges defined by start and end lines in the relevant manuscripts (here A2, the second notebook of the diary, and B, the novel). Apparatus entries use the *type* attribute to encode the level of agreement between the readings. We distinguish the levels 'identical' (i.e. text outside of the apparatus entries), 'nearly identical', 'agreeing', 'same meaning' (in different words), 'no agreement' and 'moved'. Also visible in the example is that higher-level structural differences, such as paragraph boundaries, are encoded here using milestone elements.

The variant apparatus is the basis for a display of the text relations at various levels of detail. We show a fragment of the highest-level overview in Figure 10, which displays an entry from the diary that has connections with two entries from *The Back House*.¹⁹ Clicking one of the Back House entries takes us to a parallel display of the source and target entries (not shown here); after that, the user can zoom in on a detailed

¹⁹ The screenshots show the version under development at the time of writing.

```

<ab>
  <ptr target="A2.xml#range(A19440223_L003,A19440223_L026)" />
  <ptr target="B.xml#range(B19440217_L042,B19440217_L063)" />
  <seg><milestone unit="p" />Sinds gisteren is het
  <app type="similar">
    <rdg wit="#A2"> weer heerlijk</rdg>
    <rdg wit="#B"> buiten heerlijk weer</rdg>
  </app> en ben ik
  <app type="empty">
    <rdg wit="#A2"> weer</rdg>
    <rdg wit="#B" />
  </app> helemaal opgekikkerd.
  <app type="empty">
    <rdg wit="#A2"> Mijn schrijfwerk, het fijnste wat ik bezit, schiet goed op.
    <milestone unit="p" /></rdg>
    <rdg wit="#B" />
  </app>
  <app type="related">
    <rdg wit="#A2">Vanochtend, (Donderdag) toen ik weer naar de zolder ging,</rdg>
    <rdg wit="#B"> lk ga haast elke ochtend naar de zolder</rdg>
  </app>

```

Figure 9: The variant apparatus. The type attribute in the XML corresponds to the levels mentioned in the text. Type 'similar' corresponds to level 'agreeing', type 'related' corresponds to level 'same meaning'.

representation of the differences and agreements in individual passages (Figure 11). Clicking the buttons here brings to the fore the corresponding text passages with the selected level of agreement on both sides of the screen. The various views facilitate studying the intricate connections between the different texts and gaining insights into Frank's creative writing process and her development as an author. Apart from this tool, the edition's introductory material also contains a long essay where Elli Bleeker discusses various aspects of Anne's development as a writer.

This essay, along with the other XML files representing the original content, the editorial enrichments, and the user interface, is then processed by a software pipeline described in the next section.

9 Software pipeline

To display the texts and data on the website, the XML files are pre-processed and converted into three different formats: JSON, XML and plain text (text without further structuring or formatting). JSON is used to store and display structured data, such as metadata, people, locations and themes. XML is used for text, such as the diary letters and introductions. Plain text is used to populate the full-text search index.

The pre-processing step is performed by a Node.js script. Puppeteer, a headless browser environment running on the server, is used for part of the extraction. This allows for intelligent handling of the XML using the Document Object Model (DOM) tree. The functionality is used to extract metadata and entities (persons, locations, et cetera) from the XML and to make (minor) adjustments to the XML for the purpose of visualization on the website.

After the pre-processing step, another script is run that builds an index for searching using Elasticsearch. The search index is the basis for the faceted search on the website.

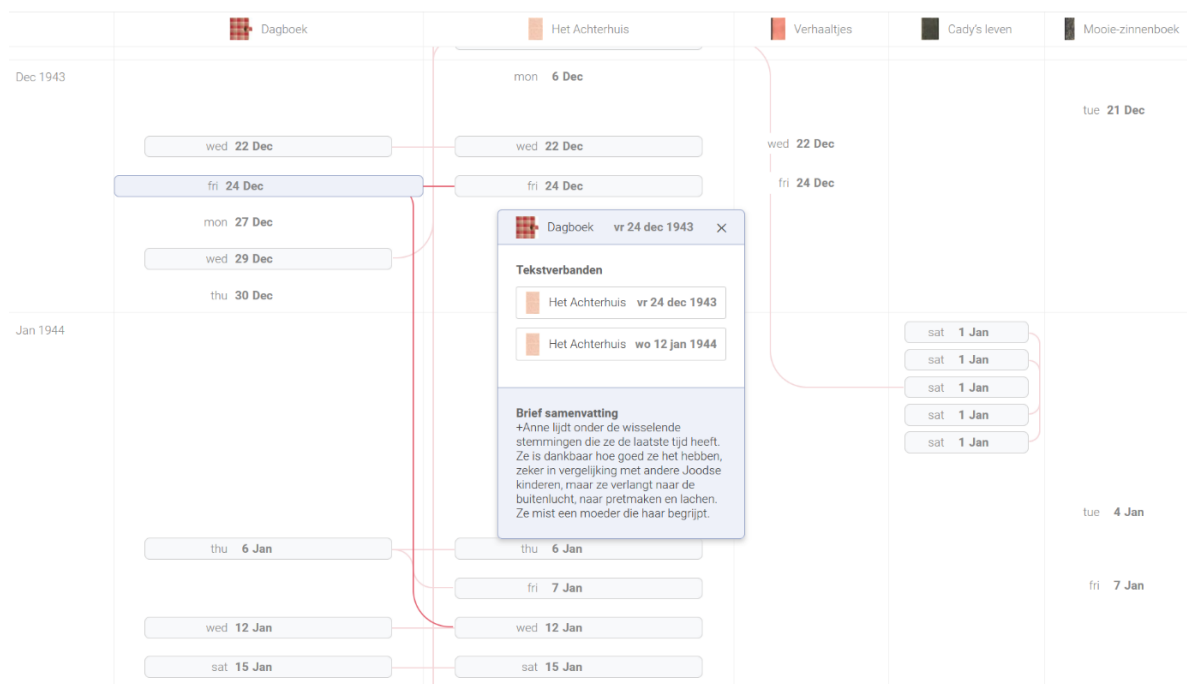


Figure 10: Overview of textual relations between the manuscripts.

Sinds gisteren is het buiten heerlijk weer en ben ik helemaal opgekikkerd. Ik ga haast elke ochtend naar de zolder om de bedompte kamerlucht uit m'n longen te laten waaien; vanuit m'n lievelingsplekje, de grond, keek ik naar de blauwe hemel, de kale kastanjeboom aan wiens takken kleine druppeltjes schitterden, naar de meeuwen en de andere vogels, die in hun scheervlucht wel van zilver leken en dat alles ontroerde en pakte ons alle twee, zodat we niet meer konden spreken. Hij stond met z'n hoofd tegen een dikke balk aangeleund, ik zat, we ademden de lucht in, keken naar buiten en voelden dat dit niets was om met woorden te onderbreken. We keken heel lang naar buiten en toen hij hout moest gaan hakken, wist ik dat hij een fijne kerel is.

Hij klom de vlieringtrap op, ik volgde en gedurende het kwartier dat hij hout

Geen relatie

Dezelfde betekenis

Overeenkomstig

Vrijwel identiek

Ongewijzigd

Verplaatst

Sluiten

Liefste Kitty,

Sinds gisteren is het weer heerlijk en ben ik weer helemaal opgekikkerd. Mijn schrijfwerk, het fijnste wat ik bezit, schiet goed op.

Vanochtend, (Donderdag) toen ik weer naar de zolder ging, was Peter aan het opruimen. Al gauw was hij klaar en terwijl ik op m'n lievelingsplekje op de grond ging zitten, kwam hij ook. We keken alle twee naar het heerlijke blauw van de hemel, naar de kale kastanjeboom aan wiens takken kleine druppeltjes schitterden, naar de meeuwen en andere vogels, die in de zon wel van zilver leken en dat alles ontroerde en pakte ons alle twee zo aan, dat we niet meer konden spreken. Hij stond en ik zat, we ademden de lucht in, keken naar buiten en voelden alle twee dat dit niets was om te onderbreken. We zaten lang zo stil bijelkander en ik durfde maar niet te beginnen, want zonder overgang kan ik toch niet over de allerintiemste dingen gaan spreken. Toen ik daar zat en ik duidelijk zag dat ook hij nadacht over alles, toen wist ik dat hij een fijne kerel is en een schat.

Figure 11: Representation of the agreements and differences between related passages from the diary and the novel.

The site's user interface (UI) uses React, a library for building user interfaces. With React, the UI can be built from reusable parts (components). Components represent a small part of the site's functionality, with associated data and interaction. CSS (Cascading Style Sheets) is used to design the site, using the Stylus pre-processor. Stylus allows the hierarchical definition of styles as well as the use of variables.

The website displays high-resolution images,²⁰ including facsimiles of the manuscripts and the map of Amsterdam. To provide the images in a user-friendly manner, we use IIIF and OpenSeadragon. IIIF (International Image Interoperability Framework) is a widely used framework for image retrieval. An IIIF service runs on the website's server. OpenSeadragon can display large images in sections and is responsible for zooming in and out. To do this, the images are requested from the IIIF service.

10 Copyrights and site access

In September 2021 the Anne Frank Manuscripts-site went live, with only Dutch transcriptions and commentary for now. As mentioned in the introduction, the site is published by a Belgian society and hosted on Belgian servers. This is due to the complicated copyright situation around Frank's manuscripts. Otto Frank bequeathed the manuscripts but not the copyrights on Anne's texts to the Dutch state. While in many jurisdictions that copyright expired seventy years after Anne Frank's death (2016), in other jurisdictions, among which the Netherlands, copyright on (parts of) her work is still protected.

For these reasons, the site cannot allow access to visitors from countries where Frank's works are still under copyright. For that purpose, the site employs a double line of defence. First, the site denies access to visitors whose IP addresses indicate they come from an area where Frank's works are copyrighted. Second, those who come from an area where Frank's works are in the public domain have to explicitly declare that they do so before they are granted entry.

All explanatory and background texts, such as the notes, introductions, biographies and essays, are available under a Creative Commons license (CC-BY). The complicated copyright situation, however, implies that what the users can do with the site's primary content, Anne's text and the images of the manuscripts, depends on their jurisdiction. In this respect we can offer no advice. It also means that we have to restrict access to the material that the edition is built on, such as the transcription XML, the page images, and access to the IIIF server. Further information is available on the site's colophon.

11 Conclusion

In this article we presented the the online edition of Anne Frank's manuscripts. We saw how Frank "mixed" her texts and how these were "remixed" in many printed editions. The online edition "unmixes" the texts and provides a sound basis for the study of Frank's manuscripts, and therefore also of her development as a writer. We described how the edition's functionality was designed to support researchers in their study of the manuscripts, their context, and their textual relations.

²⁰ The diary was digitised in 2001 according to the then current guidelines of *Metamorfoze*, the Dutch national programme for the conservation of paper heritage. See <https://www.metamorfoze.nl>.

Apart from providing a helpful environment to researchers, this edition is also making the argument that for scholarly digital editions, the affordances of the digital environment have become essential. The deeply hyperlinked character of the edition, the various presentation formats, the embedded facilities for text comparison, the text highlighting based on research interest, and the deep integration between the texts and various levels of contextualizing information provide the edition with an underlying data structure and turn it into something that print cannot even approximate.

References

- Elli Bleeker, Bram Buitendijk, and Ronald Haentjens Dekker. From graveyard to graph: Visualisation of textual collation in a digital paradigm. *International Journal of Digital Humanities*, 1:141–163, 2019.
- Anne Frank. *Het Achterhuis*. Contact, 1947.
- Anne Frank. *De dagboeken van Anne Frank*. NIOD, 1986. Edited by David Barnouw, Harry Paape, and Gerrold van der Stroom.
- Anne Frank. *Het Achterhuis. Dagboekbrieven 14 Juni 1942 - 1 Augustus 1944*. Bert Bakker, 1991. Edited by Otto Frank en Mirjam Pressler.
- Anne Frank. *The Diary Of A Young Girl : The Definitive Edition*. Doubleday, 1995. Edited by Otto Frank en Mirjam Pressler.
- Anne Frank. *Manuscripten*. Vereniging voor Onderzoek en Ontsluiting van Historische Teksten, 2021. Edited by Peter de Bruijn, with the cooperation of Elli Bleeker and Marielle Scherer. Under the auspices of the Huygens Institute for History and Culture of the Netherlands (KNAW) and the Anne Frank House. Online available on <https://annefrankmanuscripten.org> in all countries where copyright laws permit it.
- Ronald Haentjens Dekker and David Birnbaum. Collatex python v2.2, 2018. URL <https://pypi.org/project/collatex/>.
- Huygens Institute. For the first time all anne frank’s manuscripts digitised, 2021. URL <https://www.huygens.knaw.nl/en/anne-franks-digitised-manuscripts-available-in-their-entirety-for-the-first-time/>. Press release.
- Patrick Sahle. A catalog of digital scholarly editions. v.4.085, 2023. URL <https://www.digitale-edition.de/>.
- Elena Spadini. Exercises in modelling: textual variants. *International Journal of Digital Humanities*, 1:289–307, 2019.
- TEI Consortium. TEI P5: Guidelines for electronic text encoding and interchange. 2020. URL <http://www.tei-c.org/Guidelines/P5/>. Version 4.5.0. Last modified 25th October 2022.