Editing tools: transcribing and encoding

Table of Contents

Abstract

This paper deals with editing tools and editing platforms, i.e. programs used by editors in order to fulfil one or more tasks in the creation of a scholarly digital edition (SDE). Recurring themes in the field of SDEs are the standardization of XML-TEI markup and the success of documentary digital editions (DDEs) – as compared with critical and genetic editions. The existing editing tools reflect these practices: a considerable number of transcribing and encoding tools, mostly TEI compliant, are developed and used in order to produce DDEs.

A small selection of applications are analysed in detail here: T-Pen, CWRC-Writer, TextGrid, eLaborate and Ecdosis; the tools are compared in an initial summary table. A second table and recap reflect the common features of transcribing and encoding applications.

Questions concerning the entire workflow necessary for the creation of a scholarly digital edition are addressed in brief. In applying Manovich's concept of 'software culture' to the field of digital editions, editing tools emerge as crucial, insofar as they may shape investigation and, eventually, scholarly products.

0. Preliminary.

The goal of this article is to provide an overview of the field of editing tools, and in particular of transcribing and encoding tools; it does not intend to compile a list of existing tools, which will soon be obsolete. It will address digital and non-digital editors as well as the developers of such tools; basic knowledge of XML-TEI principles is required.

For the purpose of this discussion, an editing tool is defined as a web-based or stand-alone application that an editor can use to accomplish one task (or a very restricted number of tasks) during the creation of a scholarly (digital) edition; an editing environment is defined as a platform on which different editing tools work together in order to fulfil more than one task in the creation process and, if possible, to cover the whole workflow[1].

This study focuses on transcribing and encoding tools among other applications used for the preparation of a scholarly digital edition (SDE). The publication phase, for instance, will not be taken into account[2]. The article is focused on general-purpose applications, i.e. those editing tools which are not conceived specifically for one project – for one edition, but rather are intended to be used for several projects[3]. Also, only those applications specially created for producing SDEs are taken into account; XML editors, for example, which are widely used in this process [Biblissima 2013], will not be pursued here.

Overall issues concerning editing tools are addressed first; a selection of applications for transcription and encoding is analysed in the central section and their common features are summarized afterwards; finally, broader issues concerning the entire workflow for creating SDEs are briefly pursued.

1. A software culture.

In the past decades, the interaction with computing has led humanists to yield new scholarly products: not only texts, but also *things*, as images, corpora, software and platforms[4]. Understanding so-called *Generative Humanities* requires reckoning with a digital humanist not as a scholar dealing with a machine called computer, but rather with a number of digital scholarship practices related to computing, where computing is 'what people wanted computers to do and

[1] For example, CollateX is a tool, TextGrid is a platform.

[2] We are aware of the need for publishing system, especially out-of-the-box and user-friendly ones, in order to 'bring to life' files that lie idle in editors' computers [see Pape-Schöch-Wegner 2012 and its bibliography]. Moreover, considering publication a task for editors, and not for publishers, is one of the novelties of SDEs, and further efforts have to be devoted to the materiality of the representation, I.e. the interface in a digital representation [cf. Fiormonte 2009].

[3] A lot of interesting projects falls outside this study, as I shall repete below. Among them, the Bentham Transcription Desk <http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham>, Transcribo <http://transcribo.org/en/> or the Online Transcription Editor <http://wfce-ote.sourceforge.net/>. All of them may potentially be used for several editions, but so far has been strongly customized on one project.
All URLs accessed 2015-03.

[4] 'The advent of Digital Humanities implies a reinterpretation of the humanities as a generative enterprise: one in which students and faculty alike are making things as they study and perform research, generating not just texts (in the form of analysis, commentary, narration, critique) but also images, interactions, cross-media corpora, software, and platforms' Burdick 2012, p. 10.

how people designed computers to do it' [McCarty 2005, p. 14]. Modelling, prototyping and testing developments have therefore become humanist scholarly activities, even as scholars struggle to have them recognized as legitimate activities in the Humanities [Schreibman and Halon 2010, Juola 2008].

What is a digital editing tool? It is, first of all, a program, a stand-alone or web application. Software is pervasive in the digital era, hence Manovich's notion of *software society* or *software culture*. In his words, 'software has become our interface to the world, to others, to our memory and our imagination' [Manovich 2013, p. 2]. Manovich's examples vary, ranging from hospital and transport management to email providers. In order to better understand what software is, we need to analyse 'its genealogy (where it comes from), its anatomy (interfaces and operations), and its practical and theoretical effects' [*ibid.*, p. 124]. This paper will provide some targeted (and by no means comprehensive) insight into these issues.

As already mentioned, an editing tool is here defined as a piece of software used in the creation of scholarly digital editions. Word processors, concordancers or lemmatizers are programs, i.e. digital tools; dictionaries and glossaries can be computer applications or appear in print, i.e. digital or non-digital tools; a set of paper forms for bibliographical resources is a non-digital tool. 'Extending the toolkit of traditional scholarship' [Burdick 2012, p. 8] is doubtless one of the aims of Digital Humanities. The core of each discipline (and profession) can be found in its toolkit and its applicability: a tailor might not anticipate the next dress to be commissioned by her client, but with the right measuring tape, thimble, shears (etc.), s/he can produce it. Within digital media, scholars have created digital equivalents to non-digital media. However, the centrality of interpretation in the Humanities presents a challenge, as does the fact that practices are not always clearly documented and shared. How does one process irregular metre? How does one detect figures of speech? How have humanists been collating texts throughout centuries? Creating tables, using a base-text, selecting *loci critici*? As Andrews points out, the question of digital publication has attracted a lot of attention; but 'the method of production, rather than the published form that the resulting edition take, is the practice wherein lies most of the promised revolution within textual scholarship' [Andrews 2013][5].

The development of ad hoc tools for the creation of a scholarly edition is not novel. The history of editing tools remains to be written, but some pioneering projects in the field are well known: TUSTEP, for instance: a toolbox for scholarly processing textual data, designed at the Computing Center of the University of Tübingen, first implemented in the seventies and constantly upgraded until today; or Collate, a collation tool developed by Peter Robinson in the early nineties, which has only recently encountered a successor in CollateX [Dekker and Middell 2010-2014]. Theoretical reflection on digital tools for the humanities has been on the increase in past decades,

[5] Cf. Pierazzo 2011: 'Perhaps we should just stop trying to map digital editions to printed ones and instead recognize that we are producing a different type of object, one that we can perhaps call a *documentary digital edition*. This new object necessarily comprises all three components of a digital publication—the source, the output and the tools to produce and display it—and it is worth emphasizing again that all three are scholarly products that result from editorial practice' (p. 474-5) and Van Zundert-Boot 2011: 'We argued that digital scholarly editions will be composites of three types of digital web services: data sources, processing services, and interfaces' (p. 150).

fostered by landmark writings [e.g., McCarty 2005, Bradley 2002, Andrews 2013, Unsworth 2003], projects[6] and conferences[7]. On a practical note, a number of resources are available today; two main repositories of humanities applications, not focused only on editing but more generally on research tools for scholarly use, are DIRT (Digital Research Tools Directory)[8] and TAPOR (Research Tools for Textual Studies)[9]. The spreadsheet Collaborative Transcription Tools set by Ben Brumfield is a useful in-progress and collaborative resource[10].

The history of text-oriented applications swings between tools developed for specific uses and general-purpose ones[11], producing what has been called the 'tool paradox' [Pape, Schöch, Wegner 2012]: 'the conclusion would be that either “complexity is the price paid for generality” or that “limitation is the price paid for ease-of-use,” with both standing in the way of a widespread user base'. The authors continue: 'the challenge to combine expressive power with simplicity of use is faced by any developer in the domains of application software or operating systems, and it marks the long-term evolution of such endeavours'. The application of software development models to text analysis software development offers a promising path, but it does not seem to be a priority of humanists or even of digital humanists. This is changing; increasing interest is funnelled into, for example, user-centric design or web standards.

On the matter of SDEs, the importance of software is overwhelming. In Schmidt' somewhat provocative definition, 'SDE is software' and should therefore 'be governed by the procedures and principles of software development'. As this paper is focused on tools for preparing SDEs, it is worth clarifying that here the user is the editor, and not the reader of the edition.

2. TEI and DDE, or Textual Encoding Initiative and Documentary Digital Edition.

This article focuses on transcribing and encoding tools because they constitute a significant portion (at least in terms of quantity) of the existing editing tools. In this section I argue that the flourishing of transcribing and encoding tools is related to the acceptance of XML-TEI as a standard for the encoding of texts and to the predominance of documentary editions in the SDEs panorama.

[6] E.g., Project Bamboo <https://wikihub.berkeley.edu/display/pbamboo/Documentation>, Interedition <http://www.interedition.eu/>.

[7] Recently, *Easy Tools for Difficult Texts*, Cost Action IS1005 'Medioevo europeo' and Huygens ING, Den Haag, April 2013; *Research Summit on Collation of Ancient and Medieval Texts*, COST Action IS1005 'Medioevo Europeo', Münster, October 2014; *Scholarship in Software, Software as Scholarship: From Genesis to Peer Review*, Universität Bern, January 2015.

[8] <http://dirtdirectory.org/>.

[9] <http://www.tapor.ca/>.

[10] <https://docs.google.com/spreadsheets/d/1MFsRSZRGy3RRB4AUD6AFp7IQsecqcauJLyZLGVzJFWs/>.

[11] Unsworth distinguish four generations, from ad hoc programs, through reusable libraries of text-processing routines, to interactive general-purpose programs, which move over the network in the last stage [Unsworth 2003].

A recurring theme surrounding SDEs is the overwhelming presence of documentary digital editions (DDE)[12], as compared to critical and genetic editions. To better understand the relevance of DDEs in the wider SDEs panorama, we should remember that many editions that have been called Electronic Archives in fact contain documentary editions of a number of 'single documents'. This is clearly exemplified in one of the earliest SDE projects, the *Piers Plowman Electronic Archive*: on the 'The texts of *Piers Plowman*' page, the reader can find both the *Critical Edition* (for instance The B-Version Archetype) and the *Documentary Editions* of the manuscripts. Several scholarly, disciplinary and technological factors have influenced DDE's development. Among the latter, the availability of digital facsimiles, for instance, has been significant. Moreover, as Pierazzo makes clear, 'critical editions require considerable effort on many fronts, starting from the development of tools to support the many operations that are part of producing such an edition, such as transcription, collation, stemma generation, and web publishing, among others. [...] DDEs, on the other hand, require less investment'. The available editing tools reflect current practices in the creation of SDEs: transcribing tools are overwhelmingly present, if compared to, for example, collation or phylogenetic tools, just as much as documentary editions are endemic if compared to critical or genetic editions.

Observing the editing tools panorama it seems that the majority of these applications takes XML-TEI markup into account. Why do they do it? Encoding is a fundamental step: a way of putting 'intelligence in the text' [Hockey 2000, p. VI] – a kind of intelligence that computational tools can process. It therefore carries surplus value specific to digital texts. Nevertheless, there is no common procedure to follow in determining what should be encoded. A rather basic, practical and economic proposal is to encode only what will be processed[13]. XML-TEI is not the only existing data format for creating SDEs: other XML languages, markup (such as LaTex) and markdown syntax, or the code of web pages, HTML, can be used. However, XML-TEI markup, extremely rich and explicitly devoted to text encoding, has become a standard for transcribing and editing literary texts and historical sources, which are normally the objects of scholarly editions. Thus existing editing tools reflect common practices in the creation of SDEs; the development of tools follows the demand for tools; and scholars seem to demand XML-TEI compliant tools.

To summarize, available editing software reflect common practices of digital editors' work: a considerable number of transcribing and encoding tools, compliant with the XML-TEI framework, have been developed and used to produce Digital Documentary Editions, a kind of edition that is particularly successful in the SDEs' panorama.

[12] 'For the purposes of this discussion, I define a documentary edition as an edition of a text based on a single document [...] It can assume different formats, by presenting the textual content of the document as semi-diplomatic, diplomatic, ultra-diplomatic, or even facsimile editions' Pierazzo 2011.

[13] I.e., not encoding persons, if there is no plan to have an index of persons or to process in any other way the person-like tags.

3. Into the tools.

A number of editing tools and environments will be analysed here, focusing on encoding and transcribing while also considering collation for one of the environments. This selection follows a strictly empirical criterion of 'user-friendliness'. Only tools that require minimal computer literacy[14] and not to consult complex manual; only browser or portable applications, for which no installation is needed, have been selected[15]. Though this is a small selection of the available applications[16], it is worth engaging with these examples.

The tools discussed here are T-Pen and CWRC-Writer; the environments are eLaborate, TextGrid and Ecdosis. Each of these will be described and then compared in a summary table.

A variety of materials have been used to test the tools. These include modern and medieval material: a manuscript of the ancient French prose *Lancelot* (Paris, BnF fr. 1430); the three witnesses of the anonymous *reverdie Volez vos que je vous chant* (Paris, Arsenal, 5198, BnF fr. 845; BnF nouv. acq. fr. 1050), a manuscript of Giacomo da Lentini *Madonna dir vo' voglio* (Firenze, BNC, ms. Banco Rari 217) and a selection from the correspondence between Ernesto Monaci and Francisco Adolpho Coelho, textual scholars of international relevance between the 19th and 20th centuries.

3.1.Tools.

3.1.1. T-Pen.

T-Pen[17] is an open source tool developed at the Center for Digital Theology of Saint Louis University.

Its main function is managing manuscript images and transcriptions. The T-Pen Repository contains images of 4177 manuscripts; they are not the property of T-Pen, but are linked from hosting repositories: Parker Library on the Web, e-codices, Codices Electronici Ecclesiae Coloniensis, Harvard Houghton Library, SISF - Assisi et Stanford University Libraries. Furthermore, users can upload their images.

It is possible to link an already completed transcription to one of the images or to use T-Pen as a transcribing tool.

To upload an image, one can address the 'Tools for advanced users' section. A new project is created while uploading an image, and metadata can be added in the 'Active projects' page.

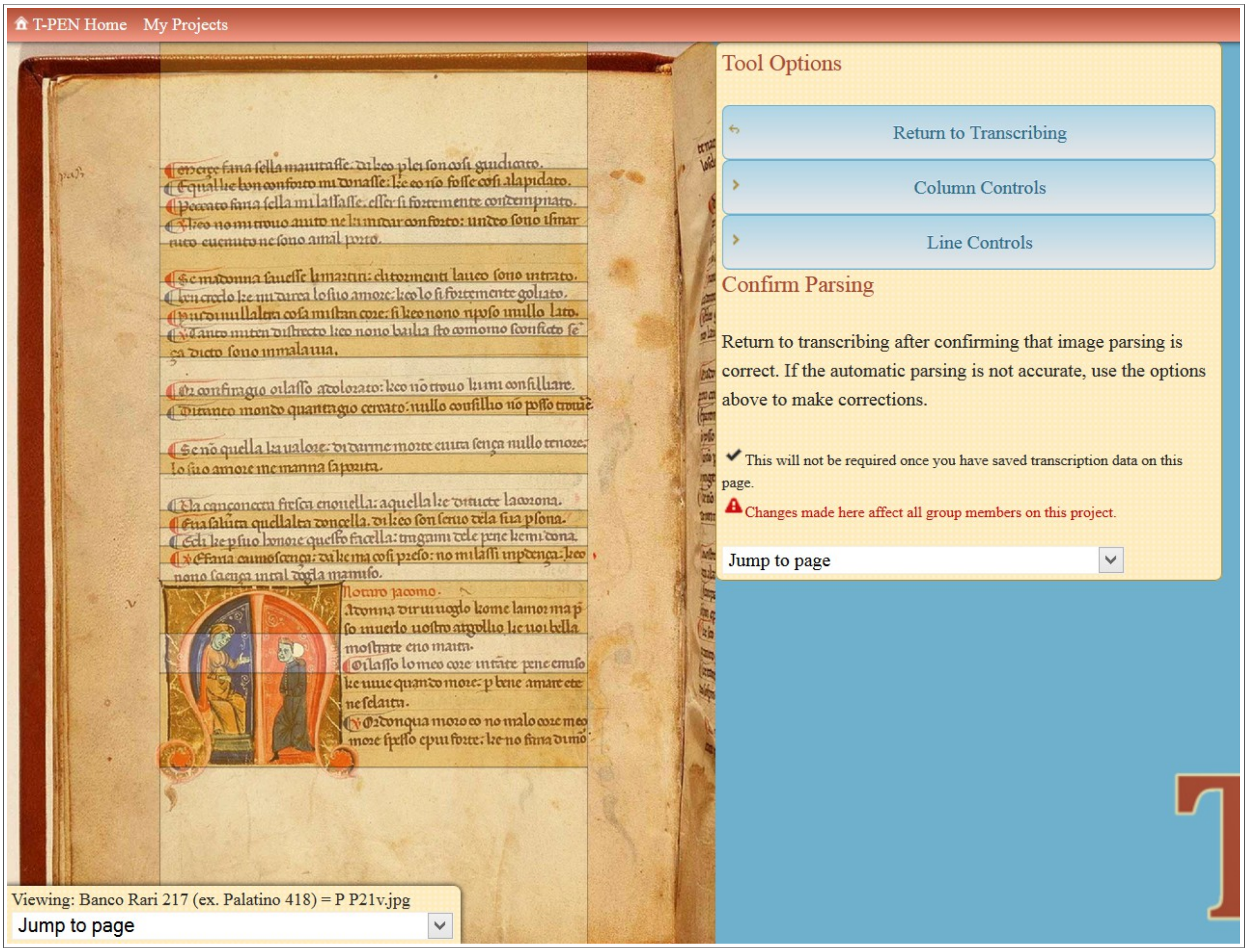
[14] As said in the Preliminary, we include in the 'minimal computer literacy' basic knowledge of XML-TEI, as it is rather common in textual scholars who are interested in Digital Humanities and more specifically in producing SDE.

[15] A tool that works in a browser is made up on web pages. Within an editor, the user interaction (typing of the text, pressing buttons, etc.) modifies the HTML source and the corresponding XML, if any, calling the proper JavaScript libraries. When the user gives the command, or automatically in time, the editor interacts with a server (or a web-worker) for saving and validating.

[16] A lot of interesting projects falls outside this presentation, for instance Tustep <http://www.tustep.uni-tuebingen.de/tustep_eng.html>, the oXygen plug-in Ediarum <http://www.bbaw.de/en/telota/software/ediarum>, the Islandora TEI Editor <https://jtei.revues.org/790>, just to mention some of them.

[17] <http://t-pen.org/TPEN/>. Further developments announced at <http://lib.slu.edu/digital-humanities/projects>.

T-Pen automatically parses the image, recognizing the manuscript columns and lines; users can then merge, delete, add or adjust them.



Confirm Parsing

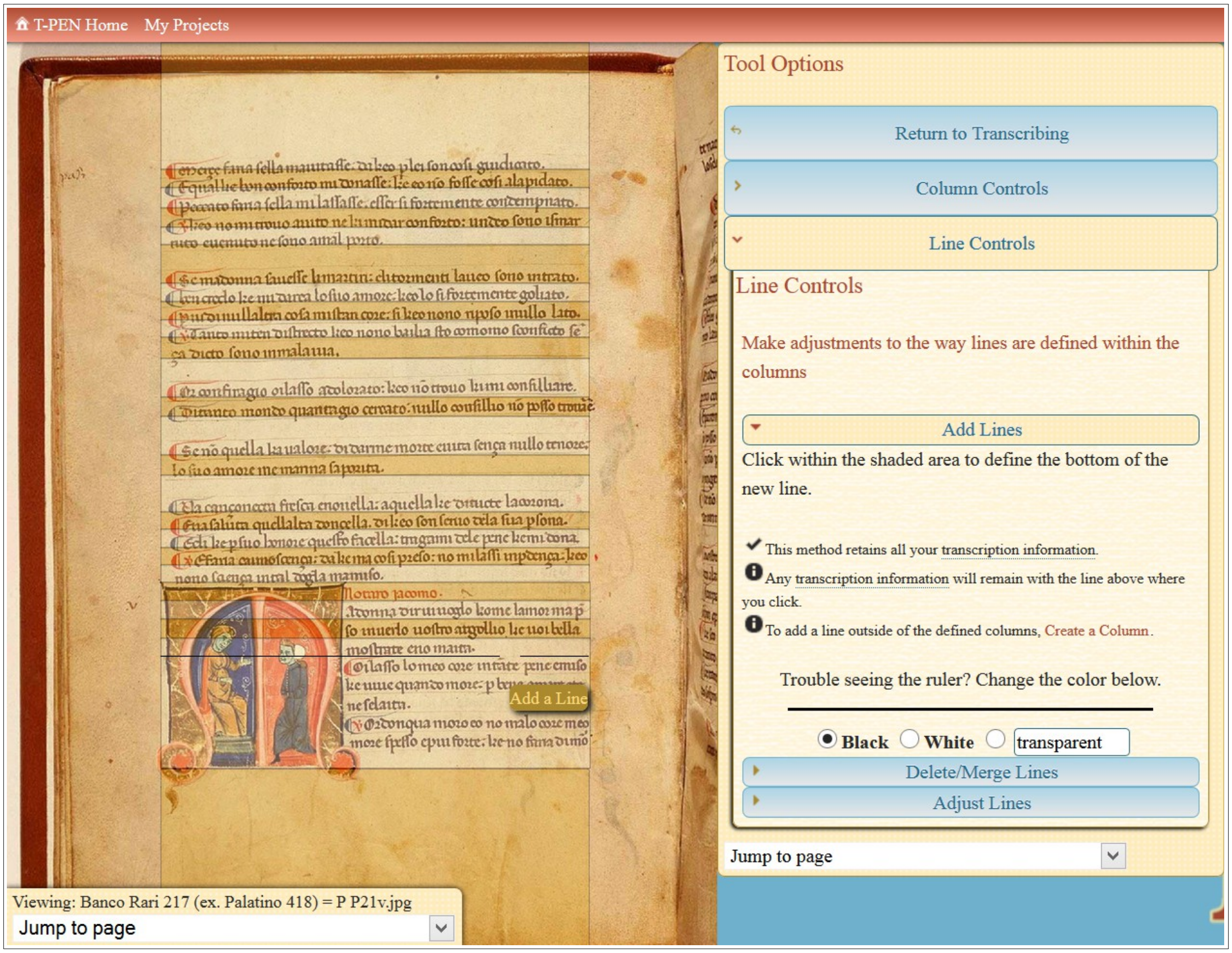
Return to transcribing after confirming that image parsing is correct. If the automatic parsing is not accurate, use the options above to make corrections.

✓ This will not be required once you have saved transcription data on this page.

⚠ Changes made here affect all group members on this project.

Viewing: Banco Rari 217 (ex. Palatino 418) = P P21v.jpg

Figure 1: T-Pen. Parser (1)



Make adjustments to the way lines are defined within the columns

Click within the shaded area to define the bottom of the new line.

✓ This method retains all your transcription information.

Any transcription information will remain with the line above where you click.

To add a line outside of the defined columns, Create a Column.

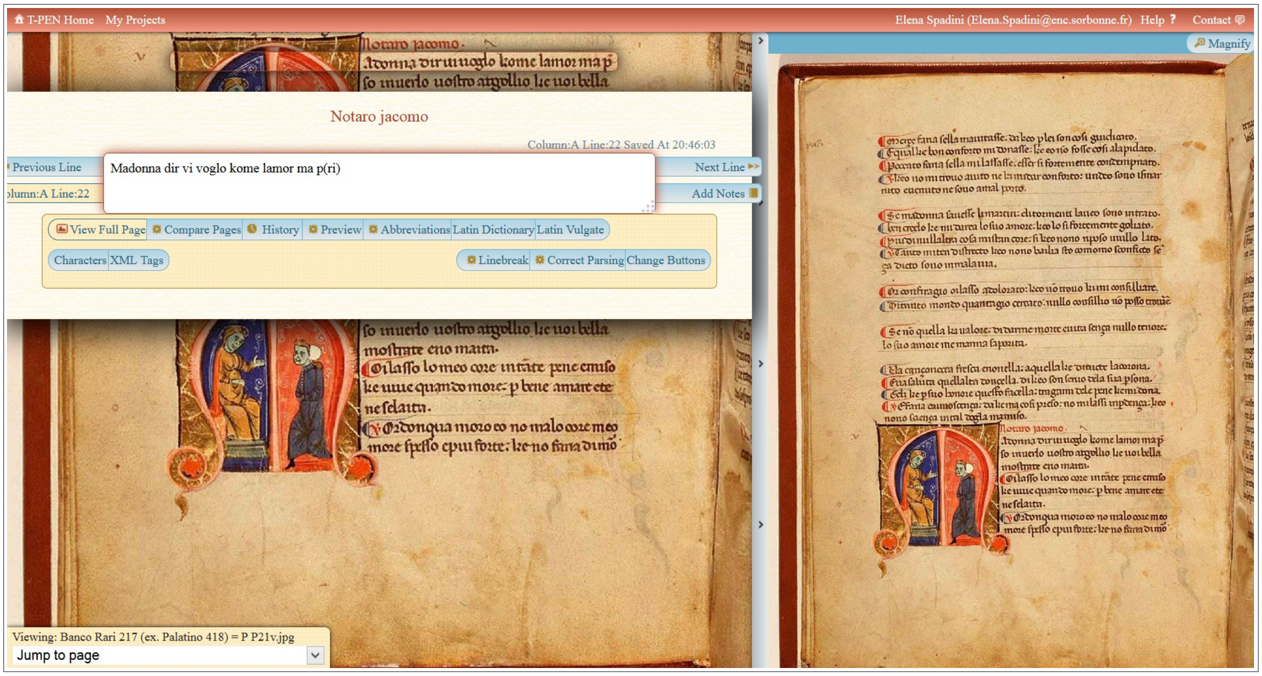
Trouble seeing the ruler? Change the color below.

● Black ○ White ○ transparent

Figure 2: T-Pen. Parser (2)

The transcription is line by line; the text of the previous line is shown on the top and the text of the whole page can be shown in a right window. Users can move from one line to another using the dedicated button, the Tab key (the most time-saving option) or by clicking on the corresponding line on the image[18].

[18] Cf. Lowe 2015: 'Although this tool is undoubtedly useful for students beginning to transcribe script, reducing as it does the likelihood of eyeskip by focusing attention on each line in turn, it proved frustrating for those practiced at transcription who found that the interface slowed the progress for touch-typers' (p. 195).



Notaro jacomo

Column:A Line:22 Saved At 20:46:03

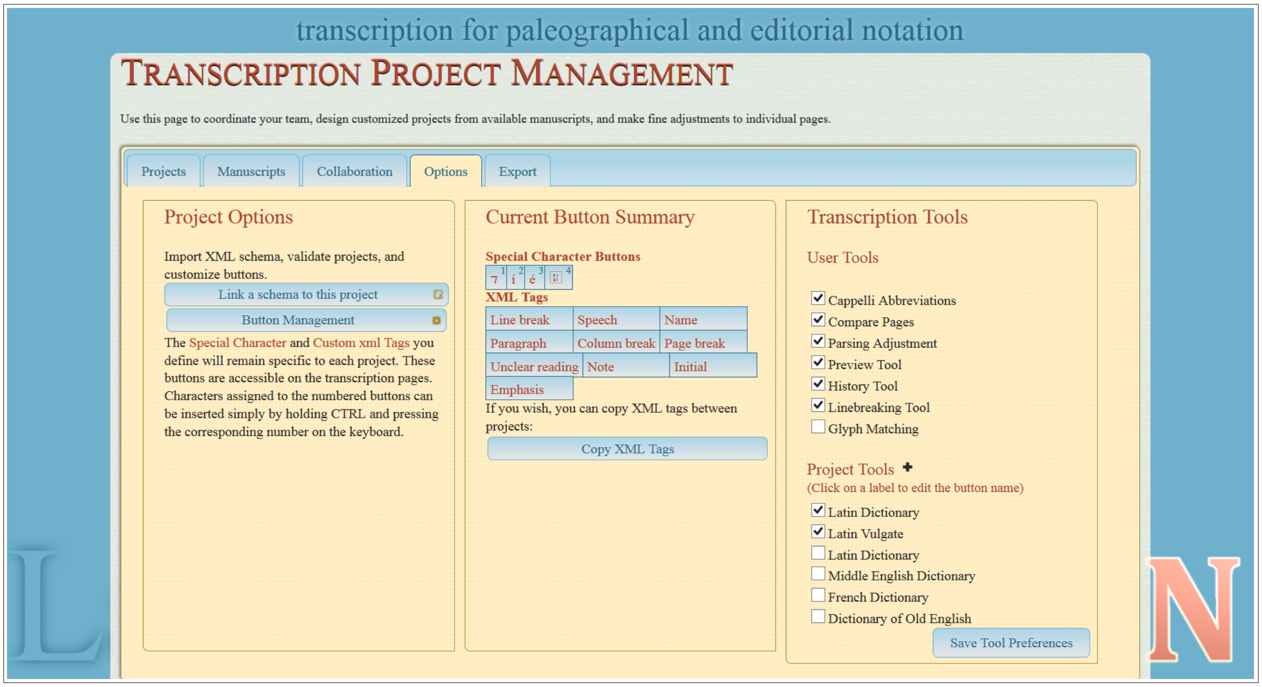
olumn:A Line:22

Madonna dir vi voglo kome lamor ma p(ri)

Viewing: Banco Rari 217 (ex. Palatino 418) = P P21v.jpg

Figure 3: T-Pen. Transcription area

Several buttons are available next to the transcription area. In the Options section of the Transcript project management page, users can determine which buttons will be displayed.



TRANSCRIPTION PROJECT MANAGEMENT

Use this page to coordinate your team, design customized projects from available manuscripts, and make fine adjustments to individual pages.

Project Options

Import XML schema, validate projects, and customize buttons.

The Special Character and Custom xml Tags you define will remain specific to each project. These buttons are accessible on the transcription pages. Characters assigned to the numbered buttons can be inserted simply by holding CTRL and pressing the corresponding number on the keyboard.

Current Button Summary

Special Character Buttons

7 i é [!]

XML Tags

Line break	Speech	Name
Paragraph	Column break	Page break
Unclear reading	Note	Initial
Emphasis		

If you wish, you can copy XML tags between projects:

Transcription Tools

User Tools

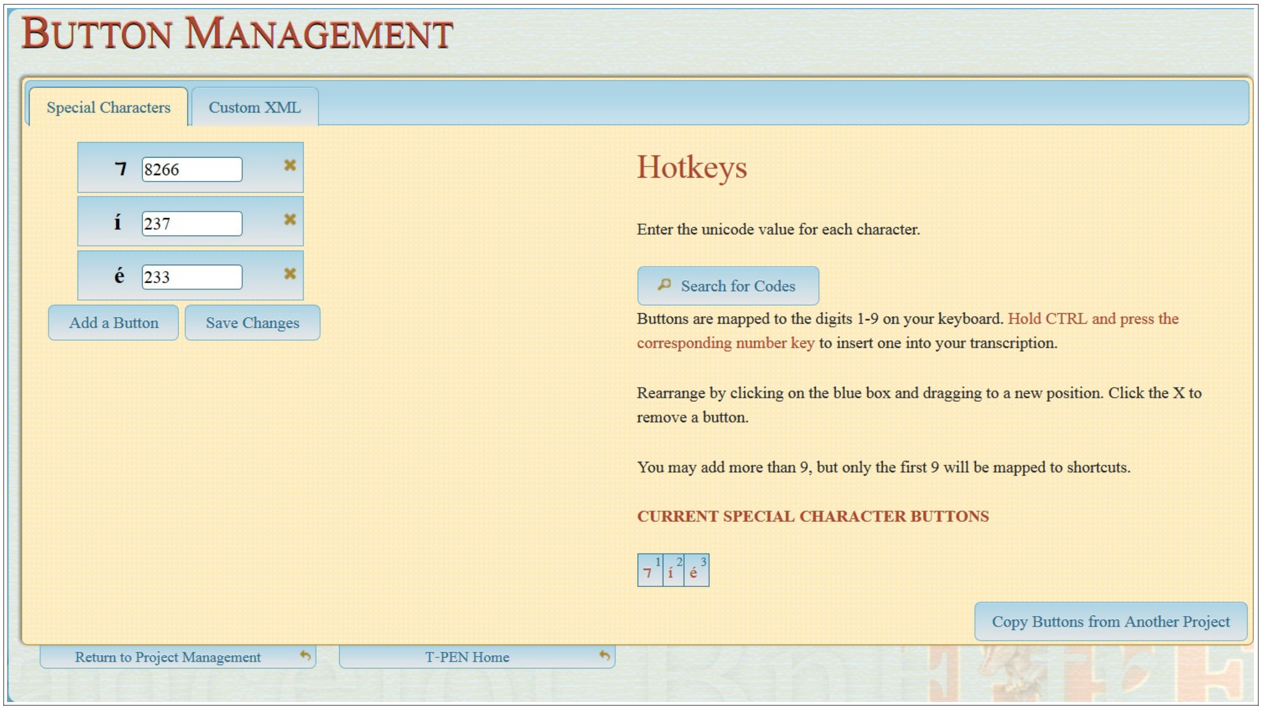
☑ Cappelli Abbreviations
☑ Compare Pages
☑ Parsing Adjustment
☑ Preview Tool
☑ History Tool
☑ Linebreaking Tool
☐ Glyph Matching

Project Tools ✚
(Click on a label to edit the button name)

☑ Latin Dictionary
☑ Latin Vulgate
☐ Latin Dictionary
☐ Middle English Dictionary
☐ French Dictionary
☐ Dictionary of Old English

Figure 4: T-Pen. Transcription project management

In the central column there is a preview of the selected buttons; by clicking on one of them in the left column, users can access the management page in order to customize buttons for inserting XML tags and special characters (keyboard shortcuts are automatically created).



Hotkeys

Enter the unicode value for each character.

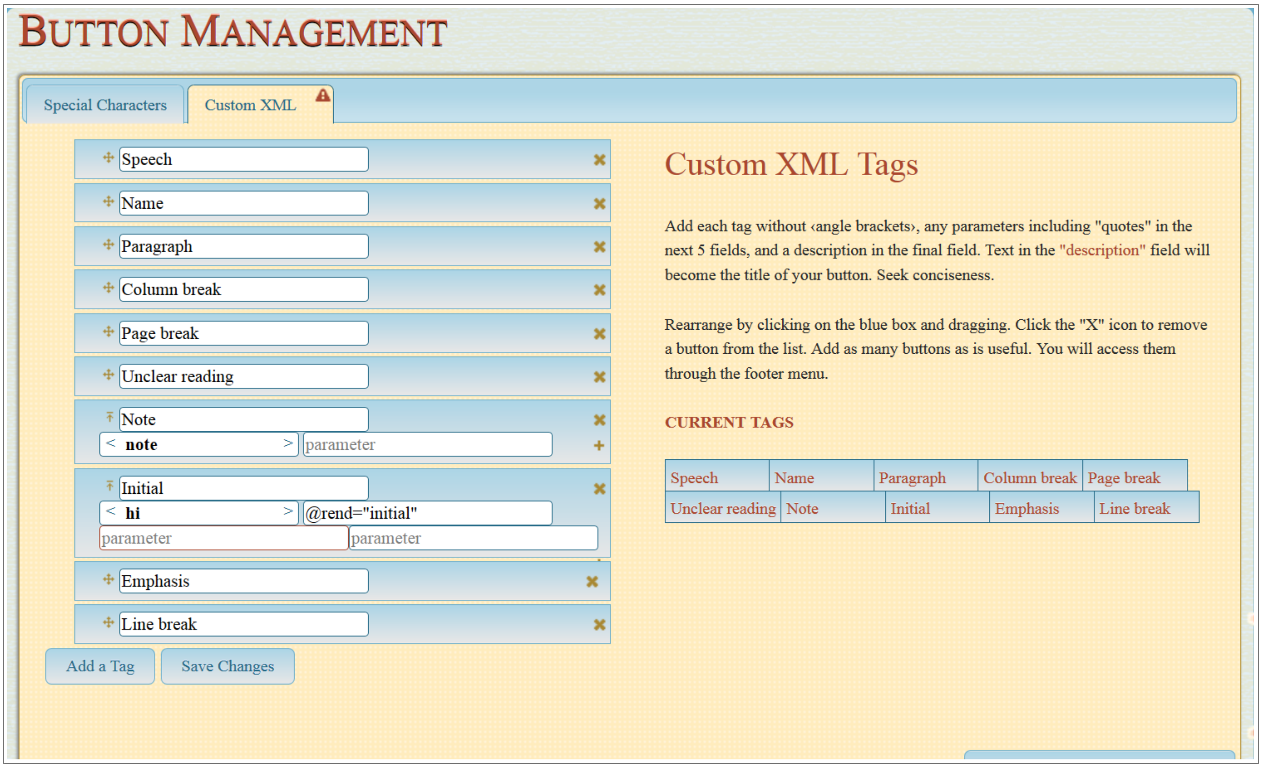
Buttons are mapped to the digits 1-9 on your keyboard. Hold CTRL and press the corresponding number key to insert one into your transcription.

Rearrange by clicking on the blue box and dragging to a new position. Click the X to remove a button.

You may add more than 9, but only the first 9 will be mapped to shortcuts.

CURRENT SPECIAL CHARACTER BUTTONS

Figure 5: T-Pen. Button management (1).



Custom XML Tags

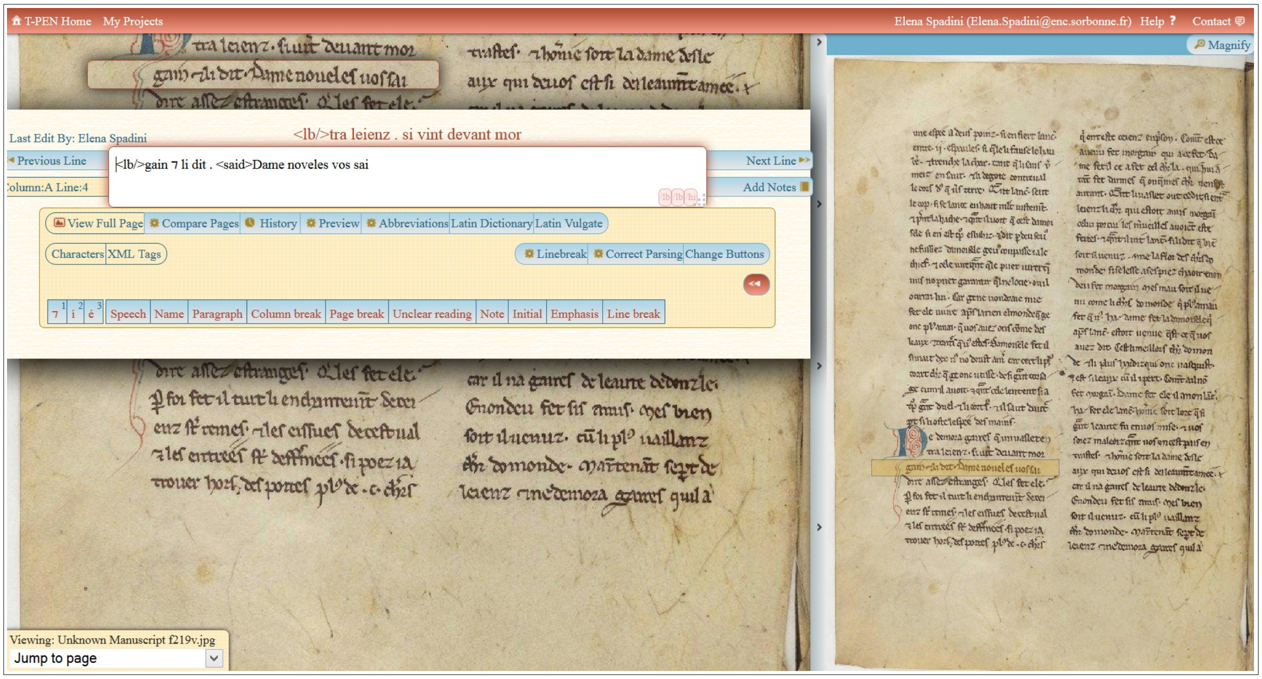
Add each tag without ‹angle brackets›, any parameters including "quotes" in the next 5 fields, and a description in the final field. Text in the "description" field will become the title of your button. Seek conciseness.

Rearrange by clicking on the blue box and dragging. Click the "X" icon to remove a button from the list. Add as many buttons as is useful. You will access them through the footer menu.

CURRENT TAGS

Speech	Name	Paragraph	Column break	Page break
Unclear reading	Note	Initial	Emphasis	Line break

Figure 6: T-Pen. Button management (2).


Last Edit By: Elena Spadini

<lb/>tra leienz . si vint devant mor

<lb/>gain 7 li dit . <said>Dame noveles vos sai

Column:A Line:4

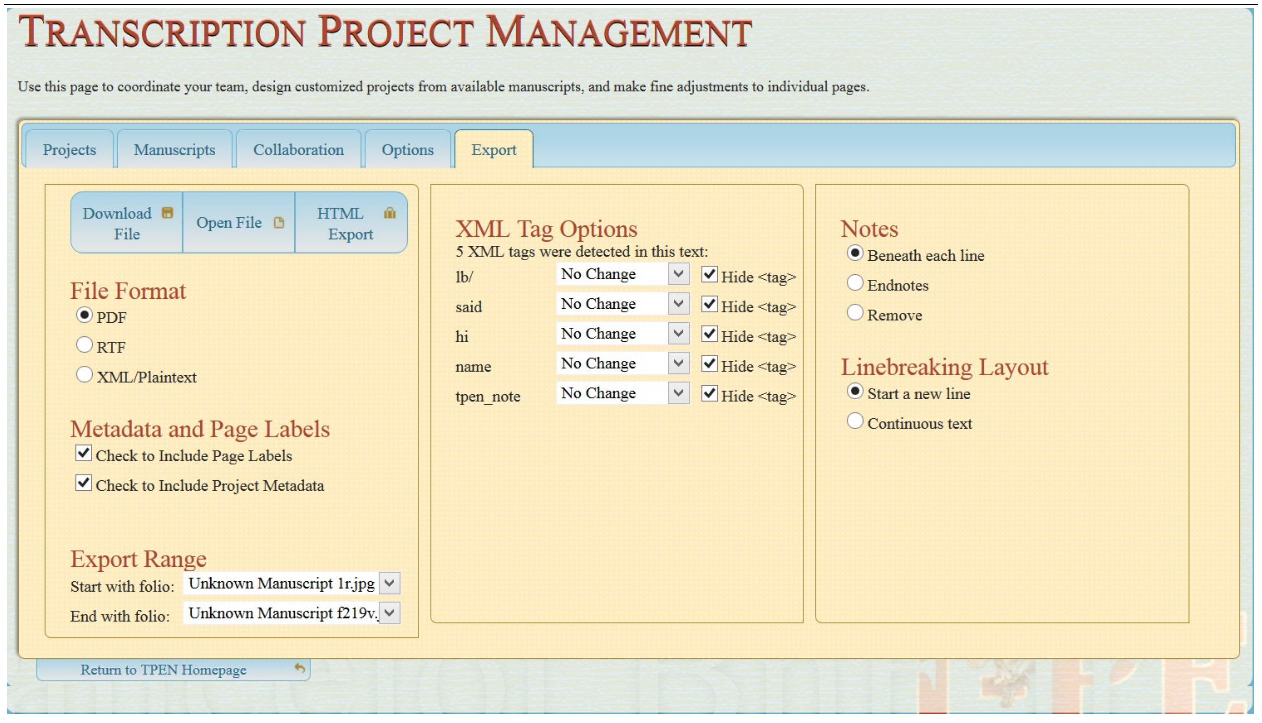
Viewing: Unknown Manuscript f219v.jpg

Figure 7: T-Pen. Transcription area.

Working on the whole page transcription and not on the line by line transcription will make it easier to insert tags that include several lines of text. Notes can be added on each line by clicking on the corresponding button.

A schema can be uploaded, pointing to a URL (no internal path is admitted).

The import and export functionalities are remarkable, allowing the user to upload and download using several formats, and offering a number of options for the HTML export.


TRANSCRIPTION PROJECT MANAGEMENT

Use this page to coordinate your team, design customized projects from available manuscripts, and make fine adjustments to individual pages.

File Format
● PDF
○ RTF
○ XML/Plaintext

Metadata and Page Labels
☑ Check to Include Page Labels
☑ Check to Include Project Metadata

Export Range
Start with folio: Unknown Manuscript 1r.jpg
End with folio: Unknown Manuscript f219v.

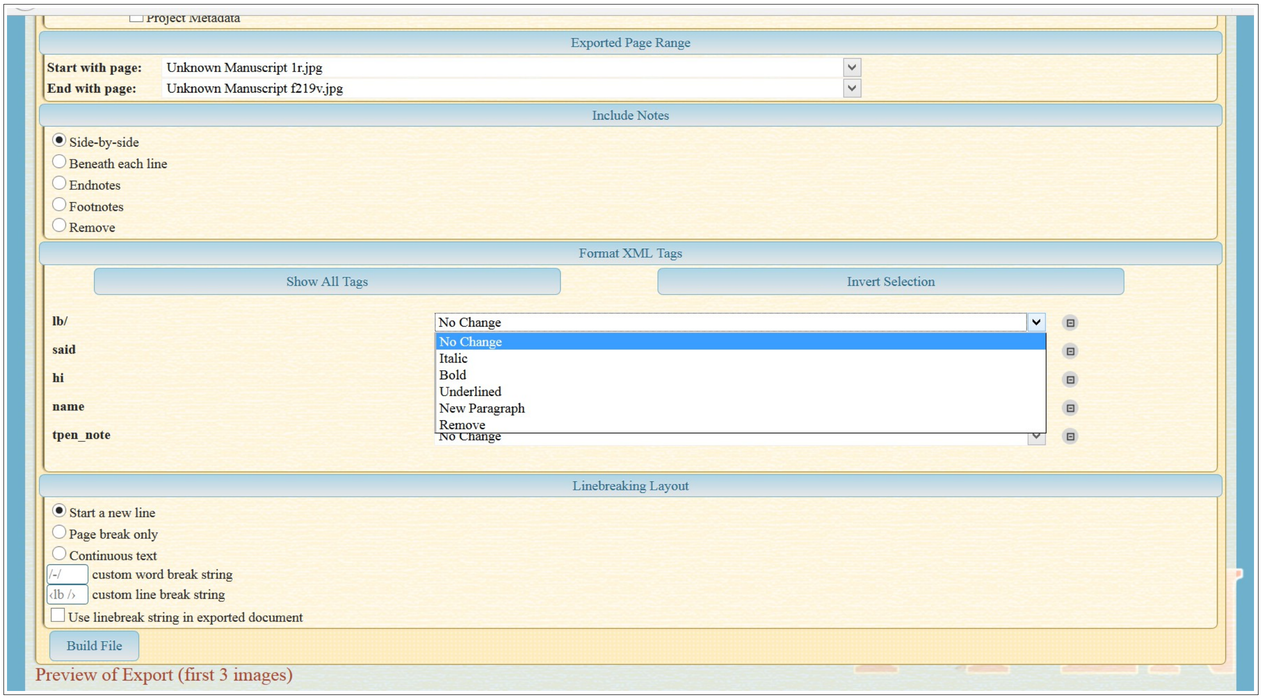
XML Tag Options
5 XML tags were detected in this text:

lb/	No Change	☑ Hide <tag>
said	No Change	☑ Hide <tag>
hi	No Change	☑ Hide <tag>
name	No Change	☑ Hide <tag>
tpen_note	No Change	☑ Hide <tag>

Notes
● Beneath each line
○ Endnotes
○ Remove

Linebreaking Layout
● Start a new line
○ Continuous text

Figure 8: T-Pen. Export (1).



□ Project Metadata

Exported Page Range

Start with page: Unknown Manuscript 1r.jpg

End with page: Unknown Manuscript f219v.jpg

Include Notes

● Side-by-side
○ Beneath each line
○ Endnotes
○ Footnotes
○ Remove

Format XML Tags

lb/ No Change
said
hi
name
tpen_note No Change

Linebreaking Layout

● Start a new line
○ Page break only
○ Continuous text
/-/ custom word break string
<lb /> custom line break string
□ Use linebreak string in exported document

Figure 9: T-Pen. Export (2).

To summarize, T-Pen is mainly a tool for managing manuscript images and transcriptions. Editing tasks such as marking abbreviations, shift of hands, normalizations, can be fulfilled through the encoding; the customizable buttons can facilitate this task, while the line by line structure may slow down it when multiple lines are included in one tag.

3.1.2. CWRC-Writer.

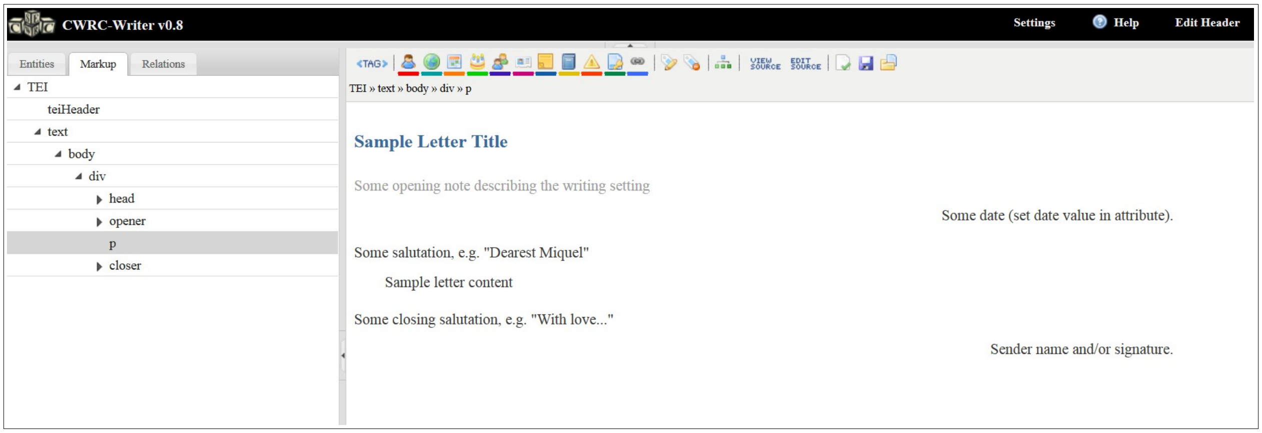
CWRC-Writer (CWRC-W)[19] is a web-based transcribing tool created by the Canadian Writing Research Collaboratory (Collaboratoire scientifique des écrits du Canada). It is under development and a prototype is available online for demonstration and testing purposes.

CWRC-W is a WYSIWYG XML editor that allows users to easily markup the text with TEI tags and to create RDF[20].

The prototype offers three templates: for letters, poems and prose texts. The first one will be used here.

[19] <http://www.cwrc.ca/projects/infrastructure-projects/technical-projects/cwrc-writer/>. For testing the tool, see <https://sites.google.com/site/cwrcwriterhelp/>.

[20] RDF (Resource Description Framework) is a Semantic Web Standard for data interchange. 'RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”)’ <http://www.w3.org/RDF/>.



Sample Letter Title

Some opening note describing the writing setting

Some date (set date value in attribute).

Some salutation, e.g. "Dearest Miquel"

Sample letter content

Some closing salutation, e.g. "With love..."

Sender name and/or signature.

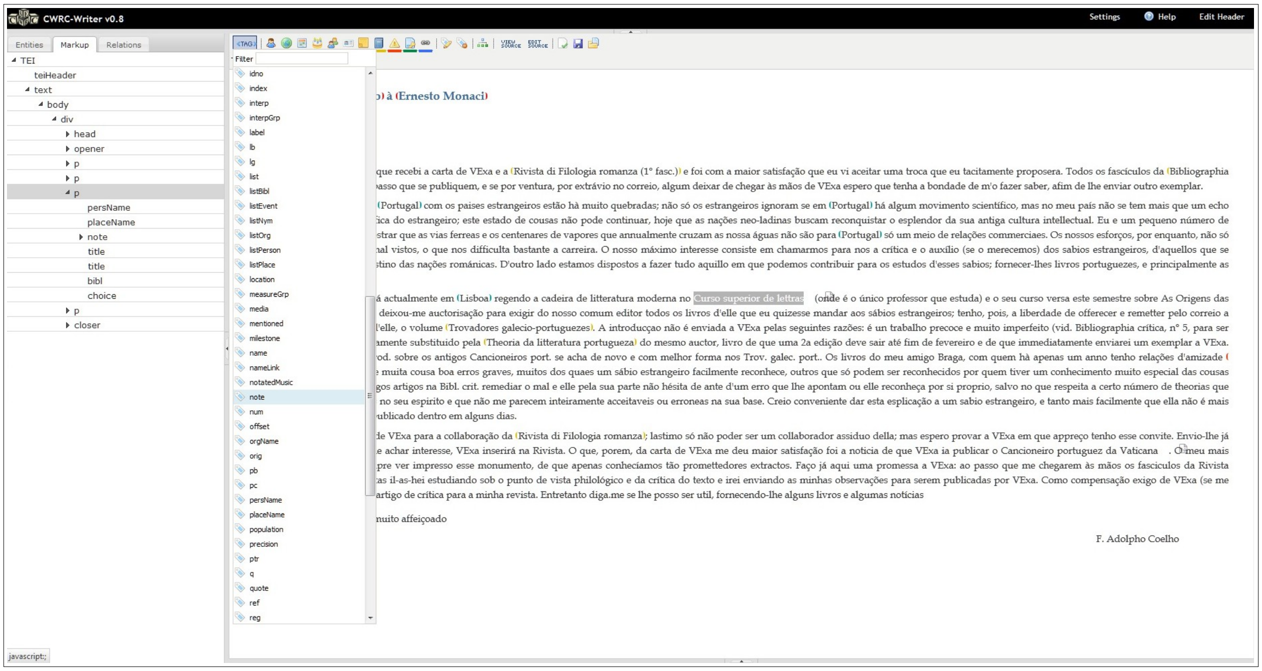
Figure 10: CWRC-Writer. Letter template.

The main elements of CWRC-W are the tags, the entities and the relations. Entities are specific tags, which can be linked in order to create a relation, i.e. a RDF triple.

The text is in the central window; on the left column the user finds the XML structure, the entities list and the relations list; on the top is a toolbar with buttons for handling tags, entities and relations, for showing or modifying the XML source and for validating, saving or uploading documents.

The letter template is suggestive of XML structure, with <opener>, text and <closer>. Each carriage return automatically creates a paragraph (<p>).

To insert a tag, the user can select a portion of the text and choose from the drop-down menu which can be opened from the toolbar or with a right click.



›) à (Ernesto Monaci)

que recebi a carta de VExa e a (Rivista di Filologia romanza (1° fasc.)) e foi com a maior satisfação que eu vi aceitar uma troca que eu tacitamente proposera. Todos os fascículos da (Bibliographia
›asso que se publiquem, e se por ventura, por extrávio no correio, algum deixar de chegar às mãos de VExa espero que tenha a bondade de m'o fazer saber, afim de lhe enviar outro exemplar.

(Portugal) com os paises estrangeiros estão hà muito quebradas; não só os estrangeiros ignoram se em (Portugal) há algum movimento scientífico, mas no meu país não se tem mais que um echo
:ica do estrangeiro; este estado de cousas não pode continuar, hoje que as nações neo-ladinas buscam reconquistar o esplendor da sua antiga cultura intellectual. Eu e um pequeno número de
strar que as vias ferreas e os centenares de vapores que annualmente cruzam as nossa águas não são para (Portugal) só um meio de relações commerciaes. Os nossos esforços, por enquanto, não só
ıal vistos, o que nos difficulta bastante a carreira. O nosso máximo interesse consiste em chamarmos para nos a crítica e o auxílio (se o merecemos) dos sabios estrangeiros, d'aquellos que se
stino das nações románicas. D'outro lado estamos dispostos a fazer tudo aquillo em que podemos contribuir para os estudos d'esses sabios; fornecer-lhes livros portuguezes, e principalmente as

á actualmente em (Lisboa) regendo a cadeira de litteratura moderna no Curso superior de lettras (onde é o único professor que estuda) e o seu curso versa este semestre sobre As Origens das
deixou-me auctorisação para exigir do nosso comum editor todos os livros d'elle que eu quizesse mandar aos sábios estrangeiros; tenho, pois, a liberdade de offerecer e remetter pelo correio a
l'elle, o volume (Trovadores galecio-portuguezes). A introducçao não é enviada a VExa pelas seguintes razões: é un trabalho precoce e muito imperfeito (vid. Bibliographia crítica, n° 5, para ser
amente substituido pela (Theoria da litteratura portugueza) do mesmo auctor, livro de que uma 2a edição deve sair até fim de fevereiro e de que immediatamente enviarei um exemplar a VExa.
od. sobre os antigos Cancioneiros port. se acha de novo e com melhor forma nos Trov. galec. port.. Os livros do meu amigo Braga, com quem hà apenas um anno tenho relações d'amizade (
: muita cousa boa erros graves, muitos dos quaes um sábio estrangeiro facilmente reconhece, outros que só podem ser reconhecidos por quem tiver um conhecimento muito especial das cousas
gos artigos na Bibl. crit. remediar o mal e elle pela sua parte não hésita de ante d'um erro que lhe apontam ou elle reconheça por si proprio, salvo no que respeita a certo número de theorias que
no seu espirito e que não me parecem inteiramente acceitaveis ou erroneas na sua base. Creio conveniente dar esta esplicação a um sabio estrangeiro, e tanto mais facilmente que ella não é mais
ublicado dentro em alguns dias.

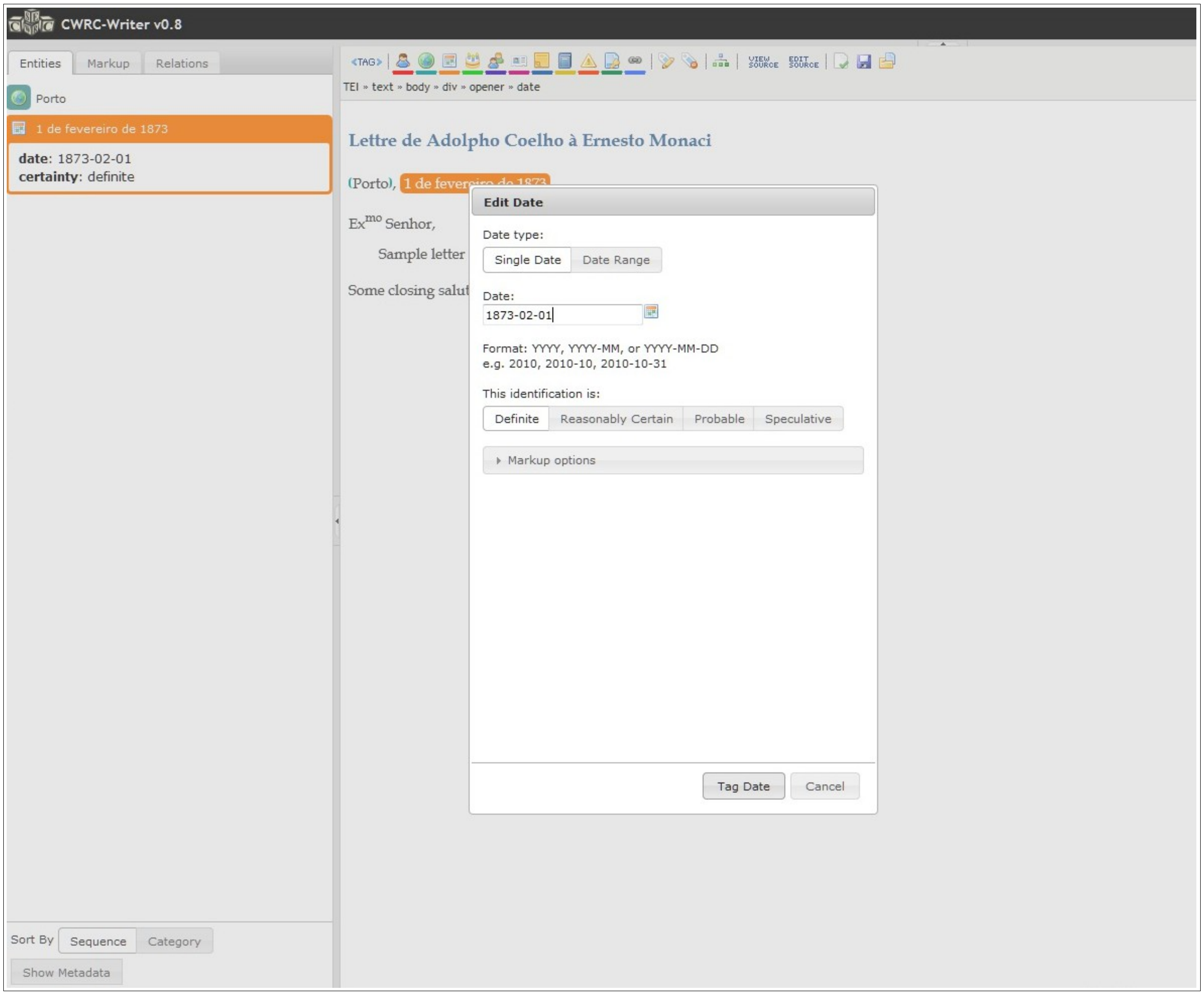
de VExa para a collaboração da (Rivista di Filologia romanza); lastimo só não poder ser um collaborador assiduo della; mas espero provar a VExa em que appreço tenho esse convite. Envio-lhe já
e achar interesse, VExa inserirá na Rivista. O que, porem, da carta de VExa me deu maior satisfação foi a noticia de que VExa ia publicar o Cancioneiro portuguez da Vaticana . O meu mais
pre ver impresso esse monumento, de que apenas conhecíamos tão promettedores extractos. Faço já aqui uma promessa a VExa: ao passo que me chegarem às mãos os fasciculos da Rivista
as il-as-hei estudiando sob o punto de vista philológico e da crítica do texto e irei enviando as minhas observações para serem publicadas por VExa. Como compensação exigo de VExa (se me
artigo de crítica para a minha revista. Entretanto diga.me se lhe posso ser util, fornecendo-lhe alguns livros e algumas notícias

nuito affeiçoado

F. Adolpho Coelho

Figure 11: CWRC-Writer. Insert a tag.

In the toolbar, there are buttons to create entities: person, place, date, event, organization, citation, note, text/title, correction, keyword, link. Users can select a part of the text and press the corresponding button: the tag will be added and the attributes defined in the dialogue box. Once an entity has been created, it appears in the Entities column on the left.



Porto

1 de fevereiro de 1873

date: 1873-02-01
certainty: definite

TEI » text » body » div » opener » date

Lettre de Adolpho Coelho à Ernesto Monaci

(Porto), 1 de fevereiro de 1873

Ex^{mo} Senhor,

Sample letter

Some closing salut

Edit Date

Date type: Single Date

Date: 1873-02-01

Format: YYYY, YYYY-MM, or YYYY-MM-DD
e.g. 2010, 2010-10, 2010-10-31

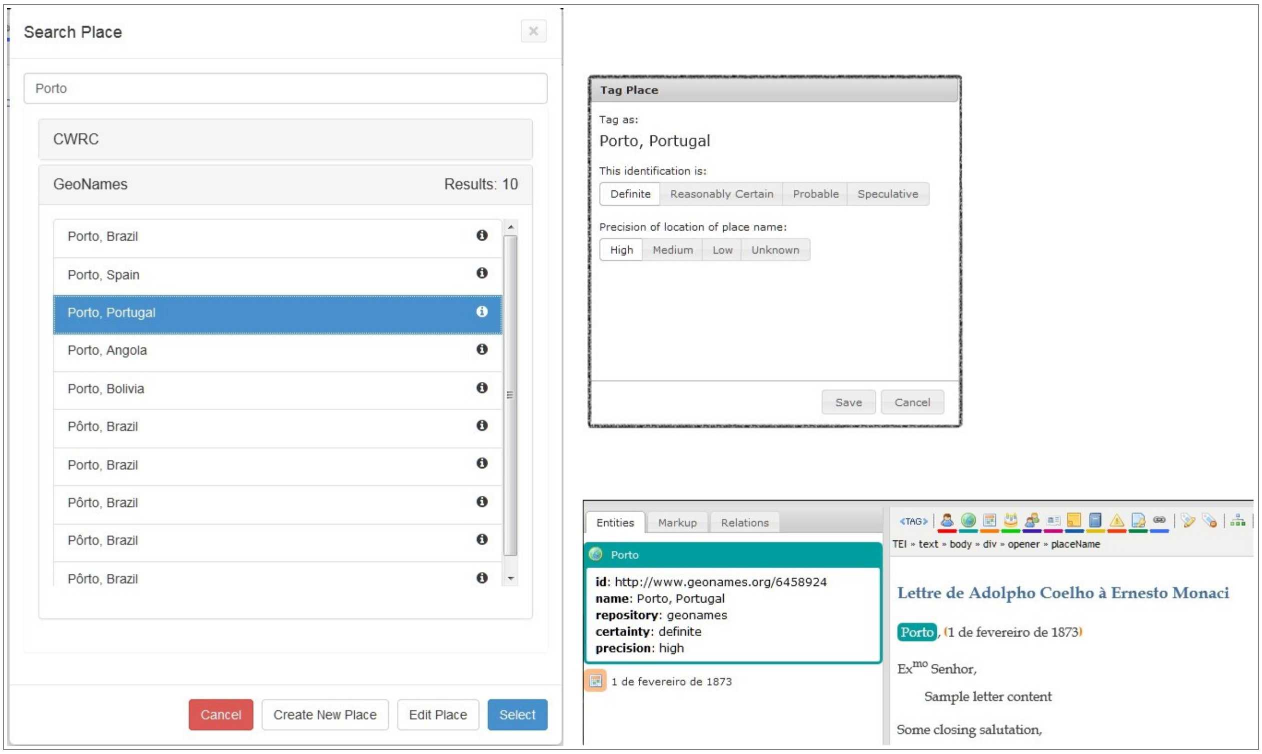
This identification is: Definite

Figure 12: CWRC-Writer. Date entity.

Authority lists are available while creating an entity: internal lists, VIAF[21] and GeoNames[22]. Any name that is not in the lists can be added to the internal lists, creating a new authority.

[21] The Virtual International Authority File <http://viaf.org/>.

[22] The GeoNames geographical database <http://www.geonames.org/>.



Search Place

Porto

CWRC

GeoNames Results: 10

- Porto, Brazil
- Porto, Spain
- Porto, Portugal
- Porto, Angola
- Porto, Bolivia
- Pôrto, Brazil
- Porto, Brazil
- Pôrto, Brazil
- Pôrto, Brazil
- Pôrto, Brazil

Tag Place

Tag as:
Porto, Portugal

This identification is: Definite | Reasonably Certain | Probable | Speculative

Precision of location of place name: High | Medium | Low | Unknown

TEI » text » body » div » opener » placeName

Porto
id: http://www.geonames.org/6458924
name: Porto, Portugal
repository: geonames
certainty: definite
precision: high

1 de fevereiro de 1873

Lettre de Adolpho Coelho à Ernesto Monaci

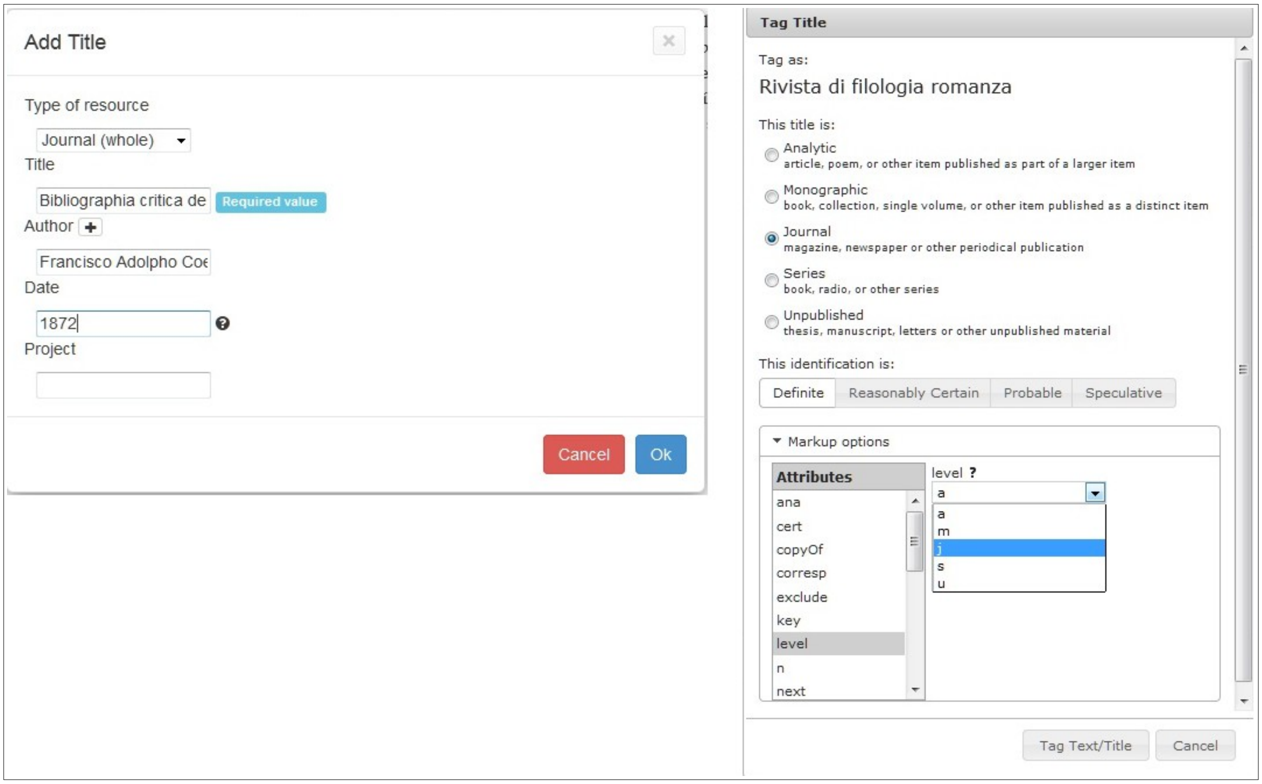
Porto, (1 de fevereiro de 1873)

Ex^{mo} Senhor,

Sample letter content

Some closing salutation,

Figure 13: CWRC-Writer. Steps for the creation of a Place entity.



Add Title

Type of resource: Journal (whole)

Title: Bibliographia critica de Required value

Author: Francisco Adolpho Coe

Date: 1872

Project:

Tag Title

Tag as:
Rivista di filologia romanza

This title is:

○ Analytic
article, poem, or other item published as part of a larger item

○ Monographic
book, collection, single volume, or other item published as a distinct item

● Journal
magazine, newspaper or other periodical publication

○ Series
book, radio, or other series

○ Unpublished
thesis, manuscript, letters or other unpublished material

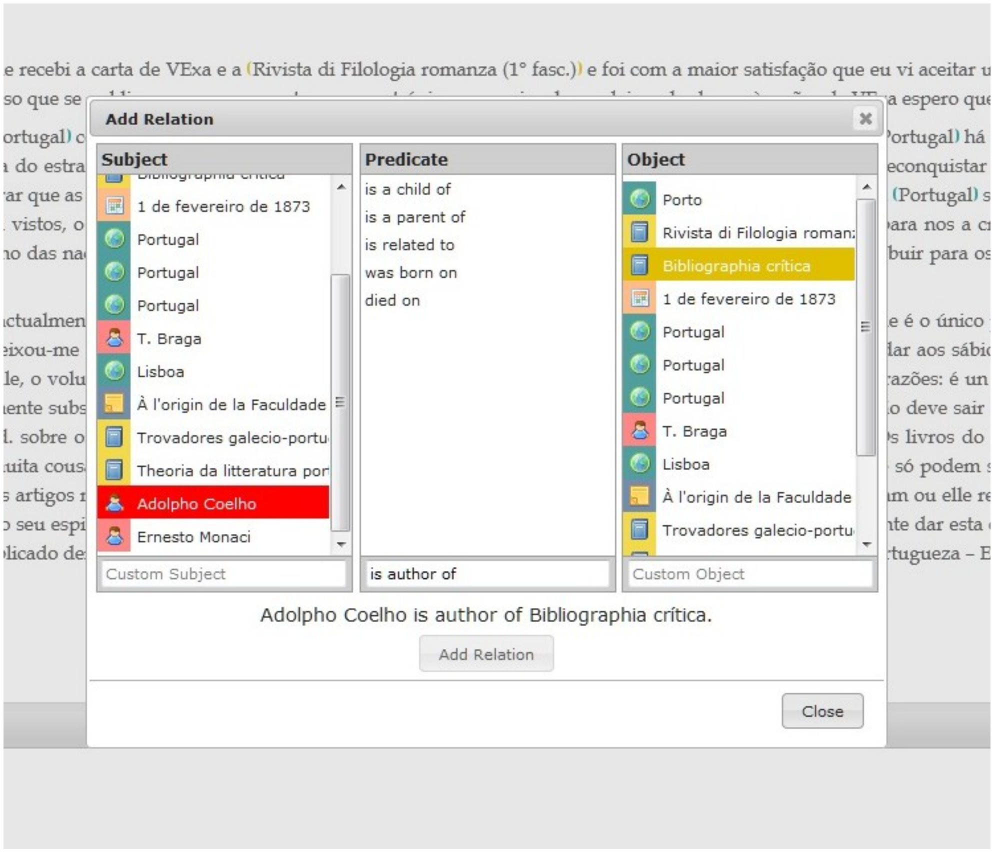
This identification is: Definite | Reasonably Certain | Probable | Speculative

Markup options

Attributes: ana, cert, copyOf, corresp, exclude, key, level, n, next

level ? a
a
m
j
s
u

Figure 14: CWRC-Writer. Creation of a new Title authority.

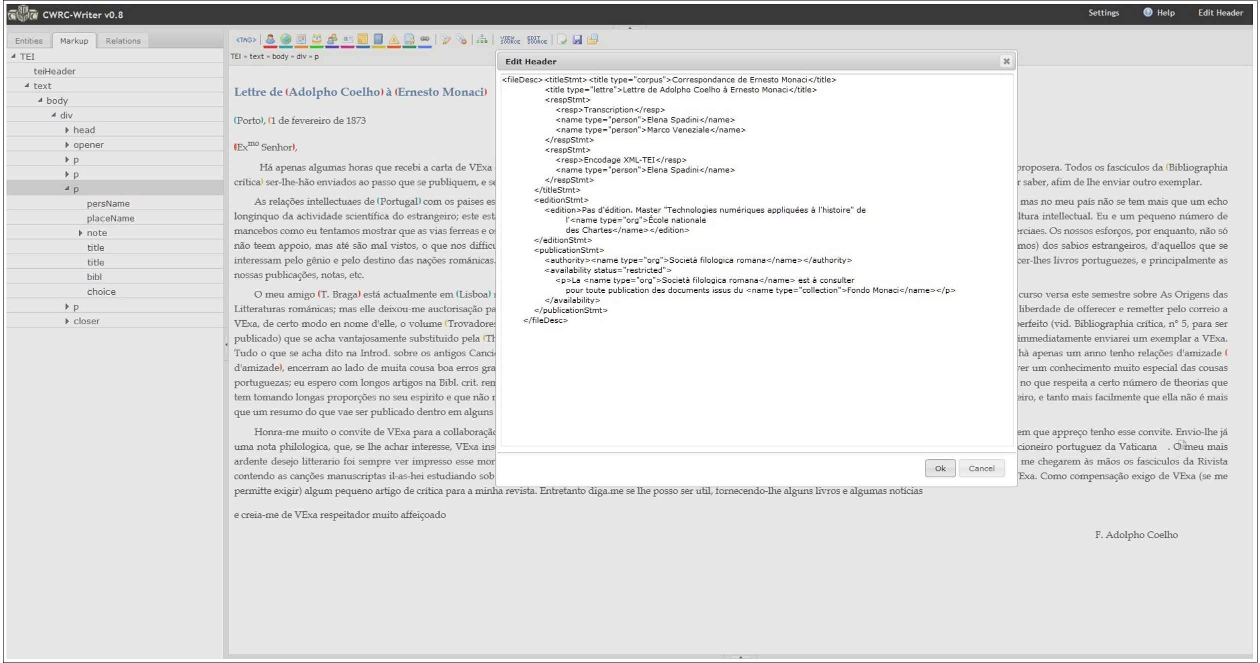


Add Relation

Subject	Predicate	Object
Bibliographia crítica	is a child of	Porto
1 de fevereiro de 1873	is a parent of	Rivista di Filologia romanz
Portugal	is related to	Bibliographia crítica
Portugal	was born on	1 de fevereiro de 1873
Portugal	died on	Portugal
T. Braga		Portugal
Lisboa		Portugal
À l'origin de la Faculdade		T. Braga
Trovadores galecio-portu		Lisboa
Theoria da litteratura por		À l'origin de la Faculdade
Adolpho Coelho		Trovadores galecio-portu
Ernesto Monaci		
Custom Subject	is author of	Custom Object

Adolpho Coelho is author of Bibliographia crítica.

Figure 15: CWRC-Writer. Add RDF.



Edit Header

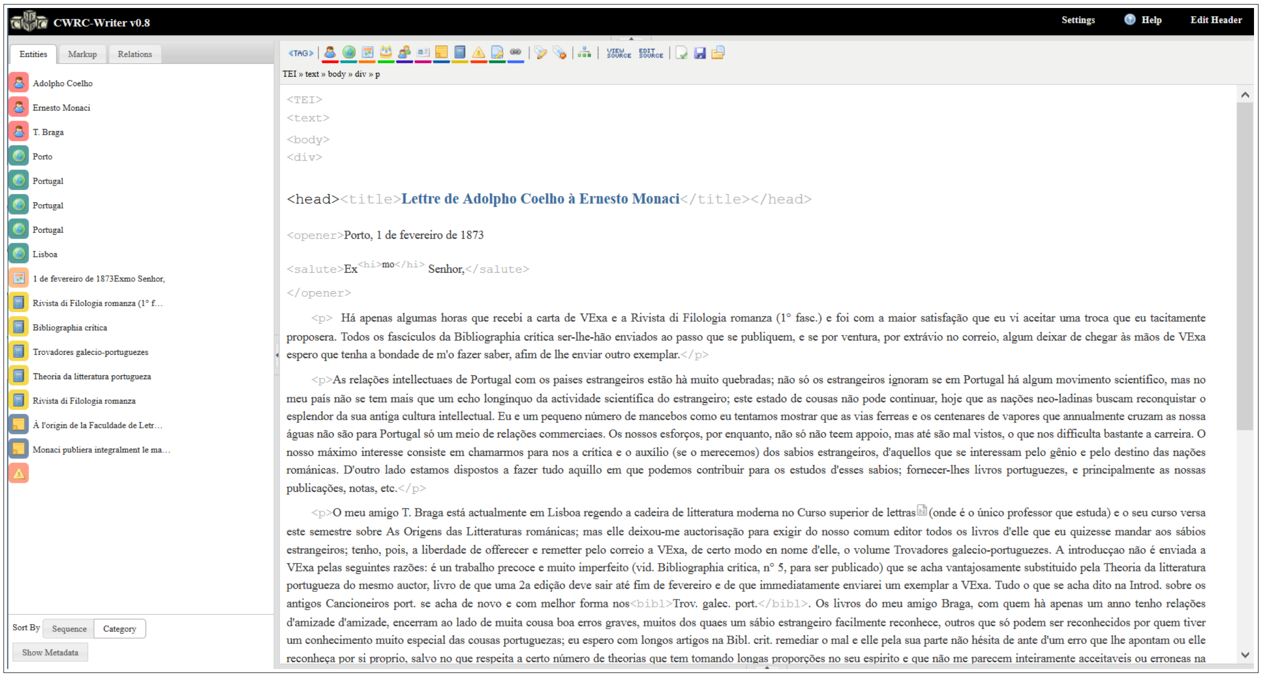
```
<fileDesc><titleStmt><title type="corpus">Correspondance de Ernesto Monaci</title>
            <title type="lettre">Lettre de Adolpho Coelho à Ernesto Monaci</title>
            <respStmt>
               <resp>Transcription</resp>
               <name type="person">Elena Spadini</name>
               <name type="person">Marco Veneziale</name>
            </respStmt>
            <respStmt>
               <resp>Encodage XML-TEI</resp>
               <name type="person">Elena Spadini</name>
            </respStmt>
         </titleStmt>
         <editionStmt>
            <edition>Pas d'édition. Master "Technologies numériques appliquées à l'histoire" de
                  l'<name type="org">École nationale
                  des Chartes</name></edition>
         </editionStmt>
         <publicationStmt>
            <authority><name type="org">Società filologica romana</name></authority>
            <availability status="restricted">
               <p>La <name type="org">Società filologica romana</name> est à consulter
                  pour toute publication des documents issus du <name type="collection">Fondo Monaci</name></p>
            </availability>
         </publicationStmt>
      </fileDesc>
```

Figure 16: CWRC-Writer. Edit the Header.

The wizard has some disadvantages: for a new title, for instance, only a limited selection of resource types are available (audio, book, correspondence, journal, manuscript, video, web resource), and an Editor field cannot replace or be added to the Author field.
Once entities have been created, users can establish relations that will be stored as RDF triples.

A dedicated dialogue box is used for editing the Header of the document.

An important functionality for advanced users is that the XML source can be shown and modified. It is possible to add an XML schema and a CSS style-sheet. One drawback is that pointing to a local resource is not allowed: like the field for associating a XML schema or a CSS style-sheet, the attribute @target for the <ref> tag, only accepts URLs.
Different views are available, with or without markup and with or without coloured brackets for entities.



Adolpho Coelho
Ernesto Monaci
T. Braga
Porto
Portugal
Portugal
Portugal
Lisboa
1 de fevereiro de 1873Exmo Senhor,
Rivista di Filologia romanza (1° f…
Bibliographia crítica
Trovadores galecio-portuguezes
Theoria da litteratura portugueza
Rivista di Filologia romanza
À l'origin de la Faculdade de Letr…
Monaci publiera integralment le ma…

```
<TEI>
<text>
<body>
<div>
```

<head><title>Lettre de Adolpho Coelho à Ernesto Monaci</title></head>

<opener>Porto, 1 de fevereiro de 1873

<salute>Ex<hi>mo</hi> Senhor,</salute>

</opener>

<p> Há apenas algumas horas que recebi a carta de VExa e a Rivista di Filologia romanza (1° fasc.) e foi com a maior satisfação que eu vi aceitar uma troca que eu tacitamente proposera. Todos os fascículos da Bibliographia crítica ser-lhe-hão enviados ao passo que se publiquem, e se por ventura, por extrávio no correio, algum deixar de chegar às mãos de VExa espero que tenha a bondade de m'o fazer saber, afim de lhe enviar outro exemplar.</p>

<p>As relações intellectuaes de Portugal com os paises estrangeiros estão hà muito quebradas; não só os estrangeiros ignoram se em Portugal há algum movimento scientífico, mas no meu país não se tem mais que um echo longínquo da actividade scientífica do estrangeiro; este estado de cousas não pode continuar, hoje que as nações neo-ladinas buscam reconquistar o esplendor da sua antiga cultura intellectual. Eu e um pequeno número de mancebos como eu tentamos mostrar que as vias ferreas e os centenares de vapores que annualmente cruzam as nossa águas não são para Portugal só um meio de relações commerciaes. Os nossos esforços, por enquanto, não só não teem appoio, mas até são mal vistos, o que nos difficulta bastante a carreira. O nosso máximo interesse consiste em chamarmos para nos a crítica e o auxílio (se o merecemos) dos sabios estrangeiros, d'aquellos que se interessam pelo gênio e pelo destino das nações románicas. D'outro lado estamos dispostos a fazer tudo aquillo em que podemos contribuir para os estudos d'esses sabios; fornecer-lhes livros portuguezes, e principalmente as nossas publicações, notas, etc.</p>

<p>O meu amigo T. Braga está actualmente em Lisboa regendo a cadeira de litteratura moderna no Curso superior de lettras (onde é o único professor que estuda) e o seu curso versa este semestre sobre As Origens das Litteraturas románicas; mas elle deixou-me auctorisação para exigir do nosso comum editor todos os livros d'elle que eu quizesse mandar aos sábios estrangeiros; tenho, pois, a liberdade de offerecer e remetter pelo correio a VExa, de certo modo en nome d'elle, o volume Trovadores galecio-portuguezes. A introducçao não é enviada a VExa pelas seguintes razões: é un trabalho precoce e muito imperfeito (vid. Bibliographia crítica, n° 5, para ser publicado) que se acha vantajosamente substituido pela Theoria da litteratura portugueza do mesmo auctor, livro de que uma 2a edição deve sair até fim de fevereiro e de que immediatamente enviarei um exemplar a VExa. Tudo o que se acha dito na Introd. sobre os antigos Cancioneiros port. se acha de novo e com melhor forma nos<bibl>Trov. galec. port.</bibl>. Os livros do meu amigo Braga, com quem hà apenas um anno tenho relações d'amizade d'amizade, encerram ao lado de muita cousa boa erros graves, muitos dos quaes um sábio estrangeiro facilmente reconhece, outros que só podem ser reconhecidos por quem tiver um conhecimento muito especial das cousas portuguezas; eu espero com longos artigos na Bibl. crit. remediar o mal e elle pela sua parte não hésita de ante d'um erro que lhe apontam ou elle reconheça por si proprio, salvo no que respeita a certo número de theorias que tem tomando longas proporções no seu espirito e que não me parecem inteiramente acceitaveis ou erroneas na

Figure 17: CWRC-Writer. Document with tags, entities in the left column.

CWRC-W does not offer an Undo button or export services so far. Facsimiles management and view are absent. Nevertheless, it stands out among the selected tools because of its powerful and user-friendly functionalities for the encoding and the semantic enrichment of the text.

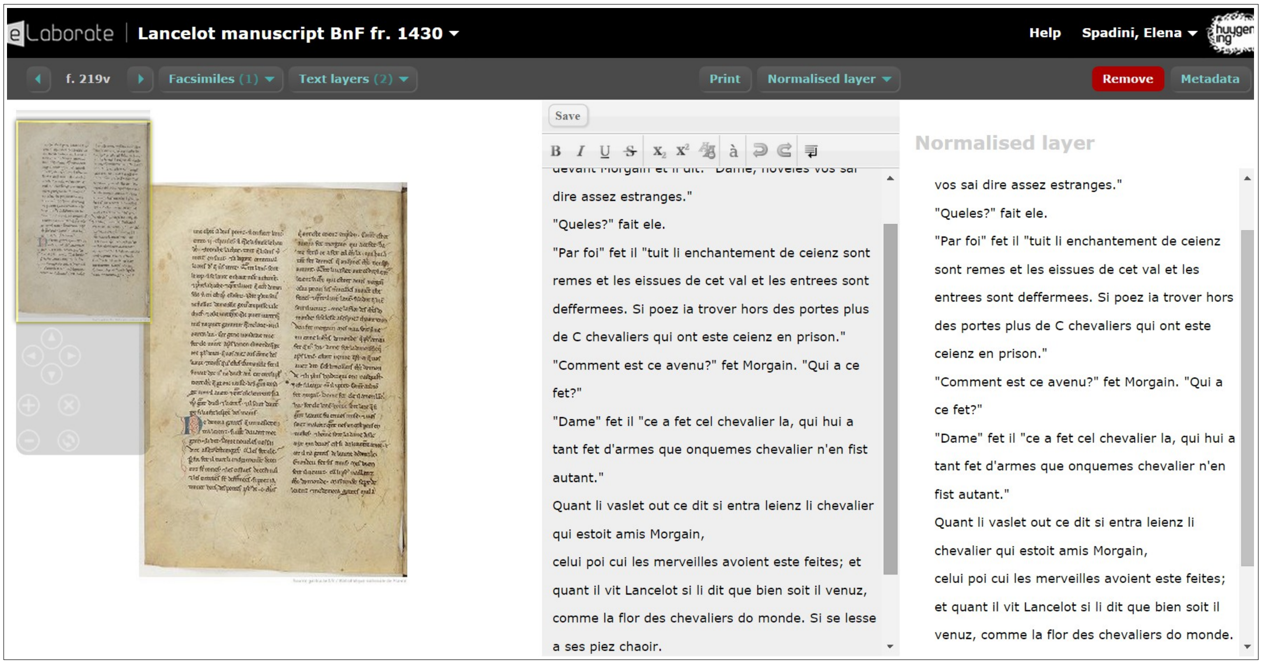
3.2. Platforms.

3.2.1. eLaborate 4.

eLaborate[23] is a web-based editing environment developed at Huygens ING – KNAW since 2003; the fourth version was released in April 2014. It allows editors with little computer literacy to create digital scholarly editions and enables collaborative editing. In the workbench, users can upload images, transcribe and annotate the text and publish the results.

In eLaborate, the text is organised in entries; each entry corresponds to the page of a manuscript or to a column or to a document (e.g. a letter).

Once an entry has been created, users can add one or more facsimiles to it, choose the number and the names of the layers (e.g. diplomatic, normalised, translation), manage metadata and insert transcriptions. Three windows for transcribing are open in the work environment: a column for the image on the left, one column for editing the transcription in the centre and, on the right, one column for a preview in which annotations can be inserted. The latter two windows are synchronized while scrolling down the text.



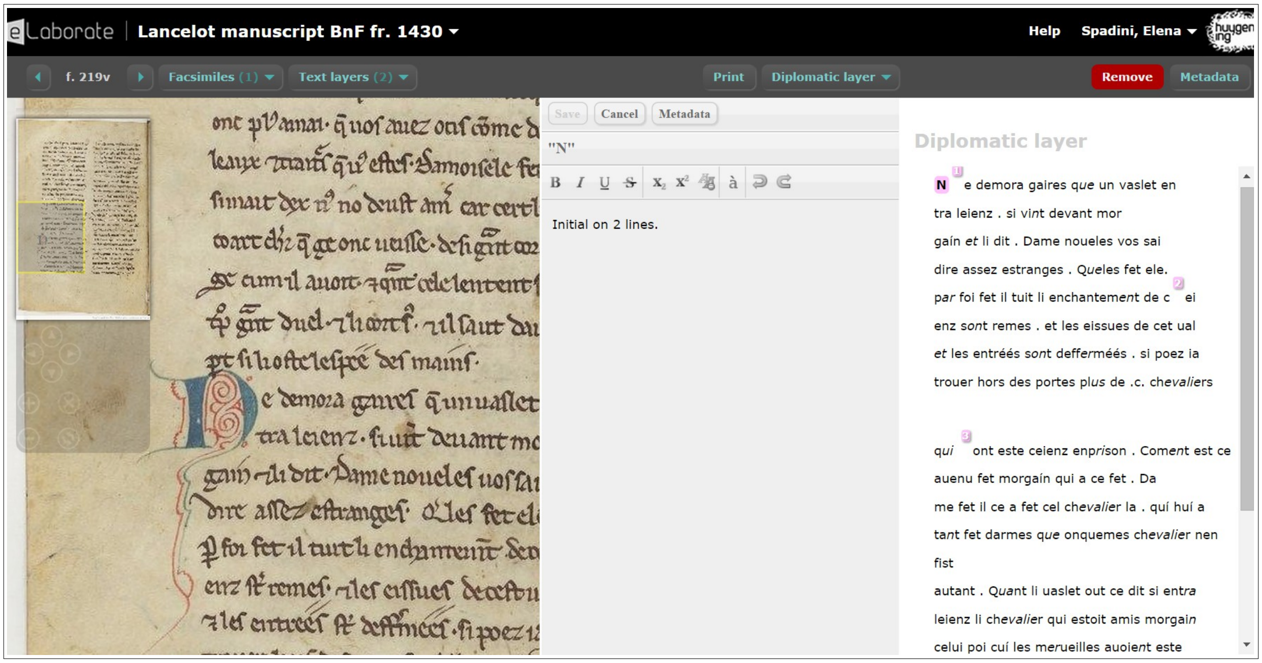
devant Morgain et li dit: "Dame, noveles vos sai
dire assez estranges."
"Queles?" fait ele.
"Par foi" fet il "tuit li enchantement de ceienz sont remes et les eissues de cet val et les entrees sont deffermees. Si poez ia trover hors des portes plus de C chevaliers qui ont este ceienz en prison."
"Comment est ce avenu?" fet Morgain. "Qui a ce fet?"
"Dame" fet il "ce a fet cel chevalier la, qui hui a tant fet d'armes que onquemes chevalier n'en fist autant."
Quant li vaslet out ce dit si entra leienz li chevalier qui estoit amis Morgain,
celui poi cui les merveilles avoient este feites; et quant il vit Lancelot si li dit que bien soit il venuz, comme la flor des chevaliers do monde. Si se lesse a ses piez chaoir.

vos sai dire assez estranges."
"Queles?" fait ele.
"Par foi" fet il "tuit li enchantement de ceienz sont remes et les eissues de cet val et les entrees sont deffermees. Si poez ia trover hors des portes plus de C chevaliers qui ont este ceienz en prison."
"Comment est ce avenu?" fet Morgain. "Qui a ce fet?"
"Dame" fet il "ce a fet cel chevalier la, qui hui a tant fet d'armes que onquemes chevalier n'en fist autant."
Quant li vaslet out ce dit si entra leienz li chevalier qui estoit amis Morgain,
celui poi cui les merveilles avoient este feites; et quant il vit Lancelot si li dit que bien soit il venuz, comme la flor des chevaliers do monde.

Figure 18: eLaborate. Transcription area.

A toolbar is available in the edit window for styles (italic, bold, underline, strike-through, subscript, superscript) and special characters, for undoing or redoing, and for wrapping the text (without changing the line breaks that the editor may have inserted manually). A double click on a word in the right column will add a note. Styles and notes allow the user to record abbreviations, deletion, shift of hands and similar features.

[23] <http://elaborate.huygens.knaw.nl/>.



"N"

Initial on 2 lines.

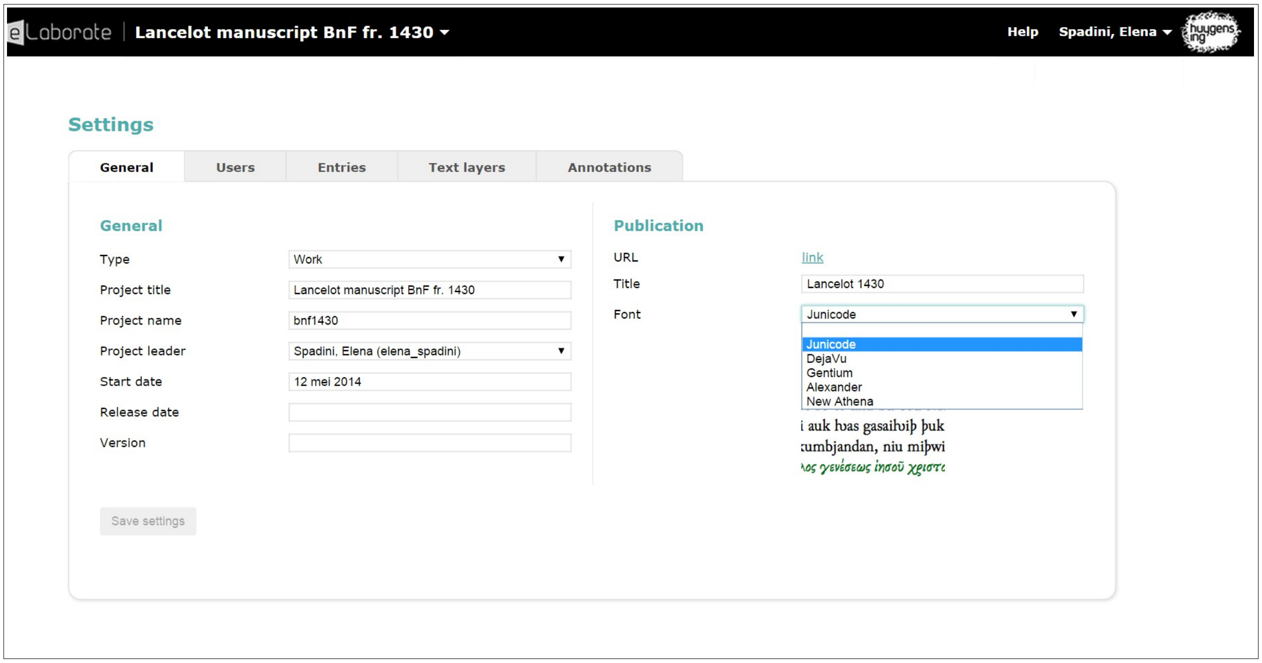
Diplomatic layer

N [1] e demora gaires que un vaslet en
tra leienz . si vint devant mor
gain *et* li dit . Dame noueles vos sai
dire assez estranges . Queles fet ele.
par foi fet il tuit li enchantement de c [2] ei
enz *sont* remes . et les eissues de cet ual
et les entréés *sont* deffermééś . si poez ia
trouer hors des portes *plus* de .c. *chevaliers*

[3]
qui ont este ceienz enpr*ison* . Coment est ce
auenu fet morgain qui a ce fet . Da
me fet il ce a fet cel *chevalier* la . qui huí a
tant fet darmes *que* onquemes *chevalier* nen
fist
autant . *Quant* li uaslet out ce dit si entra
leienz li *chevalier* qui estoit amis morgain
celui poi cuí les merueilles auoient este

Figure 19: eLaborate. Adding note.

In the Settings page, the core of the application, the project's leader inserts metadata and criteria for the publication, manages collaborators' rights and edits the parameters for the entries' metadata (name and order of the fields to fill).



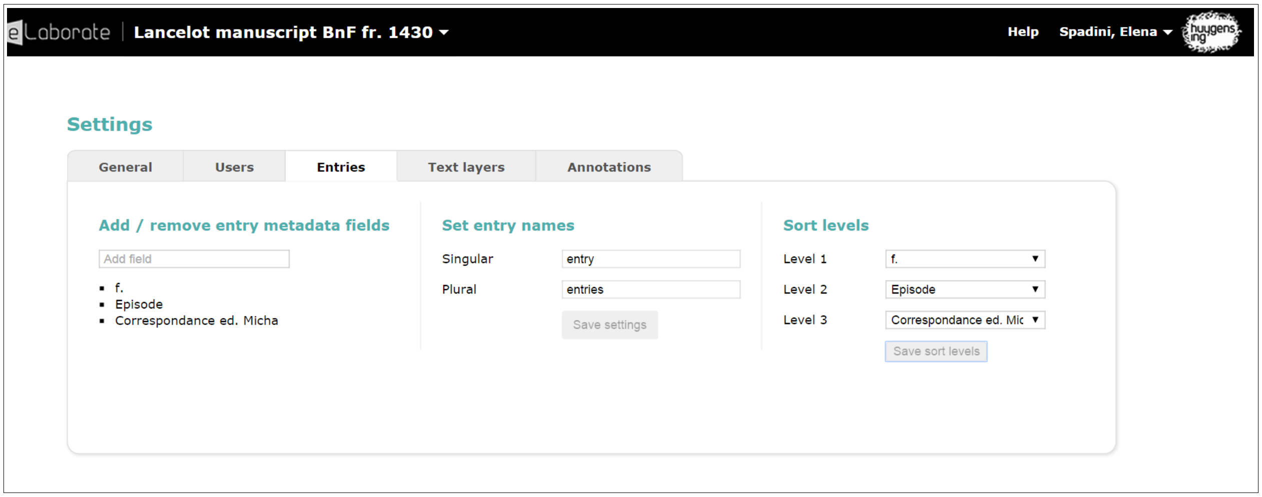
General

Type: Work
Project title: Lancelot manuscript BnF fr. 1430
Project name: bnf1430
Project leader: Spadini, Elena (elena_spadini)
Start date: 12 mei 2014
Release date:
Version:

Publication

URL: link
Title: Lancelot 1430
Font: Junicode

Figure 20: eLaborate. Settings.



Settings

Add / remove entry metadata fields

- f.
- Episode
- Correspondance ed. Micha

Set entry names

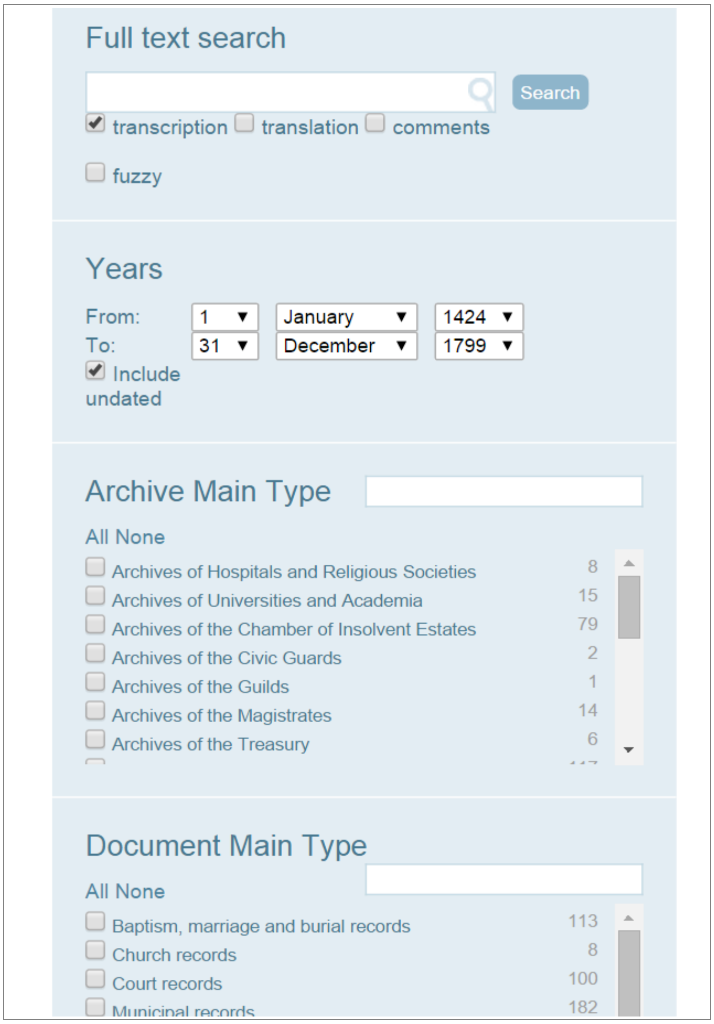
Singular	entry
Plural	entries

Sort levels

Level 1	f.
Level 2	Episode
Level 3	Correspondance ed. Mic

Figure 21: eLaborate. Entries.

Specific metadata for each entry can be very useful for the organization (and retrieval) of different materials, as shown by the *Rembrandt Documents Project*, powered by eLaborate3.



Full text search

☑ transcription ☐ translation ☐ comments

☐ fuzzy

Years

From: 1 ▾ January ▾ 1424 ▾
To: 31 ▾ December ▾ 1799 ▾
☑ Include undated

Archive Main Type

☐ Archives of Hospitals and Religious Societies	8
☐ Archives of Universities and Academia	15
☐ Archives of the Chamber of Insolvent Estates	79
☐ Archives of the Civic Guards	2
☐ Archives of the Guilds	1
☐ Archives of the Magistrates	14
☐ Archives of the Treasury	6

Document Main Type

☐ Baptism, marriage and burial records	113
☐ Church records	8
☐ Court records	100
☐ Municipal records	182

Figure 22: eLaborate. The Digital Laboratory Rembrandt Documents Project.

Again in the Settings page, users can edit the number and names of the text layers. Annotations can be organized in categories. The weak point of this functionality is that different kinds of annotations (e.g. structural, semantic, internal) are mixed on the same level.

In the History page, users will easily find all past revisions.

The application uses a WordPress environment for adding editorial materials as a forward and introduction. Users will appreciate the functionality Publish Draft, which shows a preview of the publication. No export functionality is available through the interface.

Multiple editorial tasks can be accomplished using eLaborate, from managing the facsimile to publication, though encoding is not supported. It allows editors with little technical skill to create digital editions. Its strong points are the management of metadata and collaborative work. Nevertheless, eLaborate is not suitable in certain cases: multiple witnesses and aligned synoptic edition, complex layers of annotations, need to insert non-textual objects into the annotations.

3.2.2. TextGrid.

TextGrid[24] is a joint project of ten partners, funded by the German Federal Ministry of Education and Research (BMBF) for the period between June 2012 and May 2015.

It provides an infrastructure with a strong focus on text-oriented research; it has a grid-capable architecture that assures interoperability and homogeneity throughout the complete workflow, including editing, storing and publishing; one of its strengths is that it enables collaborative work, associating roles and rights with users.

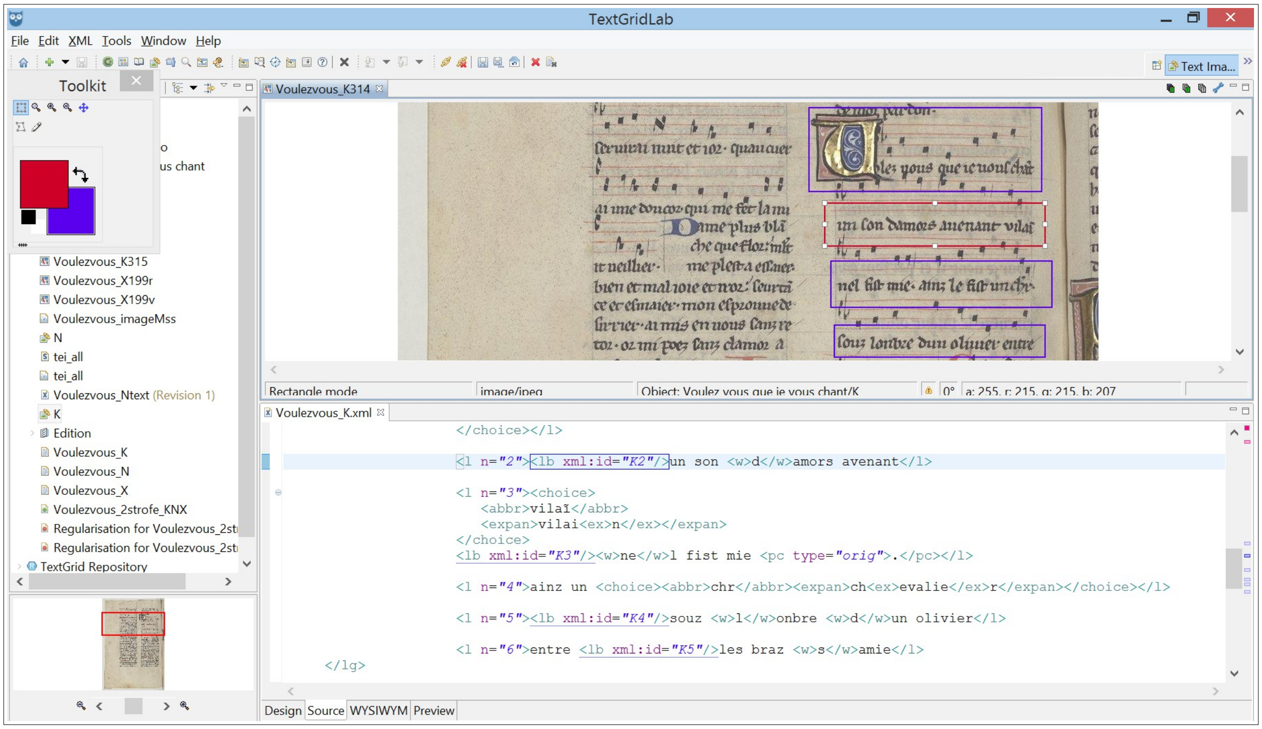
TextGrid is an open source software, developed in Eclipse. It is portable, and thus does not require any modification to the operating system; connection with the server is needed for the authentication.

The two main components are the Repository and the Laboratory. The first is a long-term archive; users can store, publish and research in the server-based Repository. It also includes an extensive collection of German texts from the Digital Library at zeno.org. The Laboratory is a work environment that combines several tools and services, such as an XML Editor, a Metadata Editor, a Text-Image-Link Editor, dictionaries, a German Lemmatizer, a Workflow Tool, the DigiLib viewer, the SADE Publish Tool, import and export facilities.

Text Grid's main elements are the objects and the relations between them (stored in RDF triples), which constitute objects too. Objects are Items, Aggregations, Works, Editions and Collections. All of them must belong to a Project, which serves as a container for the management rights. An Item is the smallest unit; it can be, for instance, a plain text or XML document, a DTD, a CSS style-sheet or an image. Some examples of the relations between objects are: a Collection may contain Editions or gather Items; an Edition is a Work's manifestation.

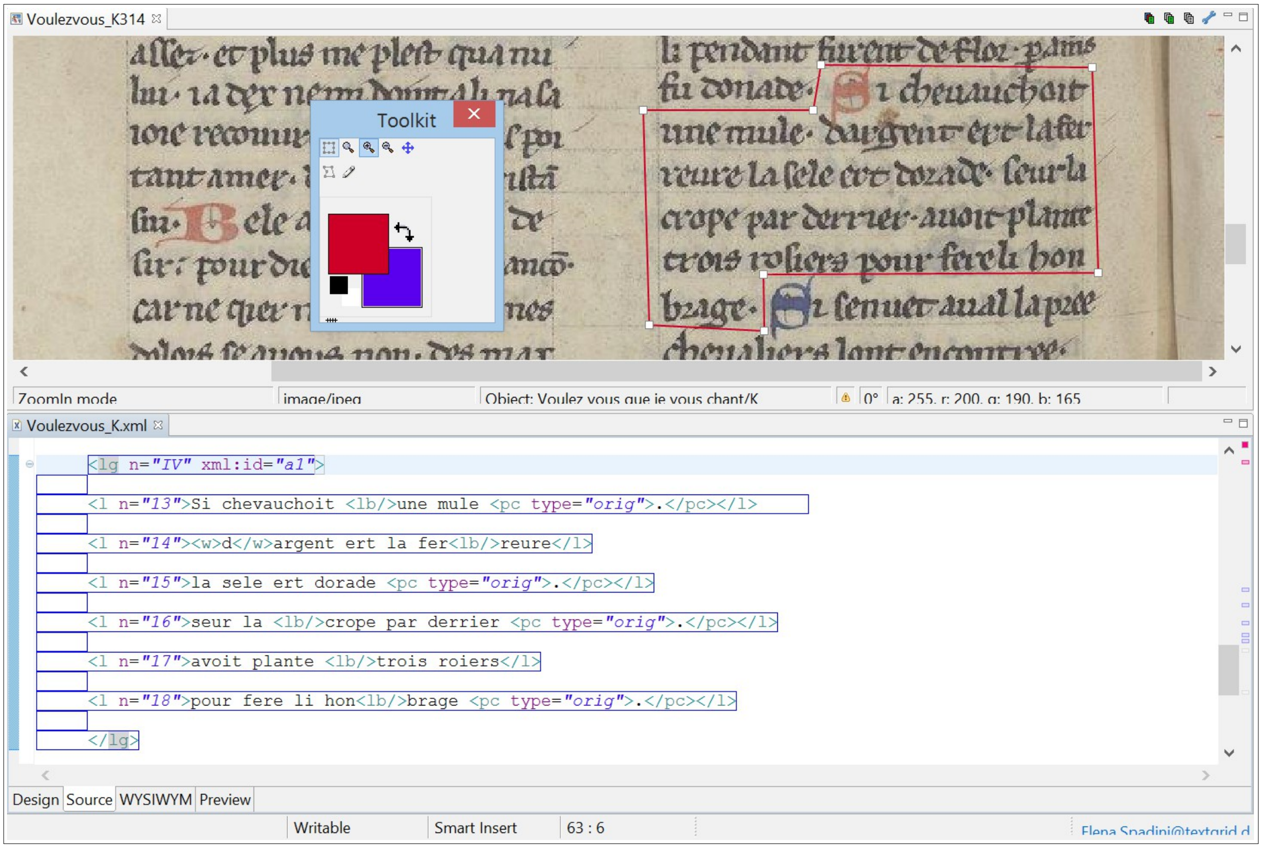
[24] <https://textgrid.de/>.

The XML Editor presents several functionalities: visualization of the XML structure, schema association and validation, content completion, WYSIWYM view (standard style-sheet) and preview (customized style-sheet). A Unicode table is available on the left column. Metadata can be inserted in the TeiHeader, or in the Metadata Editor and automatically transferred in the TeiHeader.
I will briefly describe the Text-Image-Link Editor. The chosen image is on the top part of the screen with the text below. The user selects a portion of both in order to create a link. If the selected text does not correspond to a tag, two empty <anchor> tags will be automatically created, with xml:id attributes like @xml:id="a6_start" and @xml:id="a6_end"; if the selected text does correspond to a tag, the xml:id attribute will be automatically added; if there already is an identifier, the image will be linked without any additions. The selected part of the image can have different shapes: the editor may decide to select lines, stanzas, etc. The XML document, the image and the links between them will be included while exporting a Text-Image-Link Object.



```
          </choice></l>

          <l n="2"><lb xml:id="K2"/>un son <w>d</w>amors avenant</l>

          <l n="3"><choice>
             <abbr>vilaĩ</abbr>
             <expan>vilai<ex>n</ex></expan>
          </choice>
          <lb xml:id="K3"/><w>ne</w>l fist mie <pc type="orig">.</pc></l>

          <l n="4">ainz un <choice><abbr>chr</abbr><expan>ch<ex>evalie</ex>r</expan></choice></l>

          <l n="5"><lb xml:id="K4"/>souz <w>l</w>onbre <w>d</w>un olivier</l>

          <l n="6">entre <lb xml:id="K5"/>les braz <w>s</w>amie</l>
     </lg>
```

Figure 23: TextGrid. Text-Image-Link (1).



```
<lg n="IV" xml:id="a1">
    <l n="13">Si chevauchoit <lb/>une mule <pc type="orig">.</pc></l>
    <l n="14"><w>d</w>argent ert la fer<lb/>reure</l>
    <l n="15">la sele ert dorade <pc type="orig">.</pc></l>
    <l n="16">seur la <lb/>crope par derrier <pc type="orig">.</pc></l>
    <l n="17">avoit plante <lb/>trois roiers</l>
    <l n="18">pour fere li hon<lb/>brage <pc type="orig">.</pc></l>
</lg>
```

Figure 24: TextGrid. Text-Image-Link (2)

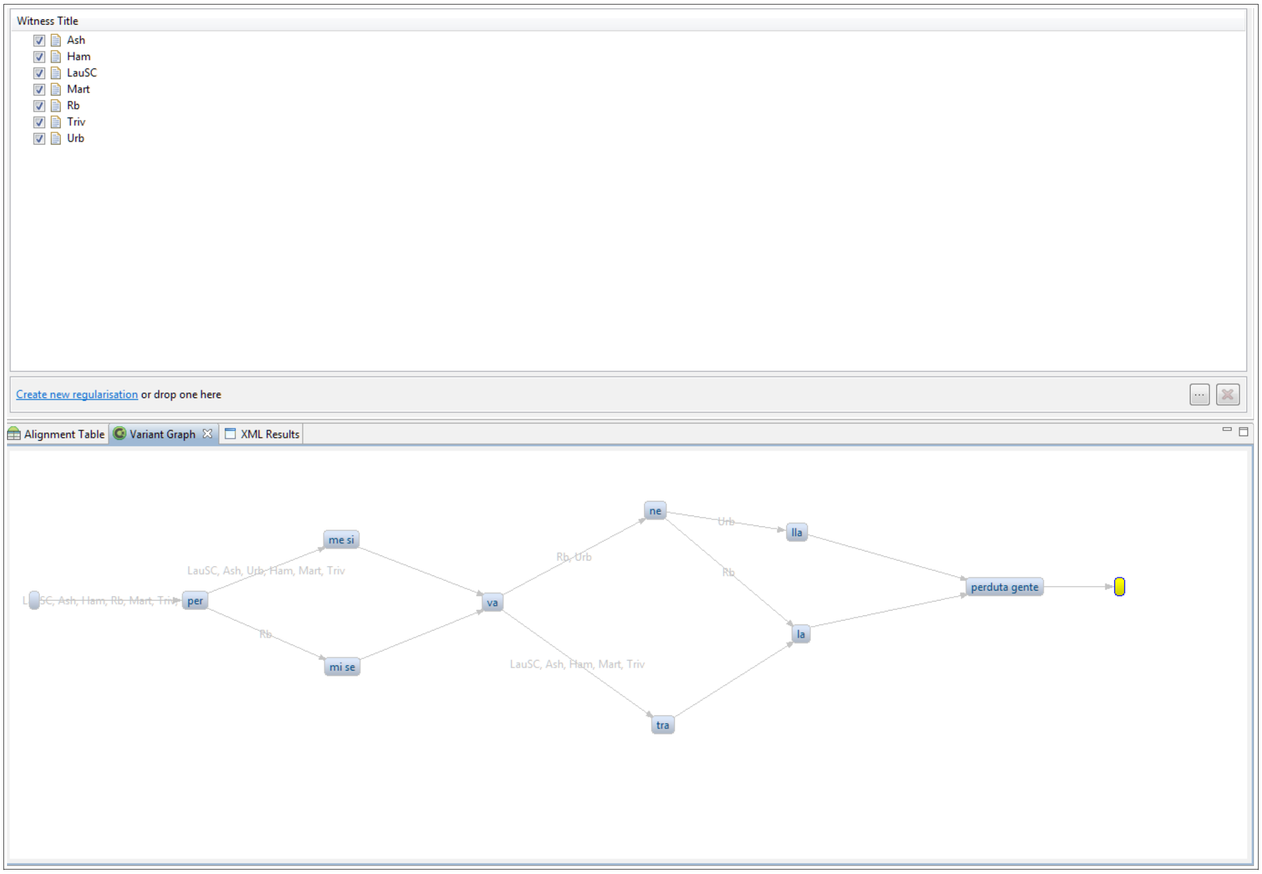
Overall, TextGrid is the most complete application among those analysed here: it includes a well-structured editing environment, linked to the storing service of the repository; export is available in plain text and XML and a publishing tool is integrated. As regards the transcription and the encoding, it merges the functionalities of a XML editor with those of the Text-Image-Link.

Beyond transcription and encoding: use CollateX in TextGrid

In addition to the basic TextGrid tools, others can be connected using the Eclipse Marketplace.

CollateX is a recently released and already popular collation tool, the designated successor of Peter Robinson's Collate. It was developed within the EU-funded initiative Interedition and under the leadership of Ronald Haentjens Dekker and Gregor Middell. It is an open source program, available in Java and Python versions.

In TextGrid, CollateX runs on a collation set, which gathers plain text documents. The results can be shown as an alignment table, a variant graph or in XML (following the guidelines of the TEI module Critical Apparatus).

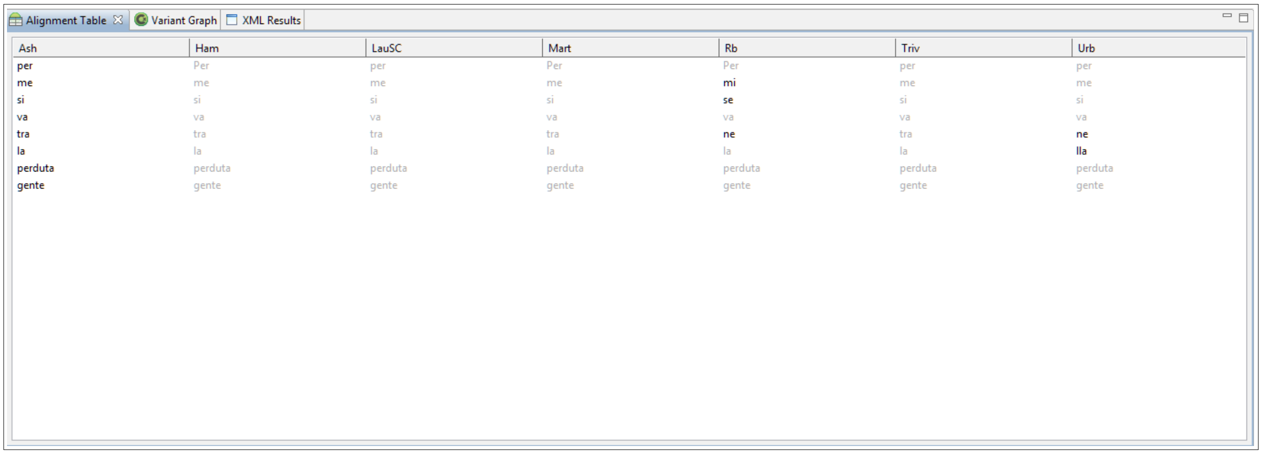


Witness Title

- ☑ Ash
- ☑ Ham
- ☑ LauSC
- ☑ Mart
- ☑ Rb
- ☑ Triv
- ☑ Urb

Create new regularisation or drop one here

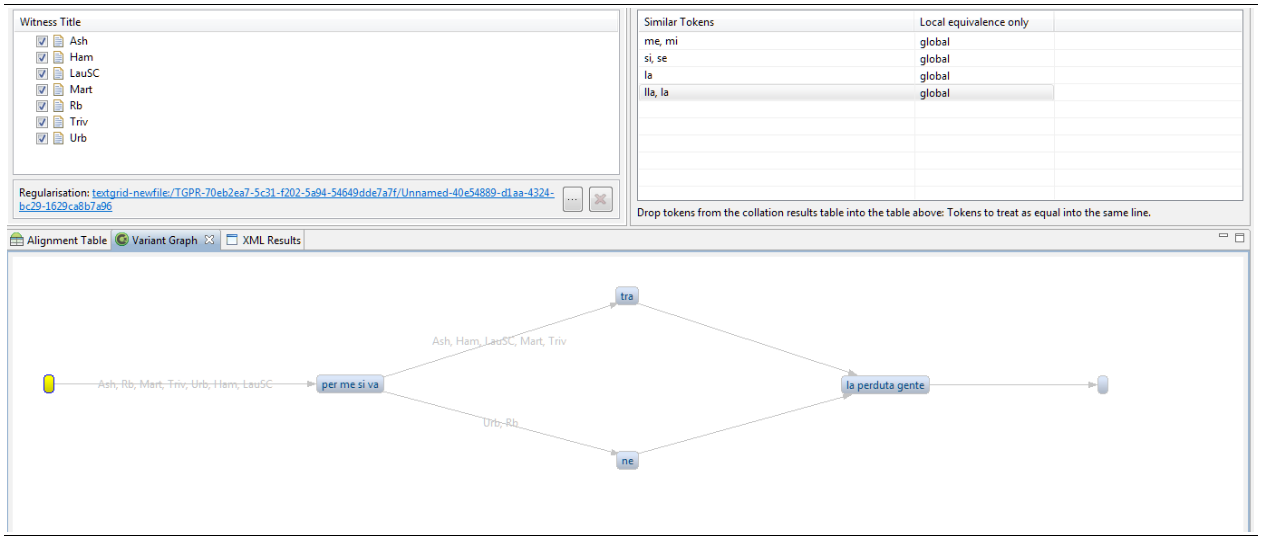
Figure 25: TextGrid. CollateX variant graph. Data from Inf., III, 3 (Dante Alighieri: *Commedia. A Digital Edition*, ed. P. Shaw).



Ash	Ham	LauSC	Mart	Rb	Triv	Urb
per	Per	per	Per	Per	per	per
me	me	me	me	mi	me	me
si	si	si	si	se	si	si
va	va	va	va	va	va	va
tra	tra	tra	tra	ne	tra	ne
la	la	la	la	la	la	lla
perduta	perduta	perduta	perduta	perduta	perduta	perduta
gente	gente	gente	gente	gente	gente	gente

Figure 26: TextGrid. CollateX alignment table.

Users can establish an equivalence set, which determines the items that should be treated as identical; variance between them will be ignored by the tool.



Witness Title

- ☑ Ash
- ☑ Ham
- ☑ LauSC
- ☑ Mart
- ☑ Rb
- ☑ Triv
- ☑ Urb

Regularisation: textgrid-newfile:/TGPR-70eb2ea7-5c31-f202-5a94-54649dde7a7f/Unnamed-40e54889-d1aa-4324-bc29-1629ca8b7a96

Similar Tokens	Local equivalence only	
me, mi	global	
si, se	global	
la	global	
lla, la	global	

Drop tokens from the collation results table into the table above: Tokens to treat as equal into the same line.

Figure 27: TextGrid. CollateX results after regularisation according to the equivalence set. Variant graph.

3.2.3. Ecdosis.

Ecdosis is an application developed by Desmond Schmidt (AustEse), in collaboration with the University of Sassari[25]. It is intended as a complete editing and publishing environment. Given Ecdosis is still under development[26], it was not possible to test it; thus I will only present its structure and some of its features. Because of its peculiarities, it is nevertheless worth including it in this overview.

The internal format of Ecdosis is the MultiVersion Document (MVD), developed in order to overcome the limitations of XML technologies together with, for instance, the overlapping problem. Schmidt and Colomb affirm: 'it will not be necessary to seek a solution to the overlapping hierarchies problem if a solution can be found to the textual variation problem. Any solution to the latter is also a solution to the former' [Schmidt-Colomb 2009]. The solution is the MVD format, which uses 'the list form of the variant graph as the basis for an encoding of a single work, in all its versions or markup perspectives, as a single digital entity. [...] It enables a work to be viewed and searched, its versions compared and edited as one file'. Therefore no additional tool for collation is needed, as in the file, 'the relationships between various parts of each version, the what-is-a-variant-of-what information, is also recorded' (*ibid.*). One of the weak points of the MVD format seems to be that there are no versions for sections of the texts (words, characters, paragraphs, etc.), but only for the whole text. When dealing, for example, with a genetic edition, it is difficult to decide whether two deletions at different points of the text belong to the same version, or layer, as this may assume that the two deletions were made at the same time. Though this might present a problem for the edition of modern manuscripts, the format

[25] <http://www.ecdosis.net/main/>.

[26] As for March 2015.

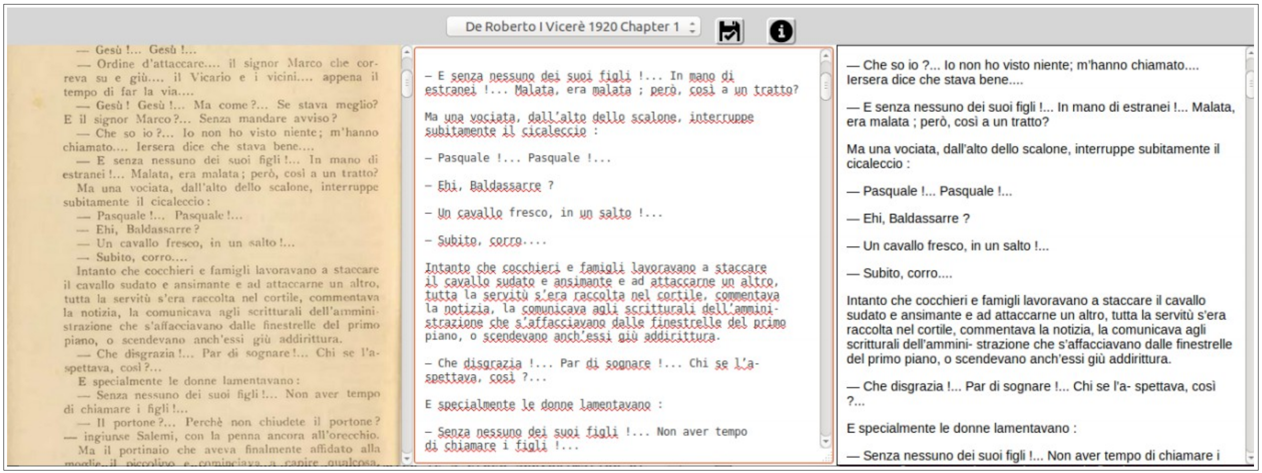
seems very suitable for recording, processing and editing works preserved in many witnesses and characterized by widespread variation, such as a number of medieval texts. A counterpart is that the usual distinction between a diplomatic and a normalized edition is hard to handle with MVD, because there is no difference between two versions of the same manuscript and two versions of the same work.

The MVD stands as the internal format of the editing platform, which accepts several input formats and provides output ones. When importing an XML document into Ecdosis, the program processes the encoding in order to create autonomous versions: each layer of nested editorial (but not structural) markup corresponds to a version.

The Ecdosis platform consist of two websites: one for the editors and one for the readers; as already mentioned, it is also a publishing platform.

In the back-ends (the editors' side), there is a MML editor, an Event editor, a TILT editor, the Doc/Image management, the Project management and a Plain Doc editor.

The Minimal Markup Editor (MML) is a Markdown editor customizable for each project; because of the separation into layers (versions), less markup is needed. The MML handles hyphenation at line-end intelligently, converts quote-marks automatically, allows user to create ‘dividers’ between sections and multiple indentation for poetry.



— Gesù !... Gesù !...

— Ordine d'attaccare.... il signor Marco che correva su e giù.... il Vicario e i vicini.... appena il tempo di far la via....

— Gesù ! Gesù !... Ma come ?... Se stava meglio? E il signor Marco ?... Senza mandare avviso ?

— Che so io ?... Io non ho visto niente; m'hanno chiamato.... Iersera dice che stava bene....

— E senza nessuno dei suoi figli !... In mano di estranei !... Malata, era malata ; però, così a un tratto?

Ma una vociata, dall'alto dello scalone, interruppe subitamente il cicaleccio :

— Pasquale !... Pasquale !...

— Ehi, Baldassarre ?

— Un cavallo fresco, in un salto !...

— Subito, corro....

Intanto che cocchieri e famigli lavoravano a staccare il cavallo sudato e ansimante e ad attaccarne un altro, tutta la servitù s'era raccolta nel cortile, commentava la notizia, la comunicava agli scritturali dell'amministrazione che s'affacciavano dalle finestrelle del primo piano, o scendevano anch'essi giù addirittura.

— Che disgrazia !... Par di sognare !... Chi se l'aspettava, così ?...

E specialmente le donne lamentavano :

— Senza nessuno dei suoi figli !... Non aver tempo di chiamare i figli !...

— Il portone ?... Perchè non chiudete il portone ? — ingiunse Salemi, con la penna ancora all'orecchio.

Ma il portinaio che aveva finalmente affidato alla moglie il piccolino e cominciava a capire qualcosa,

```
- E senza nessuno dei suoi figli !... In mano di
estranei !... Malata, era malata ; però, così a un tratto?

Ma una vociata, dall'alto dello scalone, interruppe
subitamente il cicaleccio :

- Pasquale !... Pasquale !...

- Ehi, Baldassarre ?

- Un cavallo fresco, in un salto !...

- Subito, corro....

Intanto che cocchieri e famigli lavoravano a staccare
il cavallo sudato e ansimante e ad attaccarne un altro,
tutta la servitù s'era raccolta nel cortile, commentava
la notizia, la comunicava agli scritturali dell'ammini-
strazione che s'affacciavano dalle finestrelle del primo
piano, o scendevano anch'essi giù addirittura.

- Che disgrazia !... Par di sognare !... Chi se l'a-
spettava, così ?...

E specialmente le donne lamentavano :

- Senza nessuno dei suoi figli !... Non aver tempo
di chiamare i figli !...
```

— Che so io ?... Io non ho visto niente; m'hanno chiamato.... Iersera dice che stava bene....

— E senza nessuno dei suoi figli !... In mano di estranei !... Malata, era malata ; però, così a un tratto?

Ma una vociata, dall'alto dello scalone, interruppe subitamente il cicaleccio :

— Pasquale !... Pasquale !...

— Ehi, Baldassarre ?

— Un cavallo fresco, in un salto !...

— Subito, corro....

Intanto che cocchieri e famigli lavoravano a staccare il cavallo sudato e ansimante e ad attaccarne un altro, tutta la servitù s'era raccolta nel cortile, commentava la notizia, la comunicava agli scritturali dell'ammini- strazione che s'affacciavano dalle finestrelle del primo piano, o scendevano anch'essi giù addirittura.

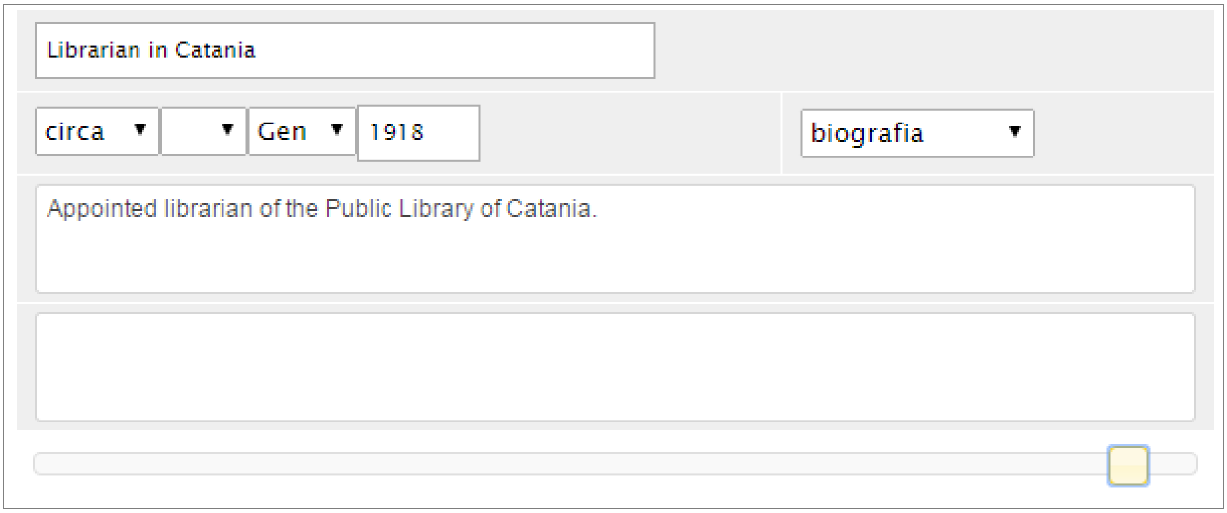
— Che disgrazia !... Par di sognare !... Chi se l'a- spettava, così ?...

E specialmente le donne lamentavano :

— Senza nessuno dei suoi figli !... Non aver tempo di chiamare i

Figure 28: Ecdosis. MML editor.

Events are biographical or publication events in the life of the author, which can be displayed in an interactive timeline or as a readable biography.



Librarian in Catania

circa ▾ | ▾ | Gen ▾ | 1918 | biografia ▾

Appointed librarian of the Public Library of Catania.

Figure 29: Ecdosis. Event creator.

TILT provides text-image links, connecting words to word-shapes semi-automatically (rather than manually as is done by other tools, e.g. EPT, TILE, TextGrid Link Editor). Further developments are expected for TILT in the British Library Labs[27].

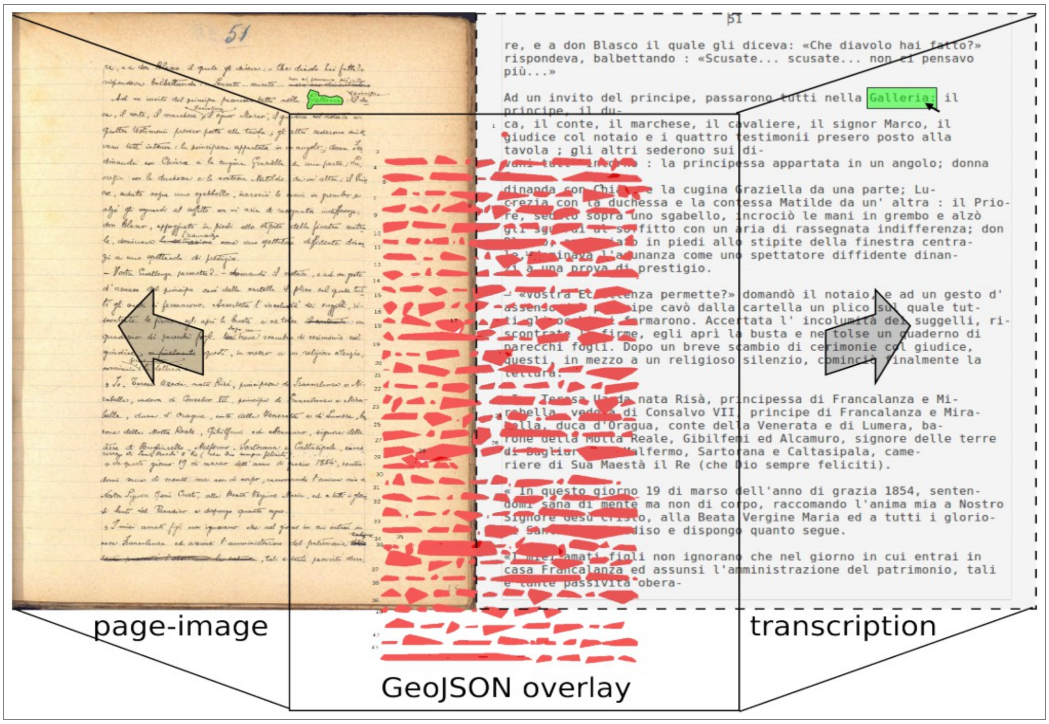


Figure 30: Ecdosis. TILT editor.

[27] <http://labs.bl.uk/>.

The Plain Doc editor is needed for writing Introductions and other paratextual materials.
Overall, Ecdosis is a complete editing environment, created with a user-oriented approach and based on an innovative data format. Its strengths and weaknesses will be better evaluated once it is officially released and tested by the community.

3.3. Comparative table

In the following table, the main features of the tools that have been analysed are summarized and compared.

	Steps of the editing process	**Annotation and markup**	**Import and export**	**Search functionalities**	**Image management**	**Online collaboration**	**Web-based ot standalone**	**Licence**	**Documentation**
T-PEN	Metadata management, transcription, encoding, annotation	Annotation. Possible TEI encoding.	Import: TXT and XML. Export: PDF, XML/PlainText, HTML	No	Advanced. Text/Image link (line level).	Yes. Leader project and users with different rights.	Web-based	Open source ECL-2.0	Users documentation
CWRC	Metadata management, transcription, encoding, semantic enrichment	TEI encoding (*embedded*) and RDF encoding (*standoff*)	Import: plain text	No	No	No.	Web-based	Open source	Users documentation
eLaborate 4	Metadata management, transcription, annotation, publication	Annotation.	Import: plain text. Export: plain text, XML (not available on the normal user interface)	In metadata: advanced and *friendly*. In text: full text search	Advanced. Text/Image link (page level).	Yes. Leader project and users with different rights.	Web-based	Open source GNU GPLv3	Users and developers documentation
TextGrid	Metadata management, transcription, encoding, collation, lexicon creation, paratext creation, publication (+ Lemmatizer for German texts)	TEI encoding.	Import: plain text, XML. Export: plain text, XML.	Full text search into the project metadata and into TextGrid Repository.	Advanced. Text/Image link (manual).	Yes. Leader project and users with different rights.	Portable (connection with the server needed)	Open source See policy at https://www.textgrid.de/en/registrationdownload/tou/	Users and developers documentation
Ecdosis	Metadata management, transcription, encoding, paratext creation, event editor, publication	Markdown encoding.	Import: plain text, XML. Export: plain text, XML.	No	Advanced. Text/Image link (word level).	Yes. Leader project and users with different rights.	Web-based	Open source	/

3.4. Common features of transcription and encoding tools.

All the tools mentioned here are or include transcription and/or encoding facilities. The following table lists the common functionalities for transcription and encoding; other features are discussed below.

Object	Action
Text	add / edit / delete insert special characters copy / paste search save import / export
Image	add / edit / delete zoom in / out link text-image (image parsing or shape recognition, link mechanism, etc.) copy / paste save import / export
Metadata	add / edit / delete may be done through a form copy / paste search save import / export
Markup	add / edit / delete code completion copy / paste search save import / export

The tools enable different views of the content: XML (and XML structure), HTML, text, text and markup.

Certain applications, such as T-Pen, offer friendly functionalities for working with images; others, like CWRC-W, are more oriented towards easy ways to markup and enrich the text.

In the case of a side-by-side view (image and text, annotated text and preview, normalised or diplomatic editions, etc.), synchronization on scroll-down is implemented.

The encoding structure may be free or fixed. In applications for archives, using for examples the EAD or CEI Guidelines, or in filling in a TeiHeader, the encoding structure is, more or less, fixed. A form can therefore be used to markup the text[28].

When there is no fixed encoding structure, it is more difficult to provide simple access to the high number of available tags, for instance TEI tags. The tool will propose the most common, using buttons or other intuitive mechanisms[29]; it is worth remembering that the more frequent tags do not correspond to the Bold, Italic and Underline buttons of most text editors: in a TEI editor, the tags <p> for paragraphs, <lb> for line breaks or <hi> for highlighted portion of text will probably be among the most used. The rest of the available tags may be accessible through menus or other graphical elements, referring to the modular structure of the TEI Guidelines[30]. Having tools that process the schema and customize the interface accordingly (for instance buttons and menus for accessing elements and attributes), may aid easier encoding mechanisms[31].

4. Towards a modular and complete editing environment?

Transcribing and encoding tools do not, of course, cover the entire editing process. What are the editing tasks, in particular in a digital context? In other words, what functionalities should an editing environment cover?

Considering several formalisations of editing tasks[32] and the peculiarities of a digital project, we can list: collection of witnesses (doc/image management and metadata), transcription, encoding, named-entity recognition, semantic enrichment, collation, analysis, constitution of the critical (or copy) text, compilation of apparatuses, compilation of indexes, preparation of paratextual material, data visualization[33]. This list, which primarily reflects a traditional workflow transposed into a digital environment, is not comprehensive. Should an editing environment offer the possibility of automatic lemmatization, of consulting dictionaries or text repositories inside it[34]? And what does analysis mean? As Andrews points out, beyond cladistic analysis based on

[28] Similar tools are LIME <http://lime.cirsfid.unibo.it/>, under development at the University of Bologna and mostly oriented towards the encoding of manuscript descriptions; and Doctored.js <http://holloway.co.nz/doctored/>, an application under development as well, which deal so far with the encoding of bibliographical references.

[29] Cf. the customizable Author View of the XML editor oXygen.

[30] See for example the developments at the Center for Textual Studies and Digital Humanities at the Loyola University Chicago <http://www.luc.edu/ctsdh/researchprojects/hrit-catt/>; demos are accessible at <http://hritwiki.ctsdh.luc.edu/demos>; the editor concept at <http://hritwiki.ctsdh.luc.edu/galleries/hrit-php-demos/tei-editor-concept>.

[31] See for example Wed, an online schema-aware XML editor, developed at the Mangalam Research Center for Buddhist Languages <http://mangalam-research.github.io/wed/>.

[32] To mention only the most influential for the present discussion, Maas 1957, West 1973, Andrews 2014, Ott 2000, Zundert and Boot 2011. I will not enter into the formalisation of different kind of scholarly editions and the related (national or trans-national) traditions.

[33] Once again, we leave outside publication.

[34] This is what TextGrid does, for German language and literature.

methods borrowed from phylogenetic to group the manuscripts, analysis may include 'any form of stylistic analysis such as authorship attribution, or even inclusion in a corpus for large-scale data mining or the application of distant reading techniques' [Andrews 2013].

As already stated, many tools handle the management of documents and images (including metadata), the transcription and the encoding. Several tools for automatic collation are available [see Andrews 2014, pp. 181-184]; the more complete editing environments deal with the comparison of versions in different ways: for instance, TextGrid integrates CollateX; Ecdosis uses the MultiVersion Document format.

The compilation of paratextual material – such as prefaces, titles, introductions – is a basic task that can be achieved with any text editor; one option is to include a plain text editor in the working environment, as in TextGrid and Ecdosis; eLaborate uses WordPress instead. More specific applications or modules may be suitable for the creation of, for instance, a glossary. If supported by consistent markup, the compilation of indexes is a rather straightforward task that can be fulfilled during the publication stage.

One of the frequently ignored tasks in editing programs is the constitution of a critical (or copy or reading) text [see Andrews 2014, pp. 190-191]; a digital tool may suggest alternative readings and ensure the consistency of the editor's choices[35].

Ideally, we will have different complete editing environments and the possibility to choose between them. If none of the (so far, only few) editing environments meets the editor's requirements, interoperability is needed between tools in order to cover the whole workflow. So long as the number of complete editing environments which reflect different approaches, traditions and disciplines remains low, a "toolflow" or personalized workflow which includes several tools is likely the most suitable environment in which to preserve the diversity of final products, i.e. the richness and variety of textual scholarship[36].

The first conference on tools for textual scholarship was probably organized by Susan Hockey and held at Princeton in 1996. The critical features identified at the Text Analysis Software Planning Meeting are still valid: modularity, professionality, integration, portability [Sperberg-McQueen 1996]. In order to achieve modularity (the system should be a 'collection of relatively independent programs, each of which offers a well-defined subset of basic operations for processing textual data'), interoperability is needed. In order to achieve interoperability, standards

[35] The TEI Critical Edition Toolbox (<http://ciham-digital.huma-num.fr/teitoolbox/index.php>) checks the consistency of the encoding, and not of the critical text. It would, for instance, control that all the witnesses have been mentioned in each <app>. This is fundamental for the constitution of the critical text.

[36] More than 70 years ago Barbi wrote about Rajna's teaching: (translation is mine) 'At Rajna's school one would not learn a system [...] There was always exercises on concrete cases, and the solution was always: - so you can see that if you go ahead rationally the problems are put in the right terms, and a satisfactory answer, more or less perfect, depending on the available data, cannot miss. - We leave with the right idea that every text has its critical problem, each problem its solution, and so that editions cannot be produced following a model and, as to say, with a machine' [Barbi 1938, p. X]. In the digital context, see e.g. Régnier 2014, p. 57: 'one is inclined to imagine that in the future that could be not just one model of scholarly edition but a whole constellation of models according to the types of objects to be edited'. Cf. Leonardi 2007, Andrews 2014, Pierazzo 2014.

for data formats and communication protocols are created and implemented[37]. As regards data formats, XML is one of the most common in Digital Humanities, in spite of the many death warrants it has received[38]. TEI is the standard XML language for textual scholars. But TEI interoperability is limited because each project uses a different subset of tags (or the same tag for different purposes). It is therefore possible to build tools that can process a fixed and restricted tag set (e.g., TeiLite), but nearly impossible to build tools that can process 'all' TEI encoded texts. Schmidt's sensible proposition for ensuring interoperability consists of an effective distinction between metadata, annotation, markup and plain text (Schmidt 2014); plain text assures interoperability. This requires at least a stand-off model for the markup, which lacks widespread implementations[39]. Another basic requirement for several tools, and possibly linked to a stand-off model, is the tokenization of texts. Also, one of the difficulty is that if we conceive of a system as a 'collection of relatively independent programs', managing the updates of each program and integrating new releases in a more complete working environment can be difficult. In all cases a long-term sustainability plan is needed.

In short, modular and comprehensive editing environments are a good objective, even if comprehensiveness can never be totally achieved (new tools will emerge and very specific tasks may be needed for particular projects) and modularity is possible only under the condition of interoperability, which is something DH has to work for.

5. Final notes.

This study provides an overview of transcribing and encoding tools in the field of editing applications. As pieces of software, the development of such tools should take into account technical standards, operability and user experience. As scholarly products, they reflect common practices in SDEs, i.e. the use of XML-TEI for text encoding and the production of DDEs among other kind of digital editions. The compared analysis of a number of applications shows how they deal with metadata, markup and images. This overview may be of use for digital editors and developers of such tools.

Editing tools and environments are flourishing around the world, from small projects developing an ad hoc application, to a consortium of universities aiming to build the ‘final’ tool. As McCarty points out, the effort to build effective applications must be paralleled by an effort to learn how to master them. As said in the beginning, modelling, prototyping and testing developments are becoming humanist scholarly activities. In order to complete these complex tasks, the

[37] The reference is to technical standards. The aim is not, as already said, a standardized way of editing a text. In this respect, flexibility and customizability, which Andrews points out to be fundamental to textual scholars, are not, as she suggests, necessary opposed to standards [Andrews 2013].

[38] See for example James Clark, *XML vs the Web* <https://web.archive.org/web/20141216130925/http://blog.jclark.com/2010/11/xml-vs-web_24.html>; cf. *Balisage Series on Markup Technologies, Vol. 14. Proceedings of the Symposium on HTML5 and XML*, Washington, DC, August 2014 <http://www.balisage.net/Proceedings/vol14/cover.html>.

[39] Markup is said to be standoff, when the markup data is placed outside the text it is meant to tag.

collaborative approach of Digital Humanities is essential in gathering expertise in technologies, standards, software development and, last but not least, textual scholarship.

Bibliography

Andrews, Tara. 2013. “The Third Way: Philology and Critical Edition in the Digital Age.” *Variants* 10: 61–76.

Andrews, Tara. 2014. “Digital Techniques for Critical Edition.” In *Armenian Philology in the Modern Era: From Manuscript to Digital Text*, Ed. V. Calzolari and M. E. Stone. Leiden: Brill.

Biblissima. 2013. Résultats du sondage sur les éditeurs XML (TEI/EAD). <http://doc.biblissima-condorcet.fr/resultats-du-sondage-sur-les-editeurs-xml-teiead>.

Barbi, Michele. 1938. *La nuova filologia e l'edizione dei nostri scrittori da Dante al Manzoni*, Firenze: Sansoni.

Bradley, John. 2002. “Tools to Augment Scholarly Activity: An Architecture to Support Text Analysis.” In *Augmenting Comprehension Digital Tools and the History of Ideas*, edited by Harold Short, Dino Buzzetti, and Guiliano Pancaldi, 19–48.

Burdick, Anne et al. 2012. *Digital Humanities*. Cambridge, Mass: MIT Press.

Burkard, Benjamin, Georg Vogeler, and Stefan Gruner. 2008. “Informatics for Historians : Tools for Medieval Document XML Markup, and Their Impact on the History-Sciences.” *Journal of Universal Computer Science* 14.2.

Dekker, Ronald, and Gregor Middell. 2010-2014. *CollateX*. <http://collatex.net>.

Drucker, Johanna, and Bethany Nowviskie. 2002. *Temporal Modelling: Conceptualization and Visualization of Temporal Relations for Humanities Scholarship*. <www2.iath.virginia.edu/time/reports/infodesign.doc>.

Fiormonte, Domenico. 2009. “Chi l’ha visto? Testo digitale, semiotica, rappresentazione. In margine a un trittico di Dino Buzzetti.” *Informatica Umanistica* 2: pp. 21-63.

Hockey, Susan. 2000. *Electronic Texts in the Humanities. Principles and Practice*. New York: Oxford University Press.

Juola, Patrick. 2008. “Killer Applications in Digital Humanities.” *Literary and Linguistic Computing* 23.1: 73–83.

Leonardi, Lino. 2007. “Filologia Elettronica Tra Conservazione E Ricostruzione.” *Digital Philology and Medieval Texts*. Ed. Arianna Ciula, Francesco Stella. Ospedaletto (Pisa): Pacini.

Lowe, Kathryn A. 2015. “Filling the silence. Shared content in four related manuscripts of Aelfric's *Catholic Homilies.*” *Digital Philology* 4.2: 190-224.

Maas, Paul. 1957. *Textkritik*. Leipzig.

Manovich, Lev. 2013. *Software Takes Command: Extending the Language of New Media*. London: Bloomsbury Publishing.

McCarty, Willard. 2005. *Humanities Computing*. Basingstoke [England]; New York: Palgrave Macmillan.

Ott, Wilhelm. 2000. “Strategies and Tools for Textual Scholarship: The Tübingen System of Text Processing Programs (TUSTEP).” *Literary and Linguistic Computing* 15, no. 1: 93–108.

Pape, Sebastian, Christof Schöch, and Lutz Wegner. 2012. “TEICHI and the Tools Paradox.” *Journal of the Text Encoding Initiative* Issue 2.

Pierazzo, Elena. 2011. "A Rationale of Digital Documentary Editions." *Literary and Linguistic Computing*. 26.4: 463-477.

Pierazzo, Elena. 2014. “Digital Documentary Editions and the Others.” *Scholarly Editing: the Annual of the Association for Documentary Editing* 35. <http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html>.

Régnier, Philippe. 2014. “Ongoing challenges for digital critical editions.” *Digital Critical Editions*. Ed. Daniel Apollon, Claire Belisle, and Philippe Régnier. Urbana: University of Illinois Press.

Schmidt, Desmond. 2014. “Towards an Interoperable Digital Scholarly Edition.” *Journal of the Text Encoding Initiative* Issue 7.

Schmidt, Desmond, and Robert Colomb. 2009. “A Data Structure for Representing Multi-Version Texts Online.” *International Journal of Human-Computer Studies* 67.6: 497–514. <http://www.sciencedirect.com/science/article/pii/S1071581909000214>.

Sperberg-McQueen, Michael. 1996. Trip Report: Text Analysis Software Planning Meeting, Princeton, New Jersey. <http://www-01.sil.org/cellar/import/teilite/ceth9605.sgm>.

Unsworth, John. 2003. “Tool-Time, or ‘Haven’t We Been Here Already?’: Ten Years in Humanities Computing.” Washington, D.C.. <http://people.brandeis.edu/~unsworth/carnegie-ninch.03.html>.

West, Martin L. 1973. *Textual Criticism and Editorial Technique*. Walter de Gruyter.

Zundert van, Joris, and Peter Boot. 2011. “The Digital Edition 2.0 and the Digital Library: Services, Not Resources.” *Digitale Edition und Forschungsbibliothek*. (Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011) 44: pp. 141–152.