Two Experiments and a New Model for the Digital Processing of Elisa Chimenti's Corpus (Naples 1883 - Tangier 1969)

Ada Desideri, PhD Student, *Sorbonne Université*, Bianca Vallarano, PhD Student, *Università di Napoli L'Orientale/Université de Lille*

***Abstract*—This paper presents two digitisation workflows and a new model proposition for the digital processing of unpublished texts by Elisa Chimenti (1883-1969). The author's corpus is extremely varied and interesting both from the point of view of academic research and for the wider public. The experiments therefore move in the area of open access digital scholarly edition. The aim is to compare the different text processing chains, in order to reflect on the pros and cons of each. Further, the new proposition aims to build a model for the digitisation of this corpus, also focusing on ethical and ecological issues that contemporary society confronts us with. The purpose is to contribute to the ongoing project of organising, preserving, and digitising the heritage of Elisa Chimenti, an extraordinary as well as forgotten author.**

***Index Terms*—Digital Scholarly Edition, Digitisation workflows, Domain Specific Language, Elisa Chimenti, XML-TEI, Text Encoding, Minimal computing, Ethics**

I Introduction

This paper presents the preliminary proceedings of a research which aims to build a model for the digital edition of Elisa Chimenti's work. We therefore discuss two different digitisation workflows we developed in 2022, trying to point out the pros and cons of each experiment. Subsequently, we propose a new model recently developed from the results of the analysis of the two experiments. All experiments aim at a scholarly open access digital edition of unpublished contemporary texts. In the first part, analysing the two workflows provides an opportunity to reflect on the broader issue of editing unpublished twentieth-century typewritten texts under copyright. As a result, we will address the issues of transcription and OCR software, the still unmodelled status of digital author philology, copyright, software ownership, ethics and sustainability of digital projects. In the second part, building on the previous reflections, we try to define a digital edition model that respects our constraints and is at the same time ecological, ethical, and sustainable.

The new workflow proposes a model which is still experimental but aims to become definitive. The aim of our intervention is to contribute to the wider discussion on digitisation workflows, tools, and infrastructures, with methodological reflections on some techniques in the field of digitisation. We will tackle this topic by experimenting on text editions, in the context of the ongoing digitisation project of the Elisa Chimenti Fund.

II Elisa Chimenti's written heritage: why a digital scholarly edition?

II.A Elisa Chimenti

Elisa Chimenti (Naples 1883-Tangier 1969) was an Italian-born writer, journalist and teacher who grew up and lived in the Maghreb. Her work focuses on the dialogue between cultures, languages, and religions from one side of the Mediterranean to the other, with a particular attention on the condition of women [6] [11] [12] [13]. Her approach promotes understanding and respect for every human and non-human being; she criticises the behaviour of European people and institutions towards Maghrebian cultures [47]. These represent pressing issues in the present context of repression against people on the move in the Mediterranean area. Whitin academic research, we furthermore can notice an increasing consideration for the cultural hybridization processes, whose Elisa Chimenti represents a peculiar example [9].

Chimenti published only five works during her lifetime: *Èves marocaines* (Tangier 1935), *Chants de femmes arabes* (Paris 1942), *Au cœur du harem* (Paris 1958), *Légendes marocaines*, (Paris 1959), *Le sortilège et autres contes séphardites* (Tangier 1964). Three works were translated: *Tales and legends of Morocco* (New York 1965), *Al cuore dell'harem* (Rome 2000), *Cuentos del Marueccos espagnol* (Madrid 2003). The works were recently collected in an anthology (Mohammedia/Casablanca 2009) [15]. Furthermore, two other texts were supposedly published by the author in her youth in

This paper was submitted on the 3rd of march, 2024.

The research was carried out by the authors within two research groups: an International Associated Laboratory established between the University of Rome La Sapienza and the University of Lille, about the writing of women's exile in the Mediterranean region (XIX-XX century), coordinated by Silvia Tatti and Camilla Cederna (2018-2022); and the "Archives of women's writing in exile" project (ALEEF, *Archives de l'écriture de l'exil au féminin,* 2023-2024), established at the University of Lille by Camilla Cederna and funded by MESHS (*Maison européenne des sciences de l'homme et de la société*, UAR 3185 of CNRS).

The article was conceived and developed by the authors in close collaboration, from the beginning till the end. The authorship of the different sections is specified below. Bianca Vallarano is responsible for Sections I, II, III.B, IV. Ada Desideri is responsible for Sections III.A, V, VI, VII. The two authors are cowriters of Section VIII.

Ada Desideri is currently PhD student at the Department of Italian Studies at Sorbonne University, Paris, France (ada.desideri@sorbonne-universite.fr).

Bianca Vallarano is currently PhD student at the Department of Literary, Linguistic and Comparative Studies at the University of Naples L'Orientale, Italy and at the Department of Roman Studies at the University of Lille, France (b.vallarano@unior.it).

Germany, but no evidence remains (probably lost during the First World War): *Meine Lieder* (Leipzig 1911), *Taitouma* (Leipzig 1913). She left an extremely large corpus of unpublished works, including novels, short stories, essays, and poems (more than 30 works). The corpus is preserved today in Tangier, Morocco [50]. An initial systematisation of the archive was carried out in the 1990s by Maria Pia Tamburlini. Since then, new documents have been found, and the work has been resumed thanks to several projects carried up by Camilla Cederna (Université de Lille) [57].

At present, the preservation conditions of the documents are precarious and access to them is very difficult. Most of them were photographed between 2018 and 2022 with personal instruments by Camilla Cederna (helped by two other researchers) who shared the images with other colleagues to ensure their preservation and allow academic research. Professional scanning was planned but could not be completed due to the inaccessibility of the fund. The images available are therefore of rather poor quality, which complicates - as we shall see - the digitisation process.

II.B A Digital edition of Chimenti's corpus

As the texts are still subject to copyright, digitisation initially focused on producing critical editions that would remain in an academic environment, solely to facilitate scholarly research on the texts [20]. Nevertheless, open access publication is important to enable the circulation and study of this heritage, which is still mostly unpublished and difficult to access, both for security reasons (documents are very fragile) and material conditions of access to the fund (the archive is not accessible, and documents are not fully archived) [14]. Currently, a cataloguing and metadata project is being undertaken with the "Archives Chimenti" project [50], promoted by the UMR Thalim (CNRS-ENS-Sorbonne Nouvelle) [57] in the EMAN digital library [52]. EMAN (*Éditions de manuscrits et d'archives numériques*) is a digital publication platform for the dissemination of manuscripts and modern archival collections. The project started in January 2023, coordinated by Camilla Cederna and carried out by her and Bianca Vallarano, with the IT support of Richard Walter (engineer at CNRS/ENS Paris). The purpose is to make available the content of the corpus, even if the texts cannot be at present accessible for the public. This can increase the interest in Elisa Chimenti by academics and other actors in publishing and in the cultural world. For those who are already working on this writer, access to metadata makes it possible to contextualise the examined object within the whole work. This is particularly interesting in this corpus, which is characterised by a dense network of cross-references and intertextual references [13].

Elisa Chimenti's corpus is extremely diverse, both in terms of genre and content of works. The author is particularly interested in the experiences of women and the intertwining of the natural and supernatural worlds, spiritualism, and societal issues, all in the Mediterranean context of exchanges between Southern Europe and Maghreb. In her corpus we can find essays on Islamic and pre-Islamic beliefs (*Les genies*, *La persistence du culte des arbres et des plantes*, *Le culte de l'eau*) as well as adventure novels of women who crossed the Mediterranean basin towards Morocco (*Khadija de l'île sarde*, *À la limite de l'ombre)*. She narrates the encounters with different societies and cultures, that the protagonists learn about, and finally embrace (*Maria Grazia et le génie*). At the fantastical extreme of this 'meeting of cultures' strand, there are actual love stories between humans and non-humans - djinns, dolls… - (*Kheira et la djenia, La tesouira et le fils du sultan…*), as well as proper science fiction texts (*Une étrange aventure)*. Another highly represented genre are the collections of short stories, where we can identify two different strands. On the one hand, tales about the daily life of Moroccan women (*La veillée du harem, Récits des cours et des terrasses, Contes berbères*). On the other hand, tales about beliefs and myths such as etiological legends and morals examples (*Contes et légendes, Légendes marocaines*). Also significant are the poetry collections (*Chants du Maghrib, Harmonies*) and specifically love poems collections (*Marra*). Further, it is important to mention the collection *Miettes de pensées*, where we find aphorisms on contemporary society, culture, and politics.

As far as the state of preservation of these texts is concerned, the situation is also varied. Whereas we can find clean, ready-to-publish texts, there are also several documents with minor to major rewrites, corrections, and variants. At one end of the spectrum, we can see the example of the *Contes berbères*, a collection of three typewritten short-stories full of handwritten corrections both in the line spacing, on the sides of the text and on the verso of the pages. Moreover, the collection is in a precarious state of preservation: the pages are extremely damaged by humidity and time. This kind of texts must be prioritised in digitisation. As we will see, they have already been the subject of some master's works of students at the University of Lille (Pauline Modolo). At the other end of the spectrum, we have the example of well preserved and cleans typewritten texts without any handwritten corrections (such as *La veillée du harem*). However, even these documents, although in a decent condition at present, are in danger of deteriorating in the future, if not well preserved.

Therefore, digitisation process is urgent, and it also changes depending on the types of texts. Regarding texts with a rich apparatus of variants, specific considerations will have to be made in the field of authorship philology. The issue of modelling heterogeneity in author philology is as open as is the discussion around the standardisation of an encoding model and set of scalable visualisation software.

Consequently, the choice of digital seems to have several advantages:

1. The possibility of preserving and studying these texts in an academic context.
2. The possibility of relating recurring themes by considering many texts, even apparently very different from each other, by the means of the creation of new paths and links.
3. The possibility of carrying out linguistic research.
4. The possibility of making the texts accessible on the Web.

Moreover, from the digital-humanities point of view, the

access to such a varied corpus and the great openness in the possibilities of experimentation represent a very fruitful framework for testing and formalising old and new workflows for the creation of digital scholarly editions.

For this paper, we will consider novels and short story collections. The texts subject of this work both have minor corrections and authorial variations. This is the case for a large part of the documents in the Chimenti corpus, and for this reason the two texts can be illustrative (even if not exhaustive of the different textual situations) and serve as a model.

III Two digitisation workflows for a digital scholarly edition of Elisa Chimenti's work

In this section, we will detail two workflows that have been applied to three Elisa Chimenti's unpublished texts. We name the workflows with the transcription format followed by the visualisation format.

1. The WORD-HTML workflow is based on text editors and XSLT transformations. It was used for digitising two texts: a short story of the collection *Contes Berbères* (*Mennouch la Rifaine*) [3] and an excerpt of about 50 pages of a science-fiction novel (*Une étrange aventure*) [4]. They are both unique copies.
2. The TXT-EVT workflow is based on TXT, XML-TEI [45] and EVT visualisation technology [39]. It was used to digitise a short story of the collection *Contes d'autrefois* (*Histoire du prince Moustapha et de la princesse Djamila*) [1]. The short story is also included in another collection, *La veillée du harem* [2], in fine copy.

Gitlab hosts the second WORD-HTML edition (*Une étrange aventure*) and the TXT-EVT edition [51]. For research purposes, they can be shared privately upon request. Regarding the manuscripts subject of the editions, unfortunately we do not have the right to share or publish the images, for copyright reasons.

III.A The WORD-HTML workflow

The firsts digital scholarly edition produced in the digitisation project of the Elisa Chimenti Fund were created in the context of Pauline Modolo's and Ada Desideri's master thesis at the University of Lille and the University of Pisa in 2022 [20]. The two texts analysed are very different from each other, both in terms of content, form, and condition of the document. Different choices were therefore made regarding the phenomena to be processed and certain aspects of the final graphic rendering. Indeed, *Mennouch la Rifaine* (n.d.) is a love short story, set during the Rif War and features numerous interventions and handwritten notes by the author. *Une étrange aventure* (n.d.), on the other hand, is in a ready-to-print version, almost without manuscript interventions, and presents a very peculiar hybridisation between supernatural aspects of Western and Moroccan cultures. An Italian version of *Une étrange aventure* exists, translated by an unknown person (L. Morsello). We didn't take it into account for establishing this edition.

The workflow applied remains the same in both cases and the work was carried out in close collaboration with Pauline Modolo. The text was initially transcribed, starting from the images of the documents, in Word format, applying paragraph and character styles. Then the XML file that allows the graphical rendering in Word was extracted. More precisely, we exported the Word file into an ODT file. We then changed the extension to ZIP and extracted the content.xml file. Afterwards, it was transformed into a second XML file, via an XSLT stylesheet produced by Nathalie Gasiglia and Luc Audrain (University of Lille). In this file, only the tags associated with the character and paragraph styles added during transcription appear. This allowed us to generate an XML-TEI file and then the HTML pages that constitute the edition interface, with further XSLT transformations (a scheme in Table I).

For the novel *Une étrange aventure*, we first tried to transcribe the text using tesseract, an open OCR software [31]. The results were not satisfactory, considering the regular form of the text, and we therefore opted for manual transcription (already chosen for *Mennouch la Rifaine*). In fact, due to the low quality of the images (less than 300 dpi) each page needs special contrast and colour adjustment to maximise the OCR result. Spending this time on manual transcription seemed to be preferable, rather than testing page by page for the best result. As a matter of fact, this allowed a closer approach to the text, which is suitable for a critical edition work. At present, due to the presence of about fifty manually transcribed pages, we can consider training OCR for the transcription of the nearly 200 remaining pages.

During transcription, errors that could be considered typing errors (inversion, addition, or missing letters, etc.) or due to the typewriter (missing accents and other symbols) were corrected. Capital letters and punctuation were also normalised.

The application of styles in Word (or in OpenOffice) is standardised in a common guide and concerns both structural phenomena (e.g., paragraphs, corrections or parts of text that cannot be read) and analytical phenomena (e.g., foreign words, passages relating to selected themes or to quotations). The conception of this guide constitutes a preliminary work of analysis of Elisa Chimenti's texts as a whole, coordinated by Camilla Cederna and Nathalie Gasiglia and in which several students participated. Actually, this guide features the encoding choices.

The annotation in this format has the advantage of being able to be created and verified by people unfamiliar with XML-based languages, but it generates several irregularities that render some of the subsequent XSLT transformations partially ineffective. In fact, these transformations required some special arrangements to achieve the desired result. We also had to intervene manually in order to correct certain errors and insert data; therefore, any integration of changes at the transcription level would not be fully automated at present.

Finally, the interface is functional but not optimal. It is a static page created independently by adapting numerous templates provided by W3C. Being a unique, typewritten, ready-to-print copy, only some images of the manuscript can be displayed, on a different page from the text. In general, priority was given to the development of functionalities for analysing

the text, based on critical notes and annotations of words and passages relating to significant themes.

This workflow allows to maximise the customisation of the interface. Indeed, some improvements have already been done from the first version in order to enhance user experience. However, modifications are not easy to handle, in particular for students and researchers who are not familiar with computer programming. So, even if we decided to start with an easy format for non-computer scientists, we finally obtain a complicated process that does not allow them to directly take charge of the modifications.

This workflow is particularly suited for exploring interfaces and hypertext fitted on the corpus. However, it becomes efficient only if it succeeds in setting up a collaboration between developers and computer scientists on one side, and researchers and students in literature on the other side, within a transdisciplinary frame. Both humanists and computer scientists are deeply interdependent for the realisation of this kind of edition.

III.B The TXT-EVT workflow

A second edition was developed in 2022, in the context of a master's work at the University of Siena, carried out by Bianca Vallarano with the supervision of Roberto Rosselli del Turco (University of Turin). The text chosen for the edition was the *Histoire du prince Moustapha et de la princesse Djamila* in its version annotated by Chimenti, contained in the collection *Contes d'autrefois* (n.d.). A number of variants - albeit not numerous - are present in the typescript of the selected text, which makes the work a matter of genetic philology as well. Another corrected version of this text exists in fine copy, in the collection *La veillée du harem* (n.d.). This case allowed us to reflect on a text whose genetic stages are explicit. We were thus able to focus on how to markup and display the genesis of the text (the erasures, additions, and substitutions) in the visualisation.

In this workflow, XML format, TEI standard [55], and EVT 2js [49] visualisation tool were used. In the first instance, the text was manually transcribed from the images of the typewritten pages in TXT format (plain text file). Human transcription was chosen because it was difficult to find open access OCR software with an acceptable text output due to the poor-quality images and damaged text in the typewritten documents, as well as due to the multilingualism (French and Moroccan Arabic) of the text. OCR software is optimised to receive only one language as input and generates a lot of errors in multilingual contexts. We did not find an open OCR software that could be set to multiple languages simultaneously; thus, setting it to French always resulted in the Arabic text being incorrect. Added to this, at a further level, is the problematic nature of the languages handled. Such software is mostly trained on the written languages and more on the politically majority Western languages. Finally, even an OCR trained on the Moroccan *darija* would not solve this problem, as Arabic words are transliterated by Chimenti into the Latin alphabet in a heterogeneous manner. We are therefore certain that words in foreign languages and especially those in Arabic and Moroccan *darija* - which make up an important percentage of Chimenti's lexicon - would not be transcribed correctly. Added to this are the considerations made for the previous workflow, i.e., that manual transcription allows a closer approach to the text, which is more useful for a critical work.

As far as the edition criteria are concerned, since these are preserved originals, we excluded a critical edition in the sense of reconstruction of the original [39]. We can thus define a genetic-interpretative edition [19] [22], which implies an accounting of authorship variants and a standardisation of formal variants (graphic, phonetic, morphological), e.g., the transliteration of Arabic terms into the Latin alphabet. Since the chosen text has handwritten notes and corrections, the base text is the corrected text - i.e., considering the author's corrections - with notes that account for the genetic aspects of those corrections. Typographical errors have been corrected, punctuation have been normalised and - where necessary - introduced, and capital letters have been added as needed.

The text was copied in an XML file and marked up according to TEI standards. Regarding the encoding choices, TEI was used to mark up the following (cfr. Table II).

1 The arborescence of the text:
   - Titles of the novels.
   - Paragraphs.
   - Direct discourses.
   - Verses (in the text, prose and poetry intersect and there are frequent inserts of poems and folk songs).
2 The annotations for exploring the text and create lists of entities:
   - Person's names (to navigate within the occurrences of the characters).
   - Place's names (to navigate the places cited in the text; this is relevant in all the Chimenti corpus, since it often focuses on exile and travel experiences. In the case of this particular short story, it is extremely relevant because the story focus on the women's exit from the harem, and her journey into the outside world).
   - Foreign languages (from the list of languages the user can see the occurrences of each language and click on the folios in which the language is used; this is relevant since Chimenti's texts are multilingual, as mentioned).
3 The apparatuses:
   - basic apparatus (additions, deletions, substitutions, author's notes, editor's notes)
   - critical apparatus (witnesses, lemmas, readings, lacunas).

The apparatuses were encoded using the TEI parallel segmentation method, according to EVT. In addition to being the one supported by EVT, the parallel segmentation method is also the most functional: it is more precise than the location-referenced method and is more intuitive than the double-end-point-attached method. Although the latter is the most effective, it is to date difficult for visualisation software to handle.

Concerning the visualisation, EVT provides a dynamic interface. To create it, EVT needs the XML-TEI file, the configuration file in JSON format, the CSS file to customise the visualisation, and the images (in JPG format) associated with

the text, for the facsimile edition (if needed). We didn't edit the CSS file. Regarding the JSON file, we chose the display modes (text-image, reading text and collation), the editions to show (critical and interpretative) and set the lists of entities. As EVT is organised, the "critical edition" mode shows the final corrected text, whereas the "interpretative" mode highlights the genesis (deletions, additions).

IV Preliminary critical assessments

IV.A Critical assessments of the two outputs

We can schematise the four steps (transcription, annotation, transformation, visualisation) of the two workflows in a comparison table (Table III).

The two products are both html pages and the main pages contains the text, which is relatively interactive. In the WORD-HTML experiment, all terms linked to a specific thematic path can light up within the text: characters, places, references, animals, plants, food, religion/magic, foreign languages, glosses, exile/travel, women. Notes may be read both in a dynamic and in a static visualisation. In the TXT-EVT experiment, genetic variants, names, and places can light up within the text. In both experiments there is the possibility of displaying thematic lists, but in the first one they are obviously more various. In the WORD-HTML experiment, the user can switch from the main page via a menu to view the images of the original, the glossary and the transcription criteria. A diplomatic transcription is also available, where one can visualise structural phenomena, such as corrections. In the TXT-EVT experiment, on the other hand, via the top menu the user can switch to the different views of the same text: the text and the image of the manuscript next to each other, the reading-text, the collation of the witnesses. Via a menu, the user can access information on the table of contents and on the lists of named entities, and find out more about the project. Also, via an "option" button you can download the XML, switch languages, and learn more about EVT.

Concerning the markup, both the experiments have their limits. In the WORD-HTML experiment, the vocabulary is extremely wide, but sometimes irrelevant and insufficient for our objectives. On the one hand, the pre-established coding was extremely detailed regarding the types of typescript corrections, which ultimately did not provide interesting and useful details for the analysis we had set. On the other hand, it did not include aspects that would have been useful and pertinent, such as an annotation indicating the equivalence between the different forms of the same name/word. As mentioned, the case of graphic variants is nevertheless extremely relevant in the corpus, given the continuous transliteration between the Arabic and Latin alphabets, as well as the relative instability of the written form of the Moroccan *darija*. Moreover, due to the different forms of the two texts to which the WORD-HTML workflow was applied, the same coding elements were in some cases used in different ways (such as notes, used in one case for author's notes and in the other for editor's notes).

In the TXT-EVT experiment, the markup was limited in part by EVT and in part by the TEI. Compulsory tag choices had to be made in order to respect EVT's vocabulary, whereas different tags would have been more suitable. For instance, foreign language terms had to be marked with the <term ref> tag instead of <foreign xml:lang>, in order to construct a list of entities. Furthermore, the TEI vocabulary itself posed limitations. For example, it does not allow the <said who> tags to be nested within each other, and thus preventing us from marking the narrative mechanism of the story within the story. The expedient is nevertheless very significant in the Chimenti corpus (the model is *The Thousand and One Nights*).

Among the two experiments of digital scholarly editions, the WORD-HTML one is more labour intensive and requires the availability of a large working group, including not only humanists, but also people with IT skills, especially for writing XSLT transformation sheets and creating the website. This process allows for the construction of digital scholarly editions that are rich in terms of hypertextual paths and leaves the field open for experimentation, but it is also time-consuming. Although, in theory, this workflow could only be carried out by a digital humanist, it is actually a work that has a strong collective vocation. The importance of a digital humanist, therefore, lies above all in facilitating communication between experts from different fields.

The TXT-EVT experiment, on the other hand, involves a simpler workflow, which can be realised by digital humanists with transversal skills in humanistic computing and basic knowledge of markup languages, moving along a path already mapped out as far as the programming part is concerned, thanks to the EVT software (which is easy to understand and for which there is an agile online manual). On the other hand, EVT, by making the work simpler, also runs the risk of limiting it, and does not allow the experimentation that is possible on a newly created HTML page. That is, unless you work directly on the code. However, it is a workflow designed to be carried out by people without such IT skills. This edition is probably more appropriate in the case of texts in which there is an important critical apparatus and in which it may be interesting to show images of the originals (e.g., for the *Contes berbères*, rich in manuscript variants).

In the end, the EVT tool is optimal for apparatus visualisation as it was developed specifically for critical editions with multiple levels of apparatuses. The HTML edition, instead, is more suitable if you want to give priority to different hypertextual paths.

IV.B Overall critical assessments

A number of issues remain unresolved. From the point of view of tools, the problem persists, as some of the software used or that could have been used - Word, Oxygen, OCR software - are proprietary and require a licence, which should not be taken for granted. On the other hand, open tools are totally available but are often difficult to handle. The issue of transcription is illustrative. In the first workflow, where an open OCR software was used and the failure was mostly due to the image quality, a long time was spent on setting and learning how to use the open software. The result was greatly time consuming. In addition, whereas using open software is not enough to ensure the full

sustainability of a project, it still increases it, thanks to available documentation about format and use.

A further aspect that remains open is the tag model for marking up aspects of author philology. In the TXT-EVT experiment, the proposed markup combines elements of the standard TEI with elements of the Critical apparatus module, in an attempt to render the two levels of criticism and genetics concisely on the text. As far as the philological aspect is concerned, the TEI markup model may be considered sufficiently detached from the visualisation software. However, it was not completely satisfactory, and proved to be limiting, as mentioned.

Another limitation of these models is that, in terms of graphical rendering, they are closely linked to the final visualisation program.

Regarding the time required to create editions, the reflection must consider the three different activities that have to be performed: the conception of the edition, the transcription and the annotation of the text, and the generation of the files that constitute the edition. The time required for the conception of the edition is difficult to assess *a priori*, because it is experimental work, which also depends on the flexibility and customisation potential of the chosen medium. The time for transcription and annotation is also very variable, as it depends on the content and form of the text, its material conditions, and the possible mark-up. In the two workflows presented, the former requires more time for transcription and conception than the latter, but quantification is very difficult since the work was carried out on a long period of time and not on a full-time basis. Instead, we can compare with more certainty the creation time of the files for the edition: once the editorial choices were made and the text was transcribed and annotated in plain text format, the creation of the files for EVT took about two weeks. The time to create the XSLT transformation files and correct the output for the first workflow presented is almost twice as long: about a month. In both cases, this is considered part-time work.

In terms of reusability, the two projects are very different. EVT provides a tool that can be used relatively easily: creating an edition with this tool can be reproduced simply by observing at how it has been used. The configuration file is set to work at a general level with texts that have author variants (even minor) and/or multiple witnesses (as is the case with most of Chimenti's corpus). Small changes are, however, extremely easy to make. Conversely, the first experiment presented can, in theory, be reproduced simply by applying XSLT transformations to the XML files, provided the text is transcribed using the same rules; in practice, the intervention of a computer scientist to correct errors and make the final interface operational is necessary. Furthermore, in this case, the workflow is adapted to Elisa Chimenti's texts, whereas EVT is much more generic.

Concerning the target audience, the WORD-HTML experiment appears more suitable for a broader audience, which may be attracted by a dynamic interface with thematic paths. The TXT-EVT experiment, on the other hand, results in a product that is more useful for the scientific community, as it is more suited to the construction of complex critical editions with multiple levels of apparatus.

At present, the issue of copyright is still open, since Chimenti's heirs own the rights. The reflection on how to address this copyright issue is ongoing. Either way, the texts should be shareable for research purposes. For this reason, we have developed a new model which aims to meet this and the other needs and requirements that have emerged in recent years.

V THE NEW MODEL

V.A Preliminary considerations

As shown, neither of the two workflows presented is particularly well suited to the Chimenti corpus. For this reason, we discussed a new proposition within the community working on Elisa Chimenti. We will test this workflow on the next experiment of digital edition, that will be carried out in May-June 2024 within an internship of a master's student from the University of Lille, Louane Morin. The internship will be directed by Camilla Cederna and the work will be carried out in close cooperation with the team working on Elisa Chimenti.

This proposal aims to take into account the problems that have arisen in the previous works that we presented in the first part of the article. Developing this new workflow, we focused on six particularly urgent aspects:

1. The difficulty of access to documents, on which the poor-quality images of manuscripts depend.
2. The variety of the corpus, both in terms of genre and state of preservation.
3. The restricted conditions of dissemination due to copyright.
4. The team's available resources (human and financial).
5. The OCR issues.
6. The ethic and ecological concerns that must be considered in digital projects.

From a technical point of view, our goal is to use non-proprietary software, to store the transcriptions in lightweight, interoperable formats, and to use technologies that are relatively easy to learn for non-IT specialists. From the point of view of textual analysis, we need to find a new annotation model. We must also consider, as mentioned, the constraints linked to the context of the project in terms of access to documents and copyright, as well as ethical and ecological concerns.

For instance, as we have already mentioned, our working group cannot count on many IT specialists who can ensure a regular presence. This is not due to a lack of will but depends more broadly on the situation of IT support centres and digital humanities centres in universities and in other cultural institutions. In fact, the large number of requests for IT support is not proportional to the services provided by these institutions (this problem, mentioned for example in [39], was also discussed at length at the DH Nord 2023 conference "Humanités numériques et questions d'éthique: débats, enjeux, pratiques", held at the MESHS in Lille on 15, 16 and 17 November 2023). What's more, we've already come up against the lack of IT support (during the first experiment) and the need to get on with our work independently. This does not mean

eliminating IT work from the project, but rather reducing it to specific moments and tasks.

Currently, in addition to the authors of this project, we have the following IT support:

- Members of DiPText-KC CLARIN Helpdesk (the Helpdesk from the CLARIN Knowledge Centre for Digital and Public Textual Scholarship) Federico Boschetti, Angelo Mario Del Grosso, and Franz Fischer are helping us to develop the new model, and in particular advising on the definition of DSLs for annotations.
- Research engineer Valentin De Craene (MESHS, Lille) is helping with the creation and management of the Chimenti project website [57].
- Research engineer Richard Walter (CNRS, ENS Paris), manager of the EMAN platform, is helping on the "Archives Chimenti" digital library project.

In the absence of sufficient long-term funding, and in the absence of an IT specialist who is a full-time member of the working group - and not wanting to give up on this research - we decided to define a workflow that would be proportional and feasible in our situation.

V.B The new workflow

In view of these numerous constraints, we developed a workflow in four steps: transcription, generic structure annotation, standoff critical annotation, and visualisation. We propose to proceed as follows.

1 Transcription. Using photographs of the manuscripts, we propose to transcribe the texts by hand, in txt format. It is not possible to format the text, but the file format is light, interoperable, and open. This makes storage and transmission easier. The transcription criteria involve a modernisation and a standardisation of the paragraphemic signs as well as of the text.

2 Generic structure annotation. Despite the impossibility of formatting texts in txt format, it is nevertheless possible to annotate them using a DSL (Domain-Specific-Language) [5]. We could either use or enhance existing systems or invent one adapted to our context. This annotation should be common to all transcribed texts and should only represent basic structural phenomena: headings and sub-headings, paragraphs, direct discourses, erasures, gaps, additions, dividers. This set of phenomena is defined in advance to avoid successive changes. The advantage of going through this phase is to have uniform, lightweight files for initial storage, before building the edition. What is more, learning such an annotation system is very quick, does not require any specific skills and many open software tools can be used. We decided to use Markdown language for this annotation, choosing basic symbols allowing us to interoperate with many Markdown flavors, in order to visualize text structure using existing software. Markdown language is supported by several applications and components (see The Markdown guide/Tools, in https://www.markdownguide.org). The symbols we choose are interoperable – not exhaustively – with gitLab, gitHub and other git software.

3 Stand-off critical annotation. The structural annotation can then be enhanced by critical annotations. Following a discussion with members of the DiPText-KC CLARIN Helpdesk, we plan to define a DSL for integrating critical notes in stand-off from the structure annotation. This has advantages in terms of work organisation, as this annotation system does not require learning a programming or mark-up language. All you must do is learn the citation system (which will probably be based on a paragraph-sentence pair) to add comments and annotations to the text. This type of work is also similar to the practice of creating a critical apparatus in philology, which may be more familiar to some of the project participants [45]. It is also possible for the two stages (structural annotation and critical annotation) to be carried out by two different people. In addition, separating structural from critical annotation seems appropriate given the wide variety of texts in the corpus. It makes it possible to diversify the critical approach, while retaining a common annotation base. Our aim is to define several annotation possibilities, depending on the kind of text and the goals of the edition. For example, one set could relate to the genetic study of the text, and another to thematic aspects. This would involve classifying texts into a number of categories, in order to standardise annotation and automate visualisation following these categories. This work of text classification has already partially been done by Maria Pia Tamburlini and Camilla Cederna (University of Lille) and should be taken up again to establish the categories according to the objectives of the various editions. Different annotations will clearly imply different visualisations.

4 Visualisation. The final visualisation will depend on the choices made during the critical annotation, the specific editing goals, and the constraints about publishing. The annotation model presented allows us to keep many possibilities open, thanks to the formats used and the diversification of annotations. Since we still need to experiment with new interfaces in order to find the solution best suited to the Chimenti corpus, we do not want to define precisely the results expected from this final stage.

V.C A focus on the OCR issue

The choice of a manual transcription might seem unnecessarily work intensive, since there are plenty of relatively accurate and reliable OCR software. Moreover, given the number of documents, the use of an OCR would certainly be preferable. In this way, transcribers would have to correct the text, rather than actually transcribing it, potentially saving time. However, there are several problems with this.

1 The images are not of a quality that allows good results. In the current situation, as we cannot access the documents, we cannot obtain images of a higher quality. What is more, storing good quality images requires a lot of space and maintenance.
2 Open or free OCRs usually need to be trained to get the best results, and people must learn how to do it. This means a lot of time and specific skills that are not at present available in our working team. Even if we could spend time on training an OCR, the variety of the documents in the corpus would imply newly training the OCR for each set of documents.
3 The difficulties of the OCR with multilingual texts would remain, generating an extremely high number of errors.
4 From a methodological point of view, if the person transcribing is the same person who will be doing the critical annotation (which is usually the case in our working group), transcription is not just mechanical work: it is part of the process of getting closer to the text and to the writer's gesture.
5 From an ethical point of view, if an artificial intelligence model works, it means that people have already corrected its output countless times. These correction processes are never totally transparent, insofar as they may have employed (directly or via platforms or companies – including GAFAM) workers in countries where labour is underpaid. Open OCR software developed in universities in France often employs trainees to correct large volumes of text, and their working conditions are not known either.
6 From an ecological point of view, these systems are highly energy consuming.

All these issues are the reason why we have decided to not invest our time in training an OCR software. Nevertheless, if in the future researchers are interested in working with our team on the development of an OCR for our case, we would clearly be open to consider it.

VI Critical assessments of the new model

At this stage of the project, this workflow has several advantages.

1 It generates a lightweight and uniform storage of transcriptions.
2 It allows us to progress in the digital processing of texts before they are published.
3 Numerous possibilities remain open for publication and/or dissemination.
4 It makes possible to differentiate the critical treatments of texts.
5 The digital tools we selected are relatively easy and quick to learn for people who are unfamiliar with computer science. Computer scientist intervention is required only for the last stages.
6 No proprietary software is strictly necessary.

We can now identify some disadvantages of this workflow, but whose consequences can be partially countered.

1 Critical annotation would be done with a stand-off DSL, probably based on paragraph-sentence citation system. It means that we can't change the paragraphs and sentences organisation after the critical annotation started. However, if we need to add paragraphs, we could define a language with an anchor system, allowing us to add paragraphs or sentences at the end of the transcription but to render them at the right place.
2 It is difficult to modify the generic annotation at a later stage. To put this problem into perspective, it is possible to define several versions of the annotation rules and make these versions interoperable and never obsolete.

VII The challenge of minimal computing: an approach to ethics and eco-sustainability?

VII.A Why minimal computing?

In addition to the technical and methodological issues relating specifically to this project, our choices are guided by ecological and social concerns. The minimal computing perspective – despite its very broad definition [36] – provides concrete avenues for making ecological assessments of the way in which a digital project is conducted. What is particularly interesting is that minimal computing develops from the statement that digital infrastructures are not uniform across the planet, and attempts to provide a political response to this inequality. The proposal of minimal computing focuses on the way we work, rather than on strictly technical issues: « the way in which we model our world […] have direct and sometimes unforeseen effects on the infrastructure needed to realise these visions, and [...] minimal computing can help us to reduce complexity » [34].

More broadly, the aim is not to compete with the GAFAM, but to think about a different model for the digital world, based on maximum appropriability and freedom (about this, see also [34]). Clearly, it means working on a smaller scale.

In the digital humanities, researchers have been working on this topic for a number of years, notably in the GO::DH (Global Outlook::Digital Humanities) working group [52] of the ADHO (Alliance of Digital Humanities Organisations) [45]. In a DHQ (Digital Humanities Quarterly) special issue about minimal computing published in 2022 [51], the editors propose four heuristic questions to guide digital humanities projects towards minimal computing [36].

1 What do we need?
2 What do we have?
3 What must we prioritize?
4 What are we willing to give up?

The new model presented has in fact been built by answering these questions, from an ecological point of view.

1 We need the availability of philologically accurate texts for study purposes. To have this, we need to transcribe, edit, and annotate the texts, and then to preserve and store them in interoperable and lightweight formats.
2 We have a small working group, composed by mostly humanists and students.
3 We prioritise easy shareability of the documents for studying purposes, sustainability of the tools and practices, transparency.

4 We are willing to give up complicated, hyper performative and proprietary software and technologies.

VII.B A step further: social and ethical issues

These heuristic questions enabled us to take ecological issues into consideration, but they are not sufficient to ensure that the responses we provide are social and ethical too. We need to refer to more explicit and binding ethical principles and value systems. The minimal computing approach provides a first step in this direction. Based on the observation of inequalities in digital access, the aim of minimal computing is to choose technologies and techniques that are functional even when access to computer equipment is discontinuous, or when the equipment performs poorly [36]. Taking a minimalist approach also means processing only the essential information, so as not to overload calculations unnecessarily.

We draw broader ethical principles from social ecology [7] [8] and the political theory of degrowth [26]. These two schools of thought see technology as the product of a certain social organisation, which in turn affect the same society. This contrasts with a view of technology as an external entity that influences (or even disrupts) a society, and with a view of technology as the sole parameter for assessing social progress. A degrowth and ecological approach to digital technology therefore considers the multiple consequences that the development of a certain technology could have on society (about the multiple consequences of digitisation see [38]), based on principles of social organisation that we believe to be ecological and egalitarian.

Social ecology provides these kinds of principles. It is based on decentralisation, small-scale organisation, valuing diversity and freedom (for humans and the rest of nature), and a rejection of hierarchy, according to which people can – and must – be directly involved in choices that concern them. Murray Bookchin reflects on the way in which technology can respect the principles of social ecology: "Human dependence on the natural world [must constitute] a visible and living part of human culture. [...] An ecological use of [...] technology could awaken in human beings a sense of their dependence on the environment; [...] reintroducing the natural world into human experience, we can contribute to the fulfilment of the fullness of the human being." [8]

VIII Conclusions: can a digital edition be ecological?

Theories of minimal computing, social ecology and degrowth allowed us to think about digital technologies that are not only less impactful, but genuinely ecological. According to Bookchin [8], for a technology to respect the values of social ecology, it must first be a conscious and creative choice on the part of the community developing or adopting it. Consequently, it must be on a scale that can be appropriated to this community. This aspect is particularly interesting for the digital humanities. Firstly, because technological choices are not self-evident, but on the contrary can be a matter of creativity. Secondly, because of the importance of the reproducibility of research and the appropriation of digital tools for humanities research.

All these concerns guided our choices within the Chimenti project. In our digital publishing work, we valued the creative and interpretative choices made at all levels by the humans who transcribe and encode the information. We sought to strike a balance between making the work as easy as possible with the minimal use of technologies. Lastly, we carefully considered the way in which the tools were produced: we favoured free or open technologies, and avoided, as far as possible, tools whose development relies (or is likely to rely) on subcontracting and on outsourcing work to countries where there is little worker's protection [10].

References

Primary sources:

1 E. Chimenti, "Histoire du prince Moustapha et de la princesse Djamila", in E. Chimenti, *Contes d'autrefois*, unpublished, n. d.
2 E. Chimenti, "Histoire du prince Moustapha et de la princesse Djamila", in E. Chimenti, *La veillée du harem*, unpublished, n d.
3 E. Chimenti, "Mennouch la Rifaine", in E. Chimenti, *Contes berbères*, unpublished, n.d.
4 E. Chimenti, *Une étrange aventure*, unpublished, n. d.

Secondary sources:

5 L. Bambaci, F. Boschetti, and R. Del Gratta, 'Qohelet Euporia: a Domain-Specific Language for the Encoding of the Critical Apparatus'. *International Journal of Information Science and Technology*, 3(5), 26–37, 2019.
6 A. Benchekroun, 'Préface', in *Anthologie*, Mohammedia/Casablanca: Éditions du Sirocco/Senso Unico Éditions, 2009, pp. 7–19.
7 M. Bookchin, *The ecology of freedom*, UK: AK Press, 2005.
8 M. Bookchin, *Vers une technologie libératrice*, Paris: Librairies Parallèle, 1974.
9 D. Budor and W. Geerts, ed., *Le texte hybride*. Paris: Presses Sorbonne Nouvelle, 2004. Accessed: Jul. 02, 2023. [Online]. Available: https://books.openedition.org/psn/10040
10 A. A. Casilli, *En attendant les robots*, Paris: Seuil, 2019.
11 C. M. Cederna, 'Elisa Chimenti : écrivaine en exil, arabophile et antifasciste', OLTREOCEANO, Rivista sulle migrazioni, no. 20, Mémoire coloniale et fractures dans les représentations culturelles d'auteures contemporaines, dir. Catherine Douzou, Alessandra Ferraro et Valeria Sperti, 2022, pp. 171-185, doi: https://doi.org/10.4000/oltreoceano.413.
12 C. M. Cederna, 'Alterità e métissage nella scrittura di esilio di Elisa Chimenti "eterna viaggiatrice nel paese delle chimere" : (Napoli 1883-Tangeri 1969)', in *Scrittrici in esilio tra Otto e Novecento : luoghi, esperienze, narrazioni*, C. Licameli and S. Tatti, eds., Roma: Quodlibet, 2022.
13 C. M. Cederna, 'L'écriture d'exil d'Elisa Chimenti, une mosaïque des voix et de créations des femmes entre métissage et transgression', *Atlante. Revue d'Études romanes*, n. 18, 2023, doi: https://doi-org.ressources-electroniques.univ-lille.fr/10.4000/atlante.28113.
14 C. M. Cederna, A. Desideri, and B. Vallarano, 'Mémoire et reconnaissance: réflexions autour d'une possible édition numérique de l'œuvre d'Elisa Chimenti', to be published.
15 E. Chimenti, *Anthologie*. Mohammedia/Casablanca: Éditions du Sirocco/Senso Unico Éditions, 2009.
16 F. Ciotti, ed., Digital Humanities. Metodi, strumenti, saperi, Roma: Carocci, 2023.
17 A. Contini, 'Il paradigma del *métissage*, fra estetica e scienze umane', *Ricerche di Pedagogia e Didattica*, n. 4, 2, 2022, pp. 1-36.
18 C. D'Agata, '«Le varianti della rosa». Per un prototipo di edizione digitale del *Nome della rosa:* interpretazione, didattica, annotazione', *AIUCD 2022. Proceedings*, 2022, pp. 105-110.
19 P. M. De Biasi, *Génétique des textes*. Paris: CNRS éditions, 2011.
20 A. Desideri, P. Modolo, 'Traitements informatiques des données, d'un traitement de texte à une structuration XML TEI pour la création d'interfaces de consultation en HTML', *Atlante. Revue d'Études romanes*, n. 18, 2023, doi: https://journals.openedition.org/atlante/27665. https://doi.org/10.4000/atlante.27665.
21 M. J. Driscoll and E. Pierazzo, ed., *Digital Scholarly Editing: Theories and Practices.* Cambridge: Open Book Publisher, 2016. Accessed: Feb. 21, 2024. [Online]. Available: https://www.openbookpublishers.com/books/10.11647/obp.0095.

22 A. Grésillon, Éléments de critique génétique: lire les manuscrits modernes (1994). Paris: CNRS éditions, 2016.
23 P. Italia and C. Bonsi, ed., *Edizioni critiche digitali/Digital Critical Editions. Edizioni a confronto/Comparing Editions*. Roma: Sapienza Università Editrice, 2016. Accessed: Feb. 21, 2024. [Online]. Available: https://www.editricesapienza.it/node/7479.
24 P. Italia, 'Filologia d'autore digitale', *Ecdotica*, n. 1, 2019, pp. 203–216.
25 P. Italia, *Editing Duemila*. Roma: Salerno, 2020.
26 F. Jarrige, *On arrête (parfois) le progrès*, Paris: L'échappée, 2022.
27 F. Laplantine and A. Nouss, *Le métissage*. Paris: Flammarion, 1997.
28 T. Mancinelli and E. Pierazzo, *Che cos'è un'edizione scientifica digitale*. Roma: Carocci, 2020.
29 F. Moretti, La letteratura vista da lontano. Torino: Einaudi, 2005.
30 F. Moretti, Falso movimento. La svolta quantitative nello studio della letteratura. Milano: Nottetempo, 2022.
31 C. du Mouza, S. Lamassé, and P. Rygiel. 'De l'histoire numérique à l'histoire données ?', *Les Cahiers de Framespa*, fasc. 42, 2023, doi: https://doi.org/10.4000/framespa.14374.
32 B. Nava, 'Siamo tutti bédieriani? Prospettive per le edizioni genetiche digitali', *Umanistica digitale*, n. 14, 2022, pp. 19-40, doi: http://doi.org/10.6092/issn.2532-8816/14949.
33 E. Pierazzo, Digital Scholarly Editing: Theories, Models and Methods. Routledge, 2014.
34 S. Quinton and J.-B. Stefani. 'Taking conviviality seriously (extended abstract)', 1st conference on Undone Science in Computer Science, Feb 2024, Nantes, France. pp. 1-4. Available: https://hal.science/hal-04448759
35 E. Radio, 'Minimal Computing and Ontologies', Minimal computing, Jul. 21, 2020, Accessed: Feb. 28, 2024. [Online]. Available: https://go-dh.github.io/mincomp/thoughts/2020/07/21/minimal-ontology/
36 R. Risam and A. Gil, 'Introduction: The Questions of Minimal Computing', *Digital Humanities Quarterly*, vol. 16, no. 2, 2022, Accessed: Feb. 27, 2024. [Online]. Available: https://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html
37 R. Rosselli Del Turco, C. Di Pietro, and C. Martignano, 'Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo', *Umanistica Digitale*, n. 7, 2019, doi: https://doi.org/10.6092/issn.2532-8816/9322.
38 G. Roussilhe, A.-L. Ligozat and S. Quinton. 'A long road ahead: a review of the state of knowledge of the environmental effects of digitization' _Current Opinion in Environmental Sustainability_, 2023, no. 62, pp. 101296. Available: https://hal.utc.fr/LIG-MFML-SPADES/hal-04448683v1
39 L. Saint-Raymond, 'Digital Viau : histoire et aléas d'un projet en humanités numériques', Les Cahiers de Framespa, no. 42, 2023, Available: DOI : https://doi.org/10.4000/framespa.14121
40 C. Schuwey, Interfaces. L'apport des humanités numériques à la littérature. Neuchâtel: Alphil, 2019.
41 N. Siddiqui, 'Hidden in Plain-TeX: Investigating Minimal Computing Workflows', *Digital Humanities Quarterly*, vol. 16, no. 2, 2022. Accessed: Feb. 27, 2024. [Online]. Available: https://www.digitalhumanities.org/dhq/vol/16/2/000588/000588.html
42 A. Stussi, *Nuovo avviamento agli studi di filologia italiana*. Bologna: Il Mulino, 1988.
43 G. Tancredi, C. Fenu, 'XML-TEI: Un modello per la filologia d'autore', in F. Ciracì, G. Miglietta, C. Gatto (ed.), AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Proceedings della 11a conferenza nazionale. Lecce, 2022, pp. 218-222.
44 B. Vallarano, 'Oltre la lingua franca. Il plurilinguismo mediterraneo di Elisa Chimenti', *Annali. Sezione romanza*, 65, 1, 2023, pp. 139-156, doi: https://doi.org/10.6093/547-2121/10715.
45 M. Venturi, A. M. Del Grosso, F. Boschetti, L. Tessarolo, A. Prontera, D. Bovet, G. Cattaneo and V. Melis. 'La 'Galassia MQDQ': un concetto di filologia tradizionale, digitale, sostenibile.' Magazén, vol. 4, no. 1, Jun. 2023. Available: https://iris.unive.it/handle/10278/5032220#?mode=simple
46 R. Vigilanti, G. del Rio Grande, N. Hernàndez, and R. De Léon, 'Open, Equitable, and Minimal: Teaching Digital Scholarly Editing North and South', *Digital Humanities Quarterly*, vol. 16, no. 2, 2022. Accessed: Feb. 27, 2024. [Online]. Available: https://www.digitalhumanities.org/dhq/vol/16/2/000591/000591.html
47 M. Zaccarello, ed., *Teoria e forme del testo digitale*. Roma: Carocci, 2019.
48 M. S. Zemmouri, "À la redécouverte d'Elisa Chimenti, écrivaine humaniste", in A. Baida (ed.), *Langue française et contacts langagiers*. Rabat: Bouregreg, 2007, pp. 35-146.

Sitography:

49 ADHO: https://adho.org/?lang=fr
50 *Archives Chimenti*: https://eman-archives.org/Chimenti/ (under construction).
51 DHQ: https://digitalhumanities.org/dhq/
52 EMAN Édition de Manuscrits et d'Archives Numériques: https://eman-archives.org/EMAN/.
53 EVT Edition Visualization Technology: http://evt.labcd.unipi.it/.
54 FMEC - https://www.elisachimenti.org/accueildef.html
55 GitLab: https://about.gitlab.com/.
56 GO::DH: http://www.globaloutlookdh.org/
57 LAI-ALEEF: https://lai-ecriture-exil-au-feminin.univ-lille.fr/accueil.
58 MESHS: https://www.meshs.fr/page/accueil
59 TEI Text Encoding Initiative: https://tei-c.org/.
60 Tesseract OCR on Github: https://github.com/tesseract-ocr/tesseract.
61 THALIM Unité mixte de recherche 7172 "Théorie et histoire des arts et des littératures de la modernité XIX-XXI siècles": https://www.thalim.cnrs.fr/

I (NOT EXHAUSTIVE) SCHEME OF THE TRANSFORMATIONS FOR THE WORD-HTML WORKFLOW.

Text	***TEI***
Page number	<pb>
Paragraph	<p>
Erasure/Restored text	<seg>
Adjustments	<choice>,<orig>,<reg>
Gloss	<gloss>
Foreign word	<foreign>
Location	<placeName>
Name or character	<persName>
Significant words	<term>
Word about significant topic	<distinct>
*Part of text about a significant topic	<milestone>
Bibliographic reference	<bibl>
*Reference symbol to external note	<ref>

II LIST OF CODING ELEMENTS FOR THE TXT-EVT WORKFLOW.

Text	***TEI***
Titles	<head>

Paragraphs	<p>
Direct discourse	<said who="">
Verses	<l>, <lg>
Personal names	<persName>
Place names	<placeName>
Foreign languages	<term ref="#">
Addition	<add>
Deletion	<del>
Substitution	<subst>
Notes	<note resp="author"> <note resp="editor">
List of witnesses	<listWit>
Witnesses	<witness xml:id="">
Critical apparatus entry	<app>
Lemma	<lem>
Variant	<rdg>
Lacuna	<lacunaStart wit=""/> <lacunaEnd wit=""/>

III COMPARISON SCHEME FOR THE TWO WORKFLOWS.

	I step: transcription	***II step: annotation***	***III step: pivot format***	***IV step: visualisation***
WORD-HTML workflow	WORD	WORD	XML-TEI	HTML
TXT-EVT workflow	TXT	XML-TEI	-	EVT