

# THE CLASSICAL TEXT EDITOR. AN ATTEMPT TO PROVIDE FOR BOTH PRINTED AND DIGITAL EDITIONS

*Stefan Hagel* \*

For a single classicist, a discussion of the principles of the editor's work among medievalists is a bit precarious. The viewpoints of both professions may differ significantly, which is basically due to the different nature of the sources and their traditions, on which we have come to build our conceptions. Furthermore, I am not even an editor myself. Most of what I know about the editors' needs has been communicated to me as functionality to be included into my software. Surprisingly little of that, however, concerned digital editing. Consequently, I shall restrain my contribution to those topics with which I am most familiar: the technical side, and the average editor's approaches to the option of a digital output. My viewpoint is also determined by my main concern in developing editorial software: to support a careful and conscious way of dealing with our most valuable resource, life time.

Since Peter Robinson's essay on making digital editions of medieval texts has been announced as a basic reading for this meeting, I shall have to start with a slight clarification as regards the nature of the *Classical Text Editor* as a piece of software. In contrast to the lack of suitable tools for digital editions, P. Robinson mentions some for print editions: «Others are based on extensions of the Microsoft Word family of software: e.g. *Imprimatur* and the *Classical Text Editor* (CTE), which seems the most fully developed and supported, and has been used by more than 16 projects and editions»<sup>1</sup>.

The CTE, however, has almost nothing to do with *Microsoft Word* – except a general orientation towards its surface, in order to facilitate the transition to another word processor for the average *Windows* user. Beneath this surface, the CTE is based only on the

---

\* stefan.hagel@oeaw.ac.at  
Austrian Academy of Sciences.

<sup>1</sup>Robinson 2005, § 21. <<http://www.digitalmedievalist.org/article.cfm?RecID=6#N103B0>>.

*MS Windows API* (and Borland's *Delphi* with its VCL components). Secondly, the CTE is not devoted to print editions only, but can produce output in typical digital edition formats. Admittedly, this is not a major concern, and if P. Robinson did not just overlook it, I am the first to understand if he did not regard the results as what he would call a digital edition. In fact, when preparing this contribution I soon understood that I had better revise part of the export routines in advance.

The reasons for this lack of attention are not to be sought in my personal preference for the book. When the basic functionality of the CTE was established back in 1998, I set out to include HTML and TEI export, although no one had asked for such features. The feedback was overwhelmingly non-existent. During the past eight years, countless additions, extensions, not to mention bug fixes, were requested by editors working on texts of all periods and various genres. Practically all of these requests regarded the book-producing aspect: I remember only one person who inquired in behalf of the XML standards. In figures: there are currently over 500 licenses in more than 250 projects. Of these, four are explicitly for electronic editions, none of them institutional. If that reflected not merely the capabilities of the CTE but came near to an actual distribution, it would point to just 1% truly digital editors, and, perhaps even more disconcerting for those concerned about the future of the digital edition, little institutional interest<sup>2</sup>. In any case, lack of feedback inevitably leads to decreasing programming efforts, which drove the 'digital' CTE into sort of a vicious circle, only presently broken by this meeting.

So what, if anything, can the CTE contribute to the targets of a conference like this? In the words of P. Robinson, «Our goal must be to ensure that any scholar able to make an edition in one medium should be able to make an edition in the other»<sup>3</sup>.

There are two possible interpretations of this statement. Perhaps it aims merely at the basic scholarly and technical expertise, in the sense that we should ensure that starting a digital edition becomes no more difficult than producing the book. Still, even if this is achieved, I doubt that it would draw the masses towards the elec-

---

<sup>2</sup> Since the CTE 'print edition' license includes the electronic export, but not vice versa, the distribution is not so easily interpreted. Large projects, of course, come up with digital editing tools of their own, as amply illustrated in this volume.

<sup>3</sup> Robinson 2005, § 25.

tronic option. The CTE follows a different track, implementing the inclusive interpretation of Robinson's words: every scholar able to make a print edition with the CTE is *at the same time* able to make a digital edition, simply by invoking a menu command and specifying a destination folder. This strategy, however, does not solve all problems; I will address its disadvantages below.

Before, I want to touch briefly the most basic requirements for editorial software. Firstly, there is the aspect of output quality. This concerns not an attractive appearance of the edition, printed or digital, however welcome that may be. Quality is, above all, reliability of content. To guarantee this, the software must ensure that the editor's concentration can remain devoted to scholarly questions. Any sort of surfacing tags or other sorts of 'code' can be detrimental in this respect. And the distraction caused by a mixture of 'text' belonging to entirely different levels is only one side, although certainly the more important one. The other relates to possible misunderstandings generated by a mismatch between the views of the editor and the user. If both have, as far as possible, the same thing before their eyes, the editor will choose unambiguous means of representation quite naturally. Often this concerns not only textual information as such, but also its graphical presentation, especially if graphical features of sources shall be reflected.

Secondly, it appears important to produce the output not by a series of conversions effected step by step through a set of more or less integrated tools, but within one and the same software environment. The reason is that it must be possible to make the most basic changes at any step of the work and have them reflected in the output without additional effort. The absence of this possibility constitutes a psychological factor that should not be undervalued – and if the most meticulous editor should overcome it entirely, it may turn into respectable amounts of wasted time.

Apart from these factors with potential influence on output quality, there is also the aspect of promoting digitality among editions. Firstly, the promulgation of the most indubitable advantages of a digital edition will have no effect, if editors are discouraged: by the need to learn and to write some kind of arcane code, or by command-line driven pieces of software, however powerful and admirable they may be. On top of this, the production of a print edition may be compulsory, if only for bureaucratic reasons. In this case, the option of producing both types of output with as little additional effort as possible proves once more advantageous, if not indispensable, for our goal.

Thus, I would adumbrate the basic requirements for the optimal editorial software as follows:

- it should consist of only one tool that produces print and digital output;
- the user must be able to produce acceptable results with minimal technical expertise, and without writing any kind of code;
- still, advanced and advancing users should be able to extend the basic functionality, for example by applying additional tags for the digital export.

The CTE tries to implement these requirements, although, as stated above, with focus on the printed edition, especially where the needs of the target media conflict with each other. On the surface, it is a graphical word processor with specialised functions, mainly for text-referenced apparatus, for the maintenance of sigla for sources and groups of sources, and for work with parallel texts. Apart from manual input, it accepts files and Clipboard/Drag&Drop data as (*Unicode*) text and Rich Text Format (cfr. Diagram 1); as regards the latter, it may be suitable to adjust the footnotes in advance for import as text-based references.

Inside the CTE, the texts can be formatted and equipped with any number of apparatus, critical or other, and notes. In addition, XML tags can be applied (which have of course no influence on the printout).

Apart from print and PDF output for book production, there is a line-based HTML export, with the notes hyperlinked to separate frames. Designed at an early stage, it must nowadays be regarded as hopelessly old-fashioned. The major browsers acquiring more and more XML capabilities, the CTE's XML export is now much to be preferred. The syntax of the resulting files is based on the definitions of the *Text Encoding Initiative* (TEI)<sup>4</sup>.

The text-referenced notes are implemented by pointers, to overcome the nesting problem posed by the simple variant encoding schemes<sup>5</sup>.

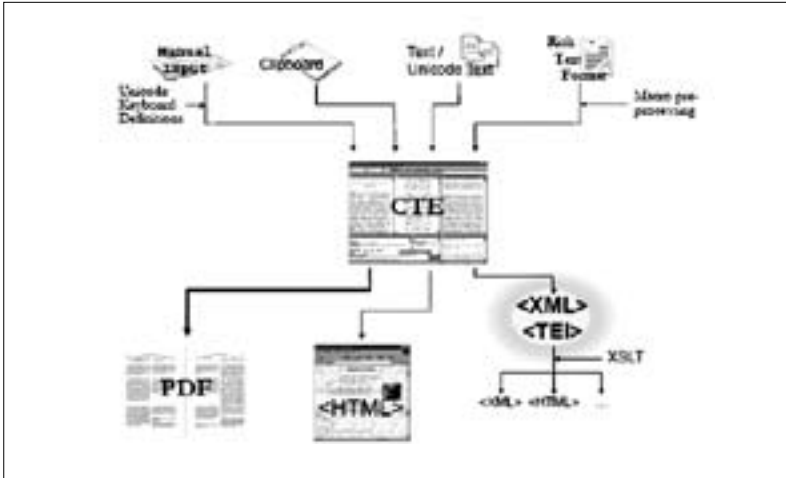
Although one can create a digital edition with the CTE, it does not support its publication any further<sup>6</sup>. Above all, no tools are included for creating stylesheets. These, however, are easily standard-

---

<sup>4</sup> <http://www.tei-c.org/>.

<sup>5</sup> Currently, the entries are put after, not before, the first word to which they belong. This ensures a comfortably interpretable display in the browser; future versions may introduce both options.

<sup>6</sup> On the problems of putting up the publication cfr. Robinson 2005, § 16f.



**Diagram 1** The CTE data flow model

ized: once suitable definitions are available, it suffices to put the file into the same directory as the CTE-generated XML file.

A more sophisticated option than the simple formatting stylesheet (such as CSS), is post-processing the XML output by XSLT<sup>7</sup>, probably at the server side, for viewing either in HTML or a different kind of XML. This technique opens exciting possibilities for the presentation of an edition. The considerable programming skills and effort that it entails are once more counterbalanced by the fact that the XSLT program can be reused in other editions.

The examples which accompany this paper illustrate merely the simple approach, namely directly formatting the CTE output by means of accompanying files. Among these are HTML files, which set up the frames and provide simple navigation tools in *JavaScript*, and CSS stylesheet files, which control the format. The sample texts with their notes have been generously provided by Pere Casanellas of the *Corpus Biblicum Catalanicum*<sup>8</sup>, to whose special needs the CTE owes several sophistications. I have prepared two presentations of very different style, oriented towards the present capabilities of two widely-used browsers, Microsoft's *Internet Explorer* and *Opera*. Both have their advantages and disadvantages, and if one opens

<sup>7</sup> <<http://www.tei-c.org/Stylesheets/>>

<sup>8</sup> <<http://www.abcat.org/cbcats/>>.

the respective presentations with the wrong browser, only part of it will work. Note, however, that only the mentioned additional files are different, which are easily copied and/or adapted to another project; the XML files, exported from the CTE, are identical.

One main shortcoming of the *Internet Explorer* (v. 6.0) is that this browser is still not able to interpret CSS attribute selectors. Thus the various formats exported in `<hi>` XML tags cannot be dealt with by means of CSS. Luckily, however, the *Internet Explorer* implements a Document Object Model in which the content of any tag – and not just of HTML tags – can be addressed as an object. Thus it is possible, albeit time-consuming, to assign the required formats programmatically by *JavaScript*, as soon as a file is loaded (or by pressing a button). In the example, the variants and notes are displayed to the right of the text<sup>9</sup>. Checkboxes are supplied to turn them on and off. In addition, a `<Go to>` button permits navigation within the text<sup>10</sup>.

The four versions of the text, which the example for the *Internet Explorer* offers alternatively, are displayed simultaneously in the presentation optimized for *Opera*. Their frames are synchronized programmatically (an option that can be turned off by deactivating a check-box). Since the text of four versions consumes much screen space, the textual variants are at first hidden, and of each note merely the first line is visible, in order to indicate its presence. The missing information is accessed by mouse-action: if the cursor is positioned above a note, its entire content is shown in a box. The critical apparatus for the individual paragraphs, on the other hand, is displayed if the user clicks on the text<sup>11</sup>.

From the foregoing it has become clear that the contribution of the CTE to the issue of promoting digital editions follows mainly one strategy: to lure the traditional editor into publishing also an electronic version. But if this editor does not set out for digital publication from the start, this affects of course the preparation of the data. Above all, a machine-readable critical apparatus will almost certainly not be maintained. Unfortunately this is in conflict with

---

<sup>9</sup> The faulty implementation of the CSS box model in the Internet Explorer ('box model bug') makes it especially troublesome to set up a presentation for XML files that displays beautifully on several browsers.

<sup>10</sup> Note that, in accordance with the quotation style used in the edition, a space must be inserted between chapter and verse number.

<sup>11</sup> Such a technique makes sense only where the paragraphs are short. Otherwise, an approach similar to that pursued for the notes might be useful, the presence of a variant being indicated by some symbol to click at or hover above.

what is perhaps the central dogma of the genuine digital editor. Once more in P. Robinson's words: «Fundamental to the model of electronic scholarly edition as it has developed over the past decade is the inclusion of full transcripts of all witnesses to the text»<sup>12</sup>.

This is probably the major sacrifice of the CTE approach: it is not to be expected that full transcripts can be extracted from data prepared with a printed critical apparatus in mind. In my opinion, though, it is better to have an imperfect electronic version to search in than to have *nothing* but the book. And those of us who in fact prefer reading a concise apparatus to hunting for the variants in a digital jungle may welcome a benefit within the deficiency: if the presentation is set up wisely, the edition will contain a human-readable apparatus<sup>13</sup>.

The 'secondary' digital edition will also fail to contain sophisticated links to graphical representations of the sources. Here, too, I hold that a little is better than nothing. On the other hand, if digitized images of the sources are available at all<sup>14</sup>, they can still be included, by adding appropriate tags right within the CTE. If undertaken, the effort should hardly be greater than in an originally digital project.

---

<sup>12</sup> Robinson 2005, § 26.

<sup>13</sup> When I created a tool for manuscript collation back in 1991, my first and natural idea was that its encoding should make it possible to retrieve the full text of each witness. This was immediately rejected by the editors who were going to use it. From the user's perspective, a full-text-of-all-witnesses edition without a concise apparatus of the crucial variants can sometimes appear as a more severe deficiency, insofar what in the case of many texts may be regarded as the editor's primary duty is missed: filtering information in which almost no one is interested from the data, in order to facilitate the life of the average user (in such matters we probably find a major point of controversy between the typical classicist and medievalist, and one must stay alert to how different views are imposed by different sorts of documents). The electronic form, of course, opened the possibility of providing variants on more than one 'level of interest' – down to the most boring orthographic variation, if this is what the editor wants to spend her or his life on. In the CTE, such differentiation is easily accomplished by distributing the variant readings between different apparatus. During the editorial process, maintenance of a full apparatus, if only for personal reference, in addition to the printed one, is a good idea, at any rate; thus two levels of variants may be available in a project even before anyone thinks about an electronic publication.

<sup>14</sup> A built-in graphics viewer encourages the CTE user to work with digitized sources (using a separate graphics program side by side with a word processor either embarrassingly constrains the available screen space or demands switching between applications not simultaneously visible. The integration within one MDI application overcomes this problem).

Let me conclude with a consideration of the circumstances in which software like the CTE may be beneficial for an edition that is conceived as electronic from the start. The relative simplicity of an all-in-one tool will of course attract the average scholar (as regards technical expertise), if no tools of comparable surface design are available that provide more definite support for the digital side of the editorial work. On the other hand, I also know of an expert in digital text processing who preferred to prepare (long) texts with the CTE for XML output: after all, this supplies a consistent, well-defined basic tagging, which can form the foundation for further processing.

Most efficient, finally, is a situation in which several 'average scholars' take advantage of one or few 'experts'. This can be done informally if stylesheets and other sorts of templates are circulated. In a bigger project, a basic tool like the CTE might be used by the staff for economical management of large textual quantities, ensuring uniform output at least at a fundamental level (features such as full text retrieval will demand further specification, of course, for whose observance the staff is responsible). All further processing, as well as the presentation, can be centralized, with the effect of separating the technical and the scholarly work as much as possible. Such a setting, in which all participants operate mainly in the field they are most familiar with, should considerably facilitate the maintenance of high quality standards.

### *References*

*Corpus Biblicum Catalanicum*, ed. Pere Casanellas <<http://www.abcat.org/cbcats/>>.

Robinson 2005 = P. Robinson, *Current issues in making digital editions of medieval texts or, do electronic scholarly editions have a future?*, in «Digital Medievalist», 1 (Spring 2005), <<http://www.digitalmedievalist.org/journal.cfm>>.

*The Text Encoding Initiative*: <<http://www.tei-c.org/>>.