

Facsimile and Document-Centric Editing

Elena Pierazzo

Digital facsimiles have become ubiquitous in editorial practices, whether the outcome is to be in digital form only or in print or both. This chapter explores the ways the TEI handles the use of facsimile within the production of digital editions, and connects them to different theoretical editorial approaches and conceptions. Readers will learn how to establish a relation between a portion of a text and a digital facsimile or a portion of it; they will learn how to *zone* an image and also how to embed a transcription within an image; they will lastly learn how to document and time the process of producing a text.

1. Facsimile and (Digital) Editing

The combination of the affordability of digital photography and the rise of digital methods in editing has brought the images of manuscripts and documents in the scholarly spotlight. The high quality digital reproductions of primary sources has represented a major improvement with respect to microfilms, and a much cheaper option with respect to photography, and it has been saluted with great enthusiasm by scholars around the world. The improvement has been so remarkable that some scholars have even been carried away by their enthusiasm, declaring that once you have a digital image “we need never see the document itself” (Twycross 2008, p. 23) and that the facsimile provide an “unmediated” access to the source (Nell Smith, 2004, p. 309), or that a facsimile edition subsumes the role of diplomatic editions (Kiernan 2006, p. 266).¹

However exciting the new phase might be, the use of facsimile and the inclusion of reproduction of documents in editions is not a new fact in textual scholarship; editors have in fact tried to provide their readers with some sort of access to the documentary sources for centuries; since the sixteenth century typesetters tried to reproduce in print the look of ancient scripts,² while engravings reproducing charters and manuscripts were firstly used by Jean Mabillon in his *Re Diplomatica* (1st ed. 1681) (Stokes 2010, 14). Such publications were nevertheless very expensive and therefore extremely rare; even when photographs replaced engravings, making the publication of representations marginally more affordable, publications including facsimile remained scarce and expensive. The easy availability and reduced cost of digital photographs have determined a more pervasive presence of reproductions within the editorial practices, not to mention the centrality they have gained in any digital edition to the point that producing a digital edition without providing access to the digitised sources has become extremely rare. Such change, which is a direct consequence of the major technological innovation that has brought us digital photography, is also coincidental with important changes in textual scholarship itself,

1. See Pierazzo 2015, pp. 93–98

2. See for instance *A testimonie of antiquity*, collected by John Joscelyn and Matthew Parker, published in 1566 by John Day.

namely with a renewed attention given to documentary evidence brought in by different scholars and scholarly approaches (Pierazzo 2104).

These combinations of technological innovation and changes in scholarly theorisation have on the one hand had a major impact on editorial workflow, and on the other hand have created new expectations on the part of the readers (and users) of digital scholarly editions (Sutherland and Pierazzo, 2012, p. 202). But there is more: in fact the combined changes brought in by the availability of reproductions of primary sources and by the new textual scholarship attention to documents are determining changes on the perceptions of textuality itself, on what text is and what it is supposed to convey to its readers (Pierazzo 2016). These new perspectives and new attention to documents have been even considered dangerous by certain scholars (Robinson, 2013c, p. 127), but while calling it dangerous is probably an overstatement, it is without a doubt that this change is unsettling and has manifold implications for scholarship and practices (Treharne 2013).

From a practical point of view, the inclusion of facsimile within a digital edition can take different forms and role, from being an accessory to illustrate the scholarly argument brought forward by the edition, to taking centre stage and becoming the focal point of the edition. This difference in weight is the result of different workflows which in turns reflects different conceptions and editorial purposes, and while a digital edition can be produced using different tools and methods, it is without a doubt that for the past thirty years the Text Encoding Initiative has provided (and still does) a very solid, scholarly-based approach to the handling of facsimile within the editorial work.

2. The TEI and the facsimiles

The TEI has been part of this pivotal change from the very beginning, but particularly in the last few years with a provision of new facilities for the handling of representations of source documents. Admittedly, the earlier version of the TEI (the so-called P3 and P4 releases) provided limited support for facsimile images, the inclusion of which was mostly mandated to the imaginative use of a handful of generic elements. However, since the release of the P5 version in 2007, the TEI has been on the front line of documentary editing. In fact this version has introduced an entire new module for the handling of facsimile and for the connection of images, or portions of images, to transcriptions or editions.³ Furthermore, since 2012 the TEI has provided a further support to documentary editing in the form of new specialised elements for the inclusion of transcription within, so-to-speak, the facsimile itself, therefore allowing for the production of hyper-diplomatic editions, able not only to reproduce the text in documents to a great extent of details, but also its layout, including also non-textual features such as blurs and doodles.

This chapter will present the different ways the TEI offers to the editor for the inclusion of facsimile into digital editions, and will be divided in two main sections discussing, respectively, the connection of an edition to an image through the use of the `<facsimile>` element, and the

3. The entire following discussion subsume the reading of chapter 11 of the TEI Guidelines (see <http://www.tei-c.org/Vault/P5/current/doc/tei-p5-doc/en/html/PH.html>)

embedding of an edition within an image, through the use of the `<sourceDoc>` element.

3. Side-by-side: editions and `<facsimile>`

There are several ways one can put in relation an image to a portion of a text within the TEI framework, most of them are connected to the use of the `<facsimile>`⁴ element which is meant to contain information about a facsimile image (or group of images) that is related in some way to a text or, more commonly, to a portion of a text (the content of a page, of a document, of a side of a stone, etc.); once that the presence of a facsimile has been declared, the actual connection between the information contained in the `<facsimile>` and the relevant portion of the text itself is then performed via the global attribute `@fac`.⁵ The purpose of this attribute is in fact to contain a reference (an URI) to an image, by the means of a reference to an identifier (established by an `@xml:id` within the `<facsimile>` element) or thanks to a link to the actual file containing the image of the facsimile itself. In fact, it is even possible to simply use the `@fac` attribute without having to use the element `<facsimile>` at all, which represents the easiest way to connect a text to an image, as in the following example which typically uses the `@fac` attribute on the `<pb>` (i.e., *page-beginning*)⁶ element; the example then connects then an image of a page of a manuscript to the element that signal the beginning of a new page in the transcribed source.

```
<body>
  <pb n="001r" facs="../images/SBK-CC1-126_001r.jpg"/>
  <div>
    <argument rend="rubric">
      <p>Incip<ex>it</ex> lib<ex>er</ex> prim<ex>us</ex> de
        inge<ex>n</ex>iosa<ex>n</ex>itatis ...</p>
    </argument>
    <p>...</p>
  </div>
</body>
```

This method presents the double advantage of being very straightforward and lightweight; however it only allows for a connection of a text (or a page of a text) to one image at the time, which might prove insufficiently detailed or nuanced in some circumstances. For instance, it might be useful to be able to connect a portion of a text to only a portion of an image; or it could be desirable for an edition to offer to the users several versions of the image containing a page of the source, for instance: a full-sized, high-resolution image for storage and/or for more advanced or registered user, a thumbnail image for browsing, a low resolution image for the side-by-side display, an infrared version for reading the underlying text of a palimpsest, a PDF version for download, and so on. For all this cases, the use of the `@fac` attribute is not sufficient and it is necessary to use the `<facsimile>` element.

4. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-facsimile.html>

5. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-att-global.facs.html>

6. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-lb.html>

3.1. One surface, many images

The `<facsimile>` element is a direct child of the root element `<TEI>`, and has to be used immediately after the `<teiHeader>` and before the `<text>` element.

```
<TEI>
  <teiHeader>
    <!-- Here metadata -->
  </teiHeader>
  <facsimile>
    <!-- Here the list of all facsimile -->
  </facsimile>
  <text>
    <!-- Here the transcription of the text -->
  </text>
</TEI>
```

The `<facsimile>` element typically contains one or many `<surface>` elements,⁷ corresponding to a page or any type of surface over which writing is found. Each of them is to contain the references of the image file(s) representing in some way the said surface. It is also good practice to give this element an `@xml:id` attribute to which then one can point to.

The link to the actual files storing the images is normally performed via a `<graphic>` element,⁸ or a sequence of them if the edition is to provide several versions of the same surface. The `<graphic>` element is normally an empty one, only providing a link to the image file via a `@url` attribute, but it can also contain a `<desc>` element to provide useful information about the usage of the image, its content or any other information that might be relevant.

In the following example, the `<facsimile>` element has been used to declare the availability of several versions of the same image for different purposes. The `<surface>` element has been given an `@xml:id` attribute, consisting in an arbitrarily chosen sequence of characters providing a unique identifier of the surface, and an `@n` attribute, providing the folio number of the surface itself. Once all the images of that surfaces have been declared and described, it is possible to refer to the surface from the transcribed and/or edited text using the `@facts` attribute that this time does not provide a direct link to an image file, but a reference to the `@xml:id` of the `<surface>` element. When the edition is ready to be published, it will be then the role of some types of scripts (typically XSLT scripts) or of an API to choose the appropriate image among the ones available for the surface and to display it according to the situation, as decided by the editors.

7. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-surface.html>

8. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-graphic.html>

```

<facsimile>
  <surface n="001r" xml:id="c001r">
    <graphic url="../images-low-res/SBK-CC1-126_001r.jpg">
      <desc>Low Resolution</desc>
    </graphic>
    <graphic url="../images-archival/SBK-CC1-126_001r.tiff">
      <desc>High Resolution</desc>
    </graphic>
    <graphic url="../images-thumb/SBK-CC1-126_small_001r.jpg">
      <desc>Thumbnail</desc>
    </graphic>
    <graphic url="../infrareds/SBK-CC1-126_infra_001r.tiff">
      <desc>Infrared version</desc>
    </graphic>
  </surface>
</facsimile>
<text>
  <body>
    <pb facs="#c001r"/>
    <div>
      <argument rend="rubric">
        <p>Incip<ex>it</ex> lib<ex>er</ex> prim<ex>us</ex> de
inge<ex>n</ex>iosa<ex>n</ex>itatis ...</p>
      </argument>
      <p>...</p>
    </div>
  </body>
</text>

```

3.2. One surface, many zones

The **<facsimile>** element can also be conveniently used to connect only portions of images to portions of texts. This could be useful if the edition is to provide detailed commentary on decoration or to relate marginal additions or comments to their actual position, or for any other circumstance when an editor wishes to discuss a particular detail of a surface. For that purpose, the **<surface>** element may contain one or more **<zone>** elements,⁹ each of them representing an arbitrarily (or better: editorially) defined area of an image. The area itself is delimited thanks to a series of attributes on the **<zone>** element containing spatial coordinates that allow to draw either a rectangular or polygonal area.

Let us suppose that we want to draw a rectangle to encase the left hand-side of the folio 1r of Cod. 126 of the Klosterneuburg, Augustiner-Chorherrenstift Library in Austria¹⁰ in order to be able to comment on the iconography of the decoration, and in particular on the archer below the drop capital. In this case we can use a series of four attributes (or better: two couples of

9. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-zone.html>

10. A description of the manuscript can be found at <http://manuscripta.at/m1/lib.php?libcode=AT5000>.

attributes) to provide the coordinates corresponding to a rectangle-shaped zone over the image. This operation is performed thanks to a familiar geometric mechanism, namely by considering the image as a bi-dimensional Cartesian space, where any point can be defined by two values, i.e., the “abscissa” (or “x” value) and the “ordinate” (or “y” value) coordinates of that particular point. It is important to note, however, that with respect to the way Cartesian coordinates normally work (i.e., with origin positioned on the lower-left corner of a bi-dimensional space), coordinates on screens are calculated from the upper-left corner.

Now, in a Cartesian space a rectangle can be ideally drawn by providing the coordinates of only two diagonally opposite corners of that rectangle, and then letting the computer to infer the coordinates of the other two remaining corners based on the ones provided. For the drawing of a rectangle, the TEI has chosen to provide the coordinates of the top left corner and the bottom-right corner, therefore the attributes offered by the TEI are:

Coordinates for the upper-left corner

- @ulx: provides the upper-left X value
- @uly: provides the upper-left Y value

Coordinates for the lower-right corner

- @lrx: provides the lower-right X value
- @lry: provides the lower-right Y value

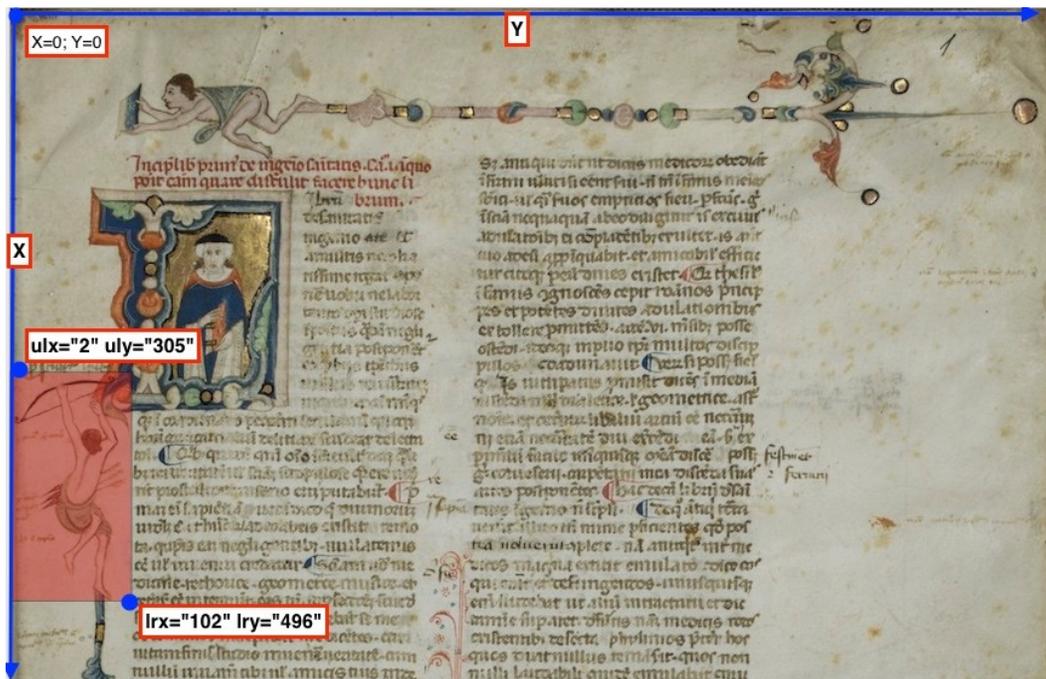


Figure 1: Rectangle drawn upon f. 1r of Cod. 126 of the Klosterneuburg, Augustiner-Chorherrenstift Library

The `<zone>` element in the snippet below is then provided with the attributes containing the coordinates for the upper-left and lower-right corner of a rectangle;¹¹ it also has an `@xml:id` attribute which is then being referenced by the `@facts` attribute in a `<note>` (but it could be a paragraph or any other appropriate element used for annotation) describing the content of the zone.

```

<facsimile>
  <surface n="001r" xml:id="c001r">
    <graphic url="SBK-CC1-126_001r.jpg" width="1481px"
height="953px"/>
    <zone ulx="2" uly="305" lrx="102" lry="496" xml:id="c001r-
archer"/>
  </surface>
</facsimile>
<text>
  <body>
    <pb facts="#c001r"/>
    <div>
      <argument rend="rubric">
        <p>Incip<ex>it</ex> lib<ex>er</ex> prim<ex>us</ex> de
        inge<ex>n</ex>io sa<ex>n</ex>itatis ...</p>
      </argument>
      <p>...</p>
      <note type="decoration" facts="#c001r-archer">Archer on the left
margin</note>
    </div>
  </body>
</text>

```

A similar approach can be adopted in all those cases in which it is considered appropriate to connect a portion of a text with a portion of an image, for annotating or documenting or providing support to the editorial arguments. It is also possible to use the `@facts` attribute from within the `<msDesc>` element; in this latter case it is particularly appropriate to use it from the `<locus>` element. For instance:

11. The coordinates have been determined thanks to the TEI Zoner, a web-based tool developed by Chris Sparks and available from the TEI Critical Apparatus Toolkit, <http://ciham-digital.huma-num.fr/teitoolbox/zoner.php>.

```

<msDesc>
  <msIdentifier>...</msIdentifier>
  <msContents>
    <msItem>
      <incipit><locus from="1r" facs="#c001r"/>Incipit Liber primus de
ingenio sanitatis</incipit>
    </msItem>
  </msContents>
  <physDesc>
    <decoDesc>
      <decoNote facs="#c001r-archer">Archer on the left
margin.</decoNote>
    </decoDesc>
  </physDesc>
</msDesc>

```

Rectangles are at times not the best way to encase decorations or jots. In these case the four attributes holding the coordinates can be substituted by a single attribute called **@points** which syntax allows to supply any number of couples of coordinates, each of them representing the X and the Y values of a point, parted by a comma; each couple is then separated from the next one by a space. The following example represents a polygon encasing a complex decoration at folio 7r of Cod. 253 of the Klosterneuburg, Augustiner-Chorherrenstift Library.



Figure 2: Polygon drawn upon f. 7r of Cod. 253 of the Klosterneuburg, Augustiner-Chorherrenstift Library

```

<surface xml:id="c007r">
  <graphic url="../images/SBK-CC1-253_7r.jpeg" width="1055px"
height="1652px"/>
  <zone points=" 352,33 474,75 546,288 580,319 574,355 616,414
739,523 762,565 738,640 747,666 736,693 757,729 757,762 696,940
777,1089 732,1191 682,1227 468,1243 141,1273 57,1209 42,993 28,714
37,514 148,355 138,213 195,96 351,33 "/>
</surface>

```

The zoning mechanism could also be used to record the exact position of a textual addition or a variant. In the following example taken from folio 1r of Cod. 110 of the Klosterneuburg, Augustiner-Chorherrenstift Library¹² the word *divine* has been added on the right margin and connected to the main text thanks to a caret.

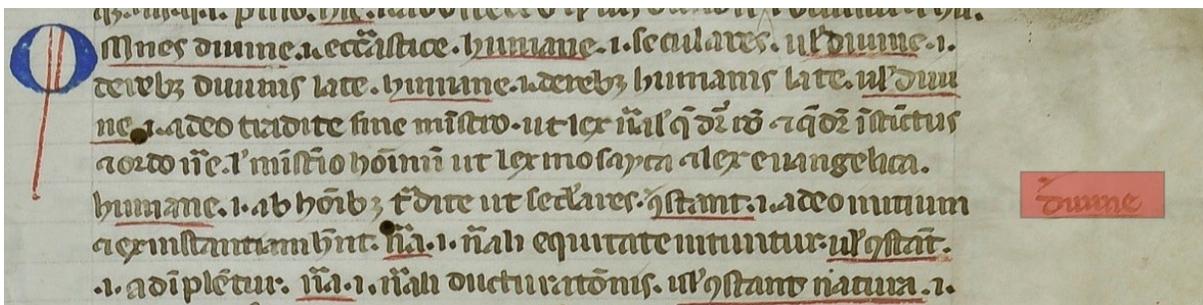


Figure 3: Addition to right margin of Cod. 110, f. 1r, of the Klosterneuburg, Augustiner-Chorherrenstift Library.

The encoding snippet below features the `<zone>` element surrounding the word *divine* within the `<facsimile>` section, as well as the `@facts` attribute used on the `<add>` element pointing at the `<zone>`.

```

<facsimile>
  <surface n="001r" xml:id="c001r">
    <graphic url="../images/SBK-CC1-110_001r.jpg" width="1312px"
height="312px"/>
    <zone ulx="1002" uly="170" lrx="1143" lry="217" xml:id="add-01"
type="marginal-addition"/>
  </surface>
</facsimile>
<text>
  <body>
    <pb facts="#c001r"/>
    <div>
      <p>...humane id est ab hominibus tradite ut sec<ex>u</ex>lares.
<metamark function="add" type="caret"/> <add facts="#add-01"
place="margin">divine</add> constant id est a deo initium </p>

```

12. A description of the manuscript can be found at http://manuscripta.at/m1/hs_detail.php?ID=131.

```
</div>  
</body>  
</text>
```

4. Embedded: editions and **<sourceDoc>**

From an editorial perspective, the use of the **<facsimile>** element as a sibling of the **<text>** element responds to a vision of the editorial work that aims at reconstructing the text of a document or of a work, or that seeks to produce a diplomatic edition of one or more witnesses of a work alongside a critical edition. All of these editorial purposes assume however that there is a text, and that the text can be determined editorially from the extant documents and that determining it is the purpose of the editions. However, there are cases when this seemingly obvious purpose is not so simple to achieve, and these are cases where the physicality of the document is overwhelmingly present (as for draft manuscripts or highly decorated ones) or where the visual component of the document has a major importance from a cultural, historical or editorial point of view.

In cases like the ones described by Pierazzo and Stokes 2010 the extraction of texts from documents determines a loss of information that could be deemed unacceptable by editors and readers. In some other cases, the purpose of editing itself may be to explore the materiality of the document or the process of creating texts and documents, studying the stratification of the ink traces left by the author or by readers. In these cases fitting the textual component of a document in a linear way, organising them in paragraphs or verses, does not seem to be the best way to represent either these scholarly purposes or the peculiarity of the documents. In these cases, the TEI offers the possibility of transcribing the words on the page following a documentary approach as opposed to a textual approach, i.e., page by page and line by line instead that paragraph by paragraph, which consists in migrating the transcription within the facsimile instead of keeping it alongside it. In these cases, instead of using the **<facsimile>** as a parent element, **<surface>** and **<zone>** are used as children of a **<sourceDoc>** element.¹³ In this latter case the element **<zone>** becomes a text holder and not only a way to draw virtual shapes over a facsimile. The **<sourceDoc>** element assumes all the functions of the **<facsimile>** element which becomes then redundant.

This new approach privileges the document over the text, therefore it does not allow to use elements that are normally used to enrich and annotate semantically a text; this is the case of elements to mark names of people or places, and also elements to mark dates and events, or any internal textual structure such a verses, paragraphs, speeches, and so on. This situation is specular to the one that only allows to mark the beginning of new lines and pages (elements **<lb>** and **<pb>**, respectively) as empty elements in the textual approach of the TEI. The reason behind the decision of limiting the content of these elements is to prevent the risk of overlapping hierarchies: in fact, when adopting a textual approach and therefore transcribing the text on a source paragraph by paragraph (for instance) a page break may happen within a paragraph, therefore, if the elements for a page was to be non-empty, it would cause an overlap between

13. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-sourceDoc.html>

paragraph and page. In the same way, if adopting a documentary approach, the name of a person, for instance, could be split between to lines, and so on. The TEI therefore only allow strictly *editorial* elements, such as, for instance, **<add>**, ****, **<gap>**, **<ex>**, **<sic>** and few others, which are essential transcriptional phenomena, and also for theses it warns about the fact that some overlapping issues will be inevitable.

The use of the **<sourceDoc>** approach determines also a different use of the **<zone>** than can have three main types of contents:

- Free text, possibly segmented by **<lb>** (line-beginning) elements .
- A sequence of **<line>** elements, each of them representing a typographical written line.
- One or more embedded **<zone>** elements.

The choice between this different contents is to some extent arbitrary, but there are situations in which one choice could be more appropriate than another. The use of the element **<line>** (which is the non-empty counterpart of the **<lb>** element)¹⁴ is recommended only in those cases where the page layout clearly shows lines of text. For instance, in case the text in a zone is not clearly structured in lines, a free text approach could be more appropriate, or if a zone shows signs of having internal structures, sub-zones could be more suitable.

The following example presents the zoning and transcription of some of the marginalia of folio 80r of the Cod. 126 seen above. The text of the first two zones is organised in clear sets of lines, while the third zone contains only one word and the forth zone is wrapped around the decoration and has a more untidy line structure.

14. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-line.html>

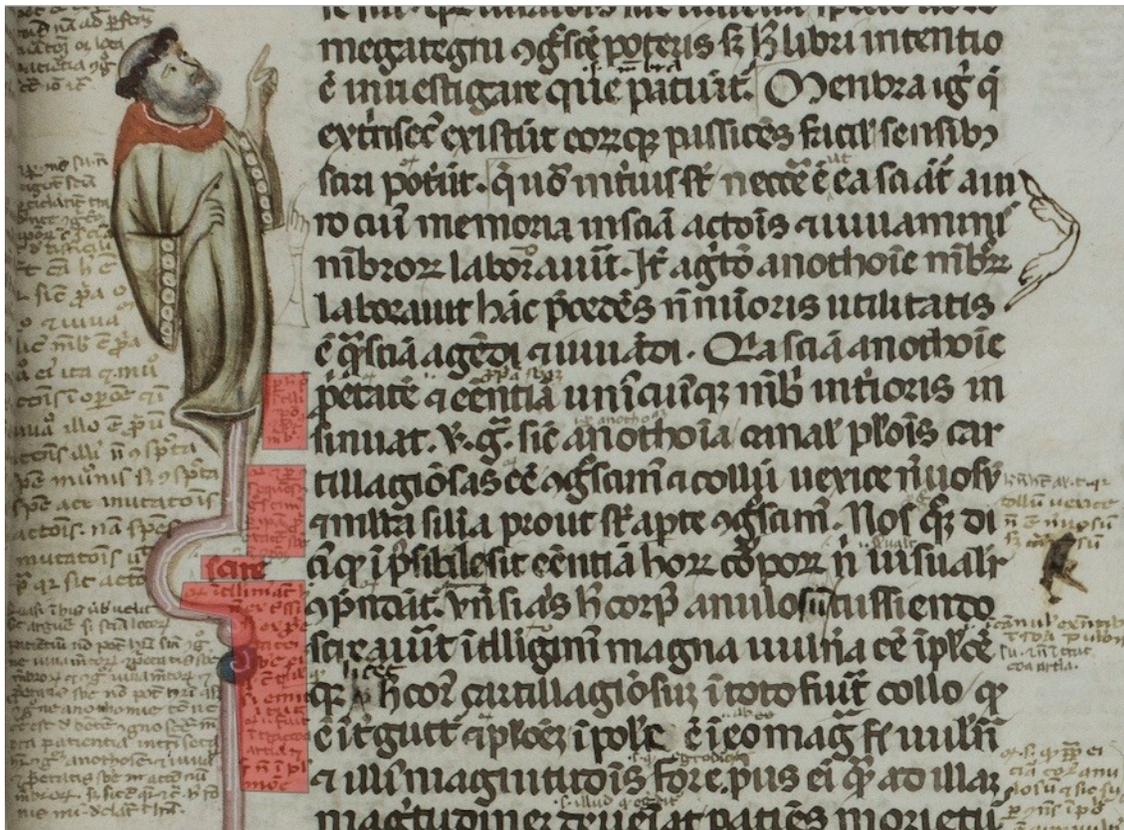


Figure 4: Zoning of cod.126 Klosterneuburg, Augustiner-Chorherrenstift Library

Because of the particularity of each zone, it may not be advisable to adopt the same approach for each of them, so in the example below the first two zones will propose a line by line transcription, while the latter two will have a free text approach, in one case because the use of `<line>` seems unnecessary, given the fact that the zone contains only one word, and in the other because of the unplanned and untidy appearance of the text for which a rigid encasing of text into `<line>` elements seems inappropriate.

```

<sourceDoc>
  <surface xml:id="c007r">
    <graphic url="../images/marginalia.png" width="1761px"
height="1349px"/>
    <zone ulx="405" uly="598" lrx="480" lry="721">
      <line>per hoc potest</line>
      <line>intelligi</line>
      <line>compositio</line>
      <line>et forma</line>
      <line>membri</line>
    </zone>
    <zone ulx="369" uly="753" lrx="477" lry="900"><line><metamark

```

```

function="reference"/>et per</line>
  <line>consequens</line>
  <line>ergo scimus</line>
  <line>per ipsam propri-</line>
  <line>-etatem substantie</line>
  <line>et autem compositionem</line></zone>
  <zone ulx="303" uly="895" lrx="421" lry="939">scire</zone>
  <zone points="271,946 351,939 484,939 472,1287 369,1285 369,1111
394,1078 387,1024 354,1015 334,987 271,979"><lb/><metamark
function="reference"/>intelligimus autem hoc <lb/>non ex crassi
<lb/>sed ex proprie<lb/>tate <lb/>substantie eius ... </zone>
</surface>
</sourceDoc>

```

Given the fact that zones are independent from one another (they are like self-contained boxes), their graphical plotting on the surface can overlap without problems, as in the example below, where the zones containing the blocks of texts overlap with the zones containing the decoration (see next page).

```

<surface>
  <graphic url=" ../images/SBK-CC1-685_074r.jpg" width="686px"
height="1019px"/>
  <zone points="36,236 288,236 258,833 27,825">
    <note>Soldier facing right</note>
  </zone>
  <zone ulx="19" uly="171" lrx="460" lry="264">
    <line>Ex geminis...</line>
  </zone>
  <zone ulx="168" uly="276" lrx="307" lry="378">
    <line>Ysidorus dixit..</line>
  </zone>
</surface>

```



Figure 5: Zoning of cod.685, f. 74r, Klosterneuburg, Augustiner-Chorherrenstift Library

Sometimes manuscripts present portion of writing that are rotated with respect to the orientation of the page; to capture this characteristic, the `<zone>` element is provided with a `@rotate` attribute which value expresses the amount of clockwise rotation that the document had to undergo in order to allow for the rotated writing to be supplied; the value of the attribute is expressed in arc degrees. In the following example there are two sets of zones on the right of the main block writing: four blocks of writing close to the spine rotated by 270 arc degrees with respect to the page orientation, and a smaller addition constituted of only two lines rotated by 90 arc degrees.

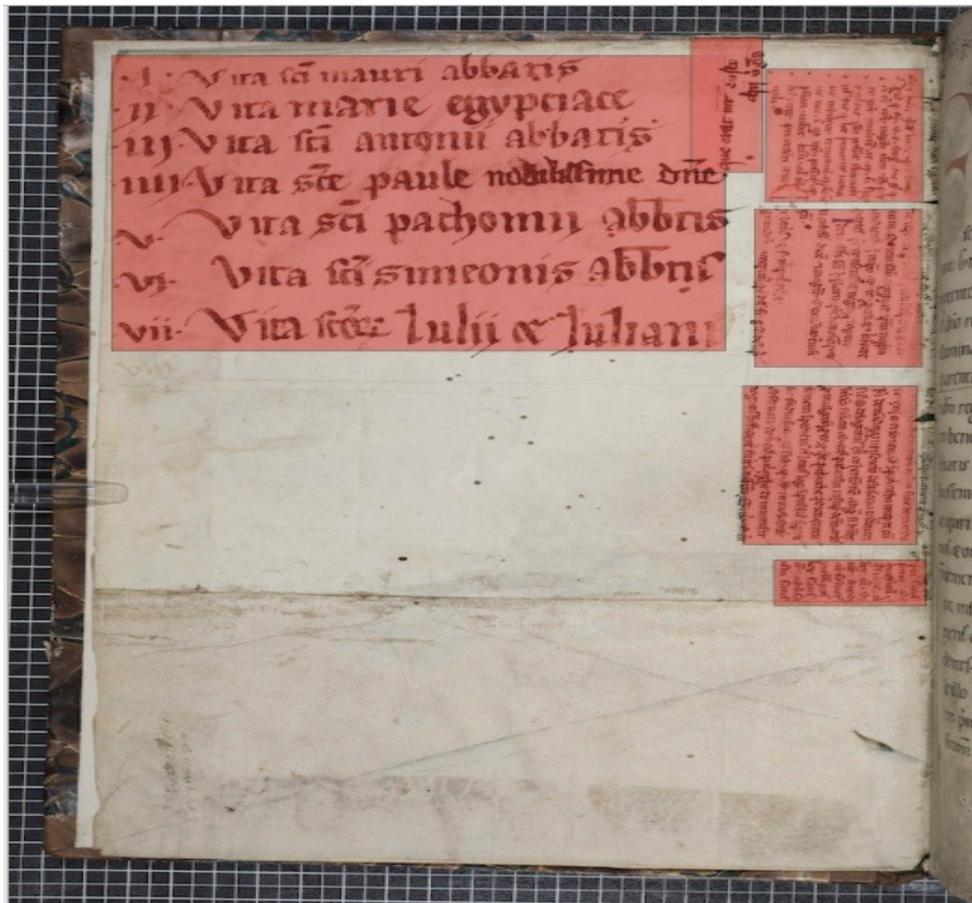


Figure 6: Zoning of cod. 705, f. Iv, Klosterneuburg, Augustiner-Chorherrenstift Library

`<surface>`

`<zone ulx="72" uly="57" lrx="484" lry="366"/>`

`<zone ulx="461" uly="38" lrx="509" lry="179" rotate="90"/>`

`<zone ulx="511" uly="71" lrx="617" lry="210" rotate="270"/>`

`<zone ulx="504" uly="218" lrx="616" lry="383" rotate="270"/>`

`<zone ulx="496" uly="403" lrx="613" lry="569" rotate="270"/>`

```
<zone ulx="517" uly="585" lrx="619" lry="633" rotate="270"/>
</surface>
```

It is clear how this type of encoding and these preoccupations are only appropriate in very special cases, since the capture of these type of details in the encoding (and not, for instance, within an editorial note) implies that they will be used to process or represent the transcription in a meaningful way. As a rule of thumb, an editor should refrain from applying unnecessary markup, where unnecessary means “not used computationally in any way.” This is because applying markup that is not used in any way distracts from the main purpose of the edition and clutters the encoding which should be otherwise kept as tidy and consistent as possible. On the other hand, only what is explicitly marked up can be later processed, either for display, or statistical analysis or any other purpose, therefore editors should always be very careful with their encoding to allow only the markup that is required by the purpose of their edition, but the one that is required should be consistently there (Pierazzo 2011).

4.1. Composite surfaces

At times it may be helpful to consider a level of analysis that goes beyond the page; for instance, if analysing the decoration of a manuscript, it could be necessary consider units such as openings; the codicology could be better studied at gathering or at bifolium level; damages or ink bleeding through the page may requires to studio the leaf, and so on. The TEI proposes one element to that purpose, **<surfaceGrp>**¹⁵ which requires one or more **<surface>**s as its children; a **@type** attribute can be used to specify the kind of grouping. In the example below two images relatives to f. 71v and f. 72r of ms. 685 of the Klosterneuburg, Augustiner-Chorherrenstift Library that presents a complex and rich iconography.

15. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-surfaceGrp.html>



Figure 7: Opening of f. 71v and f. 72r of ms. 685 of the Klosterneuburg, Augustiner-Chorherrenstift Library

The encoding of the grouping could be looking as in the example below:

```
<surfaceGrp type="opening" n="71v-72r">
  <surface n="71v">
    <graphic url="AT5000-685_71v-d.jpg"/>
    <!-- Here zones -->
  </surface>
  <surface n="72r">
    <graphic url="AT5000-685_72r-d.jpg"/>
    <!-- Here zones -->
  </surface>
</surfaceGrp>
```

This type of encoding will later allow to display the edition page by page or opening by opening, according to the preferences of the reader or of the choices of the editors. However, the mechanism does not allow for the creation of **<zone>**s trespassing the border of one surface; if the particular characteristics of a document or the purpose of an edition requires such trespassing zone to be established, the editors may consider to consider the entire opening as a **<surface>** in its own right, nesting two surfaces, as in the example below; in fact it could be arguable that if a decoration or some writing acts spread over the two faces of the opening, the opening has been considered as a single surface by the scribe or the artist.

```
<surface type="opening" n="71v-72r">
  <graphic url="opening.jpg"/>
  <!-- Here zones at opening level -->
  <surface n="71v">
    <graphic url="AT5000-685_71v-d.jpg"/>
    <!-- here zones relative to the page -->
  </surface>
  <surface n="72r">
    <graphic url="AT5000-685_72r-d.jpg"/>
    <!-- here zones relative to the page -->
  </surface>
</surface>
```

The **<sourceDoc>** approach is also helpful in cases where the writing support (i.e., the **<surface>**) is fragmentary or made from different materials, for instance when patches have been glued or stitched to the main surface. The image below (taken by a miscellaneous liturgical manuscript of the Capitular Library of Vercelli, Italy), shows a page where a cartouche has been stitched to the main page.

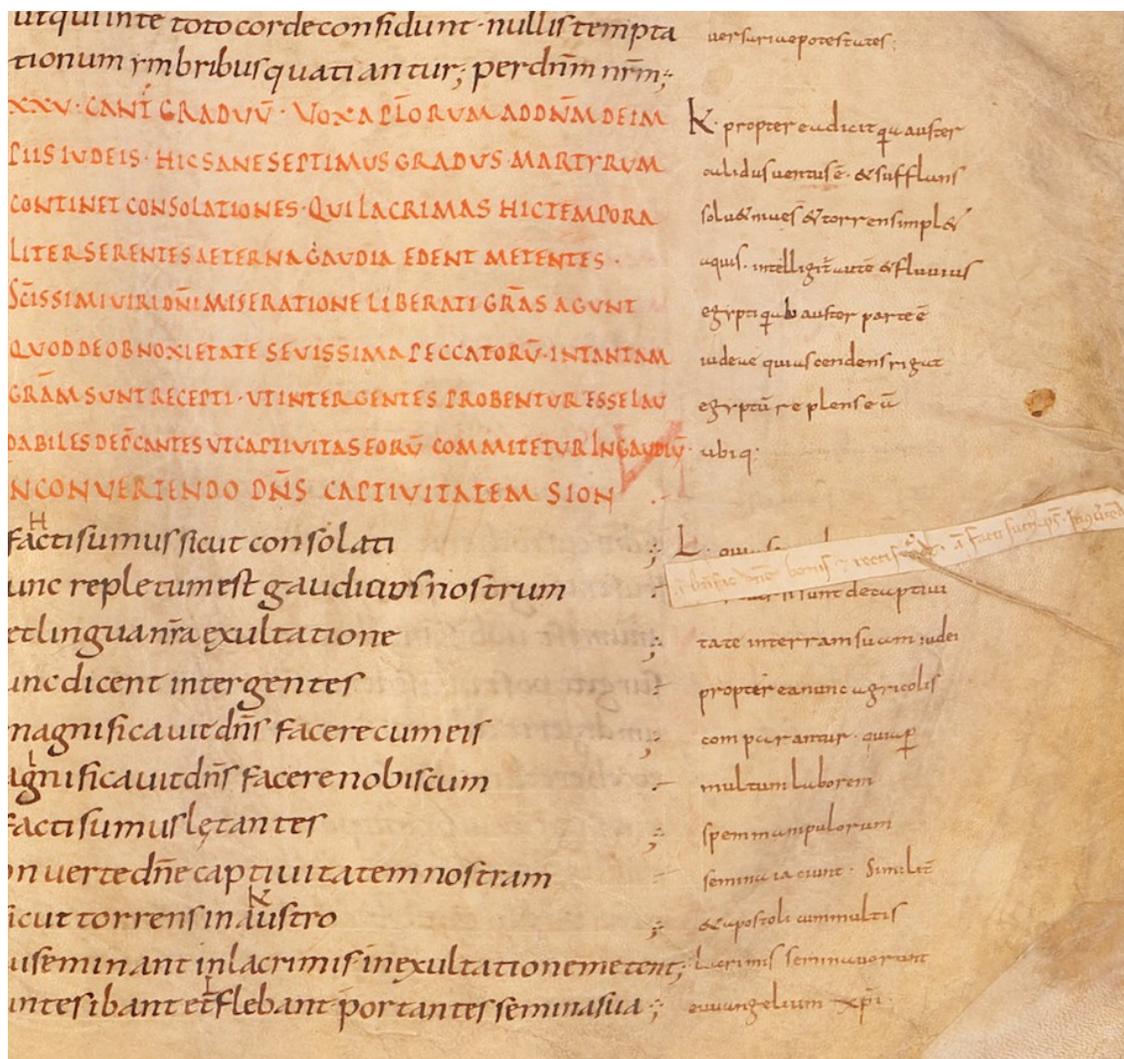


Figure 8: Zoning of cod.LXII Vercelli, Capitular Library

In these cases, since the patch (a sort of a medieval post-it note) represents a new material object, it should be encoded within a **<surface>** of its own, which can be embedded within the main one. This patch-surface carries a series of specialised attributes that allow to specify that the embedded **<surface>** is indeed a patch, that it is attached to the main one via a stitch and that the patch is flipping, i.e., its verso can be seen as well as its recto.

```

<surface type="page">
  <!-- many <zones> here -->
  <surface attachment="stitch" flipping="true" type="patch">
    <!-- reference to the image of the patch, zones, etc. here -->
  </surface>
</surface>

```

In case the editors wish to embed a transcription within the patch, a `<zone>` can be included as well; however, it should be noted that the coordinates of the embedded zone are to be given with respect to the `<surface>` relative to the patch, and not to the one relative to the page. In fact each `<surface>` element “resets” the coordinates to zero, since each `<surface>` represents an independent material object which could be (at least theoretically) be completely separate from any other.

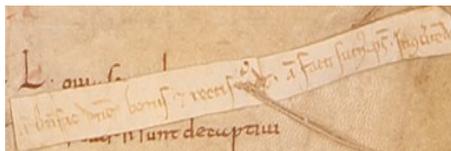


Figure 9: Detail with the patch Vercelli, Capitular Library, cod. LXII

```

<surface type="page">
  <graphic url="../images/BCVCmsLXII-f137r-crop.png" width="918px"
height="850px"/>
  <surface attachment="stitch" flipping="true" type="patch">
    <graphic url="../images/patch.png" width="385px" height="100px"/>
    <zone points="7,65 231,30 369,2 384,34 235,69 13,99">Antiphona:
Benefac, Domine, bonis et rectis corde. Antiphona: Facti sumus.
Psalmus: In convertendo</zone>
  </surface>
</surface>

```

However, the encoding above does not allow to establish where exactly the patch is located within the main `<surface>`; in order to do that, it will be necessary to wrap the patch-surface within a `<zone>` belonging to the parent `<surface>` and which coordinates establishes the exact location of that patch.

```

<surface type="page">
  <graphic url="../images/BCVCmsLXII-f137r-crop.png" width="918px"
height="850px"/>
  <zone type="patch" ulx="529" uly="381" lrx="908" lry="478">
    <surface attachment="stitch" flipping="true" type="patch">
      <graphic url="../images/patch.png" width="385px" height="100px"/>
      <zone points="7,65 231,30 369,2 384,34 235,69 13,99">Antiphona:
Benefac, Domine, bonis et rectis corde. Antiphona: Facti sumus.
Psalmus: In convertendo</zone>
    </surface>
  </zone>
</surface>

```

The complexity of this encoding reflects the complexity of unpacking a seemingly trivial phenomenon, namely the insertion of a piece of writing support (paper, parchment or similar) somewhere within another one. This operation, however, needs to be considered from different points of view: the insertion of new text (which could be an addition or a comment to existing text, or even a totally unrelated new material), the act of connecting such new surface to the existing one, the materiality of the added surface, as well as its exact location with respect to the “host” surface.

4.2. Encoding the time

As complex as the previous encoding may seem, an extra layer can be added if the editor wishes to capture not only the location, size and contents of the patch, but also the timing of its addition to the document. In fact, one of the purposes of an edition could also be to describe the process of producing the document or of writing the text, distinguishing also relative timing (i.e., something was written after or before something else) from absolute timing. This kind of preoccupations is mostly characterising the so-called genetic editing scholarly approach, which usually focusses on modern authorial draft manuscripts,¹⁶ but that can be usefully applied also to medieval manuscripts, particularly when studying readership traces, ownerships and provenance, palimpsests and so on. Editors might also want to group together certain phenomena such as a group of interlinear annotation, corrections of various sort, and so on. To do this it is necessary firstly to define and describe the criteria according to which such phenomena are being grouped and then a way to connect the single phenomenon on the surface to its description. The TEI calls such phenomena “changes.”

The first operation, namely the definition and description of changes or groups of changes, is done within the `<teiHeader>`, and more specifically within the element `<creation>`,¹⁷ a child of `<profileDesc>`. The element may contain one or more `<listChange>` element,¹⁸ which in turns contains one or more `<change>` element.¹⁹

Let us consider again the example seen in [Fig. 6](#). The image represents the verso of a flyleaf of ms. 705 of the Klosterneuburg, Augustiner-Chorherrenstift Library, which is a twelfth century manuscripts containing Lives of Saints. The flyleaf has clearly been taken from a pre-existing codex (probably a heavily glossed legal manuscript), which has been dismantled and its parts reused as it was usual for unwanted codices. Traces of the previous life of the flyleaf are well visible on the image: they are namely represented by the writing on the four blocks close to the spine, rotated by 270 arc degrees clockwise with respect to the main orientation of the page. After the insertion of the page as flyleaf, the empty space has been used to provide the newly formed codex with a table of content; to conclude the tormented life of this page, a note has been added on the right hand side of the table of content, rotated by 90 degrees clockwise. This complex history could be formally described as follow:

16. For a comprehensive introduction to *Critique Génétique*, see Grésillon 1994

17. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-creation.html>

18. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-listChange.html>

19. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-change.html>

```

<profileDesc>
  <creation>
    <listChange>
      <change xml:id="ch01" notBefore="0900"
notAfter="1100">Preparation and writing of the manuscript containing a
legal text with glosses.</change>
      <change xml:id="ch02" notBefore="1100"
notAfter="1200">Dismantling of the original manuscript and reuse of
parchment as flyleaves of a codex containing lives of saints; the
reused pages are rotated by 270 arc degrees clockwise and cut into
shapes to fit the format of the new codex</change>
      <change xml:id="ch03" notBefore="1200">Writing of a table of
content</change>
      <change xml:id="ch04">Addition of a note rotated by 90 arc
degrees</change>
    </listChange>
  </creation>
</profileDesc>

```

Here the elements `<change>` are provided with an `@xml:id` attribute, which will then allow to make references to each particular description. The *termini ante* and *post quem* are respectively encoded thanks to the `@notAfter` and `@notBefore` attributes.

Once the changes have been declared and described in the `<teiHeader>`, it is possible then to connect the transcription of the zones (as declared in the example below [Fig. 6](#)); this can be achieved thanks to the global attribute `@change`²⁰ which function is to make a reference to an `@xml:id` declared by a `<change>` element and therefore connecting the data to their metadata. The example below updates the code seen above with the references to the changes declared in the previous example.

```

<surface>
  <zone ulx="72" uly="57" lrx="484" lry="366" change="#ch03"/>
  <zone ulx="461" uly="38" lrx="509" lry="179" rotate="90"
change="#ch04"/>
  <zone ulx="511" uly="71" lrx="617" lry="210" rotate="270"
change="#ch01"/>
  <zone ulx="504" uly="218" lrx="616" lry="383" rotate="270"
change="#ch01"/>
  <zone ulx="496" uly="403" lrx="613" lry="569" rotate="270"
change="#ch01"/>
  <zone ulx="517" uly="585" lrx="619" lry="633" rotate="270"
change="#ch01"/>
</surface>

```

20. <http://www.tei-c.org/Vault/P5/3.2.0/doc/tei-p5-doc/en/html/ref-att.global.change.html>

The order of the **<change>** elements does not need to reflect the order of the **<zone>** (or any other element) carrying the **@change** attribute; furthermore, referring many times to the same **<change>** has the result of grouping together a set of zones or other elements.

The elements **<listChange>** can also nest or be repeated to present sub-ordering of **<change>**s or different types of grouping; in case the **<listChange>** is presenting a non-chronological list, it is recommended to use it with the attribute **@ordered** carrying the value of “false.”

The use of facsimiles in editing can assume different forms and different roles, which reflects different editorial theories and practices. It is certain that, whatever the reasons and the theoretical motivations behind the use of a digital facsimile within an editions, images of primary sources have had (and are likely to have in future) a major impact in the way we edit and access texts from the past. The TEI offers a solid way to handle facsimile and the different ways they can be used within an edition, but it is not by any means the only way nor it is perfect. It is nevertheless clear that the level of theoretical reflection and the modelisation that lies behind the handling of facsimiles in TEI is among the most advanced and in-depth analysis available to editors and textual scholars. ²¹Document-centric editing has taken a central role in textual scholarship, an evolution that is the result of a major technological improvement and of a modification in editorial theories; whether this is matter of a coincidence or that the change in textual scholarship is determined by the development of digital photography is not completely clear, but it seems reasonable to assume that this change will not be of short duration, and that new scholarly achievements will follow as a consequences.

Bibliography

Grésillon, Almuth 1994. *Eléments de critique génétique: lire les manuscrits modernes*, Paris: Presses universitaires de France

Kiernan, Kevin 2006. “Digital Facsimile in Editing.” In *Electronic textual editing*, Lou Burnard, Katherine O’Brien O’Keeffe and John Unsworth 262–268 New York: Modern Language Association of America http://www.tei-c.org/About/Archive_new/ETE/Preview/kiernan.xml

Nell Smith, Martha 2004. “Electronic Scholarly Editing.” In *A Companion to Digital Humanities*, Susan Schreibman, Ray Siemens and John Unsworth ?? Oxford: Blackwell <http://www.digitalhumanities.org/companion/>

Pierazzo, Elena 2011. “A rationale of digital documentary editions.” In *Literary and Linguistic Computing* 463–477 4 4. Doi: 10.1093/lc/fqr033

21. The TEI is updated twice a year by a large community of scholars, and therefore one can reasonably hope that possible drawbacks will be fixed in future (especially if the editors nothing them will take care to send an appropriate feedback to the TEI Council).

- 2014. “Digital Documentary Editions and the Others.” In *Scholarly Editing*, 35
<http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html>
- 2015. *Digital Scholarly Editing: Theories, Models and Methods*, Aldershot: Ashgate
- 2016. “Modelling Digital Scholarly Editing: From Plato to Heraclitus.” In *Digital Scholarly Editing: Theories and Practices*, Matthew James Driscoll and Elena Pierazzo, 41–58 Cambridge: Open Book Publisher
<https://www.openbookpublishers.com/product.php/483/digital-scholarly-editing--theories-and-practices>
- Pierazzo, Elena and Stokes, Peter A. 2010. “Putting the Text Back into Context: a Codicological Approach to Manuscript Transcription.” In *Kodikologie und Paläographie im Digitalen Zeitalter 2 - Codicology and palaeography in the digital age 2*, Franz Fischer, Christiane Fritze and Georg Voegler, Norderstedt: Books on Demand
- Robinson, Peter M.W. 2013. “Toward a Theory of Digital Editions.” In *Variants* 105–132 10
- Stokes, Peter A. 2010, “Scripts.” In *Handbook of Medieval Studies: Terms, Methods, Trends*, A. Classen, 1217–33 2 Berlin: A. Classen.
- Sutherland, Kathryn and Pierazzo, Elena 2012. “The Author’s Hand: from Page to Screen.” In *Collaborative Research in the Digital Humanities*, Marilyn Deegan and Willard McCarty, 191–212 Aldershot: Ashgate
- TEI Consortium 2017. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.3.0, last updated on 10th July 2017 <http://www.tei-c.org/Guidelines/P5/>
- Treharne, Eleine 2013. “Fleshing out the TEXT: the Transcendent Manuscript in the Digital Age.” In *Postmedieval*, 1–16 4 4
- Twycross, Meg 2008. “Virtual Restoration and Manuscript Archeology.” In *The Virtual Representation of the Past*, Mark Greengrass and Lorna Hughes, 23–47 Aldershot: Ashgate