Practical and Theoretical Implications of Digitizing the Middle Ages

Roberta Capelli
University of Trento

Follow this and additional works at: https://docs.lib.purdue.edu/clcweb

Recommended Citation
Capelli, Roberta. "Practical and Theoretical Implications of Digitizing the Middle Ages." CLCWeb: Comparative Literature and Culture 15.3 (2013): https://doi.org/10.7771/1481-4374.2248

This text has been double-blind peer reviewed by 2+1 experts in the field.
The above text, published by Purdue University Press ©Purdue University, has been downloaded 789 times as of 11/07/19.

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

This is an Open Access journal. This means that it uses a funding model that does not charge readers or their institutions for access. Readers may freely read, download, copy, distribute, print, search, or link to the full texts of articles. This journal is covered under the CC BY-NC-ND license.
Abstract: In her article "Practical and Theoretical Implications of Digitizing the Middle Ages" Roberta Capelli discusses scholarship about and the teaching of medieval culture in digital humanities. While every medieval manuscript is an individual entity, displaying a series of unique and unrepeatable material, structural and aesthetic characteristics, digital devices are able to generate only two-dimensional photographic reproductions. However, the digital medium brings about some major improvements in the study — and teaching — of medieval manuscripts because the hypertextual nature of its applications allows us to analyze synchronic and the diachronic dynamics simultaneously. From a theoretical point of view, the difference between the state of the manuscript taken as a material object and its status as a vector of intellectual values brings to light similarities between the notions of intellectual property and originality in traditional print and digital cultures. The increasing number of critical editions of medieval texts and manuscripts in digital form asks us to reshape the theoretical and conceptual-linguistic frameworks of textual criticism as a discipline. Capelli postulates that "hypertextual criticism" represents the study of philological theories and practices in light of new literacies and technologies.
Robert Capelli, "Practical and Theoretical Implications of Digitizing the Middle Ages" page 2 of 8
CLCWeb: Comparative Literature and Culture 15.3 (2013): <http://docs.lib.purdue.edu/clcweb/vol15/iss3/8>
Thematic Issue Literacy and Society, Culture, Media, and Education. Ed. Kris Rutten and Geert Vandermeersch

Roberta CAPELLI

Practical and Theoretical Implications of Digitizing the Middle Ages

In the Middle Ages, originality was a concept that can be made meaningful today by reference to the concept of auctoritas. That is, the concept of auctoritas asks us to reconsider the function of the author as a model of opposition between the general, trascendental anonymity of literary works and the overt authoriality displayed by scientific works. Artistic creativity is not defined as one’s capacity to individualize, culminating in genius, but as one’s ability to rely upon tradition and the reusing of textual heritage and commentarial history. In consequence, auctoritas means a wide spectrum of interrelated levels of meaning associated with authority, authorial influence, and canon, and has to do with the dialectical relationship between the author (actor), the writer (scriptor or compilator), and the written text (scriptum/scriptura). Subverting our notion of authorship and ownership, medieval literary traditions want us to get used to the idea that the original text has hardly ever survived to our days; instead, at its place we have more or less close copies and reconstructed archetypes, which give us plausible versions of the original text. The fact that neither manuscripts nor philological works will ever replace the original emphasizes the central ontological difference between original(s) and archetype(s), the first conveying the truth of the author(s) and representing the absolute referentiality of tradition, the latter conveying a truth of copyist(s) and representing the contingent referentiality of transmission. This theory of the Ur-text — presupposing a text once existed and later split into different versions — merges with the theory of the Ur-source presupposing an inventory of concurrent oral traditions and non-written materials (anthropological motifs, folk-religious customs, cultural themes, etc.).

In the article at hand I discuss three different viewpoints: 1) theoretical because the ontology of medieval studies updates the paradigms of its own disciplines around cultural and technological changes, 2) illustrative in order to examine the actual application of new technologies in the field of medieval studies, and 3) educational to find out what are and what might be the effects of digitization on the training of students and scholars in consideration of the fact that many university libraries are evolving rapidly toward the concept of a learning center (Bruckmann 13) and digital and virtual libraries are close to becoming learning platforms, where users can participate actively in the process of reading and editing books (Lucía Megías 399-40).

It is important to recall a basic principle of the philological practice that the critically edited text is not (ever) an original text. For the Middle Ages, however, it would be more proper to speak in the past: at the beginning it was the text, because — besides the more fortunate and rare cases of autography — we do not have the author's text today, but simply a copyist's text. The deconstruction of the message (and of the identity) of the author — so dear to Derrida, Barthes, and Foucault (see Landow 2) — is in progress in the very tradition of the texts we study, all of which are more or less corrupt and all of them different from an original which no longer exists. Reconstructing a version which resembles — hypothetically — the original and recomposing the information we obtain from the copies we have left (which, in turn, constitute only a small number of those produced and of the ones in existence) is the task of traditional philology which produces critical editions. Currently, the reference model for the electronic critical edition of medieval texts in romance literatures is also the first to have been made not as an experimental prototype, but as a finished product available on the market. It is Prue Shaw's edition of Dante's Monarchia on CD-ROM, coproduced by Società Dantesca Italiana and Peter Robinson's digital software work at the Institute for Textual Scholarship and Electronic Editing, University of Birmingham, responsible for the software used for supporting the edition (see Robinson, "Current" <http://www.digitalmedievalist.org/journal/1.1/robinson/>; see also Roncaglia).

I summarize the crucial points made in detail by Paolo Chiesa in 2007: in comparison to the edition on paper, the electronic edition has a photographic reproduction of all the manuscript witnesses we know of the Monarchia, their diplomatic transcription, as well as an archive with all the variants. The text can be searched using multiple search engines. However, informatics tools are not capable of automatically generating a stemma codicum realizing a hierarchized representation of the
family relations among the witnesses of tradition. Instead, there is a phylogram, a graph generated by the system thanks to a program for cladistic analysis adapted to the methods of textual criticism by Robinson ("Electronic"). The computerized phylogram singles out two large sets of manuscripts according to the relation of distance/closeness of the witnesses (= what-is-β versus what-is-not-β), but it does not match the tri-partition obtained by stemmatic reconstruction (K, α [T and A¹], β); rather, it proposes paradoxically a bipartite configuration (= β/non-β) which is the contradiction inherent in the traditional, nineteenth-century method of Karl Lachmann (see Chiesa 346-50).

Since the 1970s the discussion on the computerized classification of variants according to genealogic or statistic principles has not yet found such solutions which would make it possible to produce autonomous systems capable of replacing the editor's work completely. Recently, a team of scholars affiliated to the University of Roma "La Sapienza" have tested an "integrated method for stemma reconstruction which combines the traditional ecdotics approach with an information theory oriented methodology" (Canettieri, Loreto, Rovetta, Santini, <http://w3.uniroma1.it/cogfil/ecdotica.html>) and the software developed for use in biology has been adapted to represent phylogenetic relations among manuscripts for the Canterbury Tales Project, a project directed by Robinson aiming to investigate the textual tradition of Chaucer's masterpiece. The reasons why this topic remains debatable and debated are more logical than technological: I draw attention to the fact that all the models designed and tested until now — whether stemmatic or non-stemmatic — have given taxonomic results, that is, they prove that the documentary incompleteness of tradition, contamination, interpolation, and polygenetic phenomena are all factors which, even when resorting to mathematical calculations, do not yield exact solutions, but only approximate results. It will not be by chance, therefore, that the computerization of medieval texts has resulted in archival editions rather than critical editions. If anything, the question to ask should be as to in what cases could electronic scholarly editions be superior to anything than a print edition could ever be? The answer could come from the ideal set of characteristics any electronic edition should possess provided by Peter Shillingsburg, but still far from being fulfilled with regard to usability, transportability, archive specifications, security and order, integrity, expandability, printability, and user friendliness. Hence,

If we conceive the Middle Ages as hypersystems and medieval literature as hypertexts, we understand the potential of new media technology. For example, through electronic editing of historical documents and literary works we could set up virtually limitless e-libraries and collections, with high-definition digital interfaces and high standards of legibility; we could obtain maximum amount of storage with minimum loss of (meta)information; we would optimize the synchronous modes of e-consultation and real-time (meta)data sharing. The recreation of the past has always appeared as a multifaceted phenomenon and today medieval revivals are heavily anchored in mass and new media culture. It is not by chance that while architecture played a crucial role in disseminating the neo-gothic paradigm, the principal architectural (i.e., aesthetic and ideological) contributions to the (re)construction of today neo-medieval scenarios have to be sought in virtual reality (high-tech and video games). Following Karl Fugelsd I suggest the notion of "new mediaevalisms" as a syncretic concept to study syntactically and visually the plural manifestations of medieval literature and culture in (inter)mediality and digital humanities. Gathering many genres and media, the study of medieval literature and culture as a broad phenomenon is capable of bridging methodological gaps and overcoming disciplinary barriers. (Capelli 171-72; on [inter]mediality and digital humanities, see, e.g., Tótós de Zepetnek)

Another fundamental matter we should be concerned with is a relatively recent branch of the philological studies, génétique, which focuses on the working material of modern and contemporary authors (block notes, drafts, working copies, etc.) and that is different from what is conventional genealogy. Traditional philology and modern genetic criticism use the same tools, but with different objectives, since genetic criticism starts from an edited text — usually approved by the author — to then break down the creative process into its preparatory phases and the intermediate stages (see Grésillon), while medieval philology follows the reverse path: it does not prepare a genetic dossier, that is, made up of materials which precede the definitive text, but prepares a genealogic dossier composed of materials which derive from the definitive text and that are interpretations of that text as "secondary moments of textual production and reproduction" (McGann, Textual Criticism 192). By now it is an accepted fact and a shared concept that transcription is an act of interpretation, which obviously and heavily affects digital codification of ancient texts. This infra-disciplinary distinction is important because some digital projects carried out in the field of genetic criticism are also applicable to medieval studies, although as said pursuing different aims.
Medievalists reflect constantly upon the nature of the textual variants, considering them sometimes errors, at times innovations, and other times simply a waste of a given source. The epistemological importance of the variant involves not only literary analysis (= content of the variant), but also the psychology of composition (= cause of the variant) and the pedagogy of learning (= effects of the variant) which in toto represent the scope of privileged in-depth analysis. In this respect, an interesting project is *Digital Variants*, a digital archive project of contemporary authors created in 1996 by Domenico Fiormonte and supported by the Department of Italian Literature at the University of Edinburgh. The study of the variant, regarded as a cognitive artefact, provides material and textual information about the historical evolution of the written language and the role it has in cultural and cognitive development: "Modern textual criticism ... could surely belong to what is generally referred to as the cognitive sciences and the psychology of composition would be its natural accomplishment (and twin discipline). ... They share similar diachronic perspectives — psychology relying on experimental method, philology on historical-critical method — travelling in opposite directions along the same path" (Fiormonte <http://tecalibri.altervista.org/F/FIORMONTE-D_filologia.htm>; unless indicated otherwise, all translations are mine).

Since in medieval literature variant readings are almost without exception copyist choices and not authorial and given that we can only formulate conjectures about the degree of alteration, these variants are — in comparison — similar to the original. While they can amend the original text, worsen it, or they can convey the same message through different but equivalent words, the fact of considering textual variants as cognitive artefacts and units of meaning significant by themselves — and not only significant in the context of textual tradition from a strictly critically oriented perspective — can provide useful information about the culture of copyists and readers, about their educational background, and about literary taste at a given time and place: "After all, digital surrogates featuring high-quality color images of a writer's manuscripts offer a more ample sense of their textual conditions, including the conditions of the writing scene in which they were produced" (Nell Smith, 306). Together with the author's and the copyist's variant, I would like to mention the "technological variant": using an OCR software the attempts made so far to scan, acquire, and restore the manuscript have produced paleographic errors made by the machine which are altogether similar to those made by scribes such as erroneous grouping of adjacent letters, distortion of letters with bars, wrong recognition of spaces between letters and words, etc. (see Tomasi and Tomasi 198-200) The more exhaustive the digitized corpus of manuscripts and texts is, the more powerful our search engine is and the more specifically structured our software is in this sense, the greater — and therefore the more representative — the breadth of the sampling will prove. I leave aside here the question of what the most suitable encoding procedures should be because this aspect of digital textual editing has its own issues and bibliography, and it would bring into argument the work of the *Text Encoding Initiative* (TEI), an international community of practice in the field of digital humanities operating since the 1980s when "It became immediately apparent that the development of a text encoding scheme demanded much more than assigning tag names to features, and included looking at the conceptual structure of texts and determining the commonalities across different text types. ... Therefore, the work of participants in the TEI not only involved consideration of problems of text encoding that are likely to be with us for decades to come, but also required the development of a methodology, from scratch, for approaching these problems" (Ide and Véronis 4; see also Mordenti 55-56).

Many new strands of research could be improved by statistical inquiries covering a wide spectrum of manuscript witnesses, converted into digital formats (text, image, sound modes), and integrating heterogeneous data "for cumulative and shared scholarly work on the primary text sources in a distributed digital environment" (Deckers, Koch, Vertan 93). This line of research — which owes its methodological principles to formal language and information theories — could be applied for instance to the distribution of specific palaeographic and decorative features and textual variants in codices (or texts) compiled (or copied) in a specific epoch and geographical area. In a digital environment it has already been tested that syntactic processing could improve retrieval and that "surface structure parsers, which identify the correct structure of each sentence, would seem potentially useful for labelling the content of documents" although "they don't tell you what you need to know, do not take account of context, they may be too restrictive" (Lesk 53). The context is precisely what a
distributional philological approach could clarify. "Distributional" and "compositional semantics is a branch of linguistic philosophy which explores the idea that human conceptual knowledge is, to a considerable extent, the result of the extraction of simple distributional information from large amounts of linguistic input. Through flexible distributional models we can capture and distinguish different kinds of semantic relations, and these models are based on the assumption that the meaning of a word can be inferred from its usage, that is, its distribution in text. Today, almost all databases collecting medieval texts incorporate only their latest critically edited versions, i.e., one reconstructed version among many other versions rejected by the editor (see Leonardi 71). However, "as many computer-based studies have shown, laying open one's criteria for encoding certain textual features is of prime importance to any procedure that aspires to produce quantifiable results" (Rommel 92). By creating databases including the transcription of the manuscripts we want to digitize, we could question the known manuscript tradition about, for example, changes of meaning of denotative combinations of words (such as, for example, white hands, lance in hand, hand in hand, etc.) within genre-based, geographically-based, chronologically-based, etc. framework models.

As we have seen, it is what Prue Shaw did in her digital critical edition of Dante's Monarchia, but it is also the philosophy of The Princeton Charrette Project, a multi-media electronic archive launched by Karl Uitti in 1994 containing the manuscript tradition of Chretien de Troyes's Le Chevalier de la Charrette, and of the data base of medieval Galician Portuguese lyrical texts (Lopes Videira, Forreira, Júdice <http://cantigas.fcsh.unl.pt>) containing historical, biographical, and literary information and the manuscript tradition of all the existing texts: "A digital organization thus makes possible a significant departure from a paper-based apparatus" (McGann, "From Text" 43). This approach could improve our understanding of textual variants considered as products of human mental processes and vectors of knowledge: a composite search, borrowing additional information from the photographic reproductions of codices and from catalogues of illuminations, music, and writing styles could outline the circulation and reception of medieval texts on the basis of particular data and metadata shared by sets of homogeneous manuscript witnesses thus revealing specific cultural trends in specific social milieux.

The medieval text the reader reads and that the scholar studies is the text edited critically, that is, the result of the restoration of the manuscript tradition. It is an artificial text (the result of a philologist's work), potentially modifiable (thanks to new information and to better philologists) and unitary (it is a printed book). Originally, however, that text is in a manuscript, or rather, in a number of manuscripts, which we call "witnesses" because each contains an individual truth, has its own linguistic, morphologic, and typological features, and has a specific cultural value. The physical entity of the medieval manuscript is not — unless virtually — separable from the text it contains and vice versa, the medieval text is a text designed for a given manuscript (which in the Middle Ages is rarely "monographic" but is, for the most part, a miscellaneous book), copied from a certain copyist, and required by a certain person commissioning the work. At the International Congress La Pratique des ordinateurs dans la critique des textes, held in Paris in 1978, Cesare Segre spoke about medieval texts as "diasystems":

The image of a given text is a linguistic structure which exemplifies a system. Each copyist, in turn, has his own linguistic system which, in the course of the transcription, comes into contact with the text he is transcribing ... Every transcription produces inevitably a "Creolization" of the text ... The notion of diasystem helps us to discover a series of no less significant centripetal forces: those which hold together the diasystems existing in every manuscript. At the core of this tension is no longer the original text, but every time and every time differently, the text resulting from the compromise between systems ... Consequently, we can regard the text as the place of the real and the apparatus as the place of the virtual. (46-48).

I would like to go beyond what Segre wrote: if manuscript witnesses represent the life of the text in time and space, they are not simply linguistic diasystems, but cultural diasystems: they are the expression of precise geo-linguistic, historical and social realities within which the text exists. They are a body of — knowable — information within a galaxy of unknown and unknowable information. As suggested by Thomas G. Tanselle, "When we talk about literature ... we are inevitably referring to critically reconstructed texts" (592). Therefore, I would reverse Segre's conclusion by positing that the apparatus is the place of the real, it collects what we do have while the critical text is the place of the
virtual and it is the text we would like to have, but which we do not possess. The critical edition — with its phonetic, graphic and printing, etc. normalizations — is real only inasmuch as it constitutes a working tool.

On the one hand, the change in perspective I postulate provides a good explanation of today's prevailing tendency in the field of medieval literature to the acquisition in digital format of medieval sources and to the constitution of digital collections. On the other hand, digital formats bring to the fore the natural hyper-textual dimension of the medieval text taken as a cultural diasystem, an entity that is able to give us tangible and intangible multi-ordered, multilevel, and multi-layered information (on philological hypertexts see, e.g., Bérard and Mordenti). A single page from a manuscript may provide a series of multi-ordered information (text and paratext), multilevel information (palimpsestic writing) and multi-layered information (variants and copying activity from multiple sources). The tangible and intangible pluridimensionality of the manuscript or of the subunit folio requires synchronization of all the data and metadata which they are capable of providing and which is impossible to realize in the static and linear form of traditional printing, but which is congenial to the non-linear form of the hypertext. In a multitasking digital environment (Adamo 5), hypertexts and "other forms of electronic writing refashion or remediate the forms and genres of prints" (Boiter xi). In this way, we have true manuscripts/folio space models, whose multimodal nature (= one space in which words, images and music notation exist and intermingle with each other) manifests itself by a network of interactive links generating simultaneous visualizations in the form of multi-window interfaces. To have an idea of these complex interrogation interfaces which allow scholars to collect and combine information (from image archives, bibliographical databases, critical, diplomatic, or genetic online editions, etc.), the site designed to host the Online Editions of the École des Chartes can be a useful scholarly and educational tool because the collection of archival documents are accompanied with digital dynamic dossiers of commented and translated facsimiles (see Poupeau 30). The virtual format makes it possible to transform manuscripts into hypermedia learning environments (Nielsen 3) and makes it possible to have combinations, explanations, animations which the book product (manuscript and printed book) indicates, but cannot do.

It does not seem necessary to lay emphasis on the active role the user has in the digital format of medieval texts (see, e.g., Ciotti 213; Mordenti 143). If anything, it is important to distinguish between user-expert and user-apprentice, identifying the former with the researcher's profile, and the latter with the student's profile. This distinction is important to dispel any possible suspicion of the chaos and casualness of non-linear reading and consultation, which may represent a problem for the student who still does not have a good command of the subject (Willet 245-46). I myself cannot see how they may be a problem for the scholar accustomed to work simultaneously — literally and metaphorically — on different work desks when in fact the philologically oriented hypertext is a multi-media scriptorium (see also Cadioli 144). Besides, hypertextual didactics is and must remain a mediated kind of didactics, that is, guided. If university initiatives aimed at creating academic courses based on the use of hypertextual materials have failed because of the extremely lengthy and costly procedure of making the hypertext, the practicability of this choice by individual scholars shows that perhaps the core of the problem lies in the inadequacy of the digital skills of many humanists and in the subsequent need to organize teams with positive cooperation and human networking despite the fact that this would be more costly in terms of management and operation.

It seems to me intuitive that — as long as the moment of the conception and that of the project are entrusted to at least two different people — there would necessarily be a divide between the medievalist and the computer technician and that the comprehension and effective encoding of the manuscript would be lacking. For this reason, I think it is important to launch auxiliary courses in the humanities or even create new university specializations in digital humanities, something that is occurring in some countries, mostly in the U.S., the United Kingdom, or Germany. There is no doubt that today we should "address the field of digital humanities from disciplinary perspectives" because "computing has cut across disciplines to provide not only tools, but methodological focal points" (Schreibman, Siemens, Unsworth xxii). At the University of Trento, for example, there is a degree course in Informatics Philosophy where the traditional teaching of history of philosophy is accompanied by that of scientific subjects with the aim of introducing into the world of research and
into the job market an "organic thinker" capable of using informatics in a manner that is both direct (i.e., using of advanced technological equipment and processes) and speculative (i.e., analyzing the influence of new technologies on society and on the contemporary mindset).

In conclusion, it seems to me that the dynamics of digital presence in medieval studies develops along three major axes: storage, research, and creativity. The most substantial achievements have to do with storage activities which put information into memory, convert non-digital materials into a digital format (e.g., digital libraries, archives, data bases): "Storing information in archives or museums creates a particular form of power geometry, where access to information, and the way it is organised creates implicit hierarchies around who has a right to use information" (Miller 21). From storage to research, special collections of medieval manuscripts and early printed books have been digitized to improve query performance and text/data retrieval through advanced recovery systems (e.g., dictionaries, thesauri, concordances). The "creative" possibilities of digital applications to textual criticism and investigations related to medieval studies (e.g., digital critical editions, hypertext applications, and hypertextual environments) have not proved successful so far in terms of scholarly reliability, economic accessibility, and digital longevity. The dissemination of products and information on the internet does not give a guarantee as to their cultural contents: "We are currently in a digital dark age with respect to lack of content. Without a critical mass of information, technological capacity is a hollow structure" (Ronchi 69). Perhaps the greatest problem involved in making and publishing electronic editions of medieval texts is "the expense and difficulty of negotiating capture and permissions rights for high-quality digital images of original materials" (Robinson 11). Software and hardware technologies become rapidly obsolescent and storage media are subject to degradation (Maggioni 16). Digital preservation requires refreshing interventions, printing, or microfilming strategies, multiple copies (Chiesa 333-34). Paradoxical as it may appear, while the digitization of manuscripts aims at preserving them from physical deterioration, the conservation of born-digital materials applies to pre-digital solutions which reassert implicitly the primacy of the printed book's materiality.

Works Cited


Author's profile: Roberta Capelli teaches Romance languages and literatures at the University of Trento. Capelli's interests in research include medieval literature and the presence of the Middle Ages in modern and contemporary cultures. In addition to numerous articles, Capelli's book publications include *Sull'Eccorialense (e.III.23)*. *Problemi e proposte di edizione* (2006), *Guittone d'Arezzo. Del carnale amore* (2007), and the collected volume *Profilo delle letterature romanze medievali* (with Furio Brugnolo, 2011). E-mail: <roberta.capelli@unitn.it>