

The Case of the Bold Button: Social Shaping of Technology and the Digital Scholarly Edition

Joris J. van Zundert

Huygens Institute for the History of the Netherlands, Royal Netherlands Academy of Arts and Sciences, The Hague, The Netherlands

Abstract

The role and usage of a certain technology is not imparted wholesale on the intended user community—technology is not deterministic. Rather, a negotiation between users and the designers of the technology will result in its particular form and function. This article considers a side effect of these negotiations. When a certain known technology is used to convey a new technological concept or model, there is a risk that the paradigm associated by the users with the known technology will eclipse the new model and its affordances in part or in whole. The article presents a case study of this ‘paradigmatic regression’ centering on a transcription tool of the Huygens Institute in the Netherlands. It is argued that similar effects also come into play at a larger scale within the field of textual scholarship, inhibiting the exploration of the affordances of new models that do not adhere to the pervasive digital metaphor of the codex. An example of such an innovative model, the knowledge graph model, is briefly introduced to illustrate the point.

Correspondence: Joris J. van Zundert, Huygens Institute for the History of the Netherlands, Royal Netherlands Academy of Arts and Sciences, The Hague, The Netherlands.
E-mail: joris.van.zundert@huygens.knaw.nl

First, let us observe two things missing from almost all electronic scholarly editions made to this point. The first missing aspect is that up to now, almost without exception, no scholarly electronic edition has presented material which could not have been presented in book form, nor indeed presented this material in a manner significantly different from that which could have been managed in print.

These are words by Peter Robinson, who spoke and wrote them in 2004 (Robinson, 2004). I think little has changed in the 9 years since and the observation still more or less holds. At the time, Robinson argued vehemently for digital scholarly editions that would move decisively beyond the

realm of the possibilities of print publication. He was—and is—by no means the only one that has been advocating for such a shift. In fact, many have wondered how the digital medium, or the virtual environment, would change the nature and appearance of the scholarly edition. For that matter, grand perspectives on paradigmatic change due to medium change are not unique to textual scholarship. The introduction of a new medium or technology has always inspired great debate between advocates and antagonists of the next big thing. Self-proclaimed supporters of digital media usually advocate revolutionary changes. In the case of textual scholarship, for example, one may hear it proclaimed that the book is dead; good riddance, the advocates of ‘The Next Big Thing’ (Bod, 2013: 8)

judge, for it was a clumsy, static, institutionally bounded, difficult to use, and outdated interface. Give way to open access, process orientation, dynamic interfaces, intuitive interaction, fluid text, social editing, etc. (cf., for instance, Siemens *et al.*, 2012). With similar and undaunted zeal, luddites lament the waning of solid scholarly practice: concentration span, close reading, philological interpretation, editorial practice, and convention (Fish, 2011)—all sacrificed to the ‘Bitch goddess, QUANTIFICATION’ (sic) as Bridenbaugh once put it (Bridenbaugh, 1963).¹

The screaming and kicking of luddites aside the proponents of change do not seem really to get what they want. After many years of development of digital technology, the book is as alive as it ever was. We scarcely find digital editions, scholarly or otherwise, resembling the advanced models of dynamic, fluid, collaborative, and social texts such as those proposed by McGann (2010), Drucker (Lunenfeld *et al.*, 2012: 36), Shillingsburg (Jones *et al.*, 2010), Robinson (2004), Van Hulle (2010), Siemens (Siemens *et al.*, 2012), and myself (Boot and van Zundert, 2011). E-books are certainly impacting the market (AAP, 2010, Cain Miller and Bosman, 2011), but e-books are pure digital metaphors of the print book. Digital scholarly editions hardly have any impact (Porter, 2013), but what is more important is that they are a far cry from what many expected them to be. We could suppose that this state of affairs is due to a lack of knowledge, skills, and technology support as has been indeed suggested before (cf. Courant *et al.*, 2006). And it is probably true there are severe problems of teaching and training in our field, given that master and Ph.D. programs truly oriented on the digital humanities are only lately coming into existence. Yet, I think there might be more to the matter.

Maybe we need to answer to Borgman’s call: ‘Why is no one following digital humanities scholars around to understand their practices, in the way that scientists have been studied for the last several decades?’ (Borgman, 2009). What do we see if we step back for a while from our work as textual scholars and digital humanities researchers and look at what is happening from the social sciences, in particular of Science and Technology Studies? Science

and Technology Studies suggest *inter alia* to study technology development in its social context. In the past few years, I have studied the creation and development of the digital scholarly edition within the laboratory-like setting in the Huygens Institute for the History of the Netherlands. Here we find a relatively large—for humanities contexts in any case—IT Research and Development (R&D) group of on average sixteen persons working together with about sixty historians, textual scholars, and digital archivists. The research context consists of a dozen senior researchers, a similar amount of non-senior and associate researchers, a similar amount of Ph.D. candidates with various contracts ranging from predominantly full-time added staff to volunteer workers, and of course non-IT R&D² supporting staff.

The adoption and application of technology is as much a social as it is a technical process. These processes are inevitably intertwined: technology does not determine but operates within and is operated upon in a complex social field (Bijker *et al.*, 1987). The manifestation of such intertwined processes is directly visible in the field of digital humanities and in the development of the digital scholarly edition. Of course, the digital scholarly edition is a digital artifact brought to life in a context of heavy interaction between technology (computer science and digital humanities) and a non-technological context (textual scholarship and humanities in general). This intricate and intensive interaction is a daily practice at the Huygens Institute. One of my tasks is to guide the interaction between IT R&D, documentary editors, textual scholars, and researchers of literature and history, and to facilitate the ongoing methodological discussion between these cultures. I have had the privilege to study these processes from many angles: methodology, technology, model, role, audience, development, and so on.

As has happened in many similar research contexts, a transcription tool was developed at the Huygens Institute to support the basic work of turning non-OCR-able texts from early printed works and medieval and modern manuscripts into their digital machine-processable counterparts. The development of this tool, eLaborate (c.f. <https://www.elaborate.huygens.knaw.nl>), was based on a

strategy of encapsulating and hiding XML markup—to be transformed to TEI encoding behind the scenes—with a graphical interface. In this way, the tool was meant to present minimal barriers to transcribers who came in a variety of levels of expertise on encoding. This indeed resulted in successful participation of significant numbers of volunteers unskilled in XML over a large set of projects. Also the encapsulation of technicalities facilitated greatly the focus on community and project management (Beaulieu *et al.*, 2012).

Here I am not so much interested in the features or particulars of eLaborate. Instead I want to focus on one particular researcher–developer interaction I witnessed that, I think, stands as an example of a general and strong tendency in the scholarly community at large. The usability principle behind eLaborate is that any encoding or markup is treated as an annotation on arbitrary regions within the text. To this end, when a user has selected a certain region in the text with the mouse, a pop-up dialog appears allowing the user to enter annotative tags, comments, etc. The interface thus closely mimics a concept—using a highlighter and pen to create annotations—that is known and tangible to anyone who has basic experience in working with scholarly texts. The clear down side of this principle—if dogmatically applied—is that a user is left with an enormous number of click-and-point-and-type annotation tasks. Especially in cases of seemingly insignificant but frequent markup, such as with the indication of bold face print, this approach strikes the user as tediously pedantic.³ The result of this usability agony was a recurring and strong push in the user community to have a button labeled ‘bold’—in fact to have several such buttons for italics, underline, and other common very frequently appearing properties of text—lowering the volume of tedious annotation. I remain to this day convinced that we should not have implemented that button as we did.

The root cause for my conviction is of course that these buttons violate the rationale for XML over HTML, namely the strict and intentional separation of representational and semantic information. The most common interpretation of boldface type is that it is a material manifestation of the

concept of emphasis. Even this is not universal—many other concepts may also be expressed by the use of boldface type. Thus, the provision of a button to record that some text is in boldface type introduces principal ambiguity in a descriptive system. There is no way to tell what the function of the bold print was: it arbitrarily covers any use, without delineating which of the several possible textual concepts might apply.

More importantly, however, for my argument here is that the implementation of this simple button reveals how technology is indeed shaped through its social context. The intent of eLaborate’s approach was paradigmatic: its purpose was to allow editors of text to change from a representational paradigm to a semantic paradigm. We could have done this by forcing our users to become competent XML authors. Our users judged XML tedious and complicated, however, and complexity is a well-known ‘fail factor’ working against the adoption of any new technology (Rogers, 1983). Thus, to move our users gently into the new paradigm, we had to create an interface that offered a clear and substantial advantage over existing technology, but that at the same time did not seem overly complex. The annotation ‘highlighter’ pop-up seemed a good solution, trying to balance paradigm innovation with ease of use and compatibility with a known paradigm. However, the annotation pop-up led to a tedious routine that severely constrained ease of use. When ease of use is compromised to such an extent, the new possibilities inherent in a technology do not lead to a change of routine to accommodate the technology, and thus the adoption of a new paradigm does not occur. Instead, the perceived constraints lead to a change in the technology (Leonardi, 2011). This is exactly what happened in the interaction between developers, users, researchers, and technology in the case of eLaborate. A bold button was introduced to remedy usability constraints: social shaping of technology at work.

As an unintended consequence—as Robert Merton would have it—of this social shaping of eLaborate the paradigmatic intent of the innovation was now black boxed. This is not meant in the sense of Latour’s definition that defines a black box

according to general acceptance of the correctness of the inner mechanism (Latour, 1988), but in the sense that the innovative aspect of the new paradigm was now completely unobservable and thus effectively unknowable to its intended audience. The unobservability of such a black-box model is also a known ‘fail factor’ for innovation (Marinova and Phillimore, 2003; Rogers, 1983). This is an unintended and usually unrecognized effect I have often found interfaces to have, and it is a problem that particularly affects graphical interfaces. A graphical user interface suggests a transparency of model and paradigm that is not truly there—in fact the graphical interface is as much an opaque barrier to the internal paradigm of a system as it is a means of engaging with that very system. Analogous to Robinson (2013) and others, I would argue that software interfaces, such as the interfaces to digital text editions, are an intellectual argument about the internal model of a system rather than a direct communication of that model to any user. When (as a result of the interaction between developer and user/researcher) the interface undergoes social shaping, that is also an expression of an intellectual argument by the user about the model.

In the case of the bold button, the user has not merely molded convenience into the interface. What also happened was that the intended paradigm—that of semantically oriented XML—was expressed in a paradigm which was more familiar to most users—that of representationally oriented HTML. But this effectively prevented the user from engaging with and getting to know the new paradigm, or at least a part of it. The bold button hid a class of semantically expressive potential behind a single representational ‘wrapper’. As an extension of the Meno paradox (Nickles, 2003), not only were the users unable to negotiate new knowledge, they had shaped the technology in a way that made it now impossible to engage at all with the new paradigm. User-centered design had led to the users shaping new technology so that it was congruent with the paradigm they were familiar with. The new was expressed in the ways of the old, but also turned into something inaccessible and irrelevant. This unintended effect of an intended paradigm being encapsulated and effectively hidden by a more familiar

paradigm is caused by what I will call paradigmatic regression: the social shaping of a technological interface such that it can no longer express essential properties of an intended paradigm. The pivotal error that was made with the introduction of the ‘bold’ button was that the button does not express the digital paradigm. Instead, we did exactly the opposite: we facilitated the scholarly users’ regression toward the paradigm of the book metaphor known to them. Thereby we confirmed that nothing had changed, that print convention was still the paradigm to use. As proponents of digital scholarship, we may tend to think we are free from this sort of paradigmatic regression. But we are not. Most if not all digital scholarly editions are still solidly rooted in the book metaphors and print conventions, and I think it is exactly because of this silent regression. A brief history of humanities computing may be telltale.

The beginnings of humanities computing and the development of the digital scholarly edition are usually dated 1949 with the seminal work of Father Busa (Hockey, 2004). Roberto Busa demonstrated the first practical applications of computational text processing by automating the tasks of indexing and context retrieval. However, the result was presented in a form already well known to scholarly editing: a fifty-six-volume print publication concordance. The computational aspect was used simply to automate and scale a tedious and error-prone editorial task. The utility and sense of that of course goes without question. What interests me here, however, is that the automation was geared toward reiterating on a larger scale a scholarly task that was in essence well known and rehearsed; computational power was harnessed to produce an instrument well within the confines of the existing paradigm of print text and its scholarly applications.

The advent of the database and later the relational database prompted the curation and publication of several catalogs and indices of textual metadata, as well as the first repositories of text. This was of course a major enhancement of the capacity for discovery of texts and related metadata. Databases allowed for efficient and convenient discovery of text through the use of matching selection queries. Scholars such as Jerome McGann, Peter Robinson,

Dino Buzzetti, Manfred Thaller, and others began to envision different forms of engagement with text made possible due to the availability of full-text repositories and metadata. Despite all this, the database did not change the essential way scholars engaged with the actual texts. Even if, for instance, Buzzetti and Thaller argued that a digital edition's 'liability to processing' is the essential feature that sets it apart from conventional editions (Buzzetti and Rehbein, 1998), texts were still perceived predominantly as intentionally ordered strings of words for human interpretation. Thus, notwithstanding ideas on how to engage with text in new ways separate from the reading, commentary, and interpretation that has traditionally been handled by humans, the digital scholarly editions produced in the last part of the 20th century have again presented text to us essentially as a digitized book.

According to Hockey, in the early to mid-1990s a great deal of interest and discussion arose in the scholarly community concerning what an electronic edition might look like. However, with the 'notable exception of work carried out by Peter Robinson', few of these publications were realized in an actual implementation. Once 'theory had to be put into practice and projects were faced with the laborious work of entering and marking up text and developing software, attention began to turn elsewhere' (Hockey, 2004). As with the bold button example, we find that a new technology turned out to provide too little practical facility to lead to successful innovation. Yet there is more to the matter.

The 'Next Big Thing' of the last decade of the 20th century was the World Wide Web, founded on the technologies of the Internet and Hypertext. As Landow has pointed out, 'computer hypertext—text composed of blocks of words (or images) linked electronically by multiple paths, chains, or trails in an open-ended, perpetually unfinished textuality described by the terms link, node, network, web, and path' precisely matches Roland Barthes' ideal textuality (Landow, 2006). If we need to point to a single moment and opportunity in history when the very fabric of a new technology was made suitable to a scholarly community for the expression of relations and structures, not just within single texts but especially between texts, it was the moment of

the invention of hypertext. That the opportunity arose cannot have been surprising, as the essential mechanism of hypertext—the hyperlink—was the technological implementation of a long-standing idea that knowledge and information are interlinked. Already pioneers such as Paul Otlet in the early 20th century could contemplate information systems that would link knowledge in the form of formalized multidimensional relations between documents (Rayward, 1994). What is actually rather surprising is that such long-standing epistemological knowledge about the relation of different chunks of information within documents and congruent ideas from post-structuralist literary criticism such as Kristevas intertextual references (Mitra, 1999) found so little expression in digital scholarly editions. The expressive power of that single pivotal element of the original HTML 1.0 specification, the `http A` element with its invaluable `HREF` property, implemented by Tim Berners-Lee and itself an echo of Theodor Nelson's ideas of transclusion (Nelson, 1995) should have reverberated within the scholarly community. Here was its opportunity to give expression to the linked and intertwined natures of cultures of text, literary criticism, and (digital) textual materiality that go to the heart of the field (Van Mierlo, 2006). The hyperlink created a native digital expression for the act of referencing, an expression of knowledge very much at the core of textual description, interpretation, and criticism. Thus, here was a unique opportunity to change from a paradigm of print publication to a paradigm of interconnected texts expressing knowledge.

The scholarly editing community, however, adopted the 'markup' rather than the 'hyper' of the hypertext markup language, by developing Goldfarb's SGML eventually into the TEI-XML descriptive standard (Goldfarb, 1996 and Renear, 2004). At the time, these dialects of markup technology were used primarily to mark up texts as they are represented in books—the fact that I do not think anyone has but flippantly suggested marking up Web pages in TEI-XML may stand to prove the point. The scholarly community predominantly turned hypertext markup into a descriptive model of the book, and we have produced digital book

metaphors as digital scholarly editions ever since. As with the bold button, a new technology was not explored but rather encapsulated by a known paradigm. The hyperlink was meant not to be a descriptive tool, but to link information in different documents. Yet its foremost use in scholarly editing has been to link contents, chapter headings, and indices to pages in self-contained digital editions. Roberto Busa had ‘a vision and imagination that reach beyond the horizons of many of the current generation of practitioners who have been brought up with the Internet’. He imagined scholarly editions on the Internet combined with analysis tools (Hockey, 2004), a horizon that has been reiterated by many (cf. for instance Buzzetti, 2009). However, digital editions developed in a completely different direction. The processing involved is mostly aimed at rendering the text for consumption by human readers. To defy the intent of the hyperlink has been in my view among the most remarkable feats of paradigmatic regression in the textual scholarship community. One can wonder though whether this is a bad thing. If we accept the bilateral dynamic between audience and innovation, then why would we care when some innovations do not succeed? If the book metaphor paradigm suffices for our needs, does this not indeed suffice?

To answer that we must ask: to whose needs do digital scholarly editions actually cater? Given the designation, they should cater to scholars and researchers, but do they? The latest developments in digital scholarly editing are linked to the possibilities created for Computer-Supported Cooperative Work (CSCW)—a term that was coined by the IBM research group headed by Greif (1988) by the Internet and the rise in computer literacy. Essentially CSCW is a label that can be put on any collaborative activity that is supported by Web or Web 2.0 means. Crowdsourcing as a means of dividing large workloads has been around for a while and has been a specific implementation of CSCW ever since Web 1.0 technologies turned into Web 2.0 technologies. Many have proclaimed crowdsourcing to be the advent of the social edition—most prominently Ray Siemens (Siemens *et al.*, 2012)—which redefines the editor’s role to be that of a team leader concerned with proper workflow, quality control,

and overseeing managerial and funding aspects (Sahle, 2013), whereas concrete editorial tasks are delegated to social communities formed around specific texts. Questions have been raised about the actual effectiveness of crowd sourcing (Causser *et al.*, 2012). But more importantly, recent studies show that the old rule of thumb of the collaborative Internet—that 10% of the workforce provides 90% of the labor (cf. Brumfield *et al.*, 2012)—still holds for any open collaborative project, implying that many crowdsourced editions are not in fact truly social. Moreover, when Peter Robinson said ‘All readers may become editors too’, he was not simply referring to a cheap labor force for source transcription, to be conveniently discarded the moment a transcription phase is done (Robinson, 2004). Instead, like Ray Siemens proposed, he envisioned a ‘social edition’ that embodies the ideas of open notebook science (cf. Shaw *et al.*, 2013) and renders all aspects of the editorial process—e.g. annotation, commentary, and interpretation—open to public engagement (Siemens *et al.*, 2012). But we in the scholarly community are not at all at ease with letting go of our presumption that scholarly editing is a highly skilled practice that does not provide for easy delegation of tasks. It is challenging to truly consider the extent to which we can open up the scholarly process of creating a digital edition to leave the tedious tasks typically associated with high quality scholarly inference to the wisdom of the crowds—in the case of literary analysis, this often includes the painstaking tracing of names, annotation of plot, clarification of meaning, for instance. In current practice, however, the digital scholarly editorial tasks beyond the transcription phase remain reserved either for the single authoritative author or for a small group of qualified editors. In this way, most scholarly digital editions adhere to an authoritative publication paradigm. We use big all-encompassing words like ‘social’, ‘open’, and ‘community’, but in fact we are again regressing to authoritative processes that remain well within the paradigm of the print edition. Although on the verge of being harsh, it is nevertheless fair to state that digital scholarly editions cater to the needs of the scholarly editors, not to users and researchers as knowledge producers.

Along another tangent: Edward Vanhoutte pointed out the possibilities of targeting different audiences with different visualizations of the same editions (Vanhoutte, 2011). In his view, minimal editions should target a broader audience, while maximal editions with a far larger number of scholarly bells and whistles should provide for the needs of researchers. Several digital scholarly editions do show signs of this sort of differentiation. We can point to the *Van Gogh Letters* (Jansen *et al.*, 2009) as something of a midpoint between the minimal and maximal edition. The Samuel Beckett Digital Manuscript Project (Van Hulle and Nixon, 2011) and the pre-production version of the Digital Faust Edition (Brüning *et al.*, 2013) that I have been allowed to see certainly should qualify as maximal editions. However, these and virtually all digital scholarly editions again reiterate in the GUI metaphors of the ‘read-only’ book.

No digital scholarly editions do provide what I think is paramount for true interaction with editions or scholarly text resources: the capacity to negotiate the edition and its text as data over Web serviced Application Programming Interfaces (APIs). APIs allow for computer-to-computer negotiation of texts, opening them up to algorithmic processing and reuse. My primary reason for arguing that we need our digital scholarly editions as API accessible texts is not, as some may expect, to enable quantified computational approaches such as those that Matthew Jockers and Franco Moretti have taken (Jockers, 2013; Moretti, 2007), or the stylometric analysis desired by many others (van Dalen-Oskam and van Zundert, 2007; Kestemont, 2012). It is highly useful and convenient to have the text of scholarly editions available as open Web service, so that my computational colleagues and I can do our principal component analyses, bootstrap trees, clustering analyses, and any other analysis that can possibly be envisioned.

There is another reason, in my view more important yet overlooked, to consider anchoring digital scholarly editions on a data model that is not oriented around a book metaphor. This motivation derives from the growing and increasingly unsettling gap I find between the close reading of scholars using conventional hermeneutic approaches and

the ‘big data’ driven distant reading supported by probabilistic approaches—a discrepancy which is also signaled by others (Capurro, 2010). On the one hand, we see a conventional scholarly approach in which texts are mindfully and meticulously produced, detailed, and interpreted. On the other hand, we find a deterministic and probabilistic approach whose focus is large-scale data analysis and which is, through its statistical aspect, reductive in nature. To the hermeneutic scholar, distant reading approaches are therefore ‘lossy’, prone to discarding some of the substance, and quite incapable of capturing essential hermeneutic knowledge (cf. Ramsay, 2011). It is often the statistical outliers and not just patterns of similarities that are telltale to textual scholars and historian in their hermeneutic explorations. At present there is no model connecting these worlds of close and distant reading. Rather, the distance between them is growing, which threatens not only to set the scholarly community of textual and literary studies against itself, but also to waste the opportunity for a true and meaningful advance in our capabilities for computational-based humanities research.

If we are to close this gap, we need a model for digital text that allows for both hermeneutic and statistical approaches so that these approaches can truly inform each other. To this end we need to revisit and reconsider how we anchor digital editions on the hypertext model. The slavish adherence to the book metaphor, even in XML form, will not take us into a realm where texts and editions are published as online APIs for processing by computational means. Yet, also models of quantification fall short as they are narrowly defined for statistical methodology. Because such models are not data models, they do nothing as to expressing description, encoding, or annotation. We are in need of a model that actually provides for all of the above. That is, a model that provides for the capturing, encoding, and annotating of a text and also for processing the edited or raw resource to enable analyses by both conventional hermeneutics and quantified approaches. Lastly, this model must be recursive: it must be able to capture all resulting information from an analysis and add that information into the model itself. Only then new knowledge gained

from the model can be used ‘natively’ for a next cycle of both qualitative and quantitative analysis. Such a model captures all editorial and research aspects and outputs of scholarly activity in an encompassing lifecycle. But even more important: only such a model provides for a way to bridge the widening gap that is coming into existence between the hermeneutic tradition and new quantified means. Computational method can do far more than just counting, averaging, and comparing histograms. But currently computational approaches ignore many of the properties of text and textual materiality that are important to hermeneutic engagement. Current quantified approaches lack therefore the ability to model and computationally process the close reading aspects of text engagement.

Thus what we lack is something we could call tongue in cheek near distant or near close reading. More formally and in line with current debate, I think we should qualify what we lack as an enabler of computational heuristics for *capta* (Drucker, 2011). But arguably either ‘near close reading’ or ‘near distant reading’ both capture in their own ambiguity exactly the properties of textual scholarly data and knowledge that quantified approaches tend to overlook: extremity of sparseness, inconsistency, vagueness, ambiguity, multi-interpretability, and uncertainty. There is no readily available means for such qualitative computing. Qualitative modeling and computing are still highly explorative fields (cf. Forbus, 2008), and yet, abilities to compute and reason over qualitative data are coming into existence. As the creators and providers of the raw materials that such qualitative computational approaches should operate on, editors of digital scholarly editions should consider how text as data is to be provided.

Knowledge graphs are, I think, extremely well suited for this. Graphs are not new to us, nor to our field. The World Wide Web is a graph, a network of nodes and edges connecting information. In a sense, every digital scholarly edition put online has in fact been made part of a graph therefore. In recent years, graphs have found various more explicit applications also in the field of digital humanities, most notably as a data model for describing textual variation between different witnesses of the

same text (Schmidt and Colomb, 2009). The properties of the graph model, however, allow it to be a generic model capturing the information tied to a digital scholarly edition on all conceivable levels of granularity. Two examples may show this potential conceptually. Imagine a knowledge graph as a network with nodes and edges. In this hypothetical graph, we designate three nodes to represent texts A, B, and C. An interface to the graph allows us to add edges and nodes to this network. What is essential here is that the underlying model is a graph, the graphical display may take many forms but need not necessarily be a visual network itself. Suppose now a textual scholar X states that text A was conceived before text C. This statement can be represented as a directed relational edge (or predicate if you like) ‘precedes’ between A and C as depicted in Fig. 1.

Now assume another researcher Y at another point in time, and not necessarily even knowing anything about text A, independently of researcher X, concludes that text B was conceived after text C. This statement can be captured by putting an edge ‘precedes’ between C and B. The tiny graph as depicted in Fig. 2 now holds the accumulated knowledge. However, note that the combination of independent observations now adds up to more than just the sum of its parts, for reasoning, walking, or computing over the graph—all three verbs essentially express the same operation of inferring knowledge from the graph—gives us the added knowledge that A must have preceded B.⁴

The second example is taken from CollateX, which is a tool to automatically collate variant texts (cf. <http://collatex.net/>). The result of such comparisons can be stored as graphs, e.g. Fig. 3. Such graphs cannot be said to be quantified, they express rather the qualitative word variance between

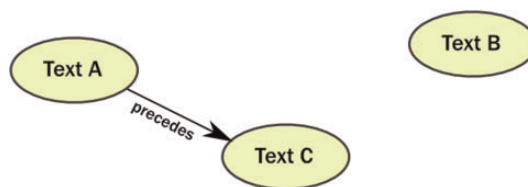


Fig. 1. Nodes in a conceptual knowledge graph

texts. But the application of the graph stretches wider. As in the previous example, we can add statements (knowledge) about this text to the graph by adding nodes and edges. The example in Fig. 4 shows two statements made by superseding nodes on partly overlapping regions of the text. They ex-

press in a hypothetical fashion how these regions should look for a reader of an EPUB serialization of the text to be read on an eReader. Note how overlap, a well-discussed problem for hierarchical models (Sperberg-McQueen, 2002), is not relevant to such a non-two-dimensional graph model.

It should be carefully pointed out that knowledge graphs as a model are not to be equated with the currently popular ideas on semantic Web and RDF. RDF can necessarily only be a static representation of a certain state of such a graph.⁵ The relation between RDF/Semantic Web and graph models is analogous to the relation between TEI and XML. A TEI conformant XML document is a singular instantiation of (a part of) the TEI model. The TEI model itself however is represented by the dynamic

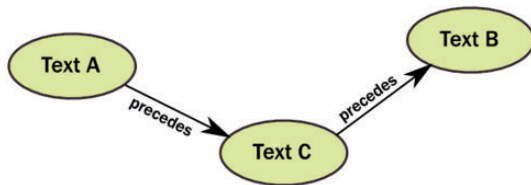


Fig. 2. Edges multiply knowledge

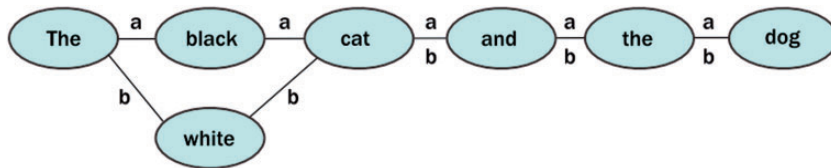


Fig. 3. Conceptual knowledge graph representing textual variation in two texts a and b

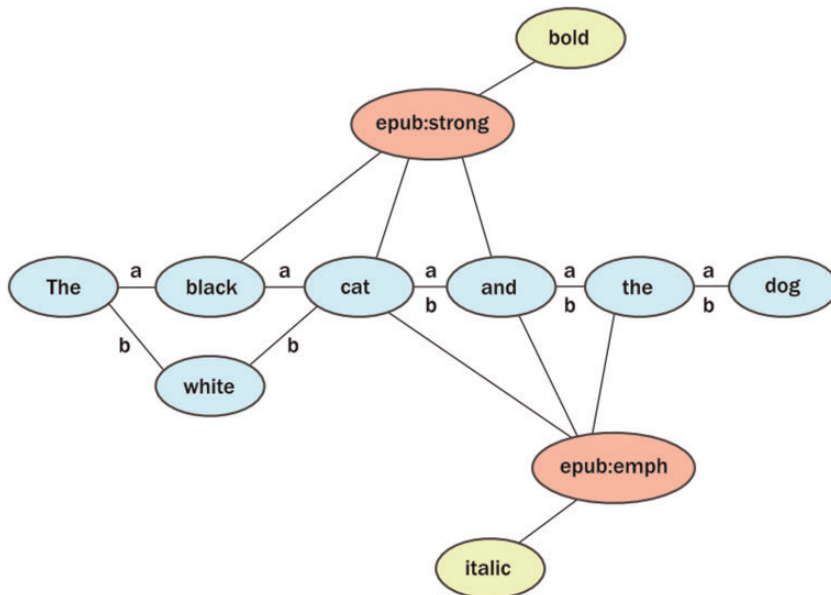


Fig. 4. Overlapping semantic and representational knowledge added to the graph of figure 3

- Takševa, T. (ed.), *Social Software and the Evolution of User Expertise: Future Trends in Knowledge Creation and Dissemination*. Hershey: IGI Global, pp. 112–129.
- Bijker, W., Hughes, T., and Pinch, T.** (eds.) (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.
- Bod, R.** (2013). Het Einde van de Geesteswetenschappen 1.0. <http://staff.science.uva.nl/~rens/OratieRens.pdf> (accessed 26 February 2013).
- Boot, P. and van Zundert, J.** (2011). The digital edition 2.0 and the digital library: services, not resources. *Bibliothek und Wissenschaft*, 44: 141–52.
- Borgman, C.** (2009). The digital future is now: a call to action for the humanities. *Digital Humanities Quarterly*, 3(4). www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html.
- Bridenbaugh, C.** (1963). The Great Mutation. *American Historical Review*, 68(2): 315–31.
- Brumfield, B., Klevan, D. and Vershow, B.** (2012). *Sharing Public History Work Using Crowdsourcing of both Data and Sources*. <http://www.ims.gov/about/webwise.aspx> (accessed 6 June 2013). Also: Brumfield, B. (2012). Crowdsourcing at IMLS WebWise 2012. Collaborative Manuscript Transcription. <http://manuscripttranscription.blogspot.nl/2012/03/crowdsourcing-at-ims-webwise-2012.html> (accessed 6 June 2013).
- Brüning, G., Henzel, K., and Pravida, D.** (2013). Multiple encoding in genetic editions: the case of “Faust”. *Journal of the Text Encoding Initiative*, 4. <http://jtei.revues.org/697>.
- Buzzetti, D. and Rehbein, M.** (1998). Textual Fluidity and Digital Editions. In Dobрева, M. (ed.), *Text Variety in the Witnesses of Medieval Texts*. Sofia: Institute of Mathematics and Informatics, pp. 14–39. <http://www.denkstaette.de/files/Buzzetti-Rehbein.pdf>.
- Buzzetti, D.** (2009). Digital Editions and Text Processing. In Deegan, M. and Sutherland, K. (eds), *Text Editing, Print and the Digital World*. Farnham/Burlington: Ashgate, pp. 45–61. http://www.academia.edu/391823/Digital_Editions_and_Text_Processing (accessed 27 September 2013).
- Causser, T., Tonra, J., and Wallace, V.** (2012). Transcription maximized; expense minimized? Crowdsourcing and editing the collected works of Jeremy Bentham. *Literary and Linguistic Computing*, 27(2): 119–37.
- Cain Miller, C. and Bosman, J.** (2011). E-Books Outsell Print Books at Amazon. *New York Times*. http://www.nytimes.com/2011/05/20/technology/20amazon.html?_r=0 (accessed 25 September 2013).
- Capurro, R.** (2010). Digital hermeneutics: an outline. *AI & Society*, 35(1): 35–42.
- Courant, P. N. et al.** (2006). *Our Cultural Commonwealth: The report of the American Council of Learned Societies’ Commission on Cyberinfrastructure for Humanities and Social Sciences*. New York: American Council of Learned Societies.
- Drucker, J.** (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1). <http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (accessed 24 August 2012).
- Fish, S.** (2011). The Old Order Changeth. *New York Times*. <http://opinionator.blogs.nytimes.com/2011/12/26/the-old-order-changeth/> (accessed 26 February 2013).
- Forbus, K. D.** (2008). Qualitative Modeling. In van Harmelen, F., Lifschitz, V., and Porter, B. (eds), *Handbook of Knowledge Representation. Foundations of Artificial Intelligence*. Amsterdam, Boston, Heidelberg etc.: Elsevier, pp. 361–94.
- Goldfarb, C. F.** (1996). *The Roots of SGML—A Personal Recollection*. <http://www.sgmlsource.com/history/roots.htm> (accessed 25 September 2013).
- Greif, I.** (ed.) (1988). *Computer-Supported Cooperative Work: A Book of Readings*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Haslhofer, B. et al.** (2011). The Open Annotation Collaboration (OAC) Model. *Arxiv*. <http://arxiv.org/abs/1106.5178> (accessed 25 September 2013).
- Hockey, S.** (2004). The History of Humanities Computing. In Schreibman, S., Ray, S., and Unsworth, J. (eds), *A Companion to Digital Humanities*. Oxford: Blackwell. <http://www.digitalhumanities.org/companion/> (accessed 25 September 2013).
- Jansen, L., Luijten, H., and Bakker, N.** (eds) (2009). *Vincent van Gogh: The Letters*. Amsterdam: Amsterdam University Press. <http://www.vangoghletters.org/> (accessed 25 September 2013).
- Jockers, M. L.** (2013). *Macroanalysis: Digital Methods and Literary History*. Urbana, Chicago, Springfield: UI Press.
- Jones, S., Shillingsburg, P., and Thiruvathukal, G.** (2010). E-Carrel: an environment for collaborative textual scholarship. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(2). <https://letterpress.uchicago.edu/index.php/jdhcs/article/view/54/65> (accessed 25 September 2013).

- Kestemont, M.** (2012). *Het gewicht van de auteur. Een onderzoek naar stylometrische auteursherkenning in de Middelnederlandse epiek*. Universiteit Antwerpen, Faculteit Letteren en Wijsbegeerte, Departementen Taal- en Letterkunde.
- Landow, G. P.** (2006). *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization*. Rev. ed. of Hypertext 2.0 1997. Baltimore: The John Hopkins University Press.
- Latour, B.** (1988). *Science in Action: How to Follow Scientists and Engineers Through Society*, Cambridge, MA: Harvard University Press.
- Leonardi, P. M.** (2011). When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. *MIS Quarterly*, 35(1): 147–67.
- Lunenfeld, P. et al.** (2012). *Digital Humanities*. Cambridge, MA/London: MIT Press. <http://mitpress.mit.edu/books/digitalhumanities-0> (accessed 27 November 2012).
- Marinova, D. and Phillimore, J.** (2003). Models of Innovation. In Shavinina, L. V. (ed.), *The International Handbook on Innovation*. Kidlington, Oxford: Elsevier Science Ltd., pp. 44–53.
- McGann, J.** (2010). Electronic archives and critical editing. *Literature Compass*, 7: 37–42.
- Mitra, A.** (1999). Characteristics of the WWW text: tracing discursive strategies. *Journal of Computer-Mediated Communication*, 5(1). <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1999.tb00330.x/full> (accessed 3 October 2013).
- Moretti, F.** (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
- Nelson, T. H.** (1995). The heart of connection: hypermedia unified by transclusion. *Communications of the ACM*, 38(8): 31–3.
- Nickles, T.** (2003). Evolutionary Models of Innovation and the Meno Problem. In Shavinina, L. V. (ed.), *The International Handbook on Innovation*. Kidlington, Oxford: Elsevier Science Ltd., pp. 54–78.
- Nowviskie, B.** (2012). too small to fail. Bethany Nowviskie. <http://nowviskie.org/2012/too-small-to-fail/> (accessed 3 October 2013).
- Porter, D.** (2013). Medievalists and the scholarly digital edition. *Scholarly Editing: The Annual of the Association for Documentary Editing*, 34. <http://www.scholarlyediting.org/2013/essays/essay.porter.html> (accessed 13 March 2013).
- Rayward, W. B.** (1994). Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. *JASIS*, 45, pp. 235–250.
- Ramsay, S.** (2011). *Reading Machines: Toward an Algorithmic Criticism (Topics in the Digital Humanities)*. Chicago: University of Illinois Press.
- Renear, A.** (2004). Text Encoding. In Schreibman, S., Ray, S., & and Unsworth, J., eds. *A Companion to Digital Humanities*. Oxford: Blackwell. <http://www.digitalhumanities.org/companion/> (accessed 25 September 2013).
- Robinson, P.** (2004). *Where We Are with Electronic Scholarly Editions, and Where We Want to Be*. <http://computerphilologie.uni-muenchen.de/jg03/robinson.html> (accessed 25 September 2013).
- Robinson, P.** (2013). Five desiderata for scholarly editions in digital form. In *Digital Humanities Conference 2013*. Lincoln, NB. <http://dh2013.unl.edu/abstracts/ab-314.html> (accessed 25 September 2013).
- Rogers, E.M.** (1983). *Diffusion of Innovations*, 2nd edn. New York, London: The Free Press.
- Sahle, P.** (2013). *Digitale Editionsformen, Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels – Befunde, Theorie und Methodik*. Norderstedt: Norderstedt Books on Demand.
- Schmidt, D. and Colomb, R.** (2009). A data structure for representing multi-version texts online. *International Journal of Human-Computer Studies*, 67(6): 497–514.
- Shaw, R., Buckland, M. and Golden, P.** (2013). Open Notebook Humanities: Promise and Problems. In *Digital Humanities Conference 2013*. Lincoln, NB. http://www.academia.edu/4030201/Open_Notebook_Humanities_Promise_and_Problems (accessed 25 September 2013).
- Siemens, R. et al.** (2012). Toward modeling the social edition: an approach to understanding the electronic scholarly edition in the context of new and emerging social media. *Literary and Linguistic Computing*, 27(4): 445–61.
- Sperberg-McQueen, C. M.** (2002). What matters? In *Proceedings of Extreme Markup Languages*. Extreme Markup Languages. Montréal, Canada. <http://conferences.idealliance.org/extreme/html/2002/CMSMcQ02/EML2002CMSMcQ02.html> (accessed 25 September 2013).
- Van Hulle D.** (2010). *Editing Samuel Beckett*. jnul.huji.ac.il/eng/docs/Israel_Interedition_NLI_2010.pdf (accessed 25 September 2013).
- Van Hulle, D. and Nixon, M.** (2011). *Samuel Beckett Digital Manuscript Project. Samuel Beckett Digital*

Manuscript Project. <http://www.beckettarchive.org/> (accessed 25 September 2013).

Van Mierlo, W. (2007). Textual scholarship and the material book. *Variants: The Journal of the European Society for Textual Scholarship*, 6. http://www.academia.edu/209627/Introduction_to_Textual_Scholarship_and_the_Material_Book (accessed 25 September 2013).

van Dalen-Oskam, K. and van Zundert, J. (2007). Delta for middle Dutch: author and copyist distinction in “Walewein.” *Literary and Linguistic Computing*, 22: 345–62.

Vanhoutte, E. (2011). *So You Think You Can Edit? The Masterchef Edition*. http://edwardvanhoutte.blogspot.nl/2011_10_01_archive.html (accessed 25 September 2013).

Winkler, M. (2013). Interpretatie en/of patroon? Over ‘Het einde van de geesteswetenschappen 1.0’ en het onderscheid tussen kritiek en wetenschap. *Vooy*, 30(1): 31–41.

Notes

1 A recent example in the Dutch literary and linguistics theatre is professor Rens Bod proclaiming the end of Humanities 1.0 (Bod, 2013), and Ph.D. student Marieke Winkler sincerely questioning that (Winkler, 2013).

2 As in many other contexts (cf. Nowwiskie, 2012), the relationship between an IT R&D group and scientific staff is some matter of internal debate in the institute. In part, this role is supporting; in part, it is collaborative at the research level.

3 There is likely a distinction to be made here between senior scholars as transcribers and non-academic volunteer ‘crowd sources’. Although I lack any statistical viable data, anecdotal evidence suggests that volunteer transcribers in fact may attach hundreds of tiny and similar annotations without complaint, but the senior researcher will feel put at odds with his experience and practice when invited to do so.

4 I kindly thank Moritz Wissenbach from Würzburg University—who is among other occupations the technical lead for the development of the digital Faust edition—for allowing me to share this example which he originally conceived.

5 Initiatives such as the Open Annotation Collaboration are proposing extensions to the World Wide Web and Semantic Web models to support annotation of linked data including temporal ‘aware’ annotations (Haslhofer et al., 2011). It is out of scope of this article to examine whether such models would provide for the needed reciprocity and dynamics for graph model-based digital scholarly editions. As the Web in its current form is not real-time read/write enabled, it is hard to imagine though how it would provide for such highly dynamic webs of knowledge interaction.