ADVANCES IN DIGITAL SCHOLARLY EDITING

ADVANCES IN DIGITAL SCHOLARLY EDITING

Sidestone Press

ADVANCES IN DIGITAL SCHOLARLY EDITING

PAPERS PRESENTED AT THE DIXIT CONFERENCES IN THE HAGUE, COLOGNE, AND ANTWERP

edited by
PETER BOOT
ANNA CAPPELLOTTO
WOUT DILLEN
FRANZ FISCHER
AODHÁN KELLY
ANDREAS MERTGENS
ANNA-MARIA SICHANI
ELENA SPADINI
DIRK VAN HULLE

© 2017 Individual authors

Published by Sidestone Press, Leiden
www.sidestone.com

Imprint: Sidestone Press

Lay-out & cover design: Sidestone Press
Cover illustration: Tessa Gengnagel

ISBN 978-90-8890-483-7 (softcover)
ISBN 978-90-8890-484-4 (hardcover)
ISBN 978-90-8890-485-1 (PDF e-book)

Contents

Welcome **11**

Preface **13**

Introduction **15**
Peter Boot, Franz Fischer & Dirk Van Hulle

WP1 Concepts, Theory, Practice

Towards a TEI model for the encoding of diplomatic charters. The charters of the County of Luna at the end of the Middle Ages **25**
Francisco Javier Álvarez Carbajal

The uncommon literary draft and its editorial representation **31**
Mateusz Antoniuk

Data vs. presentation. What is the core of a scholarly digital edition? **37**
Gioele Barabucci, Elena Spadini & Magdalena Turska

The formalization of textual criticism. Bridging the gap between automated collation and edited critical texts **47**
Gioele Barabucci & Franz Fischer

Modelling process and the process of modelling: the genesis of a modern literary text **55**
Elli Bleeker

Towards open, multi-source, and multi-authors digital scholarly editions. The Ampère platform **63**
Christine Blondel & Marco Segala

Accidental editors and the crowd **69**
Ben Brumfield

Toward a new realism for digital textuality **85**
Fabio Ciotti

Modelling textuality: a material culture framework **91**
Arianna Ciula

Multimodal literacies and continuous data publishing. Une question de rythme **99**
Claire Clivaz

Theorizing a digital scholarly edition of *Paradise Lost* **105**
Richard Cunningham

The digital libraries of James Joyce and Samuel Beckett **109**
Tom De Keyser, Vincent Neyt, Mark Nixon & Dirk Van Hulle

Editing the medical recipes in the Glasgow University Library Ferguson Collection **115**
Isabel de la Cruz-Cabanillas

The archival impulse and the editorial impulse **121**
Paul Eggert

Pessoa's editorial projects and publications. The digital edition as a multiple form of textual criticism **125**
Ulrike Henny-Krahmer & Pedro Sepúlveda

Reproducible editions **135**
Alex Speed Kjeldsen

'… but what should I put in a digital apparatus?' A not-so-obvious choice. New types of digital scholarly editions **141**
Raffaella Afferni, Alice Borgna, Maurizio Lana, Paolo Monella & Timothy Tambassi

Critical editions and the digital medium **145**
Caroline Macé

Scholarly editions of three rabbinic texts – one critical and two digital **149**
Chaim Milikowsky

From manuscript to digital edition. The challenges of editing early English alchemical texts **159**
Sara Norja

Towards a digital edition of the Minor Greek Geographers **165**
Chiara Palladino

Digital editions and materiality. A media-specific analysis of the first and the last edition of Michael Joyce's *Afternoon* **171**
Mehdy Sedaghat Payam

Challenges of a digital approach. Considerations for an edition of Pedro Homem de Mello's poetry **177**
Elsa Pereira

The born digital record of the writing process. A hands-on workshop on digital forensics, concepts of the forensic record and challenges of its representation in the DSE **183**
Thorsten Ries

Enduring distinctions in textual studies **187**
Peter Shillingsburg

Blind spots of digital editions. The case of huge text corpora in philosophy, theology and the history of sciences **191**
Andreas Speer

Data driven editing: materials, product and analysis **201**
Linda Spinazzè, Richard Hadden & Misha Broughton

Making copies **213**
Kathryn Sutherland

The Videotext project. Solutions for the new age of digital genetic reading **219**
Georgy Vekshin & Ekaterina Khomyakova

A stemmatological approach in editing the Greek New Testament. The Coherence-Based Genealogical Method **223**
Klaus Wachtel

WP2 Technology, Standards, Software

What we talk about when we talk about collation **231**
Tara L. Andrews

The growing pains of an Indic epigraphic corpus **235**
Dániel Balogh

The challenges of automated collation of manuscripts **241**
Elli Bleeker, Bram Buitendijk, Ronald Haentjens Dekker, Vincent Neyt & Dirk Van Hulle

The role of digital scholarly editors in the design of components for cooperative philology **249**
Federico Boschetti, Riccardo Del Gratta & Angelo Maria Del Grosso

Inventorying, transcribing, collating. Basic components of a virtual platform for scholarly editing, developed for the Historical-Critical Schnitzler Edition **255**
Stefan Büdenbender

Combining topic modeling and fuzzy matching techniques to build bridges between primary and secondary source materials. A test case from the King James Version Bible **261**
Mathias Coeckelbergs, Seth van Hooland & Pierre Van Hecke

The importance of being… object-oriented. Old means for new perspectives in digital textual scholarship **269**
Angelo Mario Del Grosso, Emiliano Giovannetti & Simone Marchi

Edition Visualization Technology 2.0. Affordable DSE publishing, support for critical editions, and more **275**
Chiara Di Pietro & Roberto Rosselli Del Turco

Compilation, transcription, multi-level annotation and gender-oriented analysis of a historical text corpus. Early Modern Ducal Correspondences in Central Germany **283**
Vera Faßhauer

***Hybrid scholarly edition* and the visualization of textual variants** **289**
Jiří Flaišman, Michal Kosák & Jakub Říha

Burckhardtsource.org: where scholarly edition and semantic digital library meet **293**
Costanza Giannaccini

EVI-linhd, a virtual research environment for digital scholarly editing **301**
Elena González-Blanco, Gimena del Rio, Juan José Escribano, Clara I. Martínez Cantón & Álvaro del Olmo

Critical diplomatic editing. Applying text-critical principles as algorithms **305**
Charles Li

St-G and DIN 16518, or: requirements on type classification in the Stefan George edition **311**
Frederike Neuber

Visualizing collation results **317**
Elisa Nury

The Hebrew Bible as data: text and annotations **323**
Dirk Roorda & Wido van Peursen

Full Dublin-Core Jacket. The constraints and rewards of managing a growing collection of sources on omeka.net **333**
Felicia Roșu

Of general and homemade encoding problems **341**
Daniela Schulz

The role of the base manuscript in the collation of medieval texts **345**
Elena Spadini

A tailored approach to digitally access and prepare the 1740 Dutch *Resolutions of the States General* **351**
Tuomo Toljamo

Editorial tools and their development as a mode of mediated interaction **357**
Tuomo Toljamo

TEI Simple Processing Model. Abstraction layer for XML processing **361**
Magdalena Turska

WP3 Academia, Cultural Heritage, Society

Edvard Munch's Writings. Experiences from digitising the museum **367**
Hilde Bøe

Crowdfunding the digital scholarly edition. Webcomics, tip jars, and a bowl of potato salad **375**
Misha Broughton

Editing medieval charters in the digital age **383**
Jan W. J. Burgers

Editing copyrighted materials. On sharing what you can **391**
Wout Dillen

What you c(apture) is what you get. Authenticity and quality control in digitization practices **397**
Wout Dillen

The journal al-Muqtabas between Shamela.ws, HathiTrust, and GitHub. Producing open, collaborative, and fully-referencable digital editions of early Arabic periodicals – with almost no funds **401**
Till Grallert

Digital editions of artists' writings. First Van Gogh, then Mondrian **407**
Leo Jansen

Digital editing: valorisation and diverse audiences **415**
Aodhán Kelly

Social responsibilities in digital editing – DiXiT panel. Editing and society: cultural considerations for construction, dissemination and preservation of editions **421**
Aodhán Kelly

Documenting the digital edition on film **427**
Merisa Martinez

Towards a definition of 'the social' in knowledge work **433**
Daniel Powell

Beyond Open Access. (Re)use, impact and the ethos of openness in digital editing **439**
Anna-Maria Sichani

The business logic of digital scholarly editing and the economics of scholarly publishing **449**
Anna-Maria Sichani

The social edition in the context of open social scholarship. The case of the Devonshire Manuscript (BL Add Ms 17, 492) **453**
Ray Siemens

Nowa Panorama Literatury Polskiej (New Panorama of Polish Literature). How to present knowledge in the internet (Polish specifics of the issue) **463**
Bartłomiej Szleszyński

Digital Rockaby **467**
Katerina Michalopoulou & Antonis Touloumis

Editing copyrighted materials

On sharing what you can[1, 2]

Wout Dillen[3]

Paper presented at 'Academia, Cultural Heritage, Society' DiXiT Convention, Cologne, March 14-18, 2016.

At the moment, the issue of copyright in digital scholarly editing is a big mess. In the first place, this is the case because copyright law is territorial. In other words: something that is completely legal in one country could be illegal in another. On top of that, these laws change over time, and usually only in the wrong direction (Epstein 2002; Goss 2007). And then there are cases where an especially interesting work threatens to enter the public domain, and the copyright holders start to devise mechanisms for keeping their control a little longer – see, for instance, the recent copyright debate regarding the diaries of Anna Frank (Flood 2016). As scholarly editors, we are expected to take all these issues into account while developing a digital edition, especially if we want our work to be as widely accessible as possible. But this is actually very difficult to do because the legal framework we are supposed to follow is so heterogeneous and prone to change. Rather than going into all the ins and outs of copyright law, however, this conference paper is written from a more pragmatic point of view. Say you find yourself in a situation where the materials you want to edit are still protected by copyright. What are you going to do? How

---

1 This conference presentation was based on a longer and more detailed treatment of the topic, published in *Digital Scholarship in the Humanities*. This article 'Digital Scholarly Editing within the Boundaries of Copyright Restrictions' by Wout Dillen and Vincent Neyt was first published online on 4 March 2016 by DSH. The article can be consulted at: https://doi.org/10.1093/llc/fqw011.

2 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/ under ERC grant agreement n°31360911. In addition, the author received funding to participate in this conference from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 317436.

3 wout.dillen@uantwerpen.be.

can we try to work creatively within the boundaries of those copyright restrictions, to make as much of our academic output available to the largest possible audience?

Background

By exploring how much of their data scholarly editors can still share *within* the boundaries of copyright, this paper responds to the popular sentiment in academia that copyrighting academic output is bad practice, and that the results of scientific research should be made freely available to the general public by default. Good examples of this idea are the series of blog posts and *The Guardian* columns by paleontologist Mike Taylor who claims that 'Hiding your research behind a paywall is immoral' (2013), or the decision by the UK's Higher Education Funding Bodies to introduce 'an open access requirement in the next Research Excellence Framework' (Higher Education Funding Council for England, 2015). Although these examples were written mainly with more traditional academic output formats in mind, they can be applied easily to the academic output of scholarly editors, the most recent realization of which is the digital scholarly edition.

And indeed, at the ADHO conference in Lincoln Nebraska in 2013, scholarly editor Peter Robinson posited a list of desiderata for scholarly editions, that included the dictum that 'All materials (in a scholarly edition) should, by default be available by a Creative Commons share-alike license' (2013). Commendable though Robinson's 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 this statement. While I certainly agree that Robinson's desideratum is a laudable goal that should be pursued whenever possible, the fact remains that it is often unattainable. By arguing that all digital scholarly editions 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. 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 those data. To illustrate this point, the paper zooms in on the practices of a couple of projects that work with copyrighted materials, to see how they deal with this situation.

Curation

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. Obviously, the ideal situation here would be that the texts of the 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 (see for instance Pierazzo and André 2012). But for scholarly editors working with even more recent materials, conditions are usually less favourable. 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, memory institutions, publishers, and funders.

The same holds true for the *Beckett Digital Manuscript Project* (BDMP; *Samuel Beckett. Krapp's Last Tape / La Dernière Bande: a digital genetic edition*) for example. At the moment, the BDMP has secured a contract regarding the acquisition of scans with a number of holding libraries, and hopes to strike up even more, similar collaborations in the future. 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.

This is what makes a good relationship and workable contracts between Digital Scholarly Editing projects and authors' estates (or their representatives) so important: if reports of troubled relations between scholarly editors and the Joyce Estate for instance teach us anything, it is that a project on such a scale covering an important contemporary author simply would not be possible without it. Thankfully, in the case of the BDMP, 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. That is why the Beckett Estate agreed to give the directors of the BDMP the license to publish their genetic edition of Beckett's manuscripts online, as long as this happens behind a paywall that pays for the license. Still, *limited* access is 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 is granted institutional access to the edition, this means that there is still a considerable group of people who can access the edition free of charge.

Public Access without a Public License

A different approach to this problem is the one that the *Woolf Online* project has taken (Woolf 2013). 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. 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. But we have to keep in mind that scholarly edition does not meet the requirements of Robinson's third desideratum either. 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 and may not be copied, distributed, transmitted or otherwise altered or manipulated without

the express permissions' of the copyright owners.[4] 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. But the question remains how important this final step is to enable further research. Developing a Digital Scholarly Edition, 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 not strictly necessary.

Sharing What You Can

It can be recommended, however, to make academic data that is not bound to copyright restrictions publicly available under as public a license as possible. 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. 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. Like metadata, a project's documentation, for example, can contain a wealth of harmless information about the project that may be of use for other researchers. In the case of Digital Scholarly Editions, the TEI for example already provides a standard format to share this information: the ODD (or: One Document Does it all). A file that combines a TEI-XML validation schema with human readable schema documentation in a single XML file.

Recognizing the potential of this kind of information for other scholarly editing projects, the BDMP recently put its own documentation online for consultation in the form of a digital Encoding Manual.[5] The basis of this documentation was a cheat-sheet designed by my colleague Vincent Neyt that helped the project's contributors transcribe Beckett's manuscripts. In collaboration with Vincent, I have expanded this document to include validation information, more examples and explanatory text, more information on encoding practices in general, and the BDMP's ODD validation schema. This way, the project's collectively cultivated expertise does not go to waste, but may help to support (or even initiate) other research instead.

Fair Use

The above was written from a position that takes an absolute view on copyright: namely that nothing that is not 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

4 http://www.woolfonline.com/?node=about/copyright.

5 http://uahost.uantwerpen.be/bdmp/.

the fair use doctrine. 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.[6] 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. This is a strategy we have tried to apply for the *Lexicon of Scholarly Editing* we are developing at the Centre for Manuscript Genetics.[7]

Instead of writing its own definitions, this lexicon aggregates citations in academic outputs (journal articles, monographs, etc.) that are relevant to each of its entries. 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. In order to publish these citations under the purview of the fair use doctrine, we try to make sure that: (1) the materials are used for research purposes; (2) we only use small fragments – usually only a fraction of the full text; (3) these fragments are attributed consistently to their rightful owner; (4) the use of the fragments only furthers their original aims: namely to further scholarly research; and (5) all of the above is published in a transparent, non-commercial research environment.

Conclusion

In conclusion, I would argue that scholarly editors often still have plenty of room within the boundaries of copyright restrictions for the publication their Digital Scholarly Edition, and all of the ancillary data they produce along the way. This means 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 paywall, so be it. In that case, editors will have to accept the challenge to convince their readers that their corpus is worth consulting in spite of those limitations.

6 http://www.copyright.gov/title17/92chap1.html#107.

7 http://uahost.uantwerpen.be/lse/.

References

Epstein, Richard. 2002. 'The Dubious Constitutionality of the Copyright Term Extension Act.' *Loyola of Los Angeles Law Review* 36, 123-158. Accessed April 8, 2016. http://chicagounbound.uchicago.edu/cgi/viewcontent.gi?article=2359&context=journal_articles.

Flood, Alison. 2016. 'Anne Frank's diary caught in fierce European copyright battle'. *The Guardian, January 18*. Accessed April 8, 2016. http://www.theguardian.com/books/2016/jan/18/anne-franks-diary-caught-in-fierce-european-copyright-battle.

Goss, Adrienne K. 2007. 'Codifying a Commons: Copyright, Copyleft, and the Creative Commons Project.' *Chi. -Kent Law Review* 82: 963-996. Accessed April 8, 2016. http://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=3609&context=cklawreview.

Pierazzo, Elena, and Julie André. 2012. *Autour d'une séquence et des notes du Cahier 46: Enjeu du codage dans les brouillons de Proust*. Accessed April 8, 2016. http://research.cch.kcl.ac.uk/proust_prototype/index.html.

Robinson, Peter. 2013. '5 Desiderata for Digital Editions/Digital Humanists Should Get Out Of Textual Scholarship'. *Paper presented at the DH2014, Lincoln, Nebraska*, July 19. Accessed April 8, 2016. http://www.slideshare.net/PeterRobinson10/peter-robinson-24420126.

Samuel Beckett. *Krapp's Last Tape / La Dernière Bande: a digital genetic edition*. 2015, edited by Dirk Van Hulle and Vincent Neyt (The Beckett Digital Manuscript Project, module 3). Brussels: University Press Antwerp (ASP/UPA). Accessed April 20, 2016. http://www.beckettarchive.org.

Taylor, Mike. 2013. 'Hiding your research behind a paywall is immoral'. *The Guardian, January 17*. Accessed April 8, 2016. https://www.theguardian.com/science/blog/2013/jan/17/open-access-publishing-science-paywall-immoral.

Woolf Online. 2013, edited by Julia Briggs and Peter Schillingsburg. Accessed April 20, 2016. http://www.woolfonline.com/.