Medieval Glosses as a Test Subject for the Building of Tools for Digital Critical Editions

Emmanuelle Kuhry
Medieval Glosses as a Test Subject for the Building of Tools for Digital Critical Editions

Emmanuelle Kuhry

ABSTRACT

The thirteenth-century Latin corpus of the Oxford gloss has challenging features. A digital scholarly edition of this corpus necessitates innovative solutions that on the one hand account for the complex structure and content of the text and its manuscript transmission, and on the other hand convey these data comprehensibly to the reader. There are still few user-friendly tools which allow scholars to encode a digital critical edition easily, particularly its critical apparatus, in TEI XML. In this article, I document my approach to editing a digital corpus of philosophical glosses, and explain the tools I am using and customizing to make available to a larger public of digital scholarly editors, especially in the field of medieval studies.

INDEX

Keywords: critical editions, tools, medieval manuscripts, Oxford gloss
1. The Oxford Gloss

The Oxford gloss, or glossa anglicana, is a chain of annotations copied in thirteenth-century manuscripts around the Latin translations of Aristotle’s treatises on nature.

Figure 1. First folio of Physica in manuscripts Firenze, Biblioteca Medicea Laurenziana Plut.13 sin.5, f. 1; London, British Library Royal 12 G II, f. 2; and London, British Library Harley 3487, f. 4.

The Aristotelian text was written in the middle of the page, in a large module. Most of the gloss was copied in the margins, which were designed to accommodate several columns of those notes, but a significant portion was placed between the lines. Each manuscript therefore has both marginal and interlinear annotations, which are referred to universally as glosses, no matter the length or content. Different techniques were used, even on the same page, to link a marginal gloss to its lemma, the word or phrase in the central text. Either an attachment sign was copied above the lemma and reproduced beside the gloss, or the lemma was repeated at the beginning of the gloss.

1.1 Gloss or Scholion?

The term gloss is sometimes distinguished from the term scholion or scholium. Both designate marginal or interlinear notes. For most scholars, a gloss explains a particular word through synonyms or a brief definition, whereas a scholium treats a larger part of the text and provides part of a general commentary (Tov 2020, 41–42). E. Tov lists even other types of additions, like interlinear and marginal corrections, exegetical elements, variant readings, or scribal remarks, but
in the end he admits that it is difficult to distinguish between them. Other scholars treat a *scholium* as a specific kind of gloss that is longer than average, and has an interpretative or explanatory nature. Glosses can be assembled in a separate commentary, and one can distinguish them according to a typology: prosodic, lexical, morphological, etc. (Wieland 1998, 59–60). In glossed Bibles, *scholia* would be notes added in the margins for teaching purposes, but not pertaining to the *Glosa ordinaria*, the corpus of glosses on the Biblical text standardized in the twelfth century (Morard 2017b). In the Oxford gloss, it is difficult to differentiate between glosses and *scholia* along the lines proposed by scholars for other corpora, such as position or length. Some interlinear glosses in one manuscript are marginal in another, and one version of a gloss might be short in one manuscript but in another it might be lengthy. It is nevertheless possible to distinguish types of glosses: lexical, translation-related, paraphrasing, variant readings, or commentary (some of which are lengthy, and quote from one or more other authors). Besides, as E. Tov (2020, 41) states, the term *glossa* or *gloss* is not used in the same way in Greek studies, because the Greek term *γλῶσσα* designates a word in a foreign language or dialect which needs an explanation. It seems that the meaning of explanation of an individual word, which was attached to the term *glossa* in the High Middle Ages as a continuation of the meaning inherited from the Greek, gradually dissolved into the larger meaning of “marginal and interlinear commentary” in the Late Middle Ages, as can be witnessed in the case of the *Glosa ordinaria* or of the *Oxford gloss*, which both designate a corpus of notes, both marginal and interlinear, attached to a central text, and for which the term *scholium* is seldom found in the literature. As a matter of fact, and with the meaning of marginal commentary in medieval manuscripts, *scholium* is a modern term, not used in this way in medieval lexicon and bibliotheconomy; *scholia* were then designated by the terms *notulae, glossae*, or sometimes *distinctiones* (Morard 2017b).

### 1.2 Materiality of the Oxford Gloss

Around twenty-six manuscripts have been found to contain part of the Oxford gloss. A complete textbook usually has two hundred to three hundred folios. In the manuscripts, not every treatise was glossed to the same extent. Some treatises are very sparsely glossed, and others receive glosses of another origin (for instance, Thomas Aquinas’s commentaries) that, obviously, do not belong to the corpus of the Oxford gloss.
Masters of Arts created the gloss at the University of Oxford so that students could understand Aristotelian philosophy. The gloss is believed to originate from the middle of the thirteenth century, but recent studies suggest that it was in use much earlier, maybe as early as the twelve twenties (Kuhry 2019). It reflects a lively learning environment in Oxford. Students copied glosses either during or after lectures—thus implying oral transmission—or they copied them from a manuscript model.

2. Challenging Structures

From the researcher’s point of view, this material is remarkably complex. Unlike the Biblical gloss, which achieved a high level of standardization, the Oxford gloss is quite variable across the manuscripts, which show different levels of transmission between text and gloss as well as changes in the marker-lemma which appears at the head of the gloss. There are thus at least three different possible levels of transmissions of the glossed text: at the levels of the central text, the lemma, and the gloss’s content.

The high level of variation among the manuscripts points to another factor of complexity: that the gloss is usually considered to result from oral transmission. This means that one cannot understand the text solely in terms of witnesses copied from a textual model. On the other hand, there is evidence for some transmission through copying: common errors between two witnesses, or a similar presentation of some groups of glosses in two or more manuscripts, which cannot be explained by the layout features of the manuscript or the position of the lemmas. Such features must have been arranged that way in imitation of a manuscript model.

The Oxford gloss gives us an insight, albeit indirect, to a reality which is still hard to apprehend concretely: the acts of teaching and of learning in thirteenth-century universities. Unfortunately, the location of the glosses (between the lines, in the margins), as well as their small size, make them difficult to read. Few editions have been published. The consequence is that the corpus is seldom used in academic research.
Beside the structural features, intertextual features need to be highlighted, such as the relationship between the Oxford gloss and its sources, like Averroes’s commentaries, and between the gloss and thirteenth-century literature that quoted it. As a matter of fact, careful identification shows that some later authors incorporated some of the glosses in their texts without ever distinguishing them, as if they were part of the cited Aristotelian text.

3. The “Digigloses” Project

3.1 Editorial Problems

The fluid and variant character of the glosses and of their links to the text make them difficult to edit. The few paper editions are partial (Burnett and Mendelsohn 1997; R. K. French 1997; E. J. French 1998; Galle 2008). Representing the links between the glosses and the text and the variant character of the glosses and the lemmata proves to be a problem in all of them. The most accurate edition is not very easy to read (Galle 2008). The corpus is huge, as the twenty-six or so manuscripts span a dozen different treatises. There is no electronic edition yet. My current project aims to make an electronic edition of the gloss on De plantis. The edition will provide a critical text, but also let the reader follow text and gloss as easily as a medieval reader could do, that is, text and gloss being immediately connected to one another.

Reproducing the medieval layout seems a difficult goal to reach, whether in print or digitally. The medieval annotators arranged the material to fit the space left on the page. To spare writing space they could abbreviate some places, or tighten the script. They could adapt the writing and disposition of the glosses according to the situation and context, because they could see and reflect on the text as readers. Tailoring an algorithm to resolve contextual and potentially subjective issues seems rather cumbersome, and to promise uncertain results. New solutions must be found to represent glossed corpora, other than the paper or digital editions tried so far.

The high degree of variation between witnesses, both in content and in presentation of the glosses, as well as the uncertainty about the transmission of the text (oral, written, or both) both make the idea of a stemmatic or reconstructed edition irrelevant, if not wrong-headed. Furthermore, the very instability and mutability of the manuscript tradition best helps us understand medieval teaching; differences are windows into the lively and social context which generated them.
is perhaps one of the main values of an edition based on the manuscript tradition, unlike the sole publication of a reconstructed text, as the content of the gloss is mainly a compilation of commentaries, and can probably not be retraced to an archetypal authorial version. On the contrary, the Oxford gloss is the result of gradual, collaborative work by Masters of Arts of the University of Oxford throughout the second third of the thirteenth century (Kuhry 2019). The ideal edition should thus represent the peculiarities of each of several manuscripts that carry a given treatise, to help readers identify those layers.

On the other hand, an edition which provides merely facsimiles or document-centered editions of each manuscript would not be a satisfactory scholarly edition. Born-digital critical editions are also different from print critical editions that have been digitized. They must offer features not found in a traditional critical edition, including means of reading, exploring, and analyzing the text. Document-centered editions and digitized critical editions are thus two models from which any digital critical edition should be distinguished.
3.2 Features of the Corpus

In the edition I am preparing of *De plantis*, I have determined that the corpus has the following main features:

1. From one glossed manuscript to another, the central text can be variant, with missing, variant, or additional lemmata, which means that the central text should display textual variance. The glosses being fluid, it is not possible to restrict the encoding to the lemmata: the entire Aristotelian text must also be encoded. On the other hand, the purpose is not to redo the *Aristoteles latinus*. Only the glossed manuscripts are transcribed and encoded.

2. At a given passage from *De plantis*, a gloss might be in one manuscript but not another, or may come in different versions. Because I wish to show such variation, the glosses’ variants must be encoded too.

3. A given gloss can be linked to different lemmata from one manuscript to another, or to a lemma not existing in the manuscript, which means that the design of the encoding should take into account the fluidity of the glosses, to reflect accurately the state of the manuscript tradition.

4. One gloss in one manuscript can be two or three in another.

5. One gloss can be interlinear in one manuscript but marginal in another.

6. Glosses in the same manuscript might be written by several hands.

These last two characteristics imply that any encoded gloss must allow additional levels of description, beside the text and its variants.

3.3 Editorial Principles

For this corpus, which is far less standardized than the Biblical *Glossa ordinaria*, stemmatical reconstruction of the text is irrelevant, even if it were possible, as I have already said. Before the digital edition can be published, several steps must be completed, according to a specific methodology. The manuscript witnesses must be classified according to the glosses’ typology. In the absence of precise dating for the copying of the glosses, a typology can be achieved only after significant samples have been collated for each one. Collected samples should help to identify layers of composition in the glosses’ content, and thus to define stages in the development of the glossed corpus in the thirteenth century. Then, a base manuscript containing an extended
version of the glosses is chosen and the variants of the most significant manuscripts of each type are collated against its text. This comparative activity will enable the creation of a thesaurus of the glosses, each gloss archetype—a gloss in an abstract sense, instantiated in different versions—receiving a unique identifier.

### 3.4 Creating Versatile Tools for Digital Scholarly Editions

In her paper presented at the 2015 TEI Conference, M. Burghart (2016) describes three areas in which the digital scholarly editor of medieval texts could need help: detection of human errors in the apparatus, ability to display different versions of the work-in-progress edition (corresponding to the text of different manuscripts), and handling of editions encoded with `<rdg>` or `<lem>` elements in the apparatus. Those needs are met by the TEI Critical Apparatus Toolbox, which she created.¹⁷ I would add another prior and very basic need: finding help to encode the edition. Indeed, the task of encoding is described as the main difficulty confronted by TEI users, partly because of the lack of user-friendly tools (Burghart and Rehbein 2012). One can find tools to annotate images and transcribe sources,¹⁸ but these tools do not enable the scholar to encode variants and the critical apparatus as specified in the TEI Guidelines. They are therefore purely document-centered tools, and we have seen above (see note 13) that there is a general demand for text-oriented digital scholarly editions designed to be critical. Collation of several manuscripts can also produce an automatically encoded critical apparatus using the TEI parallel segmentation method, by means of CollateX or Juxta software, provided one already has the separate transcriptions at one’s disposal.¹⁹ As such, the result is not a proper critical edition, as no selection of variants has been made to produce a critical text with an apparatus. Critical editors therefore still need a user-friendly tool which allows them to establish the edited text, select variants, and introduce all the necessary critical annotation.

¹⁷ As a matter of fact, many scholars are invited by the TEI to encode their research data but they have neither the time nor the opportunity to get proper training in TEI XML encoding. So the second part of the project consists in creating a panel of tools for the encoding of ancient textual sources in TEI XML. These tools will enhance existing software, and not create it from scratch.
The tools include XSLT stylesheets that convert styled Word or LibreOffice documents to TEI XML. Pre-encoding a text in a word processor can be useful, but frequently a deeper level of encoding is needed, which is difficult to reach working only on the text document. So a second category of tools is a series of frameworks or encoding environments made through customization of two widely used XML editors:

- **XMLmind XML Editor** (XXE), which I used to build a critical edition encoding environment. The encoding environment allows one to display the document in different views thanks to the use of CSS stylesheets, and to automate and speed up the encoding tasks with custom commands.

- **Oxygen XML Editor**, which supports right-to-left scripts, unlike the current version of XXE (however, recent steps have been made in this direction by the developers of XXE). A framework similar to XXE’s was developed for Oxygen’s Author mode, to be used by projects focusing on Hebrew or Arabic sources.

Customized frameworks in existing XML editors presents multiple advantages:

- The use of CSS stylesheets makes the appearance of the document being encoded significantly more user-friendly, which helps scholars unfamiliar with TEI encoding. Moreover, in both applications, views corresponding to different CSS configurations allow one to highlight groups of elements and attributes in different ways depending upon the work’s progress, thus facilitating the encoding tasks.

- In both applications, using custom commands and CSS stylesheets calling the commands or combined with CSS extensions enabling advanced encoding features allows the automatization of a great number of tasks, many of which are tiresome and error-prone when done by hand, like stand-off annotation, linking, or indexing.

- A framework can be adapted very easily to any other scholarly editing project. Each framework can be customized to the structure specified by the project, which is helpful for scholars with limited previous knowledge of XML and TEI, who could be confused by the myriad possibilities of the TEI All schema. Nevertheless, the raw encoding can be verified at any moment by switching to the text mode in Oxygen or to the tree view in XXE.
A framework can be configured to offer guidance for the use of elements and attributes, thanks to the use of custom commands and of features of CSS stylesheets like drop-down menus, checkboxes, and pop-ups.

For instance, the attribute values of a number of elements can be listed and explained in the `<teiHeader>` by means of several `<valItem>` elements grouped in a `<valList>`. These values are informed by the editor at the beginning of the work and can be completed (but preferably not changed) at a later stage.

**Figure 2. In the `<encodingDesc>` element of the `<teiHeader>`, defining a list of values for the `@place` attribute of the `<note>` element (Oxygen framework).**

In the Oxygen framework, inside the `<text>` element, the insertion of tags for which the typology of attribute values has been described in the `<teiHeader>` (and registered as available in drop-down menus) generates a drop-down menu listing these values, thanks to the use of XPath.

**Figure 3. In the corpus, choosing one of the defined values for the `@place` attribute of the `<note>` element (Oxygen framework).**

The same principle is used for references to witnesses in the critical apparatus, by means of the description of witnesses in the `<sourceDesc>`, and for references to particular hands described in `<handNote>`, inside a `<handDesc>`, which is a component of the `<msDesc>`.
That way, the editor can customize the semantic part of the encoding (namely, the attribute values), in a “go with the flow” mode, that is, after the formalization of the structure. This “open” characteristic is essential when it comes to adapting the framework to different projects. On the other hand, the frameworks do not prevent the user from inserting unexpected tags or from coming up with different attributes than the ones suggested. Ultimately, a restriction of the schema is desirable, notably for validation purposes (Burnard 2019).

The challenge in developing encoding frameworks is not so much technical, as they use features available in each application. Frameworks consist of command configuration and CSS files, sometimes XSLT files. They provide an economic, versatile way to provide scholars with customized tools for digital scholarly editing. The challenge is rather about ontology and modeling.
Among the difficulties or constraints are thus:

1. the need to offer tools specialized enough to fill the particular needs of each scholarly edition project, especially in the medieval field, but also generic and customizable enough to be used by other projects. A solution is to identify a core of basic needs likely common to many projects, and create a library of specialized commands for each scholar to apply as they like to their own project. The publication phase, whether in print or in digital form, can be supported by existing tools which the scholar can configure.

The tools described in this paper are meant primarily for my colleagues in medieval studies. Collaboration over several months has allowed me to collect information about the precise needs of several digital scholarly edition projects. The functionality and user-friendliness of the tools can be improved by enhancing the underlying code in accordance with the needs expressed during this test phase. Having been tested through use by scholars with little previous knowledge of digital editing, the tools can be reused by any other project.

2. the need to tailor the tools to the practice of scholars with little or no previous training in encoding and to anticipate potential issues they might confront while encoding.

Regarding this point, collaboration with my colleagues leading projects has also allowed me to draw a more precise sketch of their needs and of their behavior when switching from the creation of a traditional critical edition to that of a digital one.

The tools will be available online as soon as they are tested and stable. Currently two different frameworks are in testing phases in several projects. One of these frameworks is for critical editions and one for genetic editions, the latter being more document-centered. The user base includes all digital scholarly editors, especially in medieval studies.
3.5 Problems and Solutions

I was able to address the constraints imposed by the nature of the glossed corpus thanks to TEI encoding and to the support of frameworks. To test the selected solutions, I have encoded a small part of the text, the prologue of *De plantis*, as follows:

- The “parallel segmentation” method has been used to encode the variant readings of the central text and of the glosses.  

- Text and glosses have been separated into two files. Both are linked to one another via stand-off encoding, where the `@target` attribute on the central text’s lemmata (i.e., the words to which the glosses refer, which are encoded in a `<term>` element) points to the `@xml:id` attribute of the gloss (see figure 4). A complete stand-off approach would be more difficult because the central text is not fixed: it receives more variant readings as new manuscripts of the Oxford gloss are collated. Under such an approach, tokenization is more complicated, since the target text is constantly moving, and therefore the tokens would be as well. Besides, encoding the glosses in the same file allows one to compare the different versions of a particular gloss while encoding, thanks to the parallel segmentation, and thus to facilitate part of the philological analysis. This arrangement accommodates the fluid nature of the glosses, which cannot be encoded in a fixed point in the text. To anchor a gloss to the text while expressing its fluidity from one manuscript to another, the `<term>` element is set inside an apparatus entry: in a `<rdg>` element contained in an `<app>` element.

The consequence of encoding the lemma in a `<term>` is that the same lesson in two given manuscripts can result in two readings depending whether there is a lemma in one of them or not. This is indeed a problem because it overloads the critical apparatus, because the shifting of the lemma to another word from one manuscript to another is very common, and does not represent a variant reading in a strict sense. When encoding the lemma in a `<term>` element, which is itself hosted in a `<rdg>`, actual variant readings are mixed with lemmata, the latter being very numerous. The problem might be solved by moving the lemma information somewhere else, possibly into another type of the `<witDetail>` element mentioned below, which should be located...
in the text after the word bearing the lemma quality, as if it were a note. The `<witDetail>` can have `@wit` and `@target` attributes, which point respectively to the witness in which the word is a lemma, and to the related gloss in the glosses' thesaurus.
The framework is particularly helpful when encoding two files that must interlink.

Figure 4. The working interface of the XMLmind framework. The central pane shows the CSS-customized view of the file containing the central text. The left pane shows the—almost—raw code of the same file. The + command designed in the CSS to be available on the lemma (element `<term type="lemme">`) opens a transformed version of the glosses’ thesaurus and allows one to choose the appropriate gloss archetype toward which the `@target` attribute must point (XMLmind framework).

- In the file containing the glosses (the glosses’ library or thesaurus), each gloss archetype is encoded in an `item` element, inside a first-level apparatus entry containing two readings: the first listing the witnesses not containing the gloss and having a `@type` attribute with value "omission" (see figure 5). This feature allows one to see easily, in the framework, a list of the witnesses transmitting the gloss or not. After the second `rdg` element, which contains the gloss’s text with internal variant readings, a `witDetail` element describes the location of the gloss in each witness (marginal or interlinear), by means of @n, and to specify, if needed, the hand which copied the gloss, thanks to @corresp. The `@target` attribute identifies the `<rdg>` at issue, and `@wit` points to the precise witness(es).
Medieval Glosses as a Test Subject for the Building of Tools for Digital Critical Editions

Figure 5. Display of the glosses’ thesaurus (XMLmind framework).

- `<lem>` elements are not used because there is no putative original text. Nevertheless, offering a single base text for a gloss archetype is a future possibility. A single reconstructed text seems to be necessary in a paper edition because of layout limitations. But it is also needed in a digital critical edition, which should provide a “representative text version with a canonical work structure” (Fischer 2017, S281), allowing citation and reuse both inside and outside the project. Translations, semantic analysis, inclusion in larger corpora, and so forth all need a canonical version.

- A future encoding layer will deal with the identification of the sources, probably with stand-off markup pointing to `<milestone>` elements framing the identified segment and referred to by the `@target` of a `<note>`. Each `<note>` will contain the references to the source of the passage, possibly through `@source` on `<author>` and `<ref>`, within a `<bibl>`.
3.6 Display Possibilities

A publication prototype has been created using XSLT stylesheets and Javascript to produce interactive HTML pages. The current version allows one to

- read the reconstructed text of each manuscript;
- open each gloss in a pop-up with a click on the lemma (each pop-up can be moved where one likes);
- show interactive and optional layers of information, like textual variance (by means of a form of critical apparatus), distinction between glossing hands (by means of color), and critical notes.

The transition from print to digital form represents a real opportunity for critical editions, with the possibility of offering more, if not all, available data. But the danger of overwhelming the reader with a mass of data is real, as well as of losing the critical criterion which is a feature of scientific editions. A balance must be found between usability / readability and completeness. A selection of configurable (through display options) reading paths such as the ones provided by the digital editions published with the support of the “Pôle Document numérique” of MRSH Caen may be a solution.

4. Conclusion

Texts with complex structures allow one to test the limits of the technical possibilities available for historians not trained in computer science. In this paper, I have used the example of a corpus of medieval philosophical glosses to illustrate how TEI XML encoding can provide solutions to issues unsolved by print editions, among which are multiple layers of manuscript transmission, textual variance, and links between the central text and the glosses.

I have tried to show that it is possible to build flexible tools for digital scholarly editing, namely building frameworks within existing XML editors. Quick-tech solutions like these are critical to a future that seems to promise ever-decreasing funding in the humanities.
BIBLIOGRAPHY


NOTES

1 To read more about the Oxford gloss, see Burnett (1996); R. K. French (1997).

2 The corpus gathers the “older” Latin translations of Aristotle’s treatises on nature, which is why it is named corpus vetustius. These translations were completed between the middle of the twelfth century and the middle of the thirteenth century, before William of Moerbeke’s translations in the 1260s and 1270s (the latter are called “corpus recentius”). The former are: Physica, De celo et mundo (On the Heavens), De generatione et corruptione (On Generation and Corruption), Meteora (Meteorology), De plantis (On Plants, apocryphal), De anima (On the Soul), De memoria et reminiscencia (On Memory),
De sensu et sensato (On Sense and the Sensible), De somno et vigilia (On Sleep), De morte et vita (On Youth and Old Age, Life and Death, and Respiration), De differentiis spiritus et anime (On the Differences between the Spirit and the Soul, apocryphal), Liber de causis (Book of Causes, apocryphal).

3 On average the dimensions of the textbooks are 300 x 200 cm. See the entries for Corpus vetustius in the Aristoteles Latinus Codices (see note 16).

4 See also note 8.


6 For a typology of the concept of lemma in medieval glosses, see Morard (2017a).

7 See section 3.2 below.

8 At least two authors use quotations of the Oxford gloss in the second third of the thirteenth century: Bartholomeus Anglicus in his best-selling encyclopedia De proprietatibus rerum, and the anonymous author of the Aristotelian compilation Compendium philosophie or Compilatio de libris naturalibus Aristotelis (Ventura 2007; Kuhry 2019).

9 I give a survey of such solutions in my contribution (Kuhry 2020).

10 About the lack of “authorial” character, see note 17. Siemens et al. (2016) reach a similar conclusion about the importance of accounting for the social context of the text.

11 Identification of the sources of the Oxford gloss and of the texts of Adam of Buckfeld, an early commentator of the Aristotelian treatises on nature at Oxford, shows that Buckfeld, while being a major source for the gloss, probably himself used a primitive state of the gloss (Kuhry 2019).

12 See Pierazzo (2016b), who argues for an opposition between a Neoplatonic view of texts, which perpetuates the notion of an ideal text to be reconstructed, and a Heraclitean view, which accepts and accounts for the mutability of texts.

13 P. Robinson (2013) exposes the debate between advocates of document-centered and text-centered editions, concluding that facsimile editions are not suited to creating a dialogue between the reader, the edition, and the editor, and that a digital scholarly edition should not concentrate on the document alone, but should provide the reader with means to explore the text. F. Fischer (2017) stresses the need for digital editions to supply critical value, corresponding to critical annotation.
He also calls for ways to search corpora of digital critical editions from a single gateway in addition to the possibilities offered on each distinct digital critical edition website. F. Duval (2017) argues for editions both critical and documentary, and highlights the importance of enlightening the reader about the complexity of the transmission of texts, which necessitates some form of critical apparatus to accompany an edition (Duval 2017).

14 Digital scholarly editions cannot be printed without loss of content and functionality, and depend on a digital paradigm in their conception: see Sahle (2016).

15 See the state of the art about functionalities, including the distinction between the concepts of “scriptons” (strings as they appear to the readers) and “textons” (strings as they exist in the text), described by Pierazzo (2016a, 34–36).

16 Initiated in 1930 under the auspices of the International Union of Academies, the Aristoteles latinus project envisions publishing critical editions of all the Latin translations of Aristotle’s texts, and of the works falsely attributed to him, that circulated in the Middle Ages. To this end, a catalogue of all manuscripts containing Latin translations of Aristotle was established: as a result, the three volumes of the Aristoteles Latinus Codices were published from 1939 to 1961 by a commission directed by G. Lacombe, followed by L. Minio-Paluello (Lacombe et al. 1939, 1955; Minio-Paluello 1961), now available at https://hiw.kuleuven.be/dwmc/al/DALE. So far, thirty volumes of editions of Latin translations have been published. To this enterprise can be added the parallel project of Aristoteles Semitico-Latinus, which aims to publish the Syriac, Arabic, and Hebrew translations of Aristotle’s works, of their Latin translations, and of their medieval commentaries.

17 A. Andrée draws the same conclusion about the Glossa ordinaria and the lack of “authorial” character in this type of text (2016, 9–10).

18 I have described the methodology (Kuhry 2020).

19 An example of an interesting research question that might be answerable: can we identify glosses dating from a pre-Buckfield era? (See note 11.)

Such tools include Transkribus (from READ-COOP SCE, accessed June 3, 2021, https://transkribus.eu/Transkribus/) and Archetype (previously called the DigiPal framework, now supported by King’s Digital Laboratory, King’s College London, accessed June 3, 2021, http://archetype.ink/).


Information about directionality is not documented in TEI except in @style and @xml:lang, and should be left to CSS: Martin Holmes, “Calling all Arabic and Hebrew experts,” TEI-L mailing list archives, October 29, 2013, https://listserv.brown.edu/cgi-bin/wa?A2=TEI-L;1bf8073.1310. The CSS module Writing Modes Level 3 describes how to handle RTL scripts (W3C Recommendation, edited by Elika J. Etemad and Koji Ishii, December 10, 2019, https://www.w3.org/TR/2019/REC-css-writing-modes-3-20191210/).

Some projects have been funded to produce an online TEI-encoded edition, but the encoding is done by contractors who are usually experts in the scientific field but are unfamiliar with TEI and XML.

The Ediarum framework for Oxygen has been developed at the Berlin-Brandenburg Academy of Sciences and Humanities to help scholars create digital editions of historical documents in TEI XML. However, this framework does not enable one to encode variant readings and is therefore not designed for critical editing of medieval sources. See Mertgens (2019) and the Ediarum website: https://www.bbaw.de/en/bbaw-digital/telota/research-and-software/ediarum. The Caen team Pôle Document numérique has since published a framework for critical editing in TEI Parallel Segmentation based on XMLmind XML Editor software: see “Éditer des sources avec appareil critique,” accessed June 4, 2021, http://www.unicaen.fr/recherche/mrsh/document_numerique/ outils/apparat.

I am taking part in a collective effort between scholarly editors and scientific publishing services (in collaboration with IRHT’s publication team, the Caen team, and Ecole nationale des Chartes) to collect and analyze editorial and encoding practices when it comes to ancient sources, in order to model a common core of editorial features and to make encoding and publishing tools more interoperable and complementary.

About tools, E. Pierazzo expresses the need for sustainability and ability to meet the needs of other projects, which implies the modeling of the working methods and concepts of a potentially large community of textual scholars (2016a, 111).

This solution could be compared to T. Andrews’s and E. Pierazzo’s different conceptions of a “third way” besides “traditional” or “manual” philology and use of the computer to satisfy precise needs without involvement from the scholar (which E. Pierazzo calls “computer-assisted philology”). T. Andrews’s idea of a third way involves complete digital philology actively invested.
by the scholar in every step of the editorial task, from transcription to analysis, which eventually produces unexpected results, thanks to innovative tools and methods crafted by philologists themselves, and finally to electronic publication (Andrews 2013; Pierazzo 2016a, 109–16). E. Pierazzo doubts that philologists can acquire such a highly technical profile in the near future, with exceptions. She advocates instead for a “bricks approach”: the definition of microtasks common to most text editing projects and for which individual tools could be built, customized, and combined in different ways according to the needs of each project (2016a, 109–17). The approach described in the present paper is closer to Pierazzo’s view.

32 See, for example, the “Print an edition” tool in the TEI Critical Apparatus Toolbox by M. Burghart, in which the sample XSLT stylesheet to transform TEI XML into LaTeX, to produce a printable PDF with a traditional critical apparatus, can be downloaded and modified: http://teicat.huma-num.fr/print.php. Burghart (2016) provides a survey of transformation and display tools for scholarly editions encoded with TEI. TEI Publisher (https://teipublisher.com/index.html) and the Max display engine (http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/max) are good candidates for online publication but the average scholarly editor probably needs help using and customizing these tools, especially when creating critical editions.


35 Parallel segmentation is not devoid of issues, mainly the lack of support for overlapping structures and of transpositions, and for the expression of critical data, like editorial comments, beside the variant itself; see Buzzoni (2017).
By “complete” stand-off, I mean a working scheme in which all the encoding would be stored outside of the document. One of the goals of stand-off encoding can be to preserve the text from interpretative marking (see Spadini and Turska 2019). In this case I use stand-off encoding only to prevent overlapping structures when applying several layers of encoding that correspond to different categories of analysis and interpretation.

On the prologue of the De plantis collated from seven manuscripts, in which this piece of text barely exceeds one folio, one finds 131 variant locations (<app> elements) holding lemmata, whereas only 50 variant locations do not have lemmata but have actual variant readings.

This encoding scheme has been used on the SourcEncyMe corpus of medieval Latin encyclopedias: http://sourcencyme.irht.cnrs.fr/.

E. Pierazzo writes about the “page paradigm” that continues to constrain digital scholarly editions (2016a, 156).

See the list of projects at http://www.unicaen.fr/recherche/mrsh/document_numerique/projets.

AUTHOR

EMMANUELLE KUHRY

Emmanuelle Kuhry is a postdoctoral researcher in digital humanities applied to medieval manuscript studies at Institut de Recherche et d’Histoire des Textes, CNRS (Centre National de la Recherche Scientifique).