

Handbook with guidelines for encoding: the Petrarchive project.

Isabella Magni, *Editor and Lead Encoder*

Table of Contents

Handbook with guidelines for encoding: the Petrarchive project.	1
What is Petrarchive.....	2
<i>Rerum vulgarium fragmenta</i>	2
<i>Hyperlink to the website</i>	2
<i>Hyperlink to the Wiki page</i>	2
Basic Introduction to TEI.....	3
<i>Useful resources</i>	4
Encoding guidelines for the Petrarchive.....	5
<i>Encoding diplomatic</i>	5
<i>Encoding edited</i>	9
Entities and Special Characters.....	10
Encoding Literary Genres.....	12
* <i>Sonnets</i>	12
* <i>Madrigal</i>	12
* <i>Ballata</i>	13
* <i>Canzone</i>	14
* <i>Sestina</i>	14
Templates for encoding <i>chartae</i>	14

What is Petrarchive

The *Petrarchive* is the first open access “rich-text” digital edition of one of the masterpieces of Italian and Western literature: Francesco Petrarca’s songbook *Rerum Vulgarium Fragmenta*. The website combines different levels of visualization of the text: facsimile color high-quality images of all the chartae of the partial holograph Vaticano Latino 3195, its complete diplomatic transcriptions and edited forms. The edition also offers text and visual indices, a glossary, commentaries both in Italian and in English and an innovative interactive timeline. Special attention is paid to the poet’s “visual poetics” (Storey 1993) and his transcriptional design.

Digital tools, such as advance use of text encoding, created a more authentic visualization of both the textual and visual aspects of Petrarch’s work, allowing users to experience the scribe and the work’s original cultural contexts: the contrasting visual structures to distinguish the five different poetic genres of which the *Fragmenta* is composed (for example, the two columns horizontal reading strategy for sonnets, madrigals, ballata and canzone, as opposed to the two-column vertical reading strategy of sestina); the distribution of space and the texts on the *chartae*, forming clusters of poems virtually and materially linked together (among which the contrasting visual construct of the *chartae* including a sonnet and a sestina or simply the presence in the same “space” of poems sharing thematic or rhetorical features such as the first opening sonnets on c. 1r - see below); the continuous practice of erasure and emendations of entire poems or parts of them conducted by Petrarch himself until the very end of his life.

Rerum vulgarium fragmenta

The *Rerum vulgarium fragmenta* is one of the icons of the Italian literary tradition and it still remains one of the most studied works, mostly for its extraordinary influence on western literature: it is a songbook of 366 poems of 5 different genres (sonnet, canzone, sestina, madrigal and ballata). We possess a partial autograph (MS Vat. Lat. 3195, transcribed over – roughly – a decade: 1366ca.-1374), which was unfinished at the time of Petrarch’s death in 1374, and (we also possess) a vast and intricate early manuscript tradition.

Some key principles of Petrarch’s own work: **1.** 31-line per “carta” canvas; **2.** the contrasting visual structures to distinguish the five different poetic genres: the two columns horizontal reading strategy for sonnets, madrigals, ballata and canzone, as opposed to the two-column vertical reading strategy of sestina. This implies, for example, that the sonnet is transcribed over seven transcriptional lines (two verses per line, horizontal reading), and not the modernized fourteen lines; **3:** the functionality of space and spacing. See *Petrarchive*’s [prototypes](#).

Hyperlink to the website

<http://www.petrarchive.org>

Hyperlink to the Wiki page

<https://digitalculturelab.atlassian.net/wiki/spaces/petrarchive/overview>

Basic Introduction to TEI

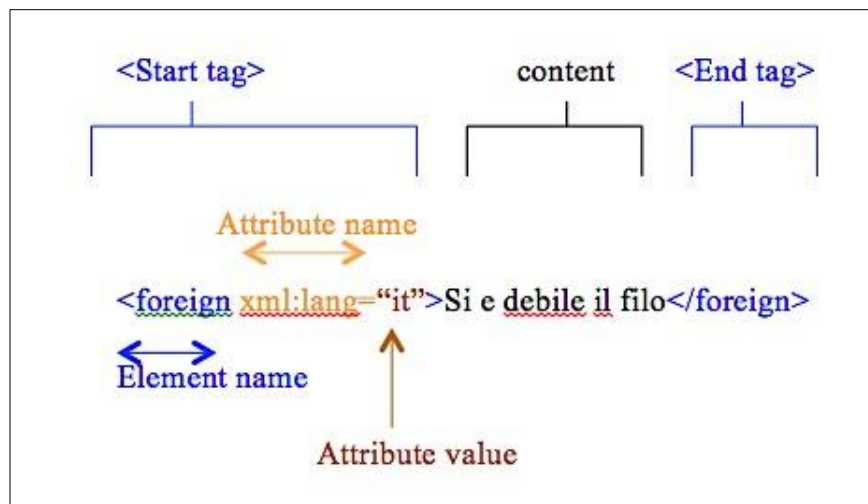
TEI is, therefore, a mark-up language, which consists in the process of adding marks or tags to a text to say specific things about specific parts of the text.

The TEI *Guidelines for Electronic Text Encoding and Interchange* define and document a **markup** language for representing the **structural**, **renditional**, and **conceptual** features of texts. They focus (though not exclusively) on the encoding of documents in the humanities and social sciences, and in particular on the representation of **primary source materials** for research and analysis. These guidelines are expressed as a **modular, extensible** XML schema, accompanied by detailed documentation, and are published under an **open-source license**. ([TEI Guidelines](#))

Being a mark-up language, TEI provides elements, attributes and schemas for encoding texts of different kinds and genres: poetry, drama, prose, dictionary entries, critical apparatus, linguistic corpora and more. Here is a basic and simple explanation that relates to these fundamental parts of a ‘language’:

- * Elements: basic, structural units of an XML document
<title>Rerum vulgarium fragmenta</title>
- * Attributes: a combination of name and value associated with an element
<lg type=“sonnet”>[...]</lg>
- * DTDs and Schemas: state the **rules** and declare elements and attributes acceptable for a particular XML document to which they are associated

Below, a visual explanation of how an element is composed; the text represented in the strip of encoding is “*Si e debile il filo*” which is marked as foreign text, specifically in Italian:



Considering TEI as a meta-language, then, elements can be seen as the ‘names’ of the encoding, attributes as the ‘adjectives’ and schemas as the ‘grammar rules’. (From Isabella Magni, *The digital reconstruction of Petrarch’s Fragmenta* – PhD Dissertation, Indiana University).

Useful resources

TEI Website: <http://www.tei-c.org/index.xml>

TEI Guidelines: <http://www.tei-c.org/Guidelines/>

Elements: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html>

Attributes: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ATTS.html>

TEI Boilerplate (John A. Walsh): <http://dcl.ils.indiana.edu/teibp/>

Text Encoding Workshop Series (John Walsh and Michelle Dalmau, Indiana University):
<http://dcl.slis.indiana.edu/teiworkshop/>

Encoding guidelines for the Petrarch archive

The single TEI-XML files (one file per *charta*/manuscript page) contain all the representations of different strata of the text: A) hyperlink to **high-resolution TIFF images** of the Vatican manuscript, B) its **diplomatic** transcription (transcription of manuscript) and C) the new **edited** texts (modernized).

Each encoder will first encode the diplomatic version of each *charta* and then at a later stage add the edited text. The texts to be encoded will be provided in .doc format and will be stored in the Project's [Wiki](#).

Encoding diplomatic

* *Charta number and poem number/genre*

Change the *charta* number in two places:

- a) in the title statement. For example, if encoding *charta* 3 recto:

```
<titleStmt>
  <title>Petrarch's <title rendition="#i">Rerum vulgarium fragmenta</title>: Charta 3
  recto</title>
  [...]
</titleStmt>
```

- b) and in the body section, in a page break. For example:

```
<text>
  <body>
    <pb n="charta 3 recto" facs="../images/vat-lat3195-f/vat-lat3195-f-003r.jpg" />
    [...]
  </text>
</body>
```

Change the poem number and the poem genre. Examples:

```
<lg xml:id="rvf017" type="sonnet" n="17">
  [...]
</lg>
```

```
<lg xml:id="rvf018" type="sonnet" n="18">
  [...]
</lg>
```

```
<lg xml:id="rvf052" type="madrigal" n="52">
  [...]
</lg>
```

```

<lg xml:id="rvf014" type="ballata_grande" n="14">
  [...]
</lg>

<lg xml:id="rvf055" type="ballata_mezzana" n="55">
  [...]
</lg>

<lg xml:id="rvf070" type="canzone_10vv" n="70">
  [...]
</lg>

<lg type="canzone_14vv" xml:id="rvf050" n="50">
  [...]
</lg>

<lg xml:id="rvf028" type="canzone_15vv" n="28">
  [...]
</lg>

<lg type="canzone_20vv" xml:id="rvf023" n="23">
  [...]
</lg>

<lg type="canzone_16vv" xml:id="rvf037" n="37">
  [...]
</lg>

<lg xml:id="rvf022" type="sestina" n="22" rendition="#columns_2">
  [...]
</lg>

```

Information on genre of each poem is included in the .doc files, next to the poem numbers. Example of sonnet *Rvf* number 21 and sestina *Rvf* number 22:

[21; s]	M ille fiate o dolce mia guerrera	Per
	Vaggio proferto il cor. ma uoi non piace	Mira
5	Et se di lui fors'altra donna spera ☒	Viue
	Mio perche sdegno cio cha uoi dispiace	Esse
	Or sio lo scaccio. et enon troua in uoi	Nel
	Ne sa star sol. ne gire oualtri il chiama.	Pori
	Che graue colpa fia dameduo noi.	Et ta
[22; sest]	A Qualunque animale alberga in terra	Lo
	Se non se alquanti chamo in odio il sole §	Pri
	Tempo da trauagliare e quanto el giorno.	O t
	Ma poi chel ciel accende le sue stelle	Las
5	Qual torna a casa. et qual sanida in selua	Ved

* *31-line canvas*

In Petrarch's personal copy of the *RVF*, he consistently uses a ruled 31-line “canvas” for the layout of the poems. There are few exceptions, where an extra line is added to the underlying canvas.

Elements that occupy a ruled manuscript line are tagged with the attribute `@corresp="#canvasline"`, where `#canvasline` references a category in a taxonomy of "graphic features" in the `<teiHeader>`¹.

Example:

```
<lg type="dblvr" corresp="#canvasline">
```

* *Double verse*

Most of the poems in Petrarch's *Fragmenta* are presented as double verse, with two verse lines appearing horizontally on one manuscript line. To identify these double verses, we will use the `<lg>` element and the `@type` attribute with the abbreviation `dblvr`. Note below how line one and line two are noted separately while still being part of the same double verse grouping.

```
<lg type="dblvr" corresp="#canvasline">
```

* *Single verse*

```
<lg type="single" corresp="#canvasline">
```

* *Triple verse*

A few of the poems in Petrarch's *Fragmenta* are presented as triple verse, with three verse lines appearing horizontally on one manuscript line. To identify these them, we will use the `<lg>` element and the `@type` attribute with the abbreviation `trplvr`.

```
<lg type="trplvr" corresp="#canvasline">
```

* *Initials and small caps*

The beginnings of poems and stanzas often use characters that are different from the rest of the text. They are a combination of capital letters (or pilcrow) and small caps, and they generally alternate between the colors red and blue. Usually, the first letter of a stanza is a capital, while the following letter is a small capital. Care should be taken when encoding so that a small capital letter is not accidentally marked as a regular capital. Encoding examples:

Initials: `<hi rendition="#red #fs24pt">x</hi>` or `<hi rendition="#blue #fs24pt">x</hi>`

Small-caps: `<hi rendition="#small-caps">x</hi>`

Pilcrow: `<hi rendition="#red #fs24pt">¶</hi>` or `<hi rendition="#blue #fs24pt">¶</hi>`

¹ Work can be checked by searching for `//@corresp[= '#canvasline']` in the XPath 2.0 search box, in the upper-right-hand corner of the Oxygen XML Editor interface. The results of the search will appear below the document, and you should get "31 items" in your search results.

Example of poem *Rvf* 17:

```
<lg xml:id="rvf017" type="sonnet" n="17">
  <lg type="octave">
    <lg type="dblvr" corresp="#canvasline">
      <l n="1"><hi rendition="#red #fs24pt">P</hi><hi rendition="#small-
        caps">i</hi>ouonmi amare lagrime dal uiso</l>
      [...]
    </lg>
  </lg>
</lg>
```

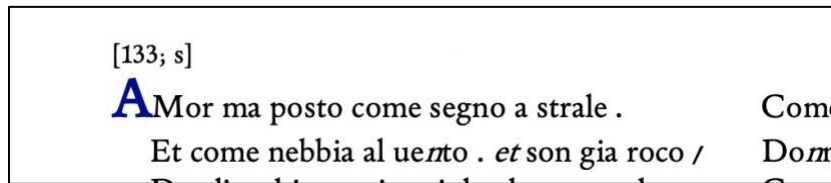
* *Poem numbers and IDs*

Each poem in the *Rvf* has a number and an ID. The number has been established by Petrarch scholars, while the ID is used only within this project. This tag should be placed directly before the beginning of each new poem. For the first poem on a page, it should be placed under the link to the facsimile image. Example:

```
<lg xml:id="rvf002" type="sonnet" n="2">
```

* *Abbreviations*

There are many scribal abbreviations used in the *Rvf*. To show abbreviations and expansions in the diplomatic text, use an <abbr> tag for the first and an <expan> tag for the latter. They must be within <choice> tags to allow users to see either option when desired. Example of how abbreviations are signaled in the .doc and how they should be encoded:



Above is an abbreviation of the word “uento”. Below is the corresponding encoding:

```
<l n="5">O <choice><abbr>uēto </abbr><expan>uento</expan></choice> donestate  
intero albergo .</l>
```

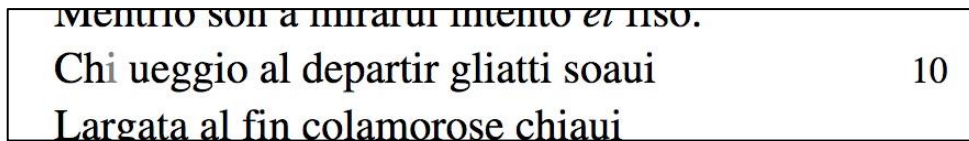
The <abbr> tag contains the abbreviated letter (the vowel preceding the abbreviated letter always has a macron placed above), the <expan> contains the expansion. In the Word document, *italic* signals the abbreviated letter.

For the *et* abbreviation and how to encode it see [Entities and Special Characters](#) section below.

* *Additions over erasures*

Sometimes text in the Rvf has been erased, or is otherwise illegible, and there is a correction or clarification written above it. To encode this, use the <subst>, , and <add> tags. The <subst> tag indicates that some text has been substituted for the original wording, while and <add> show what has been deleted and added. The tag has the type "erasure" to clarify what kind of deletion it is. The <gap reason /> tag shows that what has been erased is now illegible. If the deleted text was intact, the <gap reason /> tag would be absent and that text would be part of the encoding.

In the doc. files, erasures are signaled in grey (the letters/words in grey are the ones transcribed over erasure. In the example below, "i" is transcribed over erasure):



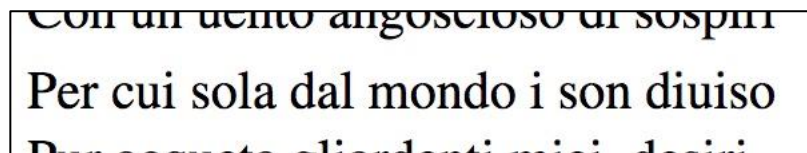
```
<l n="10">Ch<subst><del type="erasure"><gap
reason="illegible" /></del><add>i</add></subst> ueggio al departir gliatti soaui</l>
```

Encoding edited

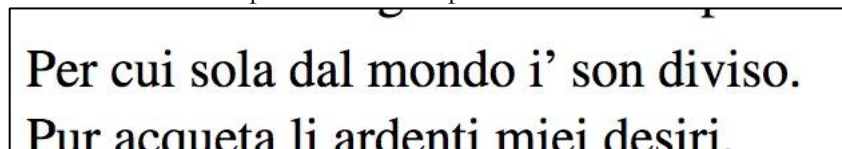
* *Original and regularized*

The diplomatic and edited versions of *Rvf* chartae are encoded in a single TEI document. When text varies between the two, the tags <orig>, denoting the diplomatic text, and <reg>, for the edited text, should be used. They must be put within a <choice> tag. If the changes are minor, such as adding or subtracting punctuation, only one tag may be needed (see below). In this case the <supplied> tag can be used on its own, but the <orig> tag must be within the <choice> tags followed by an empty <reg/> tag or when the supplied element is inside a word.

Example (charta 3 recto, *Rvf* 17, line 4), screenshot of .doc documents (diplomatic and edited) and corresponding encoding:



Diplomatic transcription in .doc file



Edited transcription in .doc file

```
<l n="4">Per cui sola dal mondo i<supplied>'</supplied> son
<choice><orig>diuiso</orig><reg>diviso</reg></choice><supplied>.</supplied></l>
```

* *Supplied*

Example (charta 3 recto, *Rvf* 17, line 4):

```
<| n="4">Per cui sola dal mondo i<supplied>'</supplied> son  
<choice><orig>diuiso</orig><reg>diviso</reg></choice><supplied>.</supplied></|>
```

* *Page breaks*

A page break, marked by the tag <pb/>, denotes the start of a new page. The tag should be placed before the encoding of the text of a charta. Occasionally, a poem will span multiple chartae but will be encoded in a single document. In this case, the tag will be placed in the middle of the text to approximate the top of the new charta. For this project, the tag will include the page number and a link to an image of the facsimile.

Entities and Special Characters

Entities were created as a shortcut for most common abbreviations (abbreviation are always signaled in *italics* in Word documents) and most common diplomatic-to-edited transformations. Here is a comprehensive list:

```
<!ENTITY amp7 "<choice><abbr>&amp;</abbr><expn>et</expn></choice>">  
<!ENTITY Amp7 "<choice><abbr>&Amp;</abbr><expn>Et</expn></choice>">  
<!ENTITY per "<choice><abbr>ꝑ</abbr><expn>per</expn></choice>">  
<!ENTITY Per "<choice><abbr>ꝑ</abbr><expn>Per</expn></choice>">  
<!ENTITY virg "/">  
<!ENTITY virg-underdot "/&#x0323;">  
<!ENTITY Non "<choice><abbr>Nō</abbr><expn>Non</expn></choice>">  
<!ENTITY impresa "<choice><abbr>īpresa</abbr><expn>impresa</expn></choice>">  
<!ENTITY Impresa "<choice><abbr>Īpresa</abbr><expn>Impresa</expn></choice>">  
<!ENTITY sempre "<choice><abbr>sēpre</abbr><expn>sempre</expn></choice>">  
<!ENTITY Sempre "<choice><abbr>Sēpre</abbr><expn>Sempre</expn></choice>">  
<!ENTITY quanto "<choice><abbr>quāto</abbr><expn>quanto</expn></choice>">  
<!ENTITY Quanto "<choice><abbr>Quāto</abbr><expn>Quanto</expn></choice>">  
<!ENTITY quando "<choice><abbr>quādo</abbr><expn>quando</expn></choice>">  
<!ENTITY Quando "<choice><abbr>Quādo</abbr><expn>Quando</expn></choice>">  
<!ENTITY giorno "<choice><abbr>giōno</abbr><expn>giorno</expn></choice>">  
<!ENTITY non "<choice><abbr>nō</abbr><expn>non</expn></choice>">  
<!ENTITY in "<choice><abbr>ī</abbr><expn>in</expn></choice>">  
<!ENTITY pre "<choice><abbr>&p-macron;</abbr><expn>pre</expn></choice>">  
<!ENTITY con "<choice><abbr>cō</abbr><expn>con</expn></choice>">  
<!ENTITY Con "<choice><abbr>cō</abbr><expn>Con</expn></choice>">  
<!ENTITY son "<choice><abbr>sō</abbr><expn>son</expn></choice>">  
<!ENTITY Son "<choice><abbr>sō</abbr><expn>Son</expn></choice>">  
<!ENTITY q-tilde "q&#x0303;">  
<!ENTITY q-macron "q&#x0304;">  
<!ENTITY p-macron "p&#x0304;">
```

```

<!ENTITY b-combtild "b&#xFE23;">
<!ENTITY p-dot "p&#x0307;">
<!ENTITY u-vert-tilde "u&#x033E;">
<!ENTITY p-vert-line "p&#x030D;">
<!ENTITY uia "<choice><orig>uia</orig><reg>via</reg></choice>">
<!ENTITY uai "<choice><orig>uai</orig><reg>vai</reg></choice>">
<!ENTITY vita "<choice><orig>uita</orig><reg>vita</reg></choice>">
<!ENTITY regapos "<reg>'</reg>">
<!ENTITY uolta "<choice><orig>uolta</orig><reg>volta</reg></choice>">
<!ENTITY uolte "<choice><orig>uolte</orig><reg>volte</reg></choice>">
<!ENTITY perche "<choice><orig>perche</orig><reg>perché</reg></choice>">
<!ENTITY egrave_reg "<choice><orig>e</orig><reg>è</reg></choice>">
<!ENTITY si "<choice><orig>si</orig><reg>si</reg></choice>">
<!ENTITY Si "<choice><orig>Si</orig><reg>Sì</reg></choice>">
<!ENTITY e "<choice><orig>e</orig><reg>è</reg></choice>">
<!ENTITY piu "<choice><orig>piu</orig><reg>più</reg></choice>">
<!ENTITY Piu "<choice><orig>Piu</orig><reg>Più</reg></choice>">
<!ENTITY quello "<choice><abbr>&q-macron;llo</abbr><expan>quello</expan></choice>">
<!ENTITY fiamma "<choice><abbr>fiāma</abbr><expan>fiamma</expan></choice>">
<!-- period "p" to "2" comma "c" -->
<!ENTITY p2c "<choice><orig>.</orig><reg>,</reg></choice>">
<!ENTITY p2colon "<choice><orig>.</orig><reg>:</reg></choice>">
<!ENTITY p2semicolon "<choice><orig>.</orig><reg>;</reg></choice>">
<!ENTITY p2e "<choice><orig>.</orig><reg>!</reg></choice>">
<!ENTITY c2p "<choice><orig>,</orig><reg>.</reg></choice>">
<!ENTITY v2c "<choice><orig> /</orig><reg>,</reg></choice>">
<!ENTITY v2q "<choice><orig> /</orig><reg>?</reg></choice>">
<!ENTITY pilcrow2section "<choice><orig>¶</orig><reg>§</reg></choice>">
<!ENTITY punctelevhback "&#xF1FA;">
<!ENTITY encodingDesc SYSTEM "encodingDesc.xml">

```

To encode press & and the name of the entity followed by a semi-column (you can also select the correct entity from the scroll-down).

Examples:

abbreviation *per* → **&per;**

when you have a punctus (.) in the diplomatic and substitute it with a comma (,) in the edited →

&p2c;

Encoding Literary Genres

* Sonnets

The most common types of poems in the *Rvf* is the sonnet. Petrarchan sonnets have 14 lines, generally consisting of an octave (8 lines) and a sestet (6 lines). Within the normal <lg> groupings, the octave and sestet should be separated. They have their own @type element tags, as seen below.

```
<lg xml:id="rvf002" type="sonnet" n="2">
  <lg type="octave">
    <lg type="dblvr" corresp="#canvasline">
      <l n="1"><hi rendition="#red #fs24pt">...</hi><hi rendition="#small-caps">
...</hi>...</l>
      <l n="2">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
      <l n="3">...</l>
      <l n="4">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
      <l n="5">...</l>
      <l n="6">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
      <l n="7">...</l>
      <l n="8">...</l>
    </lg>
  </lg>
  <lg type="sestet">
    <lg type="dblvr" corresp="#canvasline">
      <l n="9">...</l>
      <l n="10">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
      <l n="11">...</l>
      <l n="12">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
      <l n="13">...</l>
      <l n="14">...</l>
    </lg>
  </lg>
</lg>
```

* Madrigal

```
<lg xml:id="rvf052" type="madrigal" n="52">
  <lg type="dblvr" corresp="#canvasline">
```

```

    <l n="1"><hi rendition="#blue #fs24pt">...</hi><hi rendition="#small-caps">
...</hi>...</l>
    <l n="2">...</l>
</lg>
<lg type="single" corresp="#canvasline">
    <l n="3">...</l>
</lg>
<lg type="dblvr" corresp="#canvasline">
    <l n="4">...</l>
    <l n="5">...</l>
</lg>
<lg type="single" corresp="#canvasline">
    <l n="6">...</l>
</lg>
<lg type="dblvr" corresp="#canvasline">
    <l n="7">...</l>
    <l n="8">...</l>
</lg>
</lg>

```

* Ballata

```

<lg xml:id="rvf011" type="ballata_grande" n="11">
    <lg type="dblvr" corresp="#canvasline">
        <l n="1"><hi rendition="#red #fs24pt">...</hi><hi rendition="#small-caps">
...</hi>...</l>
        <l n="2">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
        <l n="3">...</l>
        <l n="4">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
        <l n="5">...</l>
        <l n="6">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
        <l n="7">...</l>
        <l n="8">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
        <l n="9">...</l>
        <l n="10">...</l>
    </lg>
    <lg type="dblvr" corresp="#canvasline">
        <l n="11">...</l>
        <l n="12">...</l>
    </lg>

```

```
<lg type="dblvr" corresp="#canvasline">
  <l n="13">...</l>
  <l n="14">...</l>
</lg>
</lg>
```

- * Canzone

- * Sestina

Templates for encoding *chartae*

Templates for 4-sonnets *chartae* and sonnet-sestina *charta* are on the Wiki under Encoding Support – [TEI Templates](#).