



Universiteit
Leiden

The Netherlands

Scholarship in interaction: case studies at the intersection of codework and textual scholarship

Zundert, J.J. van

Citation

Zundert, J. J. van. (2022, September 27). *Scholarship in interaction: case studies at the intersection of codework and textual scholarship*. Retrieved from <https://hdl.handle.net/1887/3464403>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3464403>

Note: To cite this publication please use the final published version (if applicable).

Chapter 4

Barely Beyond the Book?¹

“There is nothing deterministic about the Internet.” – David Lowery, frontman of Camper Van Beethoven and lecturer at Terry College of Business, University of Georgia.

(Timberg 2013)

This is a story about the methodological interaction between two scientific fields, that of textual scholarship and that of computer science. The names of the fields, however, only imprecisely delineate the permeable boundaries between research domains where methodologies interact – for obviously the world is much more fluid than such nouns suggest (Simon 1995:234). The interactions of interests are much more complex than the simplified image of a dynamic whereby one field donates a methodology to another. Rather than trying to reflect on the current state and the future potential of the digital scholarly edition from well inside the field of textual scholarship, let us approach the topic from the perspective of the multidisciplinary methodological interaction that has arisen to support the theoretical and practical development of the digital scholarly edition over the recent years. Textual scholarship in its digital fashion belongs to the broader field of digital humanities, itself a field built on interdisciplinarity, where many skills and theories of the

¹A previous version of this chapter was published as Van Zundert, Joris J. 2016. “Barely Beyond the Book?” In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 83–106. Cambridge (UK): Open Book Publishers. <http://www.openbookpublishers.com/reader/483>. The observations on methodological changes regarding the classic “index” in the case of the eLaborate project were expanded for clarity.

realms of computer technology and those of scholarship intersect, and thus where many new interfaces and interactions arise between those skills and the fields they are tied to (Hockey 2004). This is where digital humanities acquires its innovative power, or at least the promise of that power.

That innovative power, however, can be both exciting and confusing. The point where disciplines intersect is not a space for the calm, cool and collected exchange of technical and methodological knowledge. Rather, it is a place where the inherent social aspects of science and research are brought markedly into the foreground (Latour 1987). Take for example Jan Christoph Meister's description of the "lamented conflict between 'computationalists' and 'humanists'". This conflict, Meister (2013) states, "arises as soon as we become afraid of our own courage and shy away from jumping across these two fault lines. Let's cut through that fear. The task remains [...] to 'become capable of both – the metaphor and the formula, the verse and the calculus [...].' That's a borderline experience, no doubt, and those who prefer to pitch their tent in the comfortable centre of either laager don't run the risk of questioning their own philosophical, epistemological and ethical identity as easily." Meister's word use is notably emotive ("afraid", "fear", "courage") and at the same time vividly touches on the impact of the social dimension ("conflict", "borderline experience", "risk") of the epistemological interaction that is expressed. As Christine Borgman has suggested this is a situation where it can be useful, with respect to the design of scholarly infrastructure, to take these interactions and the behavior connected to them as the objects of study (Borgman 2009). Let us do exactly that here. Taking the digital scholarly edition as a part of the scholarly infrastructure for textual scholarship, we can try to infer what the historical interactions between textual scholarship and computer science tell us about the current state and development of the digital scholarly edition.

The field of science and technology studies (STS) offers a useful frame for critical study and reflection on what occurs at the interfaces of the various research fields within digital scholarship. When these fields intersect, it is not simply a question of objective interactions concerning technology and methodology; rather, these interfaces are also the sites of social processes that guide and steer the methodological interaction. Within STS such processes

are often referred to as the social shaping of technology – that is, the mutual interplay between technology, its developers or champions, and the users of that technology. It is this interplay that changes properties and application of the technology at hand. For example, such interplay is very prominent in software development, in which development iterations and lifecycles are a clear expression of the interaction between builders and users as they shape software until the users' requirements are satisfied (cf. e.g. Lee and Xia 2010).

In the previous chapter I argued that social shaping of technology can lead to 'paradigmatic regression'. These are acts of shaping that translate an expression of the paradigm of the new technology into an expression of a paradigm that is already known to the user. Resistance to new technologies, where the use or sophistication of the new technology is denied, can of course be a motivator of paradigmatic regression (Levallois, Steinmetz, and Wouters 2013). Not all regressions are necessarily motivated by conservatism or resistance, however. But even when users do embrace a new technology, the act of its social shaping may create a paradigmatic regression effect. An example of this effect can often be found when a metaphor is used in a graphical user interface (GUI). GUI metaphors are used to convey the processes or data underlying a particular piece of software in a manner that is meaningful or intelligible for human users. In order to help the user understand a new target domain or a new paradigm, it is expressed by way of a conceptual domain or a paradigm that is already known to the user. An obvious example is the metaphor of the desktop, which was used to communicate the functions of a personal computer (PC) to as broad an audience as possible (cf. Baecker et al. 1995). The only trouble is that such metaphors are necessarily incomplete as they conceal both the good and the bad of the deeper computational model. Inconsistencies in the model are hidden by a metaphor that suggests completeness to the user. Equally, metaphors hide useful functions and possibilities of the model that are not covered by the metaphor's originating paradigm (Ravasio and Tscherter 2004). In our example, the desktop metaphor does nothing to reveal the power of automation that a PC delivers to its user. GUI metaphors are probably best viewed as the expression of the assumptions that software developers hold about the user's interaction with the underlying model – but not, in any case, as a transparent and ef-

fective way of allowing the user to engage with the computer's raw power. Metaphors are in this respect paradoxical: what is meant to be a transparent means of interaction with new possibilities of a computational model is in fact an opaque barrier confining the user to a well-rehearsed collection of concepts and processes.

4.1 What Happens at the Intersection?

Paradigmatic regression is not only to be found in graphical user interfaces, we can observe similar dynamics at the level of methodological interaction between or even within research domains. To understand how paradigmatic regression can also occur as a result of the interaction between computer science and textual scholarship, it is useful to view this interaction through the lens of an existing analytical metaphor for such interaction: the trading zone.

The processes at the intersection of research domains (such as textual scholarship and computer science) have been compared to those in trading zones (Galison 2010). Whether they are zones of economic activity or those where methodologies of different fields are amalgamated, pidgins commonly arise in such places. As Peter Galison says: "A reduced common language, which begins with participants in a zone agreeing on shared meanings for certain terms, then progresses to a kind of pidgin and eventually to a creole, which is a new language born out of old ones" (Gorman, Jenkins, and Plowright 2012). Galison also draws attention to the possible existence of visual and mathematical creoles. Indeed, these are not hard to identify in digital humanities: a good example can be seen in the works of Franco Moretti, who has methodologically integrated quantification and visualization methods such as graphs, maps and tree heuristics into comparative literature studies (Moretti 2007). Nor is it very hard to identify current digital humanities as a whole with a new expert community as, according to Galison, they may take shape during the "creole stage" at the intersection of domains. It has been argued that, in creoles of natural language, it is the subordinate group that provides most of the syntactic structure for the creole, whereas the dominant group provides lexical items and concepts. Though Galison provides

some empirical observations, it remains an open question whether the same patterns hold for the emergence of methodological pidgins at the interface of different research domains. What interests me here is whether we can indeed observe the formation of a methodological creole in the emerging vocabulary of digital humanities, and whether hints can be found in that vocabulary of a similar regressive dynamic to that observed on the graphical user interface level. It may be that Matthew Kirschenbaum provides us with some – admittedly still anecdotal – evidence of precisely such a dynamic. In a recent article, Kirschenbaum attempts to trace the origin of the label “digital humanities”. He identifies a key moment, reported to him by John Unsworth, which seems to have been the tipping point that would propel this label towards its current status of *de facto* denominator of what then was and still is a non-homogeneous research domain. Unsworth relates the choice of “digital humanities” to a discussion surrounding the title of the Blackwell 2004 Companion to Digital Humanities: “Ray [Siemens] wanted ‘A Companion to Humanities Computing’, as that was the term commonly used at that point; the editorial and marketing folks at Blackwell wanted ‘Companion to Digitized Humanities’. I suggested ‘Companion to Digital Humanities’ to shift the emphasis away from simple digitization” (Kirschenbaum 2012). Of course we cannot take this as a *pars pro toto* for the social shaping of the dynamics for a whole field, but it is suggestive. Ray Siemens by no means stands alone in his preference for “humanities computing”. Susan Hockey, for instance, titled her contribution to this very companion “The History of Humanities Computing” (Hockey 2004). Significantly, it is the prominent authorities in the field, veritable Nestors, who consistently speak of “humanities computing” – people like Willard McCarty (2005), John Unsworth (2002), and Dino Buzzetti: “humanities computing – I still prefer this designation to digital humanities” (Buzzetti 2012). According to Siemens, the term was “commonly used at that point”, yet the publishers preferred the new term in order to broaden the appeal of the concept by choosing a metaphor that felt less challenging. This was a small but pivotal event in the history of the field, which simultaneously points to the state of digital humanities as a methodological pidgin and to an act of paradigmatic regression. The vocabulary juxtapositions in both terms are constructs of a methodological pidgin. Where “humanities computing” suggests an equal interaction or relation between two fields with a stress on computational activity, the term

“digital humanities” (purposefully or not) pushes the balance back toward the domain of humanities and subjugates the computational/digital aspect as a partial property of that field. Or in the words of Willard McCarty (2013): “Note, please, the name ‘digital humanities’ grammatically subordinates the digital [...] ‘Humanities computing’ takes advantage of the ability in English to make a noun serve as an adjective while staying a noun, and it draws upon the participle/gerund ambiguity. But it seems I’ve lost this contest!”

4.2 The Trading Zone and Digital Textual Scholarship Practice

Scholarly digital editions and the sites where they are conceived and created, virtual or concrete, are themselves methodological trading zones that materialize at two levels. There is a laboratory-like setting tied in a relatively small context to the practice of preparing and publishing a concrete digital scholarly edition – and possibly also the development of a specific technical infrastructure connected to it. At a more abstract level we find a theoretical discussion that connects to the methodological and epistemological histories of textual scholarship, knowledge representation and digital technology. The critical study of these trading zones along empirical ethnographic lines – another approach often applied in science and technology studies – would have much to tell about the methodological interaction between computer science and textual scholarship. Although such an elaborate study has yet to be undertaken, even fairly anecdotal observations nevertheless yield some intriguing insights.

The Huygens Institute for the History of the Netherlands is home to an example of a smaller-scale trading zone in a laboratory setting.² The institute encompasses a computer science and software development group that is relatively large by the standards of humanities research, numbering around sixteen professionally trained or educated IT developers. Various members of this group have distinct strengths, such as interface design, data modeling, architecture integration, and text analysis. The group works closely with at

²Cf. the information on its website <https://www.huygens.knaw.nl/?lang=en>

least three researchers who are themselves closely involved in the national and international digital humanities community. Through numerous projects, group members are also in close productive contact with most of the other researchers in the institute and with external researchers active in relevant projects. The projects themselves cover a large part of the spectrum of digital humanities undertakings, from data modeling and repository building,³ through digital scholarly editions such as the correspondence of Vincent van Gogh,⁴ to analytical tool building, of which the text collation engine CollateX⁵ is an example (Haentjens Dekker et al 2014).

The research staff of the institute originally had no particular focus on digital or computational activities. In 2005 the institute took the strategic decision to move into the domain of digital scholarly publications as well. The initiative began with the addition of a literary researcher and two developers to the institute. Staff at a related institute, later dissolved, had been developing a “collaboratory” for the curation and analysis of humanities and social science data, which today would be called a Virtual Research Environment (VRE). At the Huygens Institute the part of this environment relevant to the humanities, consisting mainly of a transcription and publication environment for historical texts, was adopted and strongly pushed forward while the social science aspect was eventually abandoned. This eventually became the eLaborate online environment, “in which scholars can upload scans, transcribe and annotate text and publish the results as an online text edition”.⁶ ELaborate is a Web-based environment where textual scholars find support for basic tasks in creating and editing a digital scholarly edition. A project in eLaborate is essentially a container for a series of scanned manuscript or print text pages that can be arranged arbitrarily in a tree structure. Fine-grained authorization allows one to arrange access or restrictions down to page level and thus to arrange for private, collaborative or fully open edition workflows. A text editor is facilitated to aid in creating diplomatic and critical transcriptions which can be layered with annotations to serve the researcher’s or reader’s needs. All data is stored and retrievable as XML. In addition eLaborate facil-

³<https://github.com/HuygensING/timbucto>

⁴<http://www.vangoghletters.org>

⁵<http://collatex.net/>

⁶<https://www.elaborate.huygens.knaw.nl/>

itates the automated publishing of Web-based editions and provides a generalized graphical interface based on “fluid” columns. Vertical areas of the screen can be arbitrarily arranged for visualizing the reading text, connected annotations, browsing in the text structure, full text search, and so forth. Given some basic training, eLaborate provides an out-of-the-box solution allowing textual scholars with only average computer skills to create basic digital scholarly editions without much need for technical support.

It is relevant to note that the IT team adopted an “agile” software development methodology. This type of software development takes a manifest user-centered and evolutionary approach to software manufacturing. Short one- or two-week iterations deliver functioning parts of software that are evaluated by the client/user. This ensures the balancing of the software production with the evolving vision and knowledge of the client (Martin 2002). Arguably this methodology feeds into the social shaping aspects of introducing new technologies and methodologies.

A case study of the methodological dynamics surrounding the development of eLaborate serves to show that the trading zone metaphor is not unproblematic. Do the dynamics and interactions in the context – the work site – where eLaborate was developed point to the emergence of a methodological pidgin? Most certainly the developers and the researcher who headed the project started exchanging terminology. The developers began to refer to concepts such as “page”, “annotation”, “transcription”. The researchers grew accustomed to using words such as “user”, “interface”, “architecture”, as well as the vocabulary that is rather typical for the agile methodology used by the developers: “planning game”, “iteration”, etc. Whether this constitutes a beginning of a methodological pidgin is debatable. The interactions that led to the exchange of vocabulary could equally be attributed to standard development practice in which there is a particular relationship between client and service provider and in which, certainly within agile methodology, the provider normally tries to understand the client’s work process and concepts in order to model them into software. The objective of the developers in that case is simply to mimic as closely as possible the concepts the client is using. Arguably this could cause a medium shift in which the researcher ends up with a digital environment that is virtually identical to his or her known analogue work process and material. Once the

work is done, the client and developer can go their separate ways, without having essentially influenced the methodologies on either side.

A clearer indicator of methodological change may be the actual loss of lexical items. During the eLaborate project it transpired that an index – in the sense of the keyword reference list in the back of a book – is not a very useful instrument to mimic in a digital environment if the texts at hand are automatically indexed and the interface includes a full-text search function that presents its result as a list of keywords in context. In various edition projects where eLaborate was deployed some friction and dissonance could be observed among users (either textual scholars or trained volunteers who transcribed manuscript material) about the lack of an index, but gradually the use of full text search as a replacement for the index became accepted, even appreciated, once the possibilities for wildcard and fuzzy search were understood. This is notwithstanding the fact that a full text search is not the epistemological equivalent of an index. Current full text indexing technology does not, for instance, facilitate named entity resolution in the same way as traditional indices may. Nevertheless, within projects based on eLaborate the concept of “index” is no longer used except for references to the past; the concept of “zoekfunctie” (search function) seems to have all but replaced it. An index covers of course more important scholarly functions than just finding topics on pages. It also serves as a small ontology to relate synonyms through entries pointing the user to other entries (usually labeled with e.g. “see” or “see also”). Replacing the index with a full text search that lacks such technology would imply a sincere reduction of scholarly function (cf. Rawson and Muñoz 2016). This ontology function that serves as a manual means for coreference resolution was replaced with a categorization function for annotations. Thus, although no scholarly functions were lost the lexical items describing them, such as “index”, did disappear. As did the traditional presentation in the case of the index. For textual scholarship and scholarly editing I would argue that the loss of the analogue representation of an index and even the lexical reference to it does indeed constitute a methodological change, as both technique and interface changed. The salient point here being not loss of scholarly function – which did not happen – but loss of terminology as an indicator of methodological change.

The same event serves to show the dynamics of social shaping and regression

in a different way. With the indexing technology used in eLaborate – first Lucene⁷ and later Solr⁸ – it is possible to generate search result lists with text context ranked by “relevance”. Although the keyword-in-context search results list eventually found unanimous adoption, the concept of “relevance” became a topic of recurring and fractious debate. Lucene applies a combination of Boolean and vector space models to determine the relevance of documents to a user’s query. The Boolean measure selects the documents that correspond to the terms the user wishes to find or ignore. A vector space model is then applied to that selection to rank the relevance of each document to the query. Formally this model determines relevance by applying a cosine measure to the vectorized document vocabulary and query (Widdows 2004). The vocabulary of any text can be expressed as a mathematical vector and the basic trigonometric function of the cosine can be applied to determine the size of the angle between two such vectors. This essentially means that the smaller the angle, the more the vocabularies of two texts are similar. In Lucene this measure is used to determine if requested search terms appear more often in a particular document than on average in the vocabulary of all documents retrieved with a specific query. The more such terms appear in a document, the higher the relevance ranking of that document. It transpired that the textual scholars and other users confronted with this technology were for the most part unimpressed with the relevance ranking, which appeared incomprehensible and alien to them. And although the feature was initially presented in the interface, most edition projects within eLaborate preferred canonical orderings such as sorting by folio number, name of author or text, shelf mark etc. As a result, word-weighted ranking is no longer offered in the editing and publication interfaces of eLaborate, and the researcher in charge of the development confirmed that in the several rounds of open testing that the software underwent, none of the trained users requested the function (Van Dalen-Oskam 2014).

The virtual disappearance of automatic ranking by relevance as a function in the current version of eLaborate is a case of social shaping of technology, and indeed of paradigmatic regression. Ranking by relevance could arguably be methodologically useful for textual scholars who must peruse a large corpus

⁷<https://lucene.apache.org/>

⁸<https://lucene.apache.org/solr/>

for occurrences of themes, words and motifs. Even if it is not the default, one would expect the option to be available. Technically there are no barriers to providing the function, as it is the default behavior of the search engine used. In fact, it took additional development effort – though admittedly not much – to provide canonical ordering. Despite all this, the functionality that is standard from the technical point of view is no longer available – a strong signal that the IT developers and the textual scholars found a barrier to knowledge exchange that they were unable to overcome. In other words, they could not create the required methodological pidgin to communicate or appreciate the possible utility of that function.

What is interesting here is not so much the disappearance of relevance-based ranking. There may be valid scholarly reasons to reject such an ordering principle – albeit that these have not been put forward by the users in this case. Rather, it serves as an example in which the pidgin, the “reduced common language” used during the interaction between developers and researchers, was not sufficient to communicate the methodological potential of a relatively straightforward, seemingly useful and non-intrusive method, and so prevented its theoretical consideration. This example shows how difficult it actually is, both for researchers and for developers, to use the trading zone for methodological gain or innovation. The textual scholars involved first needed to know of the existence of such a thing as “ranking by relevance” to be able to recognize its possible methodological potential. Next, to gauge that potential would require them ultimately to drill down to the mathematics of cosine measure for vector comparison and understand how vectors can represent documents. As it has been argued elsewhere in a similar vein, without such a detailed level of knowledge, it is difficult to assess the methodological usefulness of new technologies (Sculley and Pasanek 2004).

It should be noted additionally that this is a small example involving relatively standard digital technology. The syntactical and lexical distance that must be bridged in the case of a project such as *Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic*⁹ is significantly larger, as in that project correspondences are visualized through network analysis (Van den Heuvel 2012). A sensible understanding of what may be inferred

⁹See <http://ckcc.huylgens.knaw.nl/>

from network visualizations and what this adds in terms of methodology requires a fairly deep grasp of the mathematical models underpinning not only network modeling and analysis in general, but also the topic modeling used to generate the network data (Wittek and Ravenek 2011).

All in all, this raises the question of how much methodological interaction is actualized in a methodological trading zone in the smaller concrete context of DH projects. Some superficial vocabulary is certainly exchanged, of which some may be instrumental in future co-operation for both researchers and developers. However, there is little in the way of deep methodological trading going on. Textual scholars are not providing knowledge about theoretical notions on scholarly editing and literary criticism to developers; and, vice versa, developers are not lecturing researchers about mathematical or computational principles. The common language does no more than create an interface that answers to the perceived needs of researchers in the humanities. The interface becomes an expression of these researchers' conceptions of how the digital technology might serve their purpose.

The methodological gain in this is rather superficial: access and discovery increase in scope, but concepts and processes hardly change. There is a digital translation, but little methodological innovation. The potential or realized methodological innovation furthermore happens rather covertly. In the case of the relevance ordering in eLaborate the potential is there, but hidden – again(!) – by a graphical interface, and by an apparently suboptimal methodological exchange between researchers and developers. In the case of the Circulation of Knowledge project, the mechanics, technology and methodology are almost completely covertly integrated into the resulting digital environment by the computer scientists. A further consequence was that the main technical developer struggled with negative feelings about lack of recognition for methodological merit. The covertness of this methodological innovation is far from trivial. If, as Peter Shillingsburg has pointed out, editions are scholarly and critical arguments about what a textual record means or about how it should be read, then a digital edition is also such an argument (Shillingsburg 2013). Because both interface and model are constituents of the digital edition, they are both part of that intellectual argument. The model – i.e. the combination of the data model and the computer language logic that puts it into action – is entirely conceived by computer science ex-

perts. The interface and the view it offers on that model, including the functions of the model it exposes to or hides from the outside world, is to a very large degree conceived by developers and designers. The methodology used for this is effectively inaccessible to the textual scholars, who lack the skills to interpret and comprehend the technologies used. Given that the computer scientists create so much of the intellectual argument pertaining to a particular digital scholarly edition, it would seem that having a sufficiently broad common methodological language is pivotal to digital textual scholarship. But as we can see, our current dynamics of interaction are not helping to create it.

4.3 Trading Theory in the Larger Textual Scholarly Context

Although the trading zone between computer science or digital technology and textual scholarship seems so problematic at the smaller more concrete level, there seems to be no shortage of methodological trading on the theoretical level. Exhaustively detailing and disentangling the intricately intertwined histories of textual scholarship, knowledge representation, literary criticism, computing and digital technologies, is hardly feasible in the span of this chapter. Moreover, creating history often suggests a falsely deterministic account of cause and effect. Nevertheless, it is important to identify a number of key developments. The beginnings of the Internet and the World Wide Web are usually identified with Vannevar Bush's vision of the Memex, an imaginary system to store, track, index and retrieve any information, and – crucially – to rewrite that information and keep versioning records so as to trace the development of our thoughts (Bush 1945). Visions of such knowledge systems reach far further back, however, at the very least to the work of Paul Otlet in the early twentieth century, as has been repeatedly shown (Rayward 1994; Buckland 1997; Vanhoutte 2009; Van den Heuvel and Rayward 2011). It was Theodor Nelson who coined the term “hypertext” and constructed a theory for it, inter alia referring back to Bush (Nelson 1993[1981]). Nelson's attempts at implementing his visions failed to result in successful tools, however. Instead it was Tim Berners-Lee whose team devised – with

reference to the work of Nelson (Berners-Lee 1989) – the Hypertext Transfer Protocol (HTTP), which successfully kick-started the World Wide Web. Although sympathetic to his endeavour, Nelson deeply hates Lee's technical solution: "It is vital to point out that Tim's view of hypertext (only one-way links, invisible and not allowed to overlap) is entirely different from mine (visible, unbreaking n-way links by any parties, all content legally rewearable by anyone into new documents with paths back to the originals, and transclusions as well as links – as in Vannevar Bush's original vision)" (Nelson 2010).

Imperfect or not, HTTP technology happens to align nicely with many ideas on the nature of knowledge and text that are emerging in literary criticism, textual theory and semiotics, which increasingly problematize a linear view of text and result in more post-structuralist approaches. George Landow summarizes the convergence: "Hypertext, an information technology consisting of individual blocks of text, or lexias, and the electronic links that join them, has much in common with recent literary and critical theory. For example, like much recent work by poststructuralists, such as Roland Barthes and Jacques Derrida, hypertext reconceives conventional, long-held assumptions about authors and readers and the texts they write and read. Electronic linking, which provides one of the defining features of hypertext, also embodies Julia Kristeva's notions of intertextuality, Mikhail Bakhtin's emphasis upon multivocality, Michel Foucault's conceptions of networks of power, and Gilles Deleuze and Felix Guattari's ideas of rhizomatic, "nomad thought". The very idea of hypertextuality seems to have taken form at approximately the same time that poststructuralism developed, but their points of convergence have a closer relation than that of mere contingency, for both grow out of dissatisfaction with the related phenomena of the printed book and hierarchical thought" (Landow 1994).

Digital textual scholarship and more particularly the digital scholarly edition obviously rely on the technologies delivered by the development of the Internet and the hypertext protocol. In turn, these technologies are rooted in theory which sees the nature of knowledge, information and documents as highly interconnected and referential, or intertwined and transclusional, as Nelson would in all likelihood phrase it. Peter Robinson expresses similar views when he discusses the idea of "distributed editions", with attribution

also to Peter Shillingsburg and Paul Eggert (Robinson 2003). Robinson is interested in the volatile aspects of editions. He posits that readers may become writers too, and proposes that editions may exist in a distributed fashion in an interactive Web-based space. Each reader may have a different representation: “a manuscript transcription from one site, a layer of commentary from one scholar, textual notes and emendations from another, all on different servers around the globe. In a sentence: these will be fluid, co-operative and distributed editions, the work of many, the property of all” (Robinson 2003). According to George P. Landow, this vision is strongly associated with the Docuverse, the ideas on non-linear writing and hypertext systems described by Nelson: “Perhaps the single most important development in the world of hyper-media has been the steady development of read-write systems – of the kind of systems, in other words, that the pioneering theorists Vannevar Bush and Theodor H. Nelson envisioned. Blogs, wikis [...] all represent attempts to bring to the Web the features found in hypertext software of the 1980s that made readers into authors” (Landow 2006:xiv [1997]).

But ideas on more interactive and volatile editions also refer to another complex of theory surrounding the fundamental instability of text. This complex encompasses a post-structuralist view of text where text is not a book but a hypertext, and where hypertext stresses the volatility of text, its heterogeneous, mutable, interactive and open-ended character – ideas rather opposed to that of text as an immutable form enclosed and bound by a front and back cover in a book. This theoretical complex also borrows from ideas on the fluidity of text as expressed for example by John Bryant who calls attention to the perpetual flux texts show through preprint revisions, revised editions, and adaptations that shape literary works into forms specific to different audiences (Bryant 2002). Similarly, the importance for scholarly editing of the volatile aspects of text is expressed through what has become known as *critique génétique*, an approach to editing that focuses on the *avant-texte*, the process of writing and revision that precedes the publication of a book (Van Hulle 2004; Fiormonte and Pusceddu 2006).

The instability and process aspects of text are also important to textual scholarship and the practice of scholarly editing from the point of view of the use of editions: of what happens after publication. The ideas behind hypertext, together with those about read-write systems, also inform ideas concerning

the social aspects of text and scholarly editing. Read-write systems facilitate crowdsourcing and thus open up the process of scholarly editing to a potentially far larger source of labour by “expert amateurs” (Hayles 2012) than the individual scholar could provide for (Brumfield 2013). Crowdsourcing engages an audience of users in the scholarly process literally in the *avant-texte* phase of the creation of a scholarly edition. This potential need not be confined to, say, the transcription stage of a scholarly project. Meanwhile, ideas have been developed on the so-called social edition, which allows readers/users to add their knowledge to the edition and render its creation and use a community event under the guidance of scholarly experts (Siemens et al. 2012). Lastly, the process aspect of text is also highlighted through new computational engagements that readers/users may make with texts and scholarly editions. This aspect was already expressed as early as 1949 through what is now usually seen as the first application of humanities computing: the work of Roberto Busa, which led to the computational means necessary to derive automatically a concordance to the works of St Thomas Aquinas (Hockey 2004). This was the beginning of a long development that prefigured current computer-supported analytic engagement with literary texts such as distant reading, algorithmic reading and big data analysis (Buzzetti 2009; Moretti 2007; Ramsay 2011c; Jockers 2013).

4.4 The Shape of the Digital Edition According to Reality

In short, the interaction between digital technology and textual scholarship places the focus of methodology on both the unstable and fluid aspects of text, and on the process aspects of texts. That is the fundamental tenet that computer science brings to textual scholarship. “Hypertext, unlike print, is fundamentally process- and context-oriented. Following a basic tenet of artificial intelligence theory, it views representing and acquiring knowledge as a problem of defining and searching information spaces, and it recognizes that these spaces and search methods will vary according to the purposes and abilities of particular users” (Edwards 1994). Digital scholarly editions are indeed information spaces. But they are not often information spaces that line up

with the theoretical pidgin discussed above. The theoretical notions of textual scholarship, and the scholarly digital edition that we find in the trading zones between textual scholarship and computer science, call for an expression of text and editions through which the information contained in the edition is expressed primarily according to the principles of hypertext. Current reality, however, is very different. In textual scholarship, Internet nodes are mostly placeholders that point via a URL to a digital document or to a digital edition as a whole, as a data silo. The edition of the Van Gogh letters, for instance, sits at the node identified by <http://www.vangoghletters.org/vg/> as a fully integrated and monolithic pile of edited text from letters; the pile includes comments, annotations, translations and so on. The finest granularity presented to the network of the web is at the level of the individual letter (e.g. <http://vangoghletters.org/vg/letters/leto43/letter.html>). Even that URL identifies a compound object, that is, a meaningful set of multiple scholarly objects: two facsimiles, a transcribed text, annotations, bound together by an interface that (again following Shillingsburg) represents an editorial argument about what constitutes the digital scholarly edition of this particular letter. According to this argument, there is no need to address the transcription, the facsimile, a particular annotation, in isolation. Most of the digital scholarly editions on the Web are expressed similarly. It is hardly better than a network of nodes in which each node represents a particular edition that is offered as a PDF. This situation renders it impossible to address texts (and thus editions) beyond their graphical interface in ways compatible with a hypertext model.

Digital editions often trumpet the ability to represent text exhaustively, celebrating the fact that there is no need to make decisions on what to leave out (Price 2008). Indeed, it is an asset that digital scholarly editions may be capacious almost without limit. In the case of an important and large tradition of a particular work, this potential may allow for the presentation of all witnesses as items in an inventory, or as a digital archive. Arguably this is not just an asset because of exhaustiveness of representation, but foremost because it allows for the expression of the relations between the witnesses, and thus *inter alia* the genesis and fluidity of texts – in fact the more process-like aspects of texts – for which the hypertext model as described offers technological expressive potential. In the reality of current digital scholarly digital editions,

however, this potential seems seldom realized. A graphical interface will usually allow the user to select and view single witnesses, or perhaps to compare the texts of multiple witnesses, especially if the editor has integrated a collation or comparison tool such as Juxta.¹⁰ The inventory will probably also allow a list of witnesses to be shown in chronological order. The order of that list will in all likelihood be based on a metadata property “date” or similar in the relational database underlying the digital edition archive. The list itself is a generated GUI visualization expressing that metadata. The point here is that a list so represented is not a hypertext representation of the chronological “linkedness” of the witnesses, it is a mere list of individuated metadata. This is different from the idea of hypertext that all information is expressed as machine negotiable nodes and links, so that an expressive network of knowledge is created. This means that the chronological order of the witnesses in this case can only be inferred through human cognition from the metadata based list – it is not represented as knowledge in a computationally tractable form intrinsic to the hypertext medium. Much effort may thus be invested in gathering exhaustive representations of individual witnesses, but if the result of that effort only allows user-level navigation of relational metadata represented as a graphical interface, then the digital scholarly edition is not an effective hypertext knowledge space. Such an edition may still be valuable for the sheer wealth of information, but it remains firmly at the level of document representation for human consumption without integrating the relations between witnesses in a computationally networked representation.

4.5 Regression and Reaffirmation

There is nothing deterministic about technology, and indeed nothing much deterministic about hypertext. As a technology to express a text and to present it in the form of a digital scholarly edition, hypertext has been shaped by the scholarly community into little more than a filing cabinet for self-contained documents. Most digital scholarly editions on the Internet express the particular idea the scholar responsible for the edition

¹⁰<http://www.juxtasoftware.org/>

has about what a digital edition is or should be; normally, that idea is a re-representation of the book. We find collections of page-based facsimiles and transcriptions presented as self-contained units, wrapped up in and bound by the front matter that is the interface. There is attention for fluid aspects, and for context. The Hyperstack edition of Saint Patrick's *Confessio*,¹¹ for instance, explicitly offers its users the possibility to venture from the "centrality of the text [...]" through the dense net of textual layers and background information in answer to questions that are likely to arise in their minds" (Fischer 2011). The dense net in question is effectively a star network radiating out from the main page into leaves containing pages of metadata, facsimiles of manuscript folia, or transcriptions of entire texts. Despite the impressive density of information, the information itself is not that densely networked. The relations between the texts and the contextualizing information is described, but not expressed through the "hyper fabric" of e.g. HTTP links. Even so, the *Confessio* is rather an exception to the rule – very few of today's digital editions seem to be particularly concerned with the core ideal of hypertext as an expression of linked information, of process and context.

Most digital scholarly editions, in fact, are all but literal translations of a book into a non-book-oriented medium. Peter Robinson, writing about the distinctions of text-as-work and text-as-document, argues that in the early days of digital editions – roughly until 2005 – scholars would privilege the text-as-work perspective, focusing on the potential of digital technology to express and support the properties of text that construct its meaning (Robinson 2013a:56). In recent years, he continues, this trend has been exactly reversed. More recent digital scholarly editions harness the digital medium rather to represent the text-as-document – the faithful re-representation of a text according to its expression in the physical documents that carry it. As an example Robinson points to the online edition of Jane Austen's fiction manuscripts.¹² Elena Pierazzo, who was deeply involved with the methodological design of this edition, unsurprisingly offers a rationale for a text-as-document approach to the digital edition (Pierazzo 2011). Robinson also notes that many collaborative transcription systems are designed to record

¹¹<http://www.confessio.ie/>

¹²<http://www.janeausten.ac.uk>

text-as-document: not one of twenty-one tools listed in a survey by Ben Brumfield offers the possibility of recording text-as-work (Brumfield 2013). Indeed it is far easier to point to examples of digital scholarly editions that are in essence metaphors of the book, or in other words: translations of a print text to the digital medium, apparently for no other reason than to fulfill the same role as the print text.

Textual scholarly theory, as has been shown, embraces hypertext as a technology which enables the expression of post-structuralist ideas about information, with a focus on the fluid properties of text. It has often been suggested that the capabilities of digital technologies should become the focus and practice of digital scholarly editing. Despite all this, that ideal is not materializing in the form of concrete digital editions, and for similar reasons to those observed in the smaller context of the eLaborate project. Here, too, we find the dynamics of paradigmatic regression in the professional community surrounding the digital scholarly edition. The methodological potential of information technology is hidden by the incomplete metaphors of a paradigm that is itself reaffirmed by becoming the primary interface to the new technology. Robinson argues that there is a strong continuity of previous contemplation of print editions present in the thinking of those scholars who first conceived of the digital scholarly edition, resulting in a kind of theoretical pidgin that embraces the new technology, but uses it to express digitally a familiar form for the scholarly edition: the printed book (Robinson 2013a:60). The print edition in that digital translation is a metaphor, but one that begins to hide hypertext's native potential for expressing referential and conceptual links between texts. The graphical interfaces of digital scholarly editions almost all refer strongly to this book metaphor, reaffirming thereby the paradigm from which that metaphor springs. In the end, the use of the technology has shaped it into a tool to recreate that which is already well known. It is also worth noting that the de facto lingua franca of current digital scholarly editions, TEI-XML,¹³ is instrumental in this reaffirmation. As an encoding language it is geared fully towards describing text-as-document. Although not graphical in nature, TEI is thus an interface that, like graphical interfaces, hides many of the essential networking and process characteristics of hypertext. Instead, TEI-XML, with its text-inward orientation, print-text

¹³<http://www.tei-c.org>

paradigm and hierarchical structure focus, continuously reaffirms the view of the digital edition as representing a text-as-document.

4.6 Beyond the Book?

There is nothing deterministic about the Internet. The paradigmatic regression we currently see in the digital textual scholarship community is a clear demonstration of that. The textual scholarship community has devised a methodological pidgin that exploits a new technology to express a well-rehearsed paradigm of scholarly editing. Yet this must not be where the methodological shaping and disciplinary trading stops. The theoretical concepts pertaining to the fluidity of text are clearly important to the textual scholarly community, but they still need to be brought fully into the concrete methodological pidgin that is currently geared towards representing a text-as-document, rather than toward text-as-process. As long as scholarly editors keep producing digital metaphors of the book, this will hardly happen. Both textual theorists and computer science practitioners must intensify the methodological discourse to clarify what existing technology is needed to implement a form of hypertext that truly represents textual fluidity and text relations in a scholarly viable and computational tractable manner – a hypertext language inspired both by computer science and textual scholarship. Without that dialogue we relegate the *raison d'être* for the digital scholarly edition to that of a mere medium shift, we limit its expressiveness to that of print text, and we fail to explore the computational potential for digital text representation, analysis and interaction.

