Schriften des Instituts für Dokumentologie und Editorik — Band 3

Kodikologie und Paläographie im digitalen Zeitalter 2

Codicology and Palaeography in the Digital Age 2

herausgegeben von | edited by

Franz Fischer, Christiane Fritze, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Malte Rehbein, Patrick Sahle

2010

BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

© 2011

Online-Fassung
Herstellung und Verlag der Druckfassung: Books on Demand GmbH, Norderstedt 2010
ISBN: 978-3-8423-5032-8
Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber
Satz: Stefanie Mayer und LaTeX

Putting the Text back into Context: A Codicological Approach to Manuscript Transcription

Elena Pierazzo, Peter A. Stokes

Abstract

Textual scholars have tended to produce editions which present the text without its manuscript context. Even though digital editions now often present single-witness editions with facsimiles of the manuscripts, nevertheless the text itself is still transcribed and represented as a linguistic object rather than a physical one. Indeed, this is explicitly stated as the theoretical basis for the *de facto* standard of markup for digital texts: the *Guidelines* of the Text Encoding Initiative (TEI). These explicitly treat texts as semantic units such as paragraphs, sentences, verses and so on, rather than physical elements such as pages, openings, or surfaces, and some scholars have argued that this is the only viable model for representing texts. In contrast, this chapter presents arguments for considering the document as a physical object in the markup of texts. The theoretical arguments of what constitutes a text are first reviewed, with emphasis on those used by the TEI and other theoreticians of digital markup. A series of cases is then given in which a document-centric approach may be desirable, with both modern and medieval examples. Finally a step forward in this direction is raised, namely the results of the Genetic Edition Working Group in the Manuscript Special Interest Group of the TEI: this includes a proposed standard for documentary markup, whereby aspects of codicology and *mise en page* can be included in digital editions, putting the text back into its manuscript context.

Zusammenfassung

Im Gegensatz zu früheren wissenschaftlichen Textausgaben bieten heute digitale Editionen von singulär überlieferten Texte meist auch das Faksimile der Handschrift. Dennoch wird dabei der Text weiterhin vor allem als ein linguistisches und nicht als ein materielles Objekt transkribiert und präsentiert. In der Tat ist dies die explizit formulierte theoretische Grundlage der Richtlinien der Text Encoding Initiative (TEI), dem *de facto* Standard für die Auszeichnung digitaler Texte. Dieser Standard behandelt Texte als semantische Einheiten wie Paragraphen, Sätze, Verse usw., nicht jedoch als materielle Einheiten wie Seiten, Doppelseiten oder Oberflächen. Manche Philologen bezeichnen diese Herangehensweise sogar als einzig verlässliches Modell zur Repräsentation von Texten. Dem entgegen wird in diesem Beitrag argumentiert, die

Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2. Hrsg. Franz Fischer, Christiane Fritze, Georg Vogeler, unter Mitarbeit von Bernhard Assmann, Malte Rehbein, Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 3. Norderstedt: Books on Demand, 2010. 397–429.

Transkription von Text auf dem Dokument als einem materielles Objekt zu begründen. Hierzu werden zunächst die theoretischen Grundlagen des Textbegriffes betrachtet, wobei der Fokus auf dem Textbegriff der TEI und anderer Theorien des digitalen Markups liegt. Dann werden anhand von mittelalterlichen und modernen Beispielen eine Reihe von Gründen benannt, warum eine Herangehensweise, die das Dokument in den Mittelpunkt rückt, wünschenswert erscheint. Schließlich wird eine alternative Philosophie der Textauszeichnung diskutiert, die aus der Arbeitsgruppe "Genetische Edition" der "TEI Manuscript Special Interest Group" resultiert: ein Vorschlag zur Standardisierung dokumentarischer Textauszeichnung, die Aspekte der Kodikologie und *mise en page* in digitale Editionen integriert und damit den Text zurück in den Kontext seines materiellen Trägers bringt.

1. Introduction

In any branch of manuscript studies (editing, codicology, palaeography, art history, history) the first level of enquiry always is (or should be) the document, the physical support that lies in front of the scholar's eyes. The fact that the text was transmitted to us by means of a specific physical object which has been organised in a certain way and preserved in one place or another has all sorts of consequences in the way we understand and receive that text. To understand the text that is contained in the manuscript, a deep study of the manuscript itself is fundamental: the layout, the type of script, the type of writing support, the binding and many other aspects can tell us about when, where, how and why this particular text was written in the page. It is also worth noting that the manuscript as object is increasingly becoming the object of study itself. This is based partly on the principle that a text cannot be understood outside its context, but also that the manuscript itself can tell us things that a text cannot, particularly if one is interested in the person or people who compiled it and the intellectual *milieu* in which it was compiled. On the other hand it is also very difficult to understand how, when and where a particular manuscript was produced without understanding the text(s) that it contains and the cultural circumstances that determined its production. Texts and documents live and make sense only within each other.

Nevertheless when it comes to transcribing and editing, the text is often taken out from its physical support, its context, and considered on its own, with only little, if any, evidence retained that it was once within a specific manuscript. This is the case, for instance, with the first edition of Jane Austen's minor works, which were published by Chapman from the 1920s and collected in a single volume in 1954: from this edition (and all subsequent re-editions) the evidence that some of those texts come from heavily annotated draft manuscripts is missing, with the consequence that the texts and her

writing habits have been misunderstood by more than one reader.[1] The words used by Peter Schillingsburg about the difference in the implications and interpretation between print and digital can serve here as well:

> Meanings are generated by readers who have learned to deal with symbols and formats. Change the symbol and the meaning changes; change the format and the implications are changed; change the contexts of interactions with texts and the importance and significance of the text changes.
>
> (Schillingsburg 2006 146)

2. The TEI Guidelines and the Encoding of Documents

In the case of digital editions, this centrality of the text is encouraged by the structure and principles of the most prestigious standard for text encoding, the one produced and maintained by the Text Encoding Initiative. The approach of the TEI, in fact, forces scholars to consider the text first. The TEI certainly offers a very sophisticated way of describing manuscripts; however, when it comes to transcription, of the two main hierarchies (text and document) the TEI privileges the text, relegating topographical description to empty elements (<pb/>, <lb/>, <cb/>) or attributes (<add place="...">, <note place="...">); it is no coincidence, after all, that it is called the *Text* Encoding Initiative. The TEI does not say that documents are not relevant, but rather that they are less relevant than texts; to use a metaphor from bibliography, texts are "substantial" while documents are "accidental".[2]

By using TEI, we have learnt to distinguish how to mark a text for what it really is (using descriptive markup) from what a text will look like when it will be output in print or on the screen (using procedural markup) and we have learned how this will help us in managing and preserving our data at best. If we are transcribing and encoding a text from a primary source (be that source manuscript or print) then we have also learned to use graphical features present in the source as a way to de-code the (ambiguous) code of that source. For example, if some string in the source document is in italic, we now wonder why is it so (following to the descriptive approach): is it perhaps a title, a foreign word, or for emphasis? Again, if something is written in the margins of a manuscript page, we wonder if it is an annotation, a variant, an addition: the fact that is in the margin or, say, in the interlinear space, does not change the nature of the text in this respect. All of the above can be done without considering "*accidental* features

---

[1] One example is Virgina Woolf who, mislead by the appearance of *The Watsons* in print, imagined Austen writing very bare sentences and then coming back to add the "flesh"; in contrast the evidence of the manuscripts suggests that the process worked the other way around, with the author "scratching out" superfluous words. See Sutherland 2005 140.

[2] This terminology of substantials and accidentals is in the sense established by Greg 1951.

(such as current lineation) or arbitrarily marked regions" (Renear 2004 223). In bringing about this approach, the TEI has

> succeeded [...] [in] the development of a new data description language that substantially improves our ability to describe textual features, not just our ability to exchange descriptions based on current practice.
>
> (Renear 2004 235)

This is valuable and important if you want to encode texts. However, as will be discussed shortly, there are many reasons why we might want to record the appearance of the source: that the string is in italics, for example, as well as or instead of why it is in italics. According to Renear (2004), this means that we are using procedural instead of descriptive markup. However, when we are trying to capture what the source document looks like, it is because we believe that this is at least as meaningful as the text it contains: we are documenting our source, not formatting our output, and so our encoding is descriptive, not procedural. In such a context, markup of pages, columns, lines, spacing, and so on may indeed be descriptive, not procedural (*pace* Renear 2004 224). As a matter of fact "what the text really is" depends on whether or not we think that Sperberg-McQueen's fourth axiom ("[t]exts are linguistic objects") is more, less or equally important than his fifth ("[t]exts occur in/are realized by physical objects", 1991 37–40; see also below § 4).

TEI is based primarily on the principles of text-oriented markup, but it does make some significant concessions toward documentary markup by including elements like <space/> ("indicates the location of a significant space in the copy text": Consortium 2009 § 11.6.1) or <hi/> ("marks a word or phrase as graphically distinct from the surrounding text, *for reasons concerning which no claim is made*": Consortium 2009 § 3.3.2.2; our italics). The reason for providing such elements is that the scholar-encoder is not always able or willing to state why some textual features look the way they do.[3] But while editors have the possibility of choosing between a semantically neuter <hi/> and the interpretative <emph/>, for instance, they cannot avoid the interpretational level when transcribing interlineated manuscripts, as they are only offered elements like <add/> or <note/>. Interestingly, the TEI also includes an element to capture page features: these are mainly for printed books and include <fw/>, which "contains a running head (e.g. a header, footer), catchword, or similar material appearing on the current page" (2009 § 11.7).[4] But while the TEI offers a way to encode a header and footer, it does not provide

---

[3] The Guidelines again: "If the encoder wishes to offer no interpretation of the feature underlying the use of highlighting in the source text, then the hi element may be used, which indicates only that the text so tagged was highlighted in some way. [...] The hi element is used to mark words or phrases which are highlighted in some way, but for which identification of the intended distinction is difficult, controversial, or impossible." (2009 § 3.3.2.2). See also Sperberg-McQueen 1991 43–44.

[4] That the element is intended for the printed page is clearly suggested by its full name: "forme word".

a way to encode the pages which contain those headers and footers, only the *breaks* between pages.

In practice the elements mentioned above have proved insufficient for encoding texts within their physical context (as will be shown further below). As a result only two options have been available to scholars who wish to encode documents: either they have been convinced (or they have convinced themselves) that what they really wanted was to encode texts, perhaps also preserving some features of the original document but at a secondary level, or they have invented their own system to encode documents.

3. Why Documents

Before outlining a possible solution to these problems, it is necessary to understand why an editor might want to transcribe a text within its documentary context. Although not always recognised by the community, there are in fact very many such reasons. To list all of them is beyond the scope of this discussion; instead a necessarily short and somewhat arbitrary choice will now be presented.

3.1. The process to make the document is at least as important as the text

Scholars are not always interested in the text as a coherent flow of words: sometime they are interested in the process of production or in documenting how and why a given document was produced or a text composed. This is the case, for instance, in genetic criticism. Genetic criticism (or *critique génétique*) has characterised the French school of philology since the 1970s and is concentrated around the activities promoted by the ITEM (*L'Institut des textes et manuscrits modèrnes*). The theories and practices of genetic criticism have spread beyond France and are now considered to be fundamental scholarly approaches to the editing of any draft or working manuscript (*brouillons*). Compared to more traditional approaches to editing, genetic criticism privileges the analysis of the *process*, the stratified flow of authoring, as opposed to the "photograph" of the end result which is embodied by traditional diplomatic editions. This is one—but by no means the only—scholarly approach for which the study of the process is relevant, and any understanding of the process must surely begin with the document.

A recent facsimile and semi-diplomatic edition of a manuscript of the libretto of *Tosca* may exemplify this. A manuscript containing Puccini's working copy of the libretto of *Tosca* has been recently purchased by the Fondazione Cassa di Risparmio di Lucca which has encouraged and authorised an edition of this manuscript. While the text of *Tosca* is relatively well established and does not represent a problem in itself, it was known that Puccini was deeply involved in the composition of the libretto, together with Giulio Ricordi (a music publisher and a opera producer, to use the modern terminology) and the two *librettisti*, Luigi Illica and Giuseppe Giacosa. What was *not* known was the contribution of each of them and the way they used to work. The manuscript can give

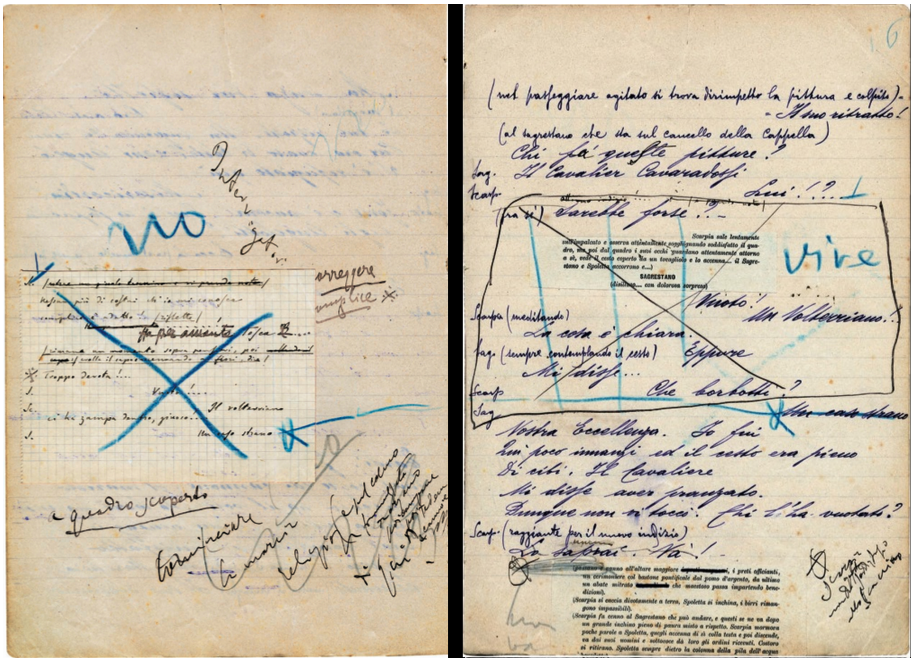

Figure 1. Tosca, pp. [38]–[39]. Reproduced by permission of Fondazione Cassa di Risparmio di Lucca.

an extraordinary insight into this matter, and the wonderful edition edited by Gabriella Biagi Ravenni is built around this principle (Sardou et al. 2009).

The working copy was most likely prepared within Ricordi's publishing house and consists of a stationery notebook, on the odd pages of which a version of the libretto has been transcribed by a professional scribe, interspersed with patches glued from an earlier printed version. The even pages were initially left blank for comments and annotations and now contain several layers of annotations by different hands which often spread onto the facing pages. The basic unit in the document is therefore the opening, the pair of facing pages as the book lies before the reader, and not the single page, as exemplified by Fig. 1.

In the transcription of the transmitted text the focus of Biagi Ravenni has been on the stratification of the hands (which she represents with different colors of ink), the temporal succession of the variants, and the disposition of words on the page. The semi-diplomatic edition reproduces the layout of the document, including the patches.

The result is not a text to be read in a traditional way, from the first to the last word, but instead each area tells a story and presents the intellectual effort of many people in producing the libretto. The document records the discussions, the thinking and the rethinking, the making and the unmaking of a process that lasted four years. The edition tries to make such a process explicit.

The fact that the edition of *Tosca* is not a text to be read should not surprise. According to Daniel Ferrer, "the draft is not a text, or a discourse, it is a protocol for making a text" (1998 261). This is demonstrated very clearly by the *Tosca* manuscript, and also by some of the pages of James Joyce which have been studied by Hans Walter Gabler. Gabler says:

> Thus, when I look at—for example—two pages from James Joyce's initial (disjunctive) draft for the "Circe" episode of Ulysses, my first concern is not to figure out what text the pages articulate. It is rather to find out how, as pages, they were successively filled. This means to analyse the patterns of ink and pencil on paper in terms of their inscriptional characteristics (that is: palaeographically), of their absolute positioning (that is: topographically), and of their relative positioning (that is: bibliographically).
>
> (Gabler 2007 201)

Another example of the process being as important as the text is given by medieval *libri vitae*. These are manuscripts in which religious houses recorded the names of people associated with that house in order to pray for their souls. In general, the surviving manuscripts often contain a core of names which were written in a single block when the manuscript was first produced, and then many further names which were subsequently added, sometimes by hundreds of scribes over many hundreds of years. One example which has been recently edited comes from Durham: it is a very complex manuscript in which hundreds of hands have been identified (Rollason and Rollason 2007). By disentangling the scribal hands and dating the stints scholars can reconstruct the genesis of the manuscript and, perhaps more importantly, witness the growth and development of a community. In this case more than 1,100 additions have been counted, ranging from the mid 9th to the 16th century. A similar example is the *Winchester Liber Vitae* which has been published in a facsimile edition, the editor of which has written the following about just one page and its opening:

> The present untidy appearance of the page is thus the end product of a highly complex process, representing the accumulation of names on many separate occasions in the eleventh and twelfth centuries. Yet the extraordinary sequence of entries on this opening [...] provides a striking reflection of the changing course of events at Winchester from the closing years of the Anglo-Danish dynasty, into the reigns of Edward the Confessor and Harold, and onwards past the Norman Conquest into the Anglo-Norman period and beyond.
>
> (Keynes 2001; cf Keynes 1996 96–98)

Not only can we see the community grow and develop, here by studying the names and noting their different ethnicities at different dates, but we can see glimpses into particular historical events as well, such as royal visits to ecclesiastical institutions which show us not only which institutions the king supported but also who he was travelling with at the time (for examples of which see Bolton 2008 *passim*).

Despite the significance of the genesis and layout of these books, the printed editions have chosen largely to ignore this and to print the names simply as one long list, with minimal attention to layout, if any. The best-known edition is that of Jan Gerchow who included a large set of *libri vitae* but preserved none of the manuscript layout at all, printing the names in continuous lines across the page as if regular prose (1988 304–20). A subsequent edition of the core of another *liber vitae*, this one from Durham, attempted to preserve the layout but ignored the genesis by excluding all additions (Dumville and Stokes 2001); the layout of even these core names is perhaps significant, however, since they were written in alternating gold and silver, a feature which was carefully preserved in the edition. Indeed, one of the principle reasons for producing this edition was precisely because the layout was felt to be so important even in the core, and because it had not been preserved in the previous editions. Finally, the most recent complete edition of the Durham manuscript is exactly two thousand pages long, in three volumes and with a digital facsimile, but it still prints the text without any reference to *mise en page*, instead printing the names continuously but ordering them first by manuscript page and then by date of scribal stint (Rollason and Rollason 2007 vol. 1).

3.2. The text is determined by the document

The physical constraints of the page

In some cases the content of a document is determined by the document itself: for example, if the author has only a limited amount of paper, he or she will probably modify his or her authoring intention to fit the space available. This is often the case with correspondence, especially modern, informal correspondence. The correspondence

of Puccini gives some clear evidence of this, such as the post card sent to Albina Magi (Puccini's mother) on the 10th of November 1880 reproduced in Fig. 2.

Puccini was in Milan to attend the local Academy of Music and was always short of money; he was using this as an excuse not to write home; therefore his mother used to give him prepaid postal cards to encourage communication. The availability of space is limited here: Puccini first writes a sort of a "normal" letter to his mother, then when he runs out of space, he starts to write telegram-like, a-grammatical sentences in every available blank space which remains, not only in the margins but even between lines. We have no idea about the temporal-logical succession of such additions, but we can assume that he would have written fully developed sentences and paragraphs if he had more space. Transcribing the text as a single flow implies an arbitrary decision on behalf of the editor about the logical succession of the parts.

The phenomenon of the physical constraints of the page limiting the text is usually associated with modern manuscript materials. This is for good reason, since it is normally only in this context that draft versions of texts survive, and so only here can we see the composition taking place. There are certainly examples in medieval manuscripts of scribes adjusting their writing to fit into the page, either as additions being crammed into spaces on the page (examples of which are legion) or the main text being compressed or indeed extended to neatly fill the available space. Despite the very many examples of these two kinds, it is often unclear whether this adjustment has any impact on the text: unambiguous evidence of scribes omitting material from their exemplar for this reason is rare. In some cases, however, we can see the production process "in action", so to speak, and sometimes here we find strong hints, if not concrete evidence, of texts being constrained by the physical space. One such example is a rare case of correspondence from the early Middle Ages which perhaps survives in the form in which it was originally drafted. The document survives as Canterbury, Dean and Chapter, C.1282, and consists of a letter written by one Earl Ordlaf to King Edward "the Elder" some time in the period between AD 897 and 901. This document has been the subject of intense scholarly interest for a long time, and this interest has included close study of the language and phrasing, not least because it is a rare example of vernacular literacy from the lay nobility (Keynes 1990 248–9; Keynes 1992; Gretsch 1994; Hough 2000; Brooks 2009). However, as some scholars have noticed (but many have not), the text seems to end fairly abruptly, and furthermore the last line is crammed onto the bottom of the sheet of parchment; the evidence seems to suggest that Earl Ordlaf ended his letter in this way simply because he ran out of space on the page that he had available. Granted this interpretation is arguable, and has indeed been argued, but the point remains that the text here may be determined by the document, and so scholars who wish to understand the text must at least be aware of this possibility and must have the evidence at their disposal to evaluate the impact this may have on their own arguments.

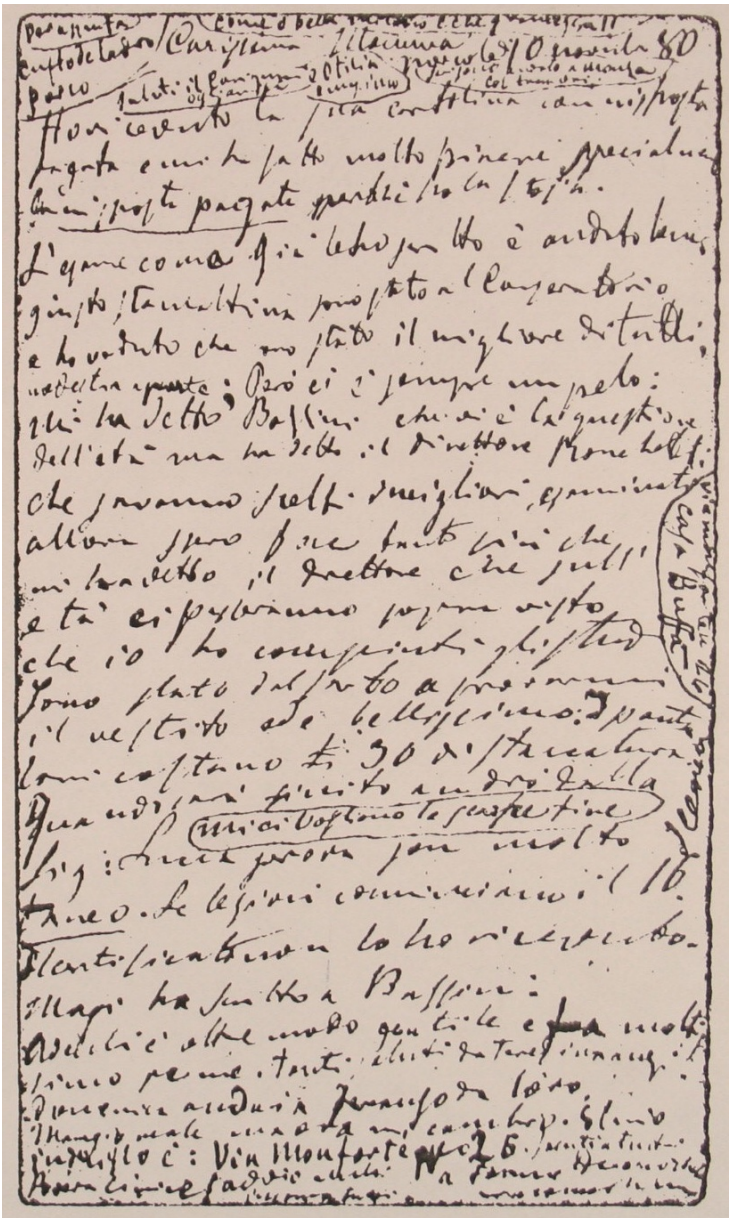

Figure 2. Post card from Puccini sent to his mother Albina Magi on the 10th of November 1880. The location of the post card is now unknown. The image is reproduced from Marchetti and Giuliani 1973 17.

Gatherings: can we understand the text without understanding the organisation of the document?

We have seen examples where the text cannot be fully understood without the document. There are other cases where the document and its codicology are required to understand the text, but also knowledge and understanding of the text is required to understand the document. One particularly well-known example of this is the so-called *Beowulf* manuscript. *Beowulf* is arguably one of the most important literary texts in English; it is an epic poem written in Old English, just over three thousands lines long, and has been the subject of almost innumerable scholarly and popular articles, books, translations, and adaptations, both written and cinematic. More recent scholarship has demonstrated that our understanding of this text is heavily determined by the codicological structure which preserves it. The only surviving copy of the poem is in a manuscript with several other works, and the relationship between these works has been debated extensively (Sisam 1953 61–96; Clement 1984; Kiernan 1996; Gerritsen 1998). However, as many scholars have failed to note, the quires of the manuscript were rearranged at some point, and there is even some evidence that *Beowulf* was once bound separately and that the manuscript as it survives today was originally conceived as two or even three separate volumes (Förster 1919 10–23 and 76; Ker 1957 281; Malone 1963 17 and 119; Clement 1984; Kiernan 1996 120–69). Similarly, another important debate relates to some damaged folios in the middle of the manuscript. The principle figures in this debate are Kevin Kiernan and Leonard Boyle: the former has suggested that this damage results from deliberate attempts by one of the scribes to erase and rewrite part of the text, whereas the latter has explained it by suggesting that the damaged pages were at a boundary between gatherings and that the gatherings were left unbound and were thereby exposed to water (Boyle 1981; Kiernan 1981; Kiernan 1996). Although this may seem like academic hair-splitting, the implications are very far-reaching, since Kiernan has used this position to argue repeatedly that the surviving manuscript represents an authorial copy of *Beowulf* and therefore that our understanding of the poem as composed in the eighth century or earlier is fundamentally wrong. He has also argued for revisions in editorial practice, since editors have tended to assume that our only surviving copy is a late and corrupt one and have therefore tended to intervene quite heavily in the text (Kiernan 1981; Kiernan 1995; Kiernan 1996 272–8 and *passim*). The implications of the codicology extend beyond *Beowulf*, too: the same manuscript also contains the only surviving copy of *Judith*, another important poem in Old English. This poem is written in the section of the manuscript which was certainly moved from its original position, and the text is now missing its beginning: how much is missing is unknown, but attempts to estmate the number of lines have been attempted based on the codicological evidence. These estimates have varied by orders of magnitude, and very different interpretations of the text have arisen as a result (Lucas 1990; Kiernan 1996 150–51).

These considerations are all very important for our understanding of these poems and therefore stand as examples of how our understanding of the text depends heavily on our understanding of the organisation of the document: in particular, discussion of these poems has been transformed by Kiernan's highly controversial interpretation of the them, but as Clement noted "[t]he collation [of the manuscript] is exceptionally important to Professor Kiernan's thesis" (Clement 1984 13). Unfortunately much of the evidence for this document's organisation and structure was destroyed when the manuscript was badly damaged by fire in 1731 (Prescott 1997); heated scholarly debate has ensued as a result, and at least six different and conflicting quire-structures have been published (Förster 1919 10–23; Dobbie 1953 xv–xvi; Ker 1957 282; Malone 1963 14–16; Boyle 1981; Kiernan 1996). For this reason, a detailed documentary edition which included full codicological evidence would be invaluable. In this case the structure of the gatherings themselves has been destroyed, as noted above, and so other forms of evidence must be preserved instead. Detailed measurements of the writing-frame, the exact distance between lines, on which side of the page the ruling was made, the arrangement of hair and flesh: all of these have been used as evidence for understanding the text, and all of them could usefully be encoded in a transcription of this manuscript. Indeed, it seems significant that Kevin Kiernan's own digital edition of Beowulf contained almost no codicological information in the transcription itself—all of this was relegated to the introduction or to his book, *Beowulf and the Beowulf Manuscript* (1996).

Another similar example is a pair of manuscripts which are now bound as one, along with fragments of a related third manuscript. These are all cartularies, that is, manuscripts containing documents which were issued originally as charters on single sheets of parchment but which were then copied into one book for administrative and organisational purposes. The three manuscripts in question were all produced at Worcester during the eleventh century: one, *Liber Wigornensis*, probably during the first or second decade of that century; the second, the "Nero-Middleton Cartulary", during the episcopate of St Wulfstan of Worcester (1062–1095); and the third, "Hemming's Cartulary", in the last decade of the eleventh century (Ker 1985; Tinti 2009 479). As Francesca Tinti has shown, the arrangement of the texts in the cartularies is significant and reveals much about the organisational and administrative practices in Worcester (Tinti 2002; Tinti 2009). *Liber Wigornensis* in particular is arranged in sections, and when the scribes finished a section they left the remaining pages in that gathering blank. In some cases these blank pages received further additions, but it also seems clear that the sections were rearranged at different times as the administrative principles changed (Tinti 2002; Baxter 2004 172–6; cf Tinti 2009 483–88). It is also worth emphasising that these issues are by no means limited only to these three manuscripts, but related questions also arise in other important sources for medieval history including Domesday Book, which is arguably the most important single source for the study of medieval

England and which also survives in two different forms, "Great" and "Little" Domesday: here even the spacing between words has proven significant in our understanding of this important pair of manuscripts (Galbraith 1961; Rumble 1985; Sawyer 1985 4). One author of this paper has been involved in discussions about a digital edition of the cartularies from Worcester, but these codicological issues leads to a number of complex scholarly requirements in any such edition. On the one hand, it is necessary to capture the current order of texts: this is a basic requirement of any edition of a single manuscript. In order to convey the different organisational principles, it is also necessary to capture the previous order (or orders) of the texts. This would then allow one to rearrange the material, presenting it in different ways according to the different arrangements. However, these prior arrangements are often difficult to establish, not least because the manuscript was damaged in the same fire as the *Beowulf* manuscript in 1731. For this reason, the editors of the proposed digital edition would like to allow scholars to rearrange the order of gatherings themselves, thereby allowing them to explore the material and test their own hypotheses. However, not all arrangements are equally likely or even possible. As with the *Beowulf* manuscript, codicological details such as ruling and the hair and flesh sides are all necessary to inform and constrain the possible arrangement of documents and quires. In this case the evidence and constraints are particularly complex, not least because they also depend on the arrangement of the text, and so the editors' ideal may not be achievable in practice, but nevertheless the framework for encoding this information is still a desideratum.

Manuscripts of homilies often come in codicological units which have been rearranged at different times: again, our understanding of homiletic practice and the homilies themselves often depends on the arrangement of texts within the manuscript, and this in turn often depends on the codicology. Pamela Robinson has demonstrated that some medieval manuscripts, particularly homiliaries, once existed as separate booklets which were unbound and designed to be carried around for preaching (Robinson 1980; see also Rumble 1985 33–35, for the application of this to Domesday Book). Although these are now bound as single manuscripts, the evidence for their previous existence as booklets often survives, and if one accepts that a text is determined in part by its presentation and use (as argued by Schillingsburg, as discussed above) then it follows that this information is important. Again, if one wishes to understand the homilies as a collection—a topic that is often discussed in the literature (a necessarily small and arbitrary sample of which is given by Cross and Tunberg 1993; Clemoes 1966; Eliason and Clemoes 1966; Loyn 1971; Sauer 2000; Da Rold 2007; Treharne 2009)—then one must understand how this collection once functioned not as a single, fixed whole but rather as a set of distinct units which were designed to be rearranged at will.

Many other examples can be presented of manuscripts in which our understanding of the text depends on our understanding of the codicology (for another detailed example see Stokes forthcoming). In most cases they are similar to *Beowulf*, insofar as scholars

have recognised that many texts depend on their manuscript context, and that context depends in turn on the codicology. It is important to note that the emphasis here need not be on the process but can focus only on the result: we may not be concerned with the process by which the Beowulf manuscript came to be arranged the way it is today, but rather in understanding how it was arranged in the eleventh century. The process in itself is certainly an important research question, as we have already established, but, as these examples show, even the original structure is often important to understanding the text and can require detailed codicological information to be preserved in the encoding.

In addition to these examples where the text and its genesis is the subject of interest, it has already been noted that the manuscript as object is also very much a legitimate object of study (see Section 3, above); in some cases, however, the codicology cannot be understood without considering the text. To give one example, Oxford, Bodleian Library, Auct F.4.32 is an extremely complex manuscript which is built up of several different units which were written at different times and places and bound together in different stages. The relationship between these units is very difficult to establish, not least because it was rebound in the modern period, and it was presumably at this time when some bifolia were inserted the wrong way around.[5] Even more complex is Cambridge, Corpus Christi College 367: M.R. James catalogued this manuscript as eight items in five distinct codicological units ranging from eleventh-century parchment to fifteenth-century paper (James 1912 II: 199–204), and again with folios misbound, reversed, different notes added at different stages, and so on (Stokes forthcoming). Our understanding of some of the texts in this manuscript depend utterly on their context: a booklist on folio 101v (*olim* 48v)[6] can be dated and localised very closely because of the two texts that it stands between; similarly, our understanding of the Vision of Leofric, an account in Old English the only copy of which is preserved in that same section of the manuscript, changes significantly when we recognise that the scribe who wrote another text in that section is the "Hemming "of *Hemming's Cartulary* (Stokes forthcoming; cf Baxter 2007 154–5 n. 6). How is one to present an edition of these manuscripts? If we present the text as it "should" be, with the folios put in their reading order, then we are not representing the manuscript. If we leave the folios as they are, then the text is unreadable. The obvious answer is: "This is a digital edition, we should present both the existing and the original arrangements." The ideal digital edition would allow one to view each of the different units separately; to view the manuscripts as they were bound at different times; to view the manuscript as it is now, and as it is but

---

5 As well as Hunt's facsimile edition (1961), the manuscript is now available online at ODL > Bodleian Library > MS. Auct. F. 4. 32.

6 I give here both the current foliation, established recently and apparently for the Parker on the Web project, and the previous foliation used in all published discussions to date, which restarts at the beginning of James's Volume II (James 1912 II: 200). For discussion see Stokes forthcoming n. 3.

with the incorrectly bound pages back in order. All of this is possible, but only with a documentary view.

3.3. The text is graphically presented

With respect to the printed page, the manuscript page (especially, but not exclusively, in the modern era) is free of constraints and develops in many ways. We have already seen examples of this, such as the *Libri vitae*, and also Puccini's correspondence, where the written words more or less anarchically stratify on the writing space (Almuth Grésillon speak about a space where "la ligne horizontale y perd bien souvent ses droits", Grésillon 1994 51). In some other cases, we find manuscripts where the unconventional layout clearly reflects intentionality which is plastic or explicitly artistic, as in the famous *calligrammes* by Apollinarie [7] or like the one by Jean Tardieu that is showing in Fig. 3.

Here the author is trying to represent with words the disposition of the mountains and hills that surround the Lake of Garda, and the reflection of those on the surface of the lake.[8] Clearly a linearised transcription of the text will irremediably loose a fundamental part of the poem's meaning.

Examples in medieval manuscripts can also be found without much difficulty. The most striking examples are perhaps found in Islamic and Jewish manuscripts, such as the Hebrew micrography which seems to have developed in the tenth century.[9] Although different in function to the modern ones, texts that are presented in graphical form are abundant in early Insular gospel-books, for example, such as the famous Lindisfarne Gospels or the Book of Kells. The opening page of each new book in the Lindisfarne Gospels is presented in a highly stylised and decorative format, so much so that the words are very difficult to read. This is illustrated by the "chi-rho" page in that manuscript, illustrated in Fig. 4, below. Many of these gospel books place special emphasis on this page, and scholars have suggested that intricate decorations like this were intended as something to be read like a text, to be meditated on and sought out in a nonlinear fashion as representation of the Godhead (Pulliam 2006 210; Brown 2003 77–8; cf the "Te igitur" pages as discussed by Suntrup 1980). Another of these gospel books is the Book of Armagh, which includes a page of readings from the Book of Revelations for which the scribe chose to arrange his text in a diamond format, illustrated in Fig. 5 below. Both the chi-rho page and the diamond-shaped one can be printed linearly, with abbreviations expanded and layout normalised, but a fundamental aspect of the page and its function is lost when this is done, and indeed it is significant that the second of the two manuscripts illustrated here was published in 1913 in an "*editio diplomatica*" which attempts to preserve the layout and some aspects of decoration

---

[7] See some examples, for instance, in Apollinaire > Textes > Calligrammes.

[8] Grésillon has provided the label of *écriture éclatée* for such type of writing (1994 57).

[9] Many examples of these can be found online; for one starting-point see JTS ([n.d.]).

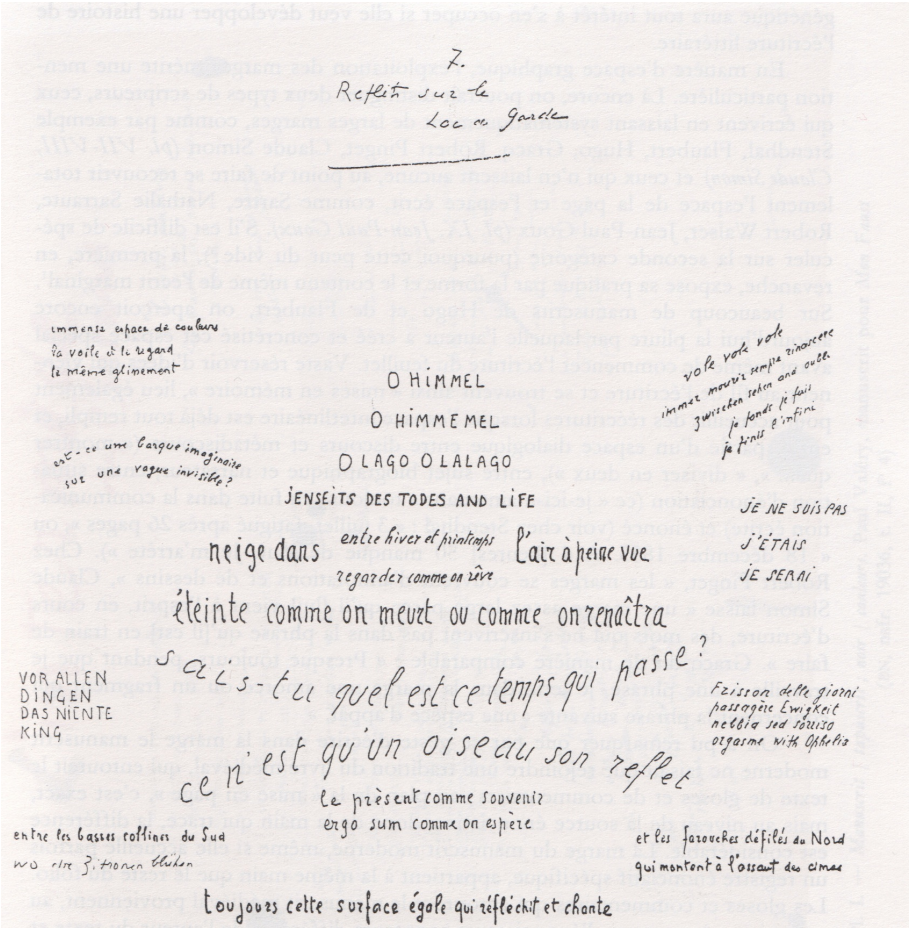

Figure 3. Jean Tardieu, *Reflet sur le lac de Garde* (reproduced from Grésillon 1994 59, which she reproduced from Tardieu 1990).

in the text (Gwynn 1913 i); a collotype facsimile of the first part of the book was also printed some years later (Gwynn 1937; for both publications see also *HyperStack*).

A further example is the set of so-called *carmina figurata* of Hrabanus Maurus, known as *De laudibus sanctae crucis* or sometimes *Opus in honorem sanctae crucis*. These survive as twenty-eight Latin poems, each of precisely thirty-six lines with thirty-six letters each which are designed to be laid out in a grid on the page; when so arranged, patterns are then formed by key letters which in turn spell out further words. An example is shown in Fig. 6, below.[10] It is therefore a very early predecessor to the *calligramme* discussed above, and it brings much the same challenges. Unlike the *calligramme*, Maurus' text does retain some meaning when printed as a conventional text, but one must ask how much is lost in such situations, and even the *Patrologia Latina* edition of 1864 printed each poem twice, first in diagrammatic form and then as conventional verse immediately afterwards (Migne 1864 col. 141a–264d).

In addition to these examples where the texts are presented graphically, there are many other cases where the physical arrangement of the words on the page is critical to understanding the text. Maps are one such case, and several projects have produced or are producing digital editions of medieval *mappae mundi*, maps of the world.[11] Another example is the Bayeux Tapestry, which again contains image and text to narrate its story and which has also been published as a digital edition (Foys 2003). Although these may be seen as primarily diagrams, nevertheless they do contain text, sometimes in significant quantities, and this must be captured in any edition. However, even these are relatively straightforward compared to works like Peter of Poitiers' *Compendium historiae in genealogia Christi*, also known as the *Genealogy of Christ*, an extremely popular work which was first written at the end of the twelfth or start of the thirteenth century (Munroe 1978; Hilpert 1985). Although relatively short, usually filling about seven or eight manuscript pages, it contains a very large amount of information presented in a sophisticated layout which is both text and diagram, incorporating the two into one. The *Genealogy* is preserved sometimes as a manuscript, sometimes as a roll, and the content is presented with varying levels of sophistication and clarity; one of many examples is shown in Fig. 7, below, but photographs of numerous others can be found online.[12] It is hard to conceive of any meaningful edition of this work which

---

[10] Further examples are Bologna, Collegio di Spagna 12 (reproduced at CIRSFID-Irnerio), Lyon, Bibliothèque Municipale 597 (reproduced at BM-Lyon) and Österreichische Nationalbibliothek, Cod. Vind. 908, fol. 3v (reproduced in Sperberg-McQueen 1991 41, Fig. 3).

[11] Example projects which are complete or in progress include the *Digital Mappa Mundi*, by Martin Foys and Asa Mitman (2009), and the *Linguistic Geographies* project which focuses on the Gough Map (Kline 2001, Gough-Map 2010).

[12] At the time of writing, these include three entries in Digital Scriptorium (searching for "Compendium historiae"); the one preserved in Harvard is reproduced in full by Harvard University Library, Page Delivery Service (HUL-PDS). Another is preserved in Oxford, Bodleian Library, Lyell 71, 17v–28r which is reproduced at LUNA (searching for "Lyell 71").



Figure 4. The Chi-Rho page of the Lindisfarne Gospels: London, British Library, Cotton Nero D.iv, 29r. Reproduced by permission of the British Library.

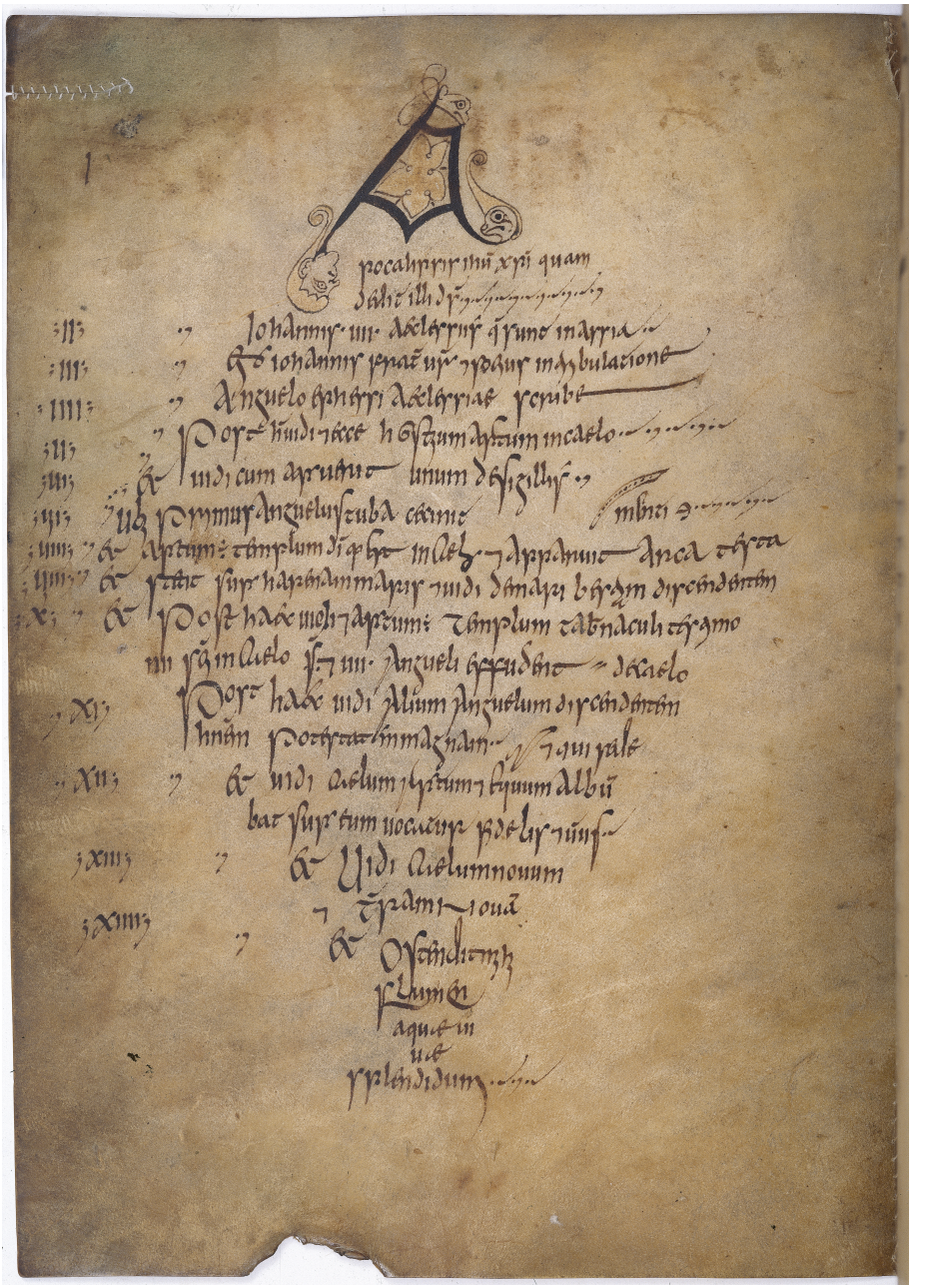

Figure 5. Readings from Revelations in the Book of Armagh: Dublin, Trinity College MS 52, 159v. Reproduced by permission of the Board of Trinity College Dublin.

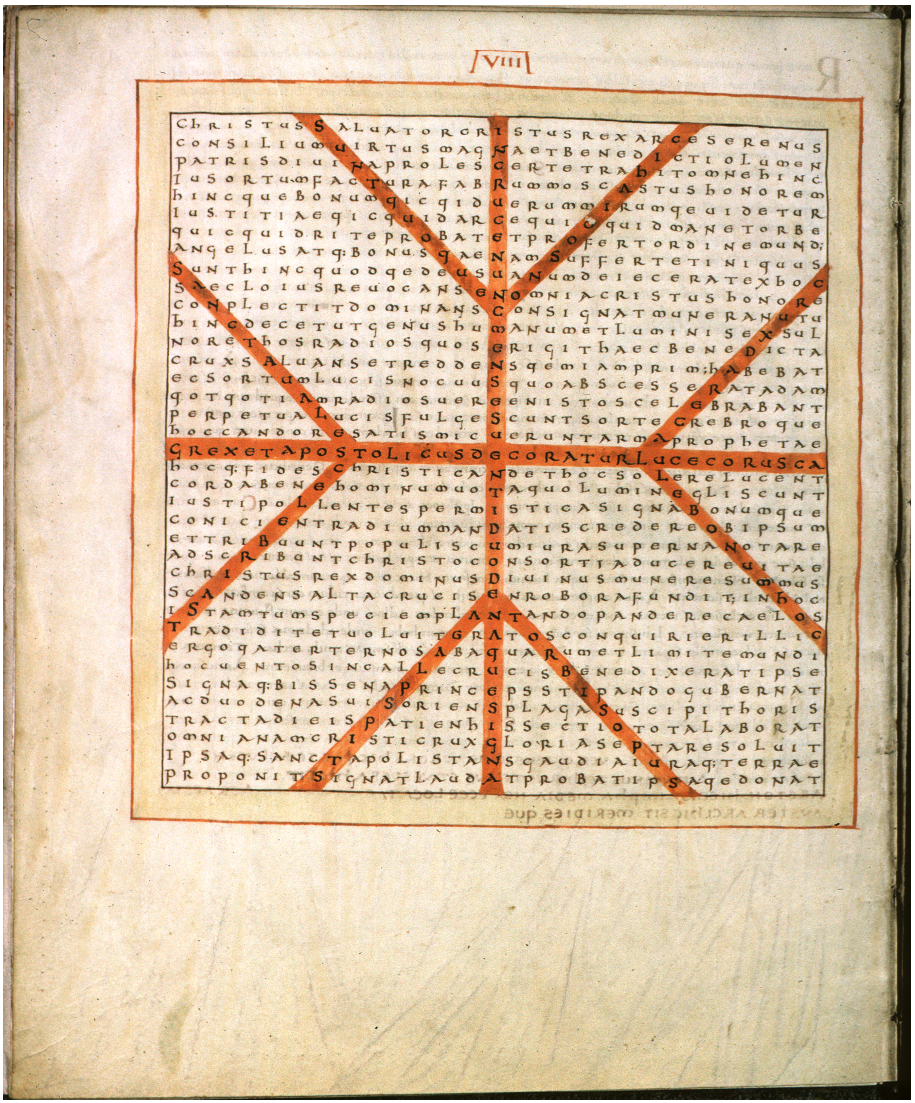

Figure 6. Hrabanus Maurus, De laudibus sanctae crucis. Bibliothèque municipale de Lyon 597, 5v. *Photograph: Bibliothèque municipale de Lyon, Didier Nicole. Reproduced by permission of the Bibliothèque municipale de Lyon.*

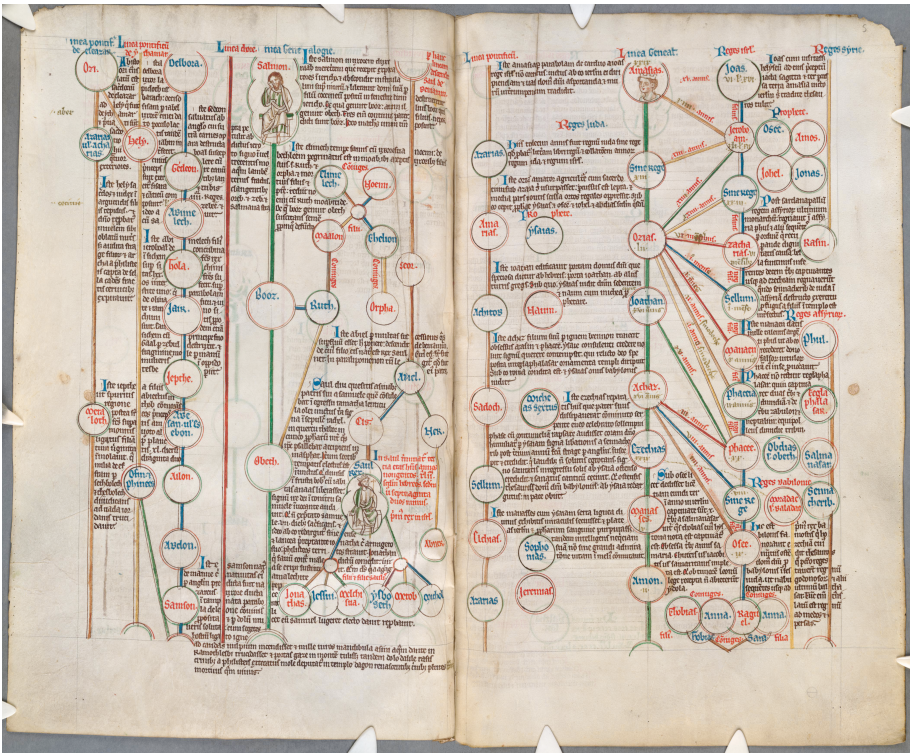

Figure 7. Peter of Poitiers, Genealogy of Christ. Cambridge, Corpus Christi College 83, 4v–5r. Reproduced by permission of the Master and Fellows, Corpus Christi College Cambridge.

does not somehow present it as both text and diagram, as neither is meaningful without the other.

Although arguably less extreme, another clear example of text which depends on layout for its meaning is the medieval gloss and commentary, in which a core text is given but with it is one or more layers of commentary. Many hundreds of manuscripts of this type survive and have been the subject of extensive study, with core texts including scripture or important and complex writers such as Boethius or Dante, or the *scholia* of Adam of Bremen. Even in the early medieval period, these manuscripts can acquire many layers of glossing, sometimes by tens of different scribes, in a complex system which includes various sets of interlinear and marginal additions. Initially these different layers glossed different aspects of the text: some might be linguistic, others

providing context, theological interpretation, and so on. In time, and particularly for biblical texts, medieval scholars started to produce commentaries of commentaries, and so the layering of glosses became more and more complex. From the twelfth century onwards, very complex and sophisticated page layouts were developed to accommodate these many interlocking texts (for examples see De Hamel 1987). The challenge, then, is how to edit these different texts while preserving the interconnections between them. Even simple linguistic glosses present problems, where (for example) the meaning of a Latin word is glossed with an alternative word written above it: as Raymond Page has reminded us (1992), more than one scholar has blundered due to editorial normalisations of these texts. If a "simple" case like this has lead to scholarly error, then what of the very complex glossed bibles like that shown in Fig. 8, below?[13] How can one accurately represent so many different texts and the relationships and connections between them without reproducing the layout of the page? In the past, some have attempted to print editions of these as simple, linearised texts (Meritt 1945; Meritt 1968), but this has resulted in significant loss of information at best, and disastrous blunders at worst.

3.4. There is no text

The final example to be considered here is draft manuscripts, in which the text is non-linear and can barely be defined text: as Daniel Ferrer reminds us, draft manuscripts are protocols, recipes to make a text (1998 261). Variations in the draft, also referred to as the *avant-text* or "pre-text", have been explored and studied principally by the French school of genetic criticism. When a revision is present on the page, it means that the text existed in at least two versions, the one before and the one after the revision; the more variations that accumulate and stratify, the more versions of the same text can be deduced. In order to disentangle the paradigmatic variation, the different possibilities offered by the written page can be made explicit, such as in the way Almuth Grésillon has presented the genesis of a verse from the poem *Une étoile tire de l'arc* by Jules Supervielle (Grésillon 1994 165-67): here we count sixteen different versions, all implied and potentially contained by the stratified draft manuscript.[14] In these cases we cannot speak of *the* text but of many possible texts, all enabled by the state of the document. A transcription should be able to offer the same possibility offered by the original manuscript, meaning that all possible readings should be present, not only the supposedly final will of the author.

---

[13] For a similar example in print, see Sperberg-McQueen 1991 45, fig. 7.

[14] For similar analyses of Giacomo Leopardi and Jane Austen, see Pierazzo 2009 182.

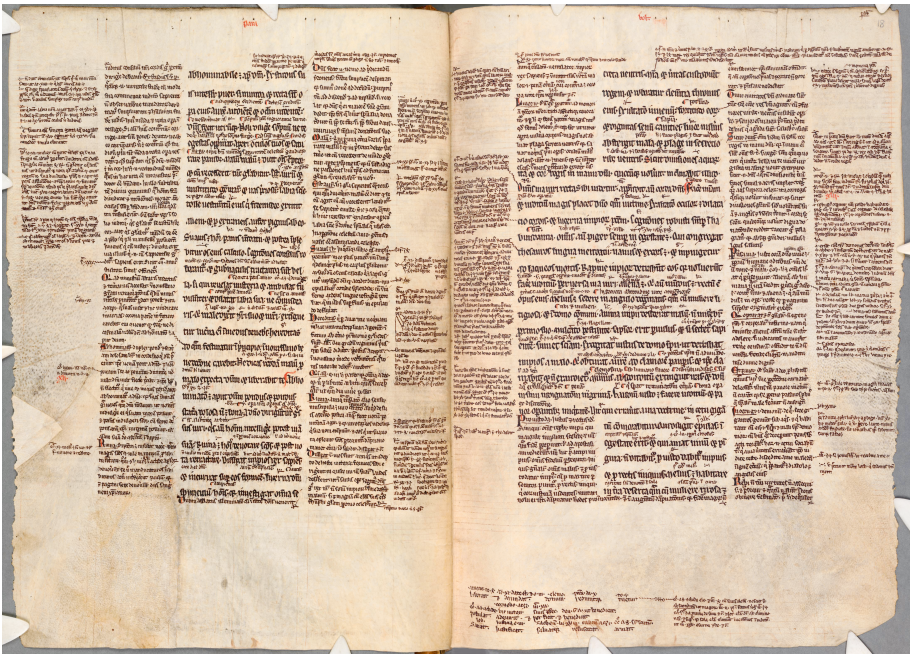

Figure 8. Glossed copy of Proverbs. Cambridge, Corpus Christi College 62, 17v–18r. Reproduced by permission of the Master and Fellows, Corpus Christi College Cambridge.

4. The OHCO View

When transcribing a text using the TEI encoding schema, one needs to take into account a few implicit theoretical assertions of "what text really is", and the implications behind the adoption of a descriptive, non-procedural markup.

According to Renear (and others before him) descriptive markup has so many advantages that it must be right.[15] More precisely, these scholars have argued that descriptive markup implies a single possible model of text—the only correct one—and that such a model postulates that texts consist of objects of a certain sort, structured in a certain way; in their view this means that texts have a linear order and they nest within each other without overlap. A text is then an "Ordered Hierarchy of Content Objects", or OHCO (DeRose et al. 1990), a view which is shared by the TEI, even if

[15] This position has been summarised by Renear 2004 225, citing his and others' work, particularly DeRose et al. 1990.

only implicitly.[16] Renear et al. derive from this statement that "A *book* for instance is a sequence of chapters, each of which is a sequence of major sections, each of which in turn is a sequence of subsections." (Renear et al. 1996, § *OHCO-1*); it is perhaps worth noticing that here in order to exemplify "what a *text* really is", the author has chosen to describe the structure of a *book*, thereby slipping from the immaterial abstraction of the text to the physical level of the document. However, as we speak about a physical entity (the book), we can also describe it as an object which is composed by a binding, a cover, and a sequence of pages which may or may not be organised in quires. This vision was not unknown to the OHCO working group, but it was rejected as not suited to scholarly purposes:

> [a] book can be divided into pages; a page into the header, the main text area (perhaps with several columns, embedded pictures, etc.), an optional footnote area, and a footer. However, even this model fails to provide the kind of text handling needed by authors and scholars. How can one find equations, poetry quotations, lines of verse, and the like?
>
> (DeRose et al. 1990 10)

But what if I don't want to find poetry quotations? What if the purpose of my research "is rather to find out how, as pages, they were successively filled", to use Gabler's words (2007 201)? Is this not also a legitimate scholarly purpose?

In our earlier discussion we have demonstrated that, when considering texts that are contained within a manuscript, in order to say "what a text really is", one must deal with the physical embodiment of that text.[17] In our opinion the OHCO view represents a highly idealised and simplified vision of the text which does not take into consideration the modalities and the circumstances of the transmission of that text. In the real world, there are fundamental layers of interpretation that are missed when the text is taken out of its context. To use a couple of metaphors (with religious connotations), text and document are like Ying and Yang, or body and soul: neither can live without the other. While these metaphors may perhaps go too far, in that we can (and indeed often do) study the document without the text or vice versa, nevertheless we inevitably loose an integral part of the whole when we do so, and this will necessarily limit our study.

---

[16] "The Guidelines of the Text Encoding Initiative exhibit a characteristically ambiguous stance: although they seem to privilege this view and benefit from its influence, they do not specifically invoke, explain, or defend it." (Renear et al. 1996, Introduction)

[17] Similar considerations can, of course, be applied to texts contained in printed books, but the case of manuscripts is perhaps even more evident for the reasons discussed above.

5. What to Do

How then can we encode both the textual and the documentary views, and how do these two views relate to each other? When we model an object to be studied we need to make sure we are not building a model that is as complex as the object of our study: a certain level of simplification and abstraction is required for the model to be useful. Decades of text encoding have shown that as soon as we try to mix texts and documents we encounter overlapping hierarchies: textual boundaries do not coincide with documentary ones, apart from some very specific cases, and to handle both at the same time is not possible, especially with a strictly hierarchical markup language like XML.

These two perspectives—textual and documentary—are therefore probably mutually exclusive in practice, but there is no reason why the former should always prevail over the latter: the choice between them should depend on the point of view of the researcher, on the nature of the document and on the intended use of the encoded material and not on the limitations of an given encoding schema. Different scholarly approaches are now paying growing attention to the physical object: to these scholars the contained text is, if not less important, certainly strictly dependant on the object in which it is preserved. These include the Critique Génétique, the New Philology, and the Textual (or Analytic) Bibliography, to mention just a few of them. Scholars who use these approaches should be offered a way to encode what they really want to encode in a standard way. At present the solution is either to invent a new encoding system from scratch or to heavily modify the TEI in order to fit the needs of the particular project, with each scholar and project doing so in a different way. In the following quotations Aurèle Crasson and Jean-Daniel Fekete discuss the need for the former, while Matt Cohen, director of the Interface Development for Static Multimedia Documents project at Duke University, for the latter.

> TEI fonctionne bien pour les manuscrits relativement propres où le texte est stable mais ne convient plus lorsque les phénomènes paratextuels prolifèrent, comme c'est le cas dans les manuscrits littéraires modernes ou dans des brouillons.
>
> (Crasson and Fekete 2004 168)

> We [...] decided to table <facsimile> and extend the existing Whitman Archive schema such that it could handle pages—surfaces—as intellectually significant structural units.
>
> (Cohen et al. 2007 2)

5.1. Encoding documents

Recently, the TEI has started to consider the possibility of encoding documents and not only texts, as we shall see shortly. For the TEI this represents a drastic new development away from the usual categories of textual analysis which have been consolidated over more than twenty years of experience with text encoding. As discussed above, TEI has been successful in the "development of a new data description language that substantially *improves* our ability to describe textual features" (Renear 2004 235); now the community needs the TEI to do the same for documents. For instance: what are documents made of? We know that, according to the TEI, texts are made fundamentally of structural divisions (chapters, sub-sections, poems, acts, scenes), and these do not contain text but further structural features (paragraphs, lists, tables, lines of poetry, speeches). But we do not yet know what documents are made of: quires? Pages? Double pages? Folios? Bifolios? Patches? And what do those units contain? Areas, regions or text?

The previous examples show how the encoding of documents should be able to address fundamental codicological questions on how the manuscript was presumably originally organised, to describe its present state and, possibly, to describe how we can go from one to another. In other words, the codicological encoding needs to be addressed not only from a descriptive point of view, but in time: it also needs to be genetic. The same examples show that the fundamental unit of transcription is not always or necessarily the page, but can be the opening, the bifolium or any surface that, according to the editor, represents the smallest meaningful subdivision of the physical object. In practice this can be almost anything. For example, the draft poem "America to Old-World Bards" was written by its author, Walt Whitman, on the back of old envelopes and letters, some of them glued together to form a bigger writing surface. In this case, everything is problematic, including the choice of what to transcribe and what not to: it is in fact worth noting that the editors have chosen to transcribe only those parts of the document that contain the poem, and not (for example) the front of the envelopes, the content of which is nevertheless used for dating the composition of the poem (Whitman 2005–2010).

How do we encode documents? Shall we just reverse the hierarchy text/document of the TEI encoding schema, regarding the document as "substantial" and semantic, textual markup as "accidental"?[18] According to Crasson et al., "un seul niveau de description ne suffit pas pour capter la structure d'un manuscrit" (2004 168), meaning that if you chose to encode either the text instead of the document or the document instead of the text you will loose some layers of meaning contained in the original object. But, as noted above, perhaps the attempt to encode the text within the document is too ambitious and can lead either to ungovernable markup or to unreconcilable overlapping hierarchies. It

[18] For a similar proposal see Pierazzo 2009 174–76.

seems that unless one level prevails, whether textual or documentary, the two levels cannot live together. This is because a scholar who is encoding page by page and line by line may wish to mark up textual features at both block level and at in-line level (examples of the former are paragraphs, stanzas, and speeches; the latter includes dates, names of people, and so on). However, these features can potentially overlap—and in practice they almost always will. From these early days in digital documentary transcription, it seems that a parallel encoding (texts and documents) is the way to go.[19]

5.2. The Proposal of the Genetic Edition Working Group

A proposal for adding a documentary view to the TEI has recently been accepted in principle by the TEI Council (April 2010): this means that the details of the elements and attributes may still be adjusted, but that the overall concept and theoretical basis has been agreed. This is part of a bigger proposal for the encoding of genetic editions which has been put together by a task force within the Manuscripts Special Interest Group.[20] The working group has recognised that it was impossible to deal with genetic criticism and modern manuscripts without first addressing the lack of support for encoding documentary features. The proposal has then been articulated in three main parts:

1. The documentary view (for which see below).
2. Transcription enhancement, which includes a set of new elements for encoding textual and para-textual features typical of working manuscripts. It includes, for instance, elements for re-writing or for functional annotation such as "move the paragraph here". It also includes a generic element able to encode any type of phenomena without implying any one particular interpretation. For instance, when an editor sees that a word has been struck through in a given document, that editor can say either that the word has been deleted (encoding it at the interpretational level) or that there is a line on top of it (encoding at the documentary level).
3. Genetic markup, which includes a group of elements for encoding evolution across time and across the different manuscripts that a work or a document has had, going from its first documented elaboration to the "finished" product which is usually the published book or the manuscript in its current state.

The documentary view allows one to transcribe texts from a documentary perspective alongside or as an alternative to the textual perspective. According to this proposal, a

---

[19] This is the choice discussed by Crasson and Fekete 2004 while presenting the *Transcripteur*, an open source editor designed to help in the encoding of modern draft manuscripts. This tool allows one to transcribe the text from two different points of view: documentary/diplomatic and textual; it also integrates images and gives the possibility of connecting the transcription to the facsimile.

[20] The Task force is chaired by Fotis Iannidis. Other members of the working group are Elena Pierazzo, Malte Rehbein and Lou Burnard, and fundamental contributions have also been made by Gregor Middell, Moritz Wissenbach and Paolo D'Iorio. See Pierazzo et al. 2010.

document can contain surfaces and surfaces can contain zones or patches (i.e. pieces of paper attached on top of the main surface). Zones can contain text, lines of text, or more zones. The terminology is deliberately generic, so "surface" could refer to a page, an opening, a face of a membrane or the side of a tapestry, according to the specific circumstance; the same applies to "zones" which could be marginal areas or any other polygonal area within a surface.

At present a dedicated way to encode codicological structure and its evolution is missing; it is nevertheless possible to use one of the new genetic structures or perhaps the generic TEI linking mechanism to group different surfaces together (for which see Consortium 2009 §16.1): the community of users has been called to use the new encoding and to create case studies and examples according to different scholarly needs. If the tools that have been offered prove inadequate then they will be encouraged to follow the example of the genetic editions working group and to propose an improvement to the TEI.

6. Conclusions

In the first volume of *Codicology and Palaeography in the Digital Age* and the associated conference, it was noted that much research on "digital" manuscript studies has been on palaeography, with relatively little attention paid to codicology (Stinson 2009 36). However, as the editors of the volume noted then, and as we hope this article has shown, codicology is crucial to the understanding of very many texts, particularly if page layout is included as part of this topic. The emphasis of most models for XML encoding, and especially the TEI, has been on modelling texts, and one may well argue that the TEI's role should be this and no more: it is, after all, the *Text* Encoding Initiative. Nevertheless, very many texts cannot be understood as "pure" text without regard to context and layout, as linear sequences of tokens (Caton 2009 80) or even as ordered hierarchical content objects (DeRose et al. 1990).

The constraints of print technology have often relegated codicology to the introductions of text editions, where editors traditionally provide a more or less superficial description of the structure of source documents. The advent of digital publication has allowed for much less constrained types of edition: texts can be presented in multiple views (diplomatic, reading, glossed, etc.), facsimiles of the source documents are much more affordable, and in general the structure of publications is much more flexible. In the same ways as genetic editions (Pierazzo 2009 171–72), these new possibilities could be used to apply codicological methods and analysis in a much more effective way to digital editions, making them the engine able to drive a scholarly interpretation which is firmly aware of the implications and consequences of text transmission.[21]

[21] Any co-authored paper normally involves close collaboration and this is no exception. Nevertheless, we shall attempt to delineate our respective contributions. Elena Pierazzo wrote sections 1, 2, 3.1, 3.4, 4, 5 and

Bibliography

Apollinaire. *Guillaume Apollinarie.* Paris: ONAC, 2007–2009. <http://www.guillaume-apollinaire.fr/>.

Austen, Jane. *Minor works now first collected and ed. from the manuscripts by R. W. Chapman, with illus. from contemporary sources.* Ed. Robert William Chapman. London: Oxford University Press, 1954.

Baxter, Stephen David. "Archbishop Wulfstan and the Administration of God's Property." *Wulfstan, Archbishop of York: the proceedings of the second Alcuin Conference.* Ed. Matt Townend. Turnhout: Brepols, 2004. 161–205.

Baxter, Stephen David. *The earls of Mercia: lordship and power in late Anglo-Saxon England.* London: Oxford University Press, 2007.

BM-Lyon. *Bibliothèque municipale de Lyon, Manuscripts Mérovingiens et Carolingiens.* Lyon: Ville de Lyon. <http://florus.bm-lyon.fr/>.

Bolton, Timothy. *The empire of Cnut the Great: conquest and the consolidation of power in Northern Europe in the early eleventh century.* Leiden: Brill, 2008.

Boyle, Leonard E. "The Nowell Codex and the Poem of Beowulf." *The Dating of Beowulf.* Ed. Colin Chase. Toronto: University of Toronto Press, 1981. 23–32.

Brooks, Nicholas. "The Fonthill Letter, Ealdorman Ordlaf and Anglo-Saxon Law in Practice." *Early medieval studies in memory of Patrick Wormald.* Eds. Stephen Baxter et al. Farnham, Aldershot: Ashgate, 2009. 301–18.

Brown, Michelle P. *The Lindisfarne Gospels: Society, Spirituality, and the Scribe.* London: British Library, 2003.

Caton, Paul. "Lost in Transcription: Types, Tokens, and Modality in Document Representation." *Digital Humanities 2009: Conference Abstracts.* Maryland (MD): Maryland Institute for Technology in the Humanities, 2009. 80–82.

CIRSFID-Irnerio. *CIRSFID Progetto Irnerio.* Bologna: University of Bologna, 2003–2007. <http://irnerio.cirsfid.unibo.it/>.

Clement, Richard W. "Codicological Considerations in the Beowulf Manuscript." *Essays in Medieval Studies,* 1 (1984): 13–27.

Clemoes, Peter. "History of the Manuscript: Origin and Contemporary Correction and Revision." In *Ælfric's First Series of Catholic Homilies: British Museum, Royal 7 C. XII, fols. 4–218.* Eds. Norman Ellsworth Eliason and Peter Clemoes. Copenhagen: Rosenkilde and Bagger, 1966. 28–35 and 24–5.

Cohen, Matt, Erika Fretwell and Kevin Webb. *White Paper: Interface Development for Static Multimedia Documents.* 2007. <http://asmodeus.ws/cohenlab/NEH_White_Paper.pdf>.

Crasson, Aurèle and Jean-Daniel Fekete. "Structuration des manuscrits: Du corpus à la région." *Proceedings of CIFED 2004.* La Rochelle (France), 2004: 162–168. <http://www.lri.fr/~fekete/ps/CrassonFeketeCifed04-final.pdf>.

Crasson, Aurèle and Jean-Daniel Fekete. *Transcripteur.* 2008 <http://tei-eclipse.gforge.inria.fr/transcripteur/>

---

the modern examples in sections 3.1.1, 3.2.1 and 3.3; Peter Stokes wrote sections 3.1.2 and 3.2.2 and the medieval examples in sections 3.1.1, 3.2.1 and 3.3.

Cross, James E. and Jennifer Morrish Tunberg. *The Copenhagen Wulfstan Collection: Copenhagen, Kongelige Bibliotek, Gl. Kgl. Sam. 1595.* Copenhagen: Rosenkilde and Bagger, 1993.

Da Rold, Orietta. "Homilies and Lives of Saints: Cambridge, University Library, Ii. 1. 33." *The Production and Use of English Manuscripts 1060 to 1220.* University of Leeds and University of Leicester, 2010. <http://www.le.ac.uk/english/em1060to1220/mss/CUL.Ii.1.33.htm>.

De Hamel, Christopher. *Glossed Books of the Bible and the Origins of the Paris Booktrade.* Woodbridge: D.S. Brewer, 1987.

DeRose, Stephen J., David G. Durand, Elli Mylonas and Allen H. Renear. "What is Text, Really." *Journal of Computing in Higher Education,* I (1990): 3–26. Online: <http://delivery.acm.org/10.1145/270000/264843/p1-derose.pdf>.

Dobbie, Elliot Van Kirk, ed. *Beowulf and Judith.* New York: Columbia University Press, 1953.

DS. *Digital Scriptorium.* Libraries Digital Program Division. New York: Columbia University Libraries. <http://scriptorium.columbia.edu/>.

Dumville,David N. and Peter Anthony Stokes, eds. *Liber Vitae Dunelmensis.* Trial Version, 2 parts. Cambridge: Dept. ASNC, 2001.

Eliason, Norman Ellsworth and Peter Clemoes, eds. *Ælfric's First Series of Catholic Homilies: British Museum, Royal 7 C. XII, fols. 4–218.* Copenhagen: Rosenkilde and Bagger, 1966.

Ferrer, Daniel. "The Open Space of the Draft Page: James Joyce and Modern Manuscripts." *The Iconic Page in Manuscript, Print and Digital Culture.* Ed. Georges Bornstein and Theresa Tinkle. Ann Arbor (MI): University of Michigan Press, 1998. 249–267. <http://www.item.ens.fr/index.php?id=23616>

Förster, Max. *Die Beowulf-Handschrift.* Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Philologisch-historische Klasse 71 (1919).

Foys, Martin K., ed. *The Bayeux Tapestry: Digital Edition.* Leicester: Scholarly Digital Editions, 2003.

Foys, Martin K. "Digital Mappaemundi: A Resource for the Study of Medieval Maps and Geographic Texts." *Library of Funded Projects.* National Endowment for the Humanities, Office of Digital Humanities, 2009. <http://www.neh.gov/ODH/Default.aspx?tabid=111&id=9>.

Gabler, Hans Walter. "The Primacy of the Document in Editing." *Ecdotica,* 4 (2007): 197–207.

Galbraith, Vivian Hunter. *The making of Domesday book.* Oxford: Clarendon Press, 1961.

Gerchow, Jan, ed. *Die Gedenküberlieferung der Angelsachsen mit einem Katalog der libri vitae und Necrologien.* Berlin: Walter de Gruyter, 1988.

Gerritsen, Johan. "The Copenhagen Wulfstan Manuscript: A Codicological Study." *English Studies,* 79 (1998): 501–511.

Gough-Map. *Linguistic Geographies: The Gough Map of Great Britain.* Belfast: Queen's University, 2010. <http://www.goughmap.org/>.

Greg, Walter Wilson. "The Rationale of Copy-Text." *Studies in Bibliography,* 3 (1950–51): 19–36.

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

Gretsch, Mechthild. "The language of the 'Fonthill Letter'". *Anglo-Saxon England,* 23 (1994): 57–102.

Gwynn, John, ed. *Liber Ardmachanus: The Book of Armagh.* Dublin: Hodges, Figgis & Co., 1913.

Gwynn, Edward, ed. *Book of Armagh: The Patrician Documents.* Dublin: Stationery Office, 1937.
Hilpert, Hans-Eberhard. "Geistliche Bildung und Laienbildung. Zur Überlieferung der Schulschrift Compendium Historiae in Genealogia Christi (Compendium veteris testamenti) des Petrus von Poitiers (1205) in England." *Journal of Medieval History,* 11 (1985): 315–31.
Hough, Carole. "Cattle-Tracking in the Fonthill Letter." *English Historical Review,* 115 (2000): 864–892.
HUL-PDS [Harvard University Library Page Delivery Service]. *Peter, of Poitiers, ca. 1130–1205. Compendium historiae in genealogia Christi: manuscript, [ca. 1200–ca. 1250]. MS Typ 216. Houghton Library, Harvard University, Cambridge, Mass.* Cambridge (MA): Harvard University Library. <http://pds.lib.harvard.edu/pds/view/12438364>.
Hunt, Richard W., ed. *Saint Dunstan's Classbook from Glastonbury: Codex Biblioth. Bodleianae Oxon. Auct. F. 4.32.* Amsterdam: North-Holland Publishing Co., 1961.
HyperStack: *Saint Patrick's Confessio Hypertext Stack Project.* Dublin: Royal Irish Academy, 2008–2010. <http://www.confessio.dmlcs.org/>.
ITEM: *Institut des Textes et Manuscrits Modèrnes.* Paris: CNRS (Centre National de Recherche Scientifique) / ENS (Ecole Normale Supérieure) – UMR 8132. <http://www.item.ens.fr/>.
James, Montague Rhodes. "The Earliest Inventory of Corpus Christi College." *Proceedings of the Cambridge Antiquaries Society,* 14 (1912): 89–104.
JTS: *Micrography. The Hebrew Word as Art.* Library of the Jewish Theological Seminary, [n.d.]. <http://www.jtsa.edu/prebuilt/exhib/microg/index.shtml>.
Ker, Neil R. *Catalogue of Manuscripts containing Anglo-Saxon.* Oxford: Clarendon Press, 1957.
Ker, Neil R. "Hemming's Cartulary: A Description of the Two Worcester Cartularies in Cotton Tiberius A.xiii." *Books, Collectors and Libraries: Studies in the Medieval Heritage.* Ed. Andrew G Watson. London: Hambledon, 1985. 31–59.
Keynes, Simon. "Royal Government and the Written Word in Late Anglo-Saxon England." *The Uses of Literacy in Early Mediaeval Europe.* Ed. Rosamond McKitterick. Cambridge: Cambridge University Press, 1990. 226–57.
Keynes, Simon. "The Fonthill Letter." *Words, Texts and Manuscripts: Studies in Anglo-Saxon Culture Presented to Helmut Gneuss on the Occasion of his Sixty-Fifth Birthday.* Ed. Michael Korhammer. Woodbridge: D.S. Brewer, 1992. 53–97.
Keynes, Simon, ed. *The Liber Vitae of The New Minster and Hyde Abbey Winchester: British Library Stowe 944: together with leaves from British Library Cotton Vespasian A. VIII and British Library Cotton Titus D. XXVII.* Copenhagen: Rosenkilde and Bagger, 1996.
Keynes, Simon. *Images of Anglo-Saxon Manuscripts 5: The Liber Vitae of the New Minster, Winchester.* Cambridge: Trinity College. <http://www.trin.cam.ac.uk/sdk13/LibVitNM.html>.
Kiernan, Kevin S. "The Eleventh Century Origin of Beowulf and the Beowulf Manuscript." *The Dating of Beowulf.* Ed. Colin Chase. Toronto: Toronto University Press, 1981. 9–21.
Kiernan, Kevin S. "The Legacy of Wiglaf: Saving a Wounded Beowulf." *Beowulf: Basic Readings.* Ed. Peter S. Baker. New York: Garland, 1995. 195–218.
Kiernan, Kevin S. *Beowulf and the Beowulf Manuscript.* Ann Arbor (MI): University of Michigan Press, 1996.
Kline, Naomi Reed. *A wheel of memory: the Hereford Mappamundi.* Ann Arbor (MI): University of Michigan Press, 2001.

Loyn, Henry. *A Wulfstan Manuscript, containing Institutes, Laws and Homilies: British Museum Cotton Nero A.I.* Copenhagen: Rosenkilde and Bagger, 1971.

Lucas, Peter J. "The Place of Judith in the Beowulf-Manuscript." *Review of English Studies,* 41 (1990): 463–78.

LUNA. Oxford: Bodleian Library, 2008-2010. <http://bodley30.bodley.ox.ac.uk:8180/luna/servlet>.

Malone, Kemp. *The Nowell Codex: British Museum Cotton Vitellius A. XV, Second MS.* Copenhagen: Rosenkilde and Bagger, 1963.

Marchetti, Arnaldo and Vittorio Giuliani, eds. *Puccini com'era.* Milano: Curci, 1973.

Meritt, Herbert Dean, ed. *Old English Glosses.* London: Oxford University Press, 1945.

Meritt, Herbert Dean. *Some of the Hardest Glosses in Old English.* Stanford (CA): Stanford University Press, 1968.

Migne, Jacques Paul, ed. B. *Rabani Mauri. Fuldensis abbatis et moguntini archiepiscopi opera omnia*, Vol. 1. Patrologia Latina 107 (Paris, 1864).

Munroe, William H. "A Roll-Manuscript of Peter of Poitiers' Compendium." *The Bulletin of the Cleveland Museum of Art,* 65 (1978): 92–107.

ODL: *Early Manuscripts at Oxford University.* Oxford: Oxford University, 2000–2001. <http://image.ox.ac.uk/>.

Page, Raymond I. "On the Feasibility of a Corpus of Anglo-Saxon Glosses: The View from the Library." *Anglo-Saxon Glossography: Papers read at the International Conference held in the Koninklijke Academie voor Wetenschappen Letteren en Schone Kunsten van België, Brussels, 8 and 9 September 1986.* Ed. R. Derolez. Brussels: Paleis der Academiën, 1992. 77–96.

Pierazzo, Elena. "Digital Genetic Editions: The Encoding of Time in Manuscript Transcription." *Text Editing, Print and the Digital World.* Ed. Marilyn Deegan and Kathryn Sutherland. Aldershot: Ashgate, 2009. 169–186.

Pierazzo, Elena and Malte Rehbein. *Documents and Genetic Criticism TEI Style.* TEI Consortium, 2010. <http://www.tei-c.org/SIG/Manuscripts/genetic.html>.

Prescott, Andrew. "'Their Present Miserable State of Cremation': The Restoration of the Cotton Library." *Sir Robert Cotton as Collector: Essays on an Early Stuart Courtier and his Legacy.* Ed. Christopher John Wright. London: British Library Publications, 1997. 391–454.

Pulliam, Heather. *Word and Image in the Book of Kells.* Dublin: Four Courts Press, 2006.

Renear, Allen, Elli Mylonas and David Durand. "Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies." *Research in Humanities Computing.* Eds. Nancy Ide and Susan Hockey. Oxford: Oxford University Press, 1996. <http://www.stg.brown.edu/resources/stg/monographs/ohco.html>.

Renear, Allen H. "Text Encoding." *A Companion to Digital Humanities.* Eds. Susan Schreibman, Ray Siemens and John Unsworth. Oxford: Blackwwell, 2004. 218–239. <http://www.digitalhumanities.org/companion/>.

Robinson, Pamela R. "The 'Booklet': A Self-Contained Unit in Composite Manuscripts." *Codicologica.* Eds. Albert Gruys and J. Peter Gumbert. 3 vols. Leiden: Brill, 1980. 46–69.

Rollason, David and Lynda Rollason, eds. *The Durham Liber Vitae ... The Complete Edition.* 3 vols. London: The British Library, 2007.

Rumble, Alexander R. "The palaeography of the Domesday manuscripts." *Domesday book: a reassessment.* Ed. Peter H. Sawyer. London: E. Arnold, 1985. 28–49.

Sardou, Victorien, Giuseppe Giacosa, Luigi Illica and Giacomo Puccini. *Tosca. Facsimile della copia di lavoro del libretto.* Ed. Gabriella Biagi Ravenni. 2 vols. Firenze: Centro Studi Giacomo Puccini – Leo Olschki Editore, 2009.

Sauer, Hans. "The Transmission and Structure of Archbishop Wulfstan's 'Commonplace Book'". *Old English Prose: Basic Readings.* Ed. Paul E. Szarmach. New York: Garland, 2000. 339–93.

Sawyer, Peter H., ed. *Domesday book: a reassessment.* London: E. Arnold, 1985.

Schillingsburg, Peter L. *From Gutenberg to Google.* Cambridge: Cambridge University Press, 2006.

Sisam, Kenneth. *Studies in the History of Old English Literature.* Oxford: Clarendon Press, 1953.

Sperberg-McQueen, C. Michael "Text in the Electronic Age: Textual Study and Text Encoding with Examples From Medieval Text." *Literary and Linguistic Computing,* 6 (1991): 32–46.

Stinson, Timothy. "Codicological Descriptions in the Digital Age." *Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age.* Eds. Malte Rehbein, Patrick Sahle and Torsten Schaßan. Norderstedt: Books on Demand, 2009. 35–51.

Stokes, Peter Anthony. "The Vision of Leofric: Manuscript, Text and Content." *Review of English Studies,* (forthcoming).

Suntrup, Rudolf. "'Te igitur'-Initialen und Kanonbilder in mittelalterlichen Sakramentarhandschriften." *Text und Bild: Aspekte des Zusammenwirkens zweier Künste in Mittelalter und früher Neuzeit.* Eds. Christel Meier and Uwe Ruberg. Wiesbaden: Ludwig Reichert, 1980. 278–382.

Sutherland, Kathryn. *Jane Austen's textual lives: from Aeschylus to Bollywood.* Oxford: Oxford University Press, 2005.

Tardieu, Jean. *Poème à voir.* Paris: Gallimard, 1990.

TEI P5: *Guidelines for Electronic Text Encoding and Interchange.* 1.7.0, 2010-07-07, TEI Consortium, 2009. <http://www.tei-c.org/Guidelines/P5/>

Tinti, Francesca. "From Episcopal Conception to Monastic Compilation: Hemming's Cartulary in Context." *Early Medieval Europe,* 11 (2002): 233–61.

Tinti, Francesca. "Si litterali memorie commendaretur: Memory and Cartularies in Eleventh-Century Worcester." *Early medieval studies in memory of Patrick Wormald.* Ed. Stephen Baxter et al. Farnham: Ashgate, 2009. 475–91.

Treharne, Elaine M. "The Bishop's Book: Leofric's Homiliary and Eleventh-Century Exeter." *Early medieval studies in memory of Patrick Wormald.* Ed. Stephen Baxter et al. Farnham: Ashgate, 2009. 521–37.

Whitman, Walt. "America to Old-World Bards (Poetry Manuscript)." *Walt Whitman Archive.* 1995–2010. Eds. Ed Folsom and Kenneth M. Price. Lincoln (NE): Center for Digital Research in the Humanities at the University of Nebraska–Lincoln, 1995–2010. <http://www.whitmanarchive.org/>.