RESEARCH ARTICLE

On edited archives and archived editions

Wout Dillen [1]

Abstract
Building on a longstanding terminological discussion in the field of textual scholarship, this essay explores the archival and editorial potential of the digital scholarly edition. Following Van Hulle and Eggert, the author argues that in the digital medium these traditionally distinct activities now find the space they need to complement and reinforce one another. By critically examining some of the early and more recent theorists and adaptors of this relatively new medium, the essay aims to shed a clearer light on some of its strengths and pitfalls. To conclude, the essay takes the discussion further by offering a broader reflection on the difficulties of providing a 'definitive' archival base transcription of especially handwritten materials, questioning if this should be something to aspire to for the edition in the first place.

Keywords Digital scholarly editing · Textual criticism · Archives · Editions

For the scholarly edition, the move to the digital medium seems to have initiated something of an identity crisis. More than 30 years after the Text Encoding Initiative was established—a milestone in the development of digital text editing—the digital scholarly edition is still regularly being defined and redefined by the scholarly editing community. What is a digital scholarly edition? How is it different from a printed edition? Is it better? What are the things we can do now (or do better) that we could not do (as well) before? This is, of course, a healthy critical reflex that can be applauded in any scholarly discipline. And with considerably less time to adjust to the new medium than its print predecessor, it is only natural that we do not have all the answers yet. Acknowledging that this evolution—if not exactly a 'revolution' (Robinson 2016)—is still ongoing and that 'the possibilities of digital technologies are in constant flux,' *Advances in Digital Scholarly Editing*—a collection of extended abstracts of the DiXiT conferences in The Hague, Cologne, and Antwerp—avoids the difficulty of trying to define what is essentially still under development by giving the floor to 'a broad selection of the community of scholars' that is shaping it instead (Boot et al. 2017: 15).

✉ Wout Dillen
wout.dillen@uantwerpen.be

[1] Centre for Manuscript Genetics, University of Antwerp, Antwerp, Belgium

Still, a look at the theoretical scholarship produced in the field allows us to discern crucial issues with which the field is struggling. One of the major themes that is returned to in these discussions on the status of the digital scholarly edition is one of terminology, and how editions self-identify. Already early on in the field's move to the digital medium it became clear that the virtual nature of the scholarly edition's new publication environment would in itself already constitute a key advantage over the print medium. Not only does the digital medium make its contents more easily adaptable and transformable (even inherently transform*ative* as we never really read the binary code the edition is written in),[1] its carrier is also much more compact, allowing the editor to widen the scope of her project significantly. Rather than constructing one text and dismissing variant readings to an apparatus, the editor can now offer high resolution images and detailed transcriptions of each relevant document in the corpus, and use that information to automatically construct (or even let the user construct) a virtually infinite number of edited texts. This way, the digital approach to scholarly editing offers the scholarly editor new possibilities for exploiting both the editorial and archival aspects of her edition. After a short introduction to the theoretical and terminological difficulties such a consolidate archival-editorial endeavour brings, this essay will compare a series of different approaches to realising this new potential in the digital scholarly edition in order to pinpoint some of its major opportunities and pitfalls.

1 What's in a name?

Over two decades ago, Peter Robinson already acknowledged that the encyclopaedic tendency of the digital scholarly edition ostensibly calls into question the essence of the editorial endeavour, as it absolves the editor of the task of selecting (and editing) one text over all others:

> By giving at least seven different views onto the fifty-eight manuscripts of the *Canterbury Tales*, we may seem to be leaving the reader drowning in a sea of variation: as if we were saying, there is no one text; there are just all these variants (1996: 110).

A few years later, Mats Dahlström would develop this idea further in a paper titled 'Drowning by Versions,' where he argued that he argued that a digital scholarly edition 'is intended to fulfil two perhaps contradictory user demands:' (1) 'the broadest possible presentation of the textual material, and (2) 'guiding [the user] through the textual mass in such a way that [she] can benefit from the editor's insights and competent judgement' (2000: §4).

Neither Robinson nor Dahlström would do away with the editor in the digital medium: the former emphasises that especially in the case of handwritten materials the transcriptions themselves are heavily edited; the latter acknowledges the editorial

---

[1] As Patrick Sahle posited the second part of his *Digitale Editionsformen*: 'Das Kennzeichen des gegenwärtigen Medienwandels ist nicht so sehr ein Wechsel des Medien, sondern vielmehr ein Transmedialisierung!' (2013: 161; see also 162).

aspect of designing the digital publication environment in such a way that it becomes comprehensible for the average user. But they both suggest that her role has drastically changed—and so we may ask ourselves: if the scholarly editor is no longer editing texts in the traditional sense of the word, can we still call the fruits of her labour scholarly editions? Or should we try to find a new, and more suitable denomination for it?[2] If our main purpose is to make a series of related documents and their transcriptions available for further research, would not the term 'archive' be a better alternative?

This idea goes back to the classic dichotomy of archive versus edition, which refers to the archive as a collection of historical documents, and the edition as an argument about those materials (see Price 2007: 345; Sahle 2005: §1–2). But while this dichotomy may have held true in the print era, the current consensus among textual scholars seems to be that the distinction is much more difficult to make in the digital age (Evenson 1999; Van Hulle 1999: §3; Sahle 2005, §9.5–9.6; Price 2007 and 2009; and Dahlström 2009 are but a few examples).

On the one hand, the large-scale digitisation projects that local and national libraries and archives are undertaking are not just highly qualitative, but also rapidly becoming more and more enhanced and contextualised for and in collaboration with researchers. Think for example of *Litteraturbanken,* the 'Swedish Literature Bank (*Litteraturbanken* 2018)'. In an impressive collaborative effort between literary and linguistic scholars, research libraries, and editorial societies and academies, this project contains wide range of digital facsimiles and their (corrected OCR based) transcriptions of documents pertaining to Swedish literary works from the middle ages to the present. Alongside their edited texts available in HTML (and, when possible, EPUB), these are contextualised further by means of scholarly introductions, presentations, other didactic materials, and even allow for basic text analysis functionalities through a collaboration with *Språkbanken*, the 'Swedish Language Bank' (see Dahlström and Dillen 2017). Acknowledging the critical effort that went into the creation of these kinds of resources, and addressing the need to hold them up to the same standards as digital scholarly editions, RIDE (the IDE's Review Journal for Digital Editions and Resources) even devoted two whole recent issues to so-called Digital Text Collections (see Henny-Krahmer and Neuber 2017; Neuber and Henny-Krahmer 2018). And indeed, in many aspects these kinds of institutions are leading the way when it comes to resource creation, and could teach scholarly editors much about dissemination, discoverability, sustainability, addressing copyright concerns, digitisation techniques and affordances, etc.

On the other hand, projects that are generally considered as digital scholarly editions often do not shy away from calling themselves archives either–think, for instance, of the *William Blake Archive*, the *Piers Plowman Electronic Archive*, the *Walt Whitman Archive*, and, more recently, the *Shelley-Godwin Archive*. As Kenneth Price argues in 'Electronic Scholarly Editions,' the fact that these projects of textual scholarship identify both with 'editions' and with 'archives' does not have to be a contradiction:

> Words take on new meanings over time, of course, and *archive* in a digital context has come to suggest something that blends features of editing and

---

[2] In 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?' Price weighed a series of alternatives against one another and makes a case for switching to the concept of 'arsenal' instead (2009).

archiving. To meld features of both — to have the care of treatment and annotation of an edition and the inclusiveness of an archive — is one of the tendencies of recent work in electronic editing (2007: 345).

A good example to demonstrate the complexity of this issue is that of the *Beckett Digital Manuscript Project* (BDMP) (2018). As its title suggests, this editorial collaboration between the Universities of Antwerp and Reading and a whole range of holding libraries across the (Western) world calls itself a 'project'. But when we visit the project's website, its URL indicates that the project is in fact also an 'archive' (www.beckettarchive.org). Still, the project's 'Series Preface'[3] explains that this digital archive is only part of the project's scope. Instead, the project consists of two parts:

(a)   a digital archive of Samuel Beckett's manuscripts, organized in 26 research modules. Each of these modules comprises digital facsimiles and transcriptions of all the extant manuscripts pertaining to an individual text, or in the case of shorter texts, a group of texts.
(b)   a series of 26 volumes, analyzing the genesis of the texts contained in the corresponding modules.

This means that the project is not just a *digital* scholarly edition, but that it is actually a *hybrid* scholarly edition: an edition with a printed part and a digital part, where each part tries to take full advantage of its medium's possibilities (Sahle 2013: 62). Furthermore, this preface introduces yet another term to our overview: that of a module. The BDMP uses this term to refer to its modular publication process, where each publication is a new module, which contains the edited texts (digital) and genetic analysis (print) of one or more related texts in Beckett's oeuvre. This means (and this is corroborated in the navigation bar on the project's homepage) that a single module can contain multiple 'genetic editions' of individual works. And finally, to make things even more complicated, the project also contains the Beckett Digital Library, which aims to reconstruct Samuel Beckett's personal library in order to expose possible links between Beckett's reading habits and his writing process. This digital environment is not counted as one of the project's modules, but it is a fully integrated aspect of the BDMP. Combining all this information, we could represent the BDMP's structure as follows (Fig. 1):

This complex structure illustrates how terms start to overlap and change meaning as Price suggested. The project's online presence is called an archive, but the whole is often referred to as a digital scholarly edition. In combination with the accompanying monograph series, it can be called a hybrid edition, but the archive refers to its sub-modular collections of manuscripts and transcriptions as individual genetic editions too. Through the project's modular structure, there is no one-on-one relation between these work-oriented genetic editions and the accompanying monographs either. And then we need to take into account that not every part of the archive belongs to a specific module: the Beckett Digital Library (which has its own accompanying monograph; Nixon and

---

[3] See: http://www.beckettarchive.org/introduction.jsp. Note the use of the word 'series' here, another term to add to the list – and one that is again perhaps more firmly rooted in print culture.
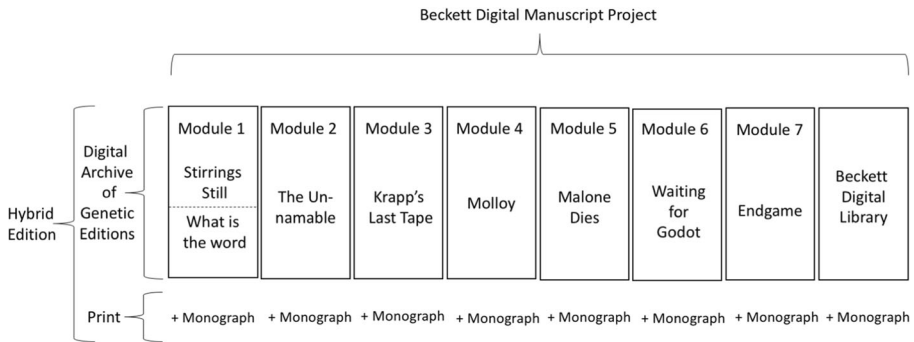
Fig. 1 Schematic representation of the BDMP's structure

Van Hulle 2017) is not labelled as a research module, while it could easily be regarded as an exogenetic *edition* of Beckett's library in itself.

2 Archival and editorial impulses

As the digital medium started to break down the borders between archives and editions, a new approach to the problem started to emerge. Rather than treating the two concepts as a strict dichotomy, Dirk Van Hulle suggested to think of them as two poles on a continuum instead, where the user can decide how to use the digital resource: as an archive of textual documents and image reproductions; as a (genetic) dossier that organises these documents and exposes their internal logic; or as an edition, a curated and edited collection of texts that informs the reader on the textual tradition of the work (2009: 177). To accomplish this, however, the resource needs to be developed in a way that combines features of both, and that is versatile enough to be queried in multiple contexts. That this is a much more productive way to think about the resources we produce can be argued by the fact that it was echoed in Paul Eggert's keynote address to the 2016 edition of the European Society for Textual Scholarship's annual conference that was held in Antwerp—where he distinguished between editorial and archival *impulses*:

> The archival impulse aims to satisfy the shared need for a reliable record of the documentary evidence; the editorial impulse to further interpret it, with the aim of reorienting it towards known or envisaged audiences and by taking their anticipated needs into account. Another way of putting this is to say that every expression of the archival impulse is to some extent editorial, and that every expression of the editorial impulse is to some extent archival. Their difference lies in the fact that they situate themselves at different positions on the slider (2017: 122).

As the archival and the editorial are inherently linked in the digital scholarly edition (and arguably in all of textual scholarship), the word 'impulse' really becomes the key to Paul Eggert's coinage. Rather than talking in absolutes, Eggert's phrase allows us to distinguish between the driving forces behind certain aspects of the digital scholarly

edition. A diplomatic transcription, for instance, is mainly driven by an archival impulse intent on preserving as many aspects of the original document as possible; but no textual scholar will deny that the making of such a diplomatic transcription involves many editorial decisions. On the other hand, the selection of a copy-text for instance, is mainly driven by an editorial impulse intent on the construction of a scholarly edited reading text; but no textual scholar will deny that this selection (as well as the construction of the reading text itself) involves a lot of archival work.

3 Two impulses, one edition?

In the last decade, we have witnessed the rise of a number of projects that tried to unleash the full potential of the digital scholarly edition, precisely by exploiting both its archival and its editorial impulses to the fullest. In what follows, I will discuss some of the different approaches that have been taken towards solving this problem.

3.1 Nietzsche Source, HyperNietzsche

The most straightforward way of dealing with both impulses, is to split the edition up, and dedicate separate parts of the edition to each of the impulses. An example of this approach is *Nietzsche Source* (2018), edited by Paulo D'Iorio, which currently hosts two individual editions: the *Digitale Kritische Gesamtausgabe Werke und Briefe* (eKGWB), which is described as a 'Digital version of the German critical edition of the complete works of Nietzsche edited by Giorgio Colli and Mazzino Montinari;' and the *Digitale Facsimile Gesamtausgabe* (DFGA), which is described as a 'Facsimile reproduction of the entire Nietzsche archive.' In this constellation, the DFGA exploits the archival impulse, by hosting digital facsimile images of all the documents in the Nietzsche estate (or at least: those that have already been scanned and uploaded). The eKGWB, on the other hand, exploits the editorial impulse, by hosting a critical edition of Nietzsche's works, in which *Textfehler* have been corrected, and editorial emendations are highlighted and justified. On *Nietzsche Source*, these two editions are rigorously separated: the eKGWB offers no image reproductions of the original documents, and the DFGA offers no transcriptions of the facsimile images. This appears to be a conscious decision on the part of the editors, as the edition's direct predecessor *HyperNietzsche* offered images, HMNL transcriptions, linear transcriptions, and diplomatic transcriptions, all easily accessible through the edition's interface.

As the *Nietzsche Source* project demonstrates, however, this approach of rigidly separating the editorial from the archival impulse can hardly be said to unlock the edition's full potential. In his review of the project for *RIDE*, Phillipp Steinkrüger remarked that this practice came at the cost of several functionalities that could be useful for textual scholars (2014: §9). The biggest drawback of the edition in its present form, is that it offers no way of comparing the edited text with the archived documents, other than opening each edition in a different browser window, and looking up a specific passage in both editions. Because the project is still under construction, however, we may hope that these aspects of the edition will be addressed (and *HyperNietzsche*'s original functionalities restored), once all of the estate's documents are put online.

3.2 The Faust Edition

A similar approach, and one that may help to bypass some of the drawbacks of the *Nietzsche Source* project, would be to split the digital scholarly edition up into two different *transcriptions*, rather than in two different *editions*. This is the approach of the *Faust Edition*, edited by Anne Bohnenkamp, Silke Henke and Fotis Jannidis. This project's parallel transcription method as described by Brünning et al. (2013) splits the archive's transcriptions up in a document-oriented transcription (which satisfies the archival impulse), and a work-oriented transcription (which satisfies the editorial impulse). When the edition will lose its beta-status later this year, both transcriptions will also be made available for further research.[4] However, while this approach certainly solves some of the problems the *Nietzsche Source* project faces, it creates new ones in the process—a rather important one being that it effectively doubles the project's transcription load.

3.3 The 'work-site'

Another attempt to satisfy both archival and editorial impulses in the same digital scholarly edition, is by means of standoff markup. This is the path that for example Paul Eggert and Peter Shillingsburg have taken. The basic idea behind this approach is simple: construct an archive of transcriptions that everyone can agree on by using as little markup as possible, and let scholarly editors edit their own text on the basis of that transcription, by building a personal layer of standoff markup on top of that transcription. That way, the archival impulse results in a permanent archive of searchable digitized documents, while the editorial impulse could result in a proliferation of scholarly editions, based on a variety of orientations to the text.

Paul Eggert's concept of the 'work-site' can be regarded as an early step in this direction. In 'Text-encoding, Theories of the Text, and the 'Work-Site', Eggert explained this concept as follows:

> The work-site is a text-construction site for the editor and expert reader; and it is the site of study of the work (of its finished textual versions and their annotation) for the first-time reader, as well as any position in between (Eggert 2005: 433).

To make this 'work-site' work, Eggert realised that there had to be some sort of common ground for editors and expert readers to construct their own edited texts from. In Eggert's proposal, this common ground is a transcription of the historical documents in so-called 'plain text'; text without markup.[5] Once this plain-text transcription is established, editors can augment and annotate this transcription through standoff markup using a technique called 'Just In Time Markup' (JITM). This markup is then stored in a JITM tagset that is kept separate from the original transcription. When users wish to read the text, they can apply any of these tagsets to the transcription on the fly

---

[4] Gerrit Brünning, one of the collaborators on the *Faust Edition* explained as much at a talk that he gave at the University of Antwerp as part of the Platform Digital Humanities Lecture Series (26 March 2018).

[5] More specifically, Eggert mentions the ISO-646 character set. This character set is a successor of ASCII (the American Standard Code of Information Interchange), and the predecessor of today's international standard character set called Unicode.

('just in time'), and even use them as the basis of their own augmented transcriptions of the documents. This allows the 'base transcription file to be annotated or augmented with analytical or structural markup, in parallel and continuously, while retaining its textual integrity' (431). Because the base transcription—like any transcription—would still be an interpretation of the original documents, Eggert recognises that the challenge of this approach is 'to establish the least objectionable form of the pre-existing text' (432)—an especially hard challenge in the case of handwritten documents, as will be discussed further below.

3.4 The 'knowledge-site'

This leads us to the chief difference between Paul Eggert's standoff markup technique and Peter Shillingsburg's. In 'Development Principles for Virtual Archives and Editions', Shillingsburg expands upon a set of principles he has helped develop for the HRIT project (Humanities Research Infrastructure & Tools). These are a broad set of principles that can be useful for 'the construction of tools and environments for developing, maintaining, and publishing scholarly textual archives / editions / commentary and pedagogical presentations' (Shillingsburg 2015: 11–12). As such, these principles could be applied to any project in those fields, but at the Center for Textual Studies and Digital Humanities (CTS&DH) at Loyola University Chicago, where HRIT is developed, they are applied to their own software development projects, such as those of the online 'Collaborative Tagging Tool' (CaTT). As Shillingsburg explains, this set of principles also specifically calls for a modular 'content management framework' that can offer 1) tools for creating archives of digital texts; 2) tools for augmenting the texts in these archives critically; and 3) an overarching architecture for displaying and navigating the resulting materials in what Shillingsburg likes to call a 'knowledge site'—a concept that shares many characteristics with Eggert's concept of the 'work-site'. More specifically, the CTS&DH has developed such a content management framework called 'Mojulem', and their first test case for this software is *Woolf Online*, a Digital Scholarly Edition of the works of Virginia Woolf.

Because modularity is such an important part of this content management framework, Shillingsburg proposes a separation of 'images', 'texts', and 'tags and commentary' (HRIT's third principle; 2015: 16), which implies a preference of standoff markup over embedded markup (HRIT's fourth principle; 2015, 23). The difference with Eggert's proposal, is that the archive of digital texts is not constructed of plain text files, but of text files that already have a minimal form of markup. This 'minimal embedded markup' would be a 'text-only' transcription of the documents to include

> everything about the sign system and its deployment that strikes the eye of the reader as having semantic force: e.g., letters, punctuation, diacritical marks, and other printed symbols, plus the meaningful deployment of white space through indentation, extra spacing, line breaks in poetry, plus certain kinds of typefont changes like the use of italics, bold, bold italics, special characters such as digraphs, and accented letters regardless of how many strokes on the keyboard it takes to render them digitally (Shillingsburg 2015: 17).

The idea behind this minimal embedded markup system is that it should be possible to tag a text's objective visual features first (the embedded markup, for instance 'italics'), and to critically analyse these features later (the standoff markup, for instance 'emphasis') – where the common practice in TEI–XML is to encode these two simultaneously (i.e. in a single element with varying attributes and attribute values, for instance <emph rend = "italics">). Shillingsburg's reason for embedding these features into the text rather than keeping them in a first, standard layer of standoff markup is not just based on a theoretical principle (as Shillingsburg believes these features to be an inherent part of the text), but also that it has the practical advantage of ensuring that they are included in the files when they are being compared by collation software and other tools. Furthermore, in order to fully exploit this feature, 'HRIT recommends the use of character-level markup' (26). By attributing markup to individual characters in the text, rather than to character ranges (as in XML; the 'range' being everything between the opening tag and the closing tag), collation tools can use this information to discover the smallest 'textual' differences between variant texts. In addition, this also solves the problem of overlapping hierarchies – which can be interpreted as a problem of conflicting character ranges. The disadvantage of this approach is that it makes the base text extremely difficult to read for humans (because individual characters will be artificially separated form one another to make room for their markup), and that it is not TEI-conformant (because it abandons the basic principles of XML). Shillingsburg does not see these problems as real drawbacks, however, because no humans would be required to read the base text, as readable (and perhaps TEI-conformant) text formats could be automatically generated from them.

What I think is a much larger problem, however, is that the content of the 'minimal embedded markup' tagset is not clearly defined and highly debatable. An example of this problem would be marginal additions in working drafts. Consider the following example of a supra linear addition:

> After writing two entire lines of text, the
> <div align="center">supralinear</div>
> author decided to make a̶n̶ |addition.

As a 'visual element with semantic force', this feature qualifies as one of Shillingburg's textual features, to be encoded in the base text.[6] At the same time, however, it is also clearly already the result of an interpretation of the text: someone who does not speak the language might just as well consider 'supralinear' to be an infralinear addition, rather than a supralinear one. Perhaps the least subjective way to encode this textual feature would be to call it an 'intralinear' addition. This example already shows that Shillingsburg's list of features is not well defined. But even if we come to a consensus and select a definitive list of textual features for the base text (something the TEI – with its 30 years of experience – has not been able to accomplish), we would still need to agree on where to place the addition in the text file. Do we use its documentary position (i.e. between 'the' and 'author'), or its textual position (i.e. between 'an' and 'addition')? Both options have their disadvantages: the documentary position would need to

---

[6] In fact, Shillingsburg's own list of these 'visual elements with semantic force' for manuscripts explicitly includes 'insertions above and below lines and in margins' (2015, 17).

introduce a line break where there isn't really one (after 'supralinear'), and it would break up the flow of the text – which makes it less useful for analysis. The textual position, on the other hand, introduces a high degree of interpretation, which is exactly what Shillingsburg's 'minimal embedded markup' strategy is trying to avoid.

Not only is HRIT's theoretical position towards this dilemma unclear, its practice at *Woolf Online* is highly contradictory. Take for instance this random page from Virginia Woolf's manuscripts for *To the Lighthouse* (Fig. 2):

When we use the website's slider tool to put the image's opacity to 0% and reveal the document's transcription, this is rendered as follows (Fig. 3):

While this is a useful visualization of the document's transcription, it loses much of the text's inherent structure. If we use the website's 'Export Options' to generate a TXT file from this page for that purpose, the first few lines of this transcription are these:

37
[strike]It[/strike] she
affected him strangely unaccustomed as he was to strangeness: moved him,
made him conscious of wishing to do something gallant, to pay some
little sum on her behalf; [strike]or[/strike] & yet puzzled him; for though she flattering
appeared so confident & even sparkling as she talked to him,
he could not help feeling [strike]doubtful[/strike] [strike]both[/strike] that this kindness to him
(& it was true;
Mrs. Ramsay
was very sorry
for this youth)
depended upon
in particular was [strike]a[/strike] an incident on the surface of some wide more
[
strike]general[/strike] & deeper capacious feeling, [strike]what[/strike] which for all her

From this exported transcript, we learn that the Mojulem software puts 'intralinear' additions in their documentary position (e.g. 'flattering' between
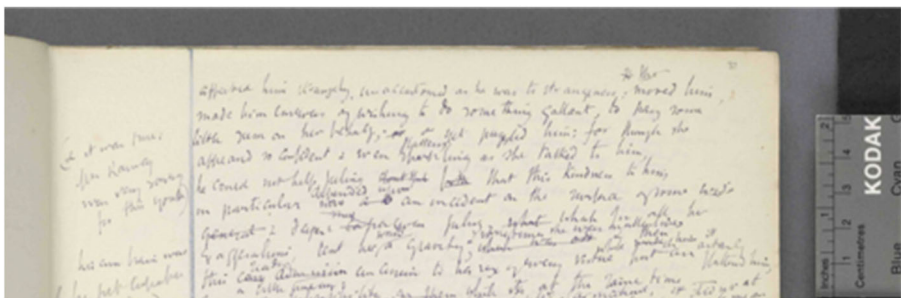

Fig. 2 Fol 37; SD. p.18 from Woolf's *To the Lighthouse*: Image View (Woolf 2013-)

Fig. 3 Fol 37; SD. p.18 from Woolf's *To the Lighthouse*: Transcription View (Woolf 2013-)

'she' and 'appeared'), while it puts marginal additions in their textual position (e.g. '(& It was true; Mrs. Ramsay was very sorry for this youth)' between 'him' and 'depended'). This inconsistency renders the file unusable, since it makes it impossible to arrive at a uniform, logical string of characters to analyse. Perhaps, in the future, the website will offer another export format that will make these transcriptions more useful for analysis, but at present that is not the case. Alternatively, because the textual reconstruction of the character sequences on these manuscripts requires a considerable interpretative effort, it may make sense to construct a (standoff) edition of the manuscript's text first, before sending it to a text analysis tool. But that would argue against the interoperability of the non-augmented base text – which was the whole reason for using standoff markup in the first place. These problems would need to be addressed before Mojulem can become a useful content management framework for storing and transcribing genetic materials.

4 The return of the definitive edition

One thing all these different approaches to digital scholarly editing seem to have in common, is the fact that they try to give their digital scholarly editions a certain permanence by exploiting their projects' archival and editorial impulses as much as possible. This brings us back to Shillingsburg's definition of the 1970s concept of the 'definitive edition', which held that 'a scholarly edition could represent definitive research which would never have to be done again, though the text itself might be re-edited according to differing principles' (1996: 174).

Although they are careful never to articulate it quite this way, Eggert's and Shillingsburg's standoff markup approach goes even further back, to the concept of the 'definitive text' as it was used in the 1960s. What are their proposed archives of 'plain text' or 'minimally marked up' transcriptions if not archives of fixed, *definitive* texts? In his 'Development Principles for Virtual Archives and Editions,' Shillingsburg explicitly argues that once a consensus has been reached that

all relevant texts have been transcribed and proofread and brought to the highest level of accuracy, and mapped onto their respective images, the door should be closed to further interference with the text files (2015: 24).[7]

By closing this door, the textual scholar effectively creates a finite archive of definitive transcriptions that can be used to create an infinite number of scholarly editions.

Of course, even though the word 'definitive' has received a negative connotation in the last half-century, editors will always strive for some sort of permanence and stability in their editions. This is the whole point behind the archival impulse in the first place: to find a long-term solution for the preservation (and dissemination) of historical documents. And at least these new 'definitive transcriptions' offer the possibility of being augmented to serve a variety of different orientations to texts, which was an important point of criticism of so-called 'definitive' texts and editions. The irony is, however, that this new type of definitive *transcription* still has the same problem the old definitive *text* from the 1960s had: the fact that everyone has a different idea of what it should look like. This is already clear in the differences between Eggert's and Shillingsburg's (otherwise quite similar) approaches. But these are just the symptoms of a much deeper problem that is rooted in the interpretative quality of the act of transcription itself.

In 'What Is Transcription?', Claus Huitfeldt and C.M. Sperberg-McQueen provide a useful definition of the concept of transcription that says that

one document (the transcription, T) is said to be the transcription of another document (the exemplar, E), if T was copied out from E with the intent, successfully achieved, of providing a faithful representation of a text as witnessed in E (2008: 296; emphasis in original).

As such, the act of 'copying T out of E' transfers a text from one document onto another, on a token per token basis. The tokens of E are not identical to those of T; if they were, T would be a copy, not a transcription. Instead, as Robinson and Solopova argued in their 'Guidelines', it transfers the text from one semiotic system into another, which always requires a certain degree of interpretation on the part of the scribe (1993: 21). This interpretation may not be apparent in the transcription of printed or digital texts, because these texts de facto use a standardised typesetting, where the exact shape of each character is predefined and stabilised. That is what makes printed texts such nice starting points for standoff markup projects: for these texts an a priori knowledge of the specific typesetting system is all that is required to come to an almost indisputable transcript of E's text (with or without Shillingsburg's minimal textual features).

For handwritten materials, however, this process quickly becomes more difficult because the shape of the tokens in E is highly unstable. Not only is handwriting personal (and prone to change over time), the actual shape of the individual tokens will usually depend on the token's context (e.g. neighboring tokens, writing tools, and

---

[7] In his paper, Shillingsburg foresees two exceptions to this rule: 'a new authoritative witness to the work or the discovery of error in the original work' (2015: 24). But the images that represent the document may need to be updated as well, if the edition wants to conform to newer and higher digital imaging standards. Such an update will invariably have a number of implications for the image-text linking tools that the content management framework uses, but it may also have consequences for the text, if the new image clarifies a textual feature the discovery of that the old image could not.

writing surfaces), and even the writer's mood. This is what prompted Peter Robinson to claim that '[a]n "i" is not an "i" because it is a stroke with a dot over it. An "i" is an "i" because we all agree that it is an "i"' (2009: 44). Although perhaps somewhat dogmatic, there is some truth to this statement: for handwritten documents, it is not the stability of the shape of individual tokens that helps us decipher a text, but rather a combination of a series of contextual features, such as our familiarity of the writer's handwriting, our knowledge of the text's place in the context of the work, and our knowledge of the language in which the text is written. Common additional difficulties for modern manuscripts in this respect are bad handwritings, the use of multiple textual layers in a single document, and the use of complex and inconsistent metamark systems. And in the case of traditional manuscripts, it may also be difficult to decide whether certain paleographical features are used to distinguish between different textual characters (what Shillingsburg would call 'a textual feature with semantic force'), or whether they should merely be interpreted to display the scribe's calligraphic artistry. Add to these concerns that it is not always possible to discern exactly how many individual characters a certain hard-to-read passage counts, or even where a specific character sequence belongs in relation to the rest of the text, and it becomes clear that transcriptions of handwritten documents often depend so much on interpretation that it becomes impossible to sustain that T contains a straightforward, objective, transcription of E, which can serve as an undisputed definitive base text for the work-site's or knowledge site's plethora of standoff markup editions.

To conclude, I would argue that that for many handwritten materials (and especially in the case of drafts and other genetic documents) locking the transcriptions down as such is a utopian endeavour. Instead, we could supply well-argumented but still debatable edited interpretations of our documents, that are still open to emendation—because this flexibility too is a great advantage of the digital medium. With the official launch of Eggert's *Charles Harpur Critical Archive*[8] on the way (which tellingly calls itself an 'archive' again), it will be interesting to see how the editorial team rose to the challenge of to establish the 'least objectionable' base transcription for their handwritten materials. Indeed, with the *Faust Edition* soon to be taken out of its beta status and Peter Robinson's *Textual Communities* project close to a public release, these are exciting times in the field of digital scholarly editing, where we will soon be able to see just how these different editing models succeed in satisfying both archival and editorial impulses in a digital environment.

Nevertheless, even as the field is still to develop fully, I think it is safe to say that this blending of the archival and the editorial will remain a key aspect of the digital scholarly edition in the near future. In a way, it creates an opportunity for the digital edition to surpass its print predecessor, in which the archival impulse was perhaps underrepresented due to the medium's technological and spatial constraints. By aggregating high quality reproductions of all the extant documents related to a specific work, the editor adds a new dimension to her digital scholarly edition, offering an archive of relevant source materials underneath the authoritative layer of edited text, while at the same time raising her accountability (and thus the trustworthiness of the edition) to a new level. Returning to the discussion of Robinson's and Dahlström's perspectives that

---

[8] The CHCA and its multi-version-document (MVD) encoding scheme are discussed in more detail elsewhere in this volume.

was referred to at the start of this paper, I would suggest that rather than shifting the responsibility of interpreting the curated materials entirely to the user, this combination of archival and editorial impulses merely makes the interpretative quality of the edited text more explicit, and encourages a more critical reading of the work—which has been the aim of the digital scholarly edition all along. Hopefully, this new direction the digital scholarly edition is heading will lead to closer collaborations between editors and archival institutions in the future—a development that the digital scholarly edition could only benefit from.

References

Beckett Digital Manuscript Project. Retrieved March 30 2018 from: www.beckettarchive.org.

Boot, P. Fischer, F. and Van Hulle, D. (2017). Introduction. In Boot, P. Cappellotto, A., Dillen, W., Fischer, F., Kelly, A., Mertgens, A., Sichani, A., Spadini, E., and Van Hulle, D., (Eds.), Advances in digital scholarly editing. Papers presented at the DiXiT conferences in the Hague, Cologne, and Antwerp (pp. 15–22). Leiden: Sidestone Press.

Brünning, G., Henzel, K., & Pravida, D. (2013). Multiple encoding in genetic editions: The case of *Faust*. *Journal of the Text Encoding Initiative, 4*, 1–12 http://jtei.revues.org/697. Accessed 23 April 2019.

Dahlström, M. (2000). Drowning by versions. *Human IT, 4*(4) http://etjanst.hb.se/bhs/ith/4-00/md.htm. Accessed 23 April 2019.

Dahlström, M. (2009). The Compleat edition. In M. Deegan & K. Sutherland (Eds.), *Text editing, print, and the digital world* (pp. 27–44). Basingstoke: Ashgate.

Dahlström, M., & Dillen, W. (2017). Review of *Litteraturbanken*: the Swedish Literature Bank. *RIDE, 6*. https://doi.org/10.18716/ride.a.6.2.

Eggert, P. (2005). Text-encoding, theories of the text, and the "work- site". *Literary and Linguistic Computing, 20*(4), 425–435.

Eggert, P. (2017). The archival impulse and the editorial impulse. In P. Boot, A. Cappellotto, W. Dillen, F. Fischer, A. Kelly, A. Mertgens, A.-M. Sichani, E. Spadini, & D. Van Hulle (Eds.), *Advances in digital scholarly editing. Papers presented at the DiXiT conferences in the Hague, Cologne, and Antwerp* (pp. 121–124). Leiden: Sidestone Press.

Evenson, J. (1999). *Electronic Archives: Creating a New Bibliographic Code. Paper presented at the ACH-AALC conference in Charlottesville*. USA: Virginia.

Faust Edition. Retrieved March 30 2018 from: http://beta.faustedition.net.

Henny-Krahmer, U., & Neuber, F. (2017). Editorial: Reviewing digital text collections. *RIDE, 6*. https://doi.org/10.18716/ride.a.6.0 Accessed on 30 March 2018.

Huitfeldt, C., & Sperberg-McQueen, C. M. (2008). What is a transcription? *Literary and Linguistic Computing, 23*(3), 295–310.

Litteraturbanken. Retrieved March 30 2018 from: https://litteraturbanken.se/start.

Neuber, F., & Henny-Krahmer, U. (2018). Editorial: Digital text collections - take two, Action! *RIDE, 8*. https://doi.org/10.18716/ride.a.8.0.

NietzcheSource. Retrieved March 30 2018 from: http://www.nietzschesource.org.

Nixon, M., & Van Hulle, D. (2017). *Samuel Beckett's library*. Cambridge: Cambridge University Press.

Price, K. (2007). Electronic scholarly editions. In S. Schreibman & R. Siemens (Eds.), *A companion to digital literary studies* (pp. 434–450). Malden: Blackwell Publishing.

Price, K. (2009). Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name? *DHQ, 3*(3) http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html. Accessed 23 April 2019.

Robinson, P. (1996). Is there a text in these variants? In R. Finneran (Ed.), *The literary text in the digital age* (pp. 99–115). Ann Arbor: University of Michigan Press.

Robinson, P. (2009). What text really is not, and why editors have to learn to swim. *Literary and Linguistic Computing, 24*(1), 41–52.

Robinson, P. (2016, 6 October). The Revolution is Coming. Paper presented at *Digital Scholarly Editing: Theory, Practice, Methods. ESTS* 2016 / *DiXiT 3.* Antwerp, Belgium.

Robinson, P., & Solopova, E. (1993). Guidelines for transcription of the manuscripts of the wife of Bath's prologue. In N. Blake & P. Robinson (Eds.), *The Canterbury Tales project occasional papers* (pp. 19–52). Oxford: Office for Humanities Communication.

Sahle, P. (2005). Digitales Archiv–Digitale Edition. Anmerkungen zur Begriffsklärung. In M. Stolz (Ed.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien*. Bern: germanistik.ch https://www.germanistik.ch/publikation.php?id=Digitales_Archiv_und_digitale_Edition. Accessed 23 April 2019.

Sahle, P. (2013). *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik*. Norderstedt: Books on Demand.

Shillingsburg, P. (1996). *Scholarly editing in the computer age. Theory and practice* (3rd ed.). Ann Arbor: The University of Michigan Press.

Shillingsburg, P. (2015). Development principles for virtual archives and editions. *Variants. The Journal of the European Society for Textual Scholarship, 11*, 11–28.

Steinkrüger, P. (2014). Review of *Nietzschesource. RIDE, 1*. https://doi.org/10.18716/ride.a.1.4.

Van Hulle, D. (1999). Authenticity or Hyperreality in hypertext editions. *Human IT, 1*, 227–244 http://etjanst.hb.se/bhs/ith/1-99/dvh.htm. Accessed 23 April 2019.

Van Hulle, D. (2009). Editie en/of Archief: modern manuscripten in een digitale architectuur. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde, 119*(2), 163–178.