

**This item is the archived peer-reviewed author-version of:**

Digital scholarly editing within the boundaries of copyright restrictions

**Reference:**

Dillen Wout, Neyt Vincent.- Digital scholarly editing within the boundaries of copyright restrictions  
Digital scholarship in the humanities : a journal of the Alliance of Digital Humanities Organizations / Alliance of Digital  
Humanities Organizations - ISSN 2055-768X - (2016), p. 1-12  
Full text (Publishers DOI): <http://dx.doi.org/doi:10.1093/lc/fqw011>

# Digital Scholarly Editing within the Boundaries of Copyright Restrictions

Revision copy

Wout Dillen & Vincent Neyt

One of the great advantages the digital medium has to offer the field of scholarly editing is that it makes its products much easier to distribute. No longer bound to a shelf, the Digital Scholarly Edition (DSE)<sup>1</sup> has the potential to reach a much wider audience than a printed edition could. To a certain extent, however, the nature of the materials textual scholars are working with dictates the perimeters within which this dissemination can take place. When working with modern manuscripts, for instance, copyright restrictions may limit the extent to which a project can distribute its resources. In an academic climate where open access is not only becoming a standard, but in some cases even a requirement for receiving funding,<sup>2</sup> such limitations may be perceived as problematic. In this paper however, we argue that even within the boundaries of copyright restrictions there can still be room to produce and distribute the results of textual scholarship. To frame this discussion, we will start by offering an example of the unfavourable reception of copyright restrictions on DSEs by referring to Peter Robinson's recent campaign against the academic use of the Creative Commons license with non-commercial restrictions, recommending its more open Share Alike license instead.

## CC BY-NC versus CC BY-SA

Both at the ‘Social, Digital, Scholarly Editing’ (SDSE) conference in Saskatoon, Saskatchewan, and at the subsequent ADHO ‘Digital Humanities 2013’ conference in Lincoln, Nebraska, Peter Robinson expressed thought-provoking claims about the future of the Digital Scholarly Edition. More specifically, he proposed the following list of five ‘Desiderata for Digital Editions’ as the first part of his presentation at the ADHO conference:

1. The text of both the document and of the work should be encoded;
2. All editorial acts should be attributed;
3. All materials should, by default, be available by a Creative Commons share-alike license;
4. All materials should be available independent of any one interface;
5. All materials should be held in a sustainable long-term storage system, such as an institutional repository

(2013)<sup>3</sup>

Commendable though this ambitious program may be, scholarly editors who work with modern manuscripts for which today’s copyright and intellectual property laws are still considerable issues may feel the need to nuance Robinson’s third desideratum. In essence, this desideratum was a criticism of the ‘Creative Commons Non-Commercial Use Only’ (CC BY-NC) license as a new standard for scholarly projects (Creative Commons, 2013). Intuitively, the CC BY-NC license makes a lot of sense: why would you allow *someone else* to earn money for work *you* did? However, as Robinson argued, this license is still too restrictive for scholarly endeavours because the ‘non-commercial’ clause is so vague that it will discourage too many people from actually using the project’s data, even if they intend to use it in a non-profit context.<sup>4</sup>

Instead, Robinson suggested using the CC BY-SA license (also called: 'Share-Alike' or 'copyleft'), which also protects the project against large-scale exploitation of the materials while at the same time allowing others to alter, transform, or build upon the work if they need to (Creative Commons, 2015b). Furthermore, Robinson posited that 'if [your project uses] the "non-commercial" restriction, you might as well lock it up and throw away the key' (2013a).

As part of a conference presentation, this last argument can be easily dismissed as a provocative statement designed to trigger an interesting debate during the subsequent round of questions - which it inevitably did. But it seems that the roots of this statement lie deeper: they form the premise of the *Textual Communities* project Robinson is developing at the University of Saskatchewan, and can be traced back to a paper he had written in 2003. In 'Where We Are With Electronic Scholarly Editions, And Where We Want To Be', Robinson had already argued that '[t]he best guarantee that an electronic edition should remain usable is that it should be used', and suggested that the only way to achieve this goal is by putting the edition 'on the internet in a manner that allows it to be appropriated by others, augmented, corrected, infinitely reshaped' (2003). It is probably no coincidence that this reasoning resembles that of the open source movement that had gradually become more prominent through the Free Software Foundation (founded by Richard Stallman in 1985) and its development of the GNU General Public License (GNU GPL) that would in turn inspire the foundation of Creative Commons in 2001, which would publish the first version of its licenses in 2002 - i.e. around the time Robinson was writing his paper on electronic scholarly editions. And as the rising interest in open access in academia has already shown, applying Stallman's vision of free software<sup>5</sup> to academic research paints an attractive picture indeed.

But while we certainly agree that Robinson's desiderata are laudable goals that should be pursued whenever possible, the fact remains that they are often unattainable. By arguing that all DSEs should *by default* be made available under a CC-BY SA license, Robinson neglects the fact that this license is not necessarily the editor's to give. Antithetical as they may seem, copyright and copyleft are not mutually exclusive. Quite the contrary, as Severine Dusollier argues in 'Open Source and Copyleft: Authorship Reconsidered': 'the principle [of copyleft] rests entirely on copyright' (Dusollier 2002, 286). While CC BY-SA may prevent the user from restricting the license of her source materials further, copyright will still prevent her from doing the opposite.<sup>6</sup> In the case of contemporary literary works, this inevitably means that the decision who to give the rights to the materials under which license remains in the hands of the authors and the executors of their estates, until the day those rights expire and the materials enter the public domain. Should these limitations stop us from building digital scholarly editions around their works? For Robinson, the answer seems to be: yes. Expanding on the background of the project on the *Textual Communities* website, Robinson relates how two experiences have shaped the project's firm stance towards copyright: (1) a disagreement between collaborators that resulted in the inability to publish a large body of work on the *Canterbury Tales*; and (2) Hans Walter Gabler's unsuccessful attempts to win the favour of the Joyce estate, which rendered him unable to continue his work on a digital edition of Joyce's *Ulysses* (Robinson 2012-2013). 'Events like these decided us', Robinson explains, 'We would never work on materials where someone else could, by fiat, render all our work worthless just by refusing publication permission' (ibid.).

Perhaps even more so than other editions, DSEs that are obliged to restrict access to their work (for whatever reason) have to prove their worth to their potential

readers: they have to prove that the edition is worth consulting despite those restrictions. If they do not, the editions will not be used – and being used is one of the main goals of every digital edition. But there are different kinds of uses a DSE can pursue. Different users, with different interests, will want to use the DSE in different ways. At the most basic level of interest, users are looking for simple browsing functionalities. To satisfy these users, editors will want to present the materials within an attractive and intuitive interface. At a more advanced level of interest, users will want to research the materials the DSE has to offer, and access them in non-linear ways. To reach those users, editors will need to provide indexes, advanced search options, advanced textual comparison options, to open the corpus up for analysis in a standardized format, etc. Finally, at the highest level of interest, there are meta-users, who want to re-use the DSE's data for their own purposes: to write their own transcriptions of the DSE's facsimiles (and publish the results), to build their own interface around the data the DSE provides, or to perform functionalities the DSE does not (yet) offer (and publish the results). In his paper at the 2013 SDSE conference, Peter Robinson focused on the needs of the meta-users when he suggested that editors have to 'give up control' over the materials they are editing:

Editors, hear this: if you want your digital edition to be used by others, to become the centre of other people's work, to be transfigured over and over into bright new forms of scholarship, [...] you have to release it without restriction, to permit both commercial and non-commercial work, to license the making of any and all derivative works.

(2013c, 2)

We agree that to satisfy the needs of all possible users, releasing a DSE without restrictions is the best course of action. But in many cases, editors are not allowed to

give up control in the first place. While this may prevent the data from being re-used outside the walls of the edition, it should not diminish the value of the data, nor of the functionality that was built around that data. To illustrate this point, we will focus on a series of academic projects that work with copyright materials, and discuss how they try to creatively work within the boundaries their material binds them to.

### **Curation of Source Materials**

The first step in the creation of any scholarly edition – be it digital or in print – is to find an interesting corpus of texts to analyse. The digital turn has certainly facilitated this process: never before did locating historical documents, contacting the institutions that hold them, and sharing facsimile images for research purposes require so little effort. The problem is, however, that although those corpora are usually made accessible for scientific research, not all of them permit their free distribution in the form of a DSE. For a scholarly editor, the ideal situation would be that the texts of her source materials have entered the public domain, and that high-resolution scans of these materials have been made freely available under a Creative Commons license. This is the case, for instance, for the manuscripts of Marcel Proust, facsimiles of which have been made available by the Bibliothèque nationale de France (BnF). This allowed Elena Pierazzo and Julie André to publish the transcriptions of their 'Proust Prototype' under a CC BY-NC 3.0 license (2012), as long as they referred back to the BnF website for the copyright of the facsimile images.<sup>7</sup> But for scholarly editors working with even more recent materials, conditions are usually less favourable, which may cause the process of acquiring the license to publish their work to require a considerable effort. When copyright restrictions come into play, careful negotiations

with authors or the executors of their estates become a crucial aspect of the contractual agreements between what is already a large group of people with a wide range of (commercial and non-commercial) interests: scholars, holding libraries, publishers, and funders. The same holds true for this paper's first case study: the *Beckett Digital Manuscript Project* (BDMP).

The BDMP ([www.beckettarchive.org](http://www.beckettarchive.org)) is an international collaboration under the direction of Dirk Van Hulle (University of Antwerp) and Mark Nixon (University of Reading), between the Centre for Manuscript Genetics at the University of Antwerp, the Beckett International Foundation at the University of Reading, and the Harry Ransom Humanities Research Center (HRC) at the University of Austin, Texas, with the kind permission of the Estate of Samuel Beckett. As its name implies, the BDMP aims to reunite and make publicly accessible all manuscripts of Samuel Beckett's works, the physical documents of which are located in different holding libraries around the world. This goal will be realized in the form of a hybrid genetic edition of Beckett's works: *hybrid* because it combines a digital archive of the manuscripts organized in twenty-six research modules with a series of twenty-six accompanying interpretative volumes; *genetic* because it aims to analyse and reconstruct these works' individual writing processes (or: 'geneses'). As may be expected, the curation of all the extant manuscripts on which the edition is based necessitated a long (and still on-going) process of collaboration between the directors of the BDMP and all other parties involved. At the moment of writing, the BDMP has secured a contract regarding the acquisition of scans with the holding libraries of the Harry Ransom Center, the University of Reading, Washington University St Louis, Dartmouth College, Trinity College Dublin, Ohio State University, Harvard University, and Boston College. Similar contracts are being negotiated with other

libraries that hold relevant manuscripts. As the project advances and more manuscripts surface, the BDMP hopes to collaborate with the libraries that acquire those manuscripts as well.

These agreements allow the project to request high resolution scans of the necessary documents for scientific purposes, and to incorporate them into the edition. As with most digital editions, these facsimiles form a crucial part of the BDMP's publication. But what is of course even more important for the edition is the text those facsimiles contain. And the copyright of those texts currently belongs to the Beckett Estate – where it will remain for quite some time to come. Because copyright law is territorial, the exact date when Beckett's works will enter the public domain will differ from country to country.<sup>8</sup> And for unpublished works such as manuscripts, letters, diary entries, etc., the waters are particularly muddy. In Canada, for instance, copyright of unpublished materials expires 50 years after the calendar year of the author's death, which in Beckett's case (who died in 1989) would mean 1 January 2040. In the USA, this period is extended to the author's life plus 70 years – here the year 2060. In the UK, on the other hand, "[w]orks that were unpublished at the author's death and remained so until 1 August 1989 [i.e. five months *before* Beckett's death] are protected by copyright [...] for 50 years from 1 January 1990" (James Joyce Foundation 2012) – in other words until the year 2040. And in Australia, copyright can be enforced for 70 years after the unpublished work has been 'disclosed', meaning that it will differ from work to work, and on the purview of what it means to legally 'disclose' an unpublished work.

This is what makes a good relationship and workable contracts between DSE projects and authors' estates (or their representatives) so important: if the troubled relation between scholarly editors and the Joyce Estate teaches us anything, it is that a

project on such a scale covering an important contemporary author would simply not be possible without it. Thankfully, in the case of the *Beckett Digital Manuscript Project*, all parties involved have realized that the future of scholarly editing is digital, and that the scholarly augmentation of Beckett's legacy will only increase the interest in his works – academic or otherwise. For this reason, the Beckett Estate agreed to give the directors of the BDMP the license to publish their genetic edition of Beckett's manuscripts, as long as this happens behind a pay-wall. This is of course not an ideal situation, because the pay-wall inevitably prevents a number of potential users from accessing the edition. Although the prices to access the edition are not excessive, the user (be it an individual or an institution) will still need to be convinced of the edition's worth *before* being allowed to fully appreciate its functionalities. But *limited access* is of course better than *no access*; and because the BDMP's contractual agreement with holding libraries and the Beckett Estate stipulates that each of the collaborating institutions are granted institutional access to the edition, a considerable group of people may access the edition free of charge.

### **Public Access without a Public License**

A different approach to this problem is the one the *Woolf Online* project has taken. Under the auspices of the Loyola University Chicago's Center for Textual Studies and Digital Humanities, this project aims to offer a genetic edition of Virginia Woolf's *To the Lighthouse* in the form of a digital archive that combines transcribed facsimiles of Woolf's manuscripts, a number of the work's editions, extracts from diaries and letters, photo albums, critical reviews, etc. – all of which can be accessed by anyone who visits the project's website ([www.woolfonline.com](http://www.woolfonline.com)). Obtaining the license to publish these

materials is in itself already quite an achievement, since the edition incorporates a range of different kinds of materials belonging to a number of different copyright holders. Let us focus on what is arguably the project's core material: the so-called 'Virginia Woolf Material' in the form of a collection of unpublished works (a manuscript, a typescript, corrected proofs, and extracts from diaries, letters, etc.) and five editions that were published between 1927 and 1939. Since Virginia Woolf died in 1941, the unpublished materials have already entered the public domain in many countries: in Canada since 1992, and in the USA (and countries that follow their lead) since 2012. But this is not a global phenomenon: in the UK for instance (as with Beckett), they will only enter the public domain in 2040. In many cases, the procedure for published works is similar to those of unpublished works: in Canada and Australia copyright expires 50 years after the author's death (i.e. 1992), and in many E.U. countries (including the UK and the Republic of Ireland) this becomes 70 years (i.e. 2012; see Spoo 2009, 229). But in the USA, for example, copyright of books that were published after 1923 (like *To the Lighthouse*) expires 95 years after the work was first published – meaning that they will remain in copyright until the year 2022. Still, because the project is limited to a single work, and because published and unpublished versions of the work have already started to enter the public domain in a wide range of countries, acquiring a reasonably priced license to make these texts publicly available is arguably a lot more realistic than in the case of the BDMP, for instance, where the unpublished works in its corpus will only start to enter the public domain in at least another 25 to 45 years – granting that copyright will not have been extended even further by that time.

It should be noted, however, that neither of these scholarly editions meet the requirements of Robinson's third desideratum. Although the *Woolf Online* project has

acquired a license to offer its users free access to its materials, those materials are not allowed to leave the project's website. As it is clearly stated on the project's Copyright Notice, the 'material is provisioned for online publication and reading only at Woolf Online ([www.woolfonline.com](http://www.woolfonline.com)) and may not be copied, distributed, transmitted or otherwise altered or manipulated without the express permissions' of the copyright owners.<sup>9</sup> This means that the edition's users will not be able to publish the project's data in a new interface, as in the reusable future Robinson envisions. Adding a Creative Commons license of any kind will remain impossible until all of its source materials (which includes both the texts and their images) have entered the public domain. But the question remains how important this final step is to enable further research. Developing a DSE, scholarly editors will want to build the best possible environment for their readers to fully appreciate the nuances of the materials they are editing, and to distribute it in such a way that it may be a useful foundation for others to build their research on. To achieve this goal, Creative Commons licenses are useful, but far from necessary. Instead, the editor's first concern will be to distribute her edition to as many people as her source materials will allow – which, as has been discussed, will depend on their copyright owners. And if copyright restrictions prevent the editor from licensing her data under a CC BY-SA license, this should not be held against her. It is not because the images of Proust's manuscript materials, for example, are bound to a non-commercial clause on behalf of the BnF, that Pierazzo and André's Proust Prototype is any less valuable as an object of academic research.

### **Sharing Metadata and Ancillary Data**

It could be recommended, however, to make academic data that is not bound to copyright restrictions publicly available under as public a license as possible. This is

perhaps an area where the *Woolf Online* project might still improve. As stated on the website's Copyright Notice, all materials that do not belong to its list of copyrighted materials are shared under a CC BY-NC-ND license. This license is the same as the CC BY-NC license, except that it also prohibits the user from making derivative works – meaning that the user is not allowed to transform the edition's own materials, even in a non-commercial environment. From this copyright notice, it is not entirely clear to which data this license applies. The wording of the texts on the home page, project description page etc. perhaps? If that is all, we (admittedly as outsiders to the project) see no reason why this license could not be exchanged for a CC BY-NC, or even a CC BY-SA license – both of which still require proper attribution to the author of the appropriated materials.

But there is of course more data to share that is not subjected to copyright restrictions. The most obvious example in this category is metadata. Short of copying their contents, researchers are allowed (and could be encouraged) to describe the resources they are studying in as much detail as they desire, and to share their findings – preferably in a standardized format like RDF (Resource Description Framework). The success of a project like NINES that aggregates this kind of metadata for digital scholarship relating to the 19<sup>th</sup> century proves the relevance of this kind of data to the rest of the academic research community, and offers a great example of how it can be used to further the dissemination of these projects. Similar projects have already been developed to aggregate the metadata of digital projects relating to medieval studies (MESA), 18<sup>th</sup> century studies (18thConnect), and radicalism studies (SiRO), and Loyola University Chicago's Center for Textual Studies and Digital Humanities is working on an analogous project concerning modernist literature and culture studies called ModNets (Modernist Networks). When this latest addition to the ARC

(Advanced Research Consortium) catalogue is launched, this will become a welcome hub for scholarly editors who are working with modernist authors (like both *Woolf Online* and the *Beckett Digital Manuscript Project*), who will finally have a centralized place to share rich metadata regarding their source materials with the rest of the scholarly community.

But we can do even more. For researchers who are working with similar data, the steps we take to achieve our results can be just as valuable as the results themselves. The fact that Digital Humanities has a history of encouraging the documentation of failure,<sup>10</sup> for instance, indicates that the research community values the process of research as much as its product. And there are many ways in which the ancillary products of academic research can be shared without infringing on the source material's copyright. Like metadata, a project's documentation, for example, can contain a wealth of information about the project that may be of use to other researchers. In the case of Digital Scholarly Editions, the TEI already provides a standard format to share this information, in the form of an ODD (One Document Does it all) file that combines a TEI-XML validation schema with human readable schema documentation in a single XML file (Burnard 2014, 90). Unlike the edition's transcribed XML-files themselves, these documentation files contain no sensitive information whatsoever, and may be shared freely under any license without legal consequence. This means that once a scholarly editor has generated a suitable TEI customization, she can easily export this schema into a comprehensible format that other contributors or interested researchers may use to find out which TEI-XML tags the project uses, and how.<sup>11</sup> As such, the project's collectively cultivated expertise does not go to waste, but may help to support (or even initiate) other research instead.

Recognizing the potential of this kind of information for other scholarly

editing projects, the *Beckett Digital Manuscript Project* has recently put its own documentation online for consultation in the form of a digital Encoding Manual. Initially designed as a Microsoft Word format cheat-sheet that helped the project's contributors transcribe Beckett's manuscripts, the *BDMP Encoding Manual* was expanded to include validation information, more examples and explanatory text, and more information on encoding practices in general, in the form of a WordPress website ([www.beckettarchive.org/encodingmanual](http://www.beckettarchive.org/encodingmanual)). Before this website could be made public, some minor changes had to be made for copyright reasons (e.g. replacing screenshots in the examples with more generic homemade image files), but these did not impinge on the Manual's functionality in any way and are not perceived as a great loss by the project's contributors. That this kind of information may be useful for other scholarly editors who are looking to jumpstart the development of their scholarly editions can be illustrated by pointing to the work of DiXiT<sup>12</sup> fellow Elli Bleeker, who is currently using the *BDMP Encoding Manual* as a starting point for the encoding schema of a new genetic edition of Raymond Brulez' *Scheherazade*.

### **The Fair Use Doctrine**

The discussion of copyright restrictions above has been written from a position that takes a rather absolute view on copyright: namely that nothing that is protected through copyright may be copied and distributed in any way without the express permission of the copyright holder. But that is of course not exactly true: there are exceptions to the rule that make it possible to share copyrighted materials (to a certain extent), notably by means of the fair use doctrine. Originating in the United States, this doctrine is described in §107 of the US Copyright law, which states that the

reproduction of copyrighted materials 'for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.<sup>13</sup> The problem with this exception, however, is that it is open to interpretation and assessed on a case by case basis, taking a non-exhaustive list of vague criteria into account that include 'the character of the use' (e.g. commercial vs. non-commercial use), 'the nature of the copyrighted work', 'the amount and substantiality of the portion used to the copyrighted work as a whole', and 'the effect of the use upon the potential market value for or value of the copyrighted work' (§107). As Richard Stim argues in *Getting Permission: How to License & Clear Copyrighted Materials Online & Off*, this means that there is no guarantee that the user's interpretation of 'fair use' will not be questioned in a court of law, a procedure that (even when ruled in favour of the defendant) 'may well outweigh any benefit of using the material in the first place' (2013, 301).<sup>14</sup>

Still, there are some basic rules of conduct that can be followed to help minimize the risk of litigation to an acceptable degree. For example, using only the most relevant passages of a copyrighted work strictly for research purposes in a non-commercial environment can already go a long way towards convincing copyright holders that the odds would be against them if they decided to pursue the matter legally.<sup>15</sup> In addition, the US courts have recently become more inclined to rule in favour of the defendant in a copyright infringement case, if the use is deemed to be 'transformative' (i.e. when it considerably alters or augments the work by situating it in a new context), ever since a precedent for this ruling was established in 1994 in *Campbell v. Acuff-Rose Music*. As Laura A. Heymann reveals in her paper titled 'Everything is Transformative', the court in question 'derived the term "transformative" as a relevant criterion for determining fair use' largely from Pierre N.

Neval's article titled 'Toward a Fair Use Standard' that had appeared in the *Harvard Law Review* four years before (Heymann 2012, 451). In that article, Leval had posited that to be considered as 'transformative' – and therefore as fair use – the challenged use

must be productive and must employ the quoted matter in a different manner or for a different purpose from the original. A quotation of copyrighted material that merely repackages or republishes the original is unlikely to pass the test [...]. If, on the other hand, the secondary use adds value to the original – if the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings – this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society.

(Leval 1990, 1111)

Of course, like the other criteria for determining fair use, the transformative quality of the quoted material is open for interpretation: when does re-use stop to be repackaging, and start to add sufficient value to be considered transformative? And how does this transformative quality weigh up against the substantiality of the portion of the material that was used, for example? Although there are no straightforward answers to these questions, this development in copyright law (in the US at least) may give researchers a little more leeway when they need to re-use copyrighted materials for their research.

At the Centre for Manuscript Genetics, we are working on a project that hopes to achieve exactly this kind of transformation: the *Lexicon of Scholarly Editing* ([www.uantwerpen.be/lexicon-scholarly-editing](http://www.uantwerpen.be/lexicon-scholarly-editing)). Initiated by Dirk Van Hulle and developed by Wout Dillen as a project of the European Society for Textual Scholarship (ESTS),<sup>16</sup> this *Lexicon* aims to be a useful resource for anyone who is interested in Textual Criticism in general – be they students, early career scholars, or

experts in the field. Rather than writing new definitions for important concepts in Textual Criticism, the *Lexicon* gathers definitions from academic source materials, such as journals, monographs, and the like. On the website, these definitions are grouped together under an English lemma, but are rendered chronologically in their original language. As such, the *Lexicon* aims to reveal the lively multilingual debate these concepts have spurred in the field. This is where the use of these citations arguably becomes 'transformative': displaying its quoted definitions as if in a multilingual discussion with one another (which, in some cases, they quite literally are), the *Lexicon* aims to enrich the scholarly knowledge it encompasses by repurposing it in a new contextual environment. This way, the passages quoted in the *Lexicon* can be used to develop a better understanding of certain problematic concepts; to discover what those concepts are called in different languages; and to help textual scholars develop more nuanced arguments in their own writing.

As may be expected, the *Lexicon's* fundamental premise as a citation archive presupposes the public dissemination of what are almost exclusively copyrighted materials. The *Lexicon's* use of its source materials could qualify as fair use, as (1) the materials are used for research purposes; (2) the *Lexicon* only uses those fragments that are strictly relevant for its argument – typically only a fraction of the size of the complete work; (3) these fragments are consistently attributed to their rightful authors according to the relevant accepted citation practices; (4) the use of these fragments only furthers the original nature and aim of the copyrighted material – i.e. the advancement of academic knowledge; and (5) all of the above is published in a transparent, non-commercial environment. In addition, the *Lexicon's* 'Credits' page explicitly invites anyone who feels that the *Lexicon* infringes on their rights as copyright holders to step forward so that the editors can promptly remove the challenged

content from the website. This makes the *Lexicon* a good example of how the fair use doctrine can be used to work creatively within the boundaries of copyright restrictions.

Even though it may not be advisable to hide behind the fair use doctrine when building a Digital Scholarly Edition, there could even be some room for creativity here. A good example of the fair use of a complete text of a literary work would be Ralph Slepon's impressive *Finnegans Wake Extensible Elucidation Treasury* (or: FWEET; see [www.fweet.org](http://www.fweet.org)). This website aggregates a total of 83,984 notes (and counting) on Joyce's last work *Finnegans Wake*, gathered from a wide array of academic source materials. These notes (called 'elucidations' on the website) constitute annotations on fragments of the text of the work – often offering a reading of (or intertextual reference behind) a phrase, word, or even part of a word in *Finnegans Wake*. The website's elucidations are themselves of course already copyrighted materials, that are used more or less in the same way as the citations used in the *Lexicon of Scholarly Editing*: as part of an academic research tool that builds on earlier academic research (that is properly attributed and never constitutes a substantial part of the original work) Slepon's use of these elucidations can be considered fair use. What may be more problematic, however, is Slepon's use of the text of *Finnegans Wake* itself. Because the database links every line in the *Wake* to its relevant annotations, this means that the entire text of the work can be freely consulted by anyone with an Internet connection. But the purpose of the website is not to offer a digital reading text of the work – it is to aggregate a vast collection of scholarly research. As a result, the website was designed to be used *alongside* (rather than *instead of*) the user's personal copy of the work. Only presenting one line of the *Wake* at a time (without a clear way to navigate between subsequent lines), the website's interface obfuscates the work's text to such a degree that it becomes almost impossible to use without a personal copy at hand. This may

help explain why FWEET remains one of the few scholarly resources using Joyce's work that was never taken down by the executors of Joyce's estate. Since the website was first published online in 2005 – long before the *Wake's* text entered the public domain – this supports the argument that the fair use doctrine can be employed to creatively pursue the increase of scholarly access to copyrighted materials.

## **Conclusion**

In this paper, we have argued that the limitations posed on copyrighted materials should not be a reason for academics to stop using those materials. Quite to the contrary: they should be an incentive to make them available for further research by any means possible. There is no reason to put off trying to answer the research questions we face today – especially if we already have the technology to do so. If that means forgoing a CC BY-SA license, or setting up a pay-wall, so be it. In that case, editors will have to accept the challenge to convince their readers that their corpus is worth consulting despite those limitations. Because even though doing that might make the process of building a digital scholarly edition more difficult, it does not make it impossible. As our case studies of the *Beckett Digital Manuscript Project*, *Woolf Online*, *ModNets*, the *BDMP Encoding Manual*, the *Lexicon for Scholarly Editing*, and *FWEET* have shown, there is often still plenty of room within the boundaries of copyright restrictions for the publication of Digital Scholarly Editions, and all of the ancillary data their scholarly editors produce along the way. At the same time, we believe it is important for editors to make their own intellectual contributions as open as possible. This would not only benefit current research (as it would allow other scholars to access and use the project's ancillary data, for example); it would also help the editors move

restricted materials to the edition's unrestricted space when their copyright expires. By favouring shared access over restricted access as much as possible, we stand a better chance of safeguarding these materials for the benefit of future scholars.

## **Notes**

The authors wish to thank Dirk Van Hulle and Geert Lernout for their comments on earlier versions of this essay. Thanks also to our reviewers, whose suggestions were greatly appreciated.

## Funding

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007- 2013) / ERC grant agreement n° 31360911.

## References

**Beckett, Samuel** (2011). *Stirrings Still / Soubresauts and Comment Dire / what is the word: an electronic genetic edition (Series 'The Beckett Digital Manuscript Project' module 1)*, edited by Dirk Van Hulle and Vincent Neyt. Brussels: University Press Antwerp (ASP/UPA).  
<http://www.beckettarchive.org> (accessed on 6 March 2014).

**Beckett, Samuel** (2014). *L'Innommable / The Unnamable: an electronic genetic edition (Series 'The Beckett Digital Manuscript Project' module 2)*, edited by Dirk Van Hulle, Shane Weller and Vincent Neyt. Brussels: University Press Antwerp (ASP/UPA).  
<http://www.beckettarchive.org> (accessed on March 6, 2014).

**BDMP Encoding Manual** (2014-). *BDMP Encoding Manual*.  
<http://www.beckettarchive.org/encodingmanual> (accessed on April 8, 2015).

**Burnard, Lou** (2014). *What is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources. New Edition [online]*. Marseille: OpenEdition Press. Open Access version: <http://books.openedition.org/oep/426> (accessed on May 13 2015).

**Burke, John** (2014). The Expansion of Fair Use in Response to Copyright's Growing Protection. *Law School Student Scholarship*. Paper 631.  
[http://scholarship.shu.edu/student\\_scholarship/631](http://scholarship.shu.edu/student_scholarship/631) (accessed on May 13, 2015).

**Creative Commons** (2001-). *Creative Commons*. <http://creativecommons.org> (accessed on March 13, 2015).

**Creative Commons** (2015a). Attribution-NonCommercial 4.0 Unported (CC BY-NC 3.0). *Creative Commons*. <https://creativecommons.org/licenses/by-nc/4.0/>

(accessed on May 15, 2015).

**Creative Commons** (2015b). Attribution-ShareAlike 4.0 Unported (CC BY-SA 3.0).

*Creative Commons*. <http://creativecommons.org/licenses/by-sa/4.0> (accessed on May 15, 2015).

**Dusollier, Severine** (2002). Open Source and Copyleft: Authorship Reconsidered.

*Colum. J.L. & Arts* 26: 281-296.

<http://heinonline.org/HOL/LandingPage?>

[handle=hein.journals/cjla26&div=18&id=&page=](http://heinonline.org/HOL/LandingPage?handle=hein.journals/cjla26&div=18&id=&page=) (accessed on March 13, 2015).

**Free Software Foundation** (1985-). *Free Software Foundation*. <http://fsf.org> (accessed on March 13, 2015).

**Gabler, Hans Walter** (2010). Theorizing the Digital Scholarly Edition. *Literature*

*Compass* 7.2: 43-56.

**Gerber, Anna, Andrew Hyland, and Jane Hunter** (2010). *A Collaborative Scholarly*

*Annotation System for Dynamic Web Documents-A Literary Case Study*. Berlin: Springer.

**Heymann, Laura A.** (2008). Everything is transformative: Fair use and reader response. *Columbia Journal of Law & the Arts* 31 08-06.

<http://ssrn.com/abstract=1148379> (accessed on May 15, 2015).

**James Joyce Foundation** (2012). Joyce & Copyright. *James Joyce Foundation*.

<https://joycefoundation.osu.edu/joyce-copyright> (accessed on May 5, 2015; last updated in April 2012).

**Leval, Pierre N.** (1990). Toward a fair use standard. *Harvard Law Review* 1105-1136.

[https://www.law.berkeley.edu/files/Leval\\_-\\_Fair\\_Use.pdf](https://www.law.berkeley.edu/files/Leval_-_Fair_Use.pdf) (accessed on May 15, 2015).

**Möller, Erik** (2005). Creative Commons -NC Considered Harmful. *Kuro5hin*.

<http://www.kuro5hin.org/story/2005/9/11/16331/0655> (accessed on September

30, 2013).

**Murray Michael D.** (2012). 'What is Transformative? An Explanatory Synthesis of the Convergence of Transformation and Predominant Purpose in Copyright Fair Use Law'. *Chicago-Kent Journal of Intellectual Property Law, Forthcoming* 12–09.

<http://ssrn.com/abstract=2088164> (accessed on May 15, 2015).

**Pierazzo, Elena** (forthcoming). *Digital Scholarly Editing*. Surrey: Ashgate.

**Pierazzo, Elena and Julie André** (2012). Autour d'une séquence et des notes du Cahier 46: enjeu du codage dans les brouillons de Proust.

[http://research.cch.kcl.ac.uk/proust\\_prototype/index.html](http://research.cch.kcl.ac.uk/proust_prototype/index.html) (accessed on May 6, 2015).

**Robinson, Peter** (2000-). Scholarly Digital Editions. *sd-editions*. <http://www.sd-editions.com/> (accessed on March 12, 2015).

**Robinson, Peter** (2003). Where We Are with Electronic Scholarly Editions, and Where We Want To Be *Jarbuch für Computerphilologie Online* 5: 123-143.

<http://computerphilologie.uni-muenchen.de/jg03/robinson.html> (accessed on March 13, 2015).

**Robinson, Peter** (2012). The Principles of the Textual Communities Project. *Textual Communities*. <http://www.textualcommunities.usask.ca/web/textual-community/wiki/-/wiki/Main/The+Principles+of+the+Textual+Communities+Project> (accessed on March 13, 2015).

**Robinson, Peter** (2012-2013). The Background to the Textual Communities Project. *Textual Communities*.

<http://www.textualcommunities.usask.ca/web/textual-community/wiki/-/wiki/Main/The+background+to+the+Textual+Communities+Project> (accessed on March 13, 2015).

**Robinson, Peter** (2013a). Desiderata for Digital Editions. *Slideshare*.

<http://www.slideshare.net/PeterRobinson10/peter-robinson-24420126> (accessed on September 30, 2013).

**Robinson, Peter** (2013b). Five Desiderata for Scholarly Editions in Digital Form.

Paper presented at 'Digital Humanities 2013', University of Lincoln, Nebraska, July 16-19, 2013.

<http://dh2013.unl.edu/abstracts/ab-314.html> (accessed on September 30, 2013).

**Robinson, Peter** (2013c). SDSE 2013: Why Digital Humanists Should Get out of Textual Scholarship. *Academia.edu*.

[http://www.academia.edu/4124828/SDSE\\_2013\\_why\\_digital\\_humanists\\_should\\_get\\_out\\_of\\_textual\\_scholarship](http://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship) (accessed on September 30, 2013).

**Robinson, Peter** (2013d). Why Digital Humanists Should Get Out Of Textual Scholarship. And If They Don't, Why We Textual Scholars Should Throw Them Out. *Scholarly Digital Editions*.

<http://scholarlydigitaleditions.blogspot.be/2013/07/why-digital-humanists-should-get-out-of.html> (accessed on September 30, 2013).

**Sahle, Patrick** (2012). What is a Scholarly Digital Edition? Paper presented at the 'NeDiMAH Meeting' in The Hague, The Netherlands, November 11, 2012.

<https://prezi.com/mdt8efbe3o3a/patrick-sahle-what-is-a-scholarly-digital-edition> (accessed on March 13, 2015).

**Slepon, Richard** (2005). *Finnegans Wake Extensible Elucidation Treasury*.

<http://fweet.org/> (accessed on April 8, 2015).

**Spiro, Lisa** (2012). "This Is Why We Fight": Defining the Values of the Digital Humanities. *Debates in the Digital Humanities* 16. [http://kompetenzzentrum.univ-trier.de/files/9113/9696/2499/Values\\_in\\_the\\_DH\\_Lisa\\_Spiro.pdf](http://kompetenzzentrum.univ-trier.de/files/9113/9696/2499/Values_in_the_DH_Lisa_Spiro.pdf) (accessed on May

15, 2015).

**Spo, Robert.** (2009). "For God's sake, publish; only be sure of your rights": Virginia Woolf, Copyright, and Scholarship. In *Woolf Editing/Editing Woolf*, edited by Wayne Chapman. Clemson: Clemson University Press, pages 227–231.

[http://digitalcommons.law.utulsa.edu/cgi/viewcontent.cgi?article=1358&context=fac\\_pub](http://digitalcommons.law.utulsa.edu/cgi/viewcontent.cgi?article=1358&context=fac_pub) (accessed May 6, 2015).

**Stallman, Richard** (2002). *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Boston: Free Software Foundation. <http://www.gnu.org/doc/fsfs-ii-2.pdf> (accessed on March 13, 2015).

**Stim, Richard** (2013). *Getting permission: how to license & clear copyrighted materials online & off*. Nolo.

**Textual Communities** (2013). Home. *Textual Communities*. <http://www.textualcommunities.usask.ca> (accessed on September 30, 2013).

**Tushnet, Rebecca** (2004). Copy this essay: How fair use doctrine harms free speech and how copying serves it. *Yale law journal* 535–590. <http://ssrn.com/abstract=1793181> (accessed on May 15, 2015).

**Unsworth, John** (1997). Documenting the Reinvention of Text: The Importance of Failure. *Journal of Electronic Publishing* 3.2 <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0003.201> (accessed on May 15, 2015).

**Woolf Online** (2006-) Copyright Notice. *Woolf Online*. <http://www.woolfonline.com/?node=about/copyright> (accessed on April 8, 2015).

<sup>1</sup> In Textual Criticism, digital editorial projects that are the result of textual scholarship are sometimes called 'digital scholarly editions' (e.g. Gabler 2010; Pierazzo forthcoming), and sometimes 'scholarly digital editions' (e.g. Gerber Hyland and Hunter 2010; Sahle 2012). To avoid any confusion with Scholarly Digital Editions – an electronic publishing house founded by Peter Robinson in 2000 (Robinson 2000-; see also Robinson 2013) we chose to use the former denominator throughout this article.

<sup>2</sup> In the UK, for instance, the four Higher Education Funding Bodies (HEFCE, SFC, HEFCW, and DELNI) have recently 'introduced an open access requirement in the post-2014 Research Excellence Framework' (Higher Education Funding Council for England 2014). As long as they do not hinder the research itself, initiatives like these should be applauded. Thankfully, this policy does not apply to 'creative or practice-based research outputs, or data', meaning that it does not pose a threat to proposals for digital scholarly editions of copyrighted materials.

<sup>3</sup> This list was copied from the slides Robinson shared on slideshare (2013a). More information about each of these desiderata can be found in Robinson's abstract for the presentation (2013b). As may be expected, these desiderata correlate to a large extent with the six principles of the *Textual Communities* project he is currently working on (2012).

<sup>4</sup> For more information on the advantages of the CC BY-SA license over the CC BY-NC license, Robinson referred to an online article by Erik Möller titled 'Creative Commons –NC License Considered Harmful' (2005).

<sup>5</sup> For a more detailed account of Stallman's argument for open source software, see Stallman 2002.

<sup>6</sup> 'Should the second author decide to use the work outside conditions specified in the license – for example, to modify, distribute, or copy the original work without sharing the source code or the modifications made – copyright takes back its power' (Dusollier 2002, 286).

<sup>7</sup>As is stated on the BnF website, non-commercial reuse of the library's materials is allowed as long as their source is properly attributed (see: <http://gallica.bnf.fr/html/conditions-dutilisation-des-contenus-de-gallica>). The notebook Pierazzo and André used for their prototype can be found here: <http://gallica.bnf.fr/ark:/12148/btv1b6000131k>.

<sup>8</sup>For a general overview of how long it takes different types of publications to enter the public domain in different countries, the James Joyce Foundation's webpages on 'Joyce & Copyright' can be highly recommended, where some of these figures were taken from (<https://joycefoundation.osu.edu/joyce-copyright>). Another especially interesting reference in this context is Robert Spoo's contribution to *Woolf Editing/Editing Woolf* titled "'For God' s sake, publish; only be sure of your rights': Virginia Woolf, Copyright, and Scholarship" that will be referenced further down in this paper. The Society of Authors (which acts as the representative of the Woolf Estate, among others) also has a useful collection

of guides and articles on copyright issues – mostly related to the situation in the UK (see <http://www.societyofauthors.org/guides-and-articles>). For a detailed and regularly updated summary of when works enter the public domain in the USA, Cornell University's 'Copyright Term and the Public Domain in the United States' can be a useful resource (see: <https://copyright.cornell.edu/resources/publicdomain.cfm>).

<sup>9</sup> For more information, see: <http://www.woolfonline.com/?node=about/copyright>.

<sup>10</sup> See, for instance John Unsworth's 1997 article 'Documenting the Reinvention of Text: The Importance of Failure', or – more recently – Lisa Spiro's "'This Is Why We Fight": Defining the Values of the Digital Humanities' (2012). The fact that 'Best Exploration of Failure in DH' was a category in the most recent edition of the DH Awards also points in this direction (see: <http://dhawards.org/dhawards2014/voting/>).

<sup>11</sup> ODD files can be exported into different file formats (such as HTML and PDF) through Roma (<http://www.tei-c.org/Roma>) or OxGarage (<http://www.tei-c.org/oxgarage/#>), for example.

<sup>12</sup> DiXiT (an acronym for DIgital sCHolarly editing Initial Training network) is an international network of high-profile institutions from the public and the private sector that are actively involved in the creation and publication of digital scholarly editions. For more information, please visit the network's website: <http://dixit.uni-koeln.de>.

<sup>13</sup> See <http://www.copyright.gov/title17/92chap1.html#107>. Building on this description, some other jurisdictions (such as those of the Commonwealth) use the more limited term 'fair dealing' instead (see, for instance, §29 of the UK's 'Copyright Designs and Patents Act, 1988': <http://www.legislation.gov.uk/ukpga/1988/48/section/29>). In his essay on copyright restrictions for scholarship pertaining to the works of Virginia Woolf, Robert Spoo reveals that one of the ways in which the 'fair dealing' doctrine is more restrictive than the 'fair use' doctrine is that it 'is sometimes treated as *not* applying to unpublished writings' (2009, 228; emphasis in original). §107 of the US Copyright law, in contrast, explicitly states that '[t]he fact that a work is unpublished shall not itself bar a finding of fair use'.

<sup>14</sup> Sections of Richard Stim's *Getting Permission* (including the one on fair use from which this citation was taken) can be freely consulted in the 'Copyright Overview' on the 'Copyright & Fair Use' section of Stanford University's website (see: <http://fairuse.stanford.edu/overview/>).

<sup>15</sup> Some suggestions in this direction can be found in Richard Stim's *Getting Permission*, more specifically in the book's section on 'Disagreements Over Fair Use: When Are You Likely to Get Sued?' (for an online version of this section, see: <http://fairuse.stanford.edu/overview/fair-use/disagreements-over-fair-use-when-are-you-likely-to-get-sued/>). A more elaborate assessment of the different factors that come into play to determine whether or not a use of copyrighted materials can be counted as 'fair' can be found on the University of Texas' 'Copyright Crash Course' page on fair use (see: <http://copyright.lib.utexas.edu/copypol2.html>).

<sup>16</sup> For more information on ESTS, please visit [www.textualscholarship.eu](http://www.textualscholarship.eu).