Transformissions in Trade Digitization and Shadow Libraries

A case study of Vladimir Nabokov's *Pale Fire*

Simon Rowberry

ABSTRACT

The field of Critical Digitization Studies primarily focuses on projects conducted by enthusiasts, pirates, or large technology companies. In this article, I introduce the concept of 'trade digitization', encapsulating the processes involved in creating an ebook or web-based copy of a print publication for the mass market. Through a case study of Vladimir Nabokov's Pale Fire*, I demonstrate how an increased degree of editorial decision making within trade digitization can introduce the proliferation of new variants and errors in digital publications.*

TRADE BOOK DIGITIZATION LAGS BEHIND OTHER MEDIA INDUSTRIES. MUSIC, television, and film had largely adapted to the material conditions of digital technology during the 2000s. The sharp shock of both Google Book's launch in 2004 and Amazon's release of the Kindle in 2007 demonstrated trade publishers' lack of preparedness for large-scale structural change.[1] In the interim period mass digitization projects such as HathiTrust and the success of self-publishing platforms such as Kindle Direct Publishing increased acceptance of digital publishing but the digitization of trade publications has stalled. For example, in my study of the availability of British books published in 1989 via the Kindle Store I found that only 2,697 (9%) of 30,940 titles were available as ebooks in 2018 (ROWBERRY

1. Outside of trade publishing, the industry was more prepared. Academic publishing made the transition to digital distribution earlier with projects such as JSTOR, founded in 1995, leading the way: see SCHONFELD 2003. Similarly, by the 1990s, journals were already experimenting with online open access distribution: see MOORE 2020, 856–66.

2022, 89). Trade digitization has received less scholarly attention than mass digitization despite its central role in book publishing.

In this article, I present a case study of the digitization of *Pale Fire* by Vladimir Nabokov, an author who had a complex relationship with publishing technology through his six-decade writing career. His 1962 novel *Pale Fire* has long been associated with digital publishing, dating back to Ted Nelson's interest in the novel as a proto-hypertext in the late 1960s (Nelson 1987, 2). Rather than expand upon the parallels between the novel's forms and hypertext structures, I instead focus on the transformative potential of digitization that occurs through both official and illicit channels.[2] A case study of a single novel enables us to focus on the smaller details of digitization's material impact rather than the more abstract level of a mass digitization project. Through this case study, I examine the tensions between the watershed moment of mass digitization projects and the under-scrutinized area of trade digitization. This approach complements existing research on mass digitization while showing the different material conditions of publication due for public consumption rather than other secondary purposes including, most prominently, text mining.

These tensions shape the core research question of this project: What effect does the digitization process of trade publications still protected by copyright have on the digital edition's textual and material form? In order to answer this question, I zoom in on two categories — folk and trade digitization — from a broader taxonomy of digitization summarised in Table 1. While there are a number of important differences between these types of digitization, I focus here on the specifics of the conversion and output process, whereby both trade and folk digitization focus on producing texts that are human readable through processing small batches of text rather than entire libraries. This approach emphasizes accuracy and requires a range of editorial interventions to convert the raw material of the scanned images or manually typed text into a viable published product.

2. See Bodó 2018, 53–78 and Thylstrup 2019, 79–100.

Table 1. A Taxonomy of Digitization

Type of Digitization	Definition	Examples
Mass	Large-scale projects that have a "relative lack of selectivity of materials" which are digitized at high speed.[3] As a result, the underlying quality of any scans can be low or only available as facsimiles.	Google Books HathiTrust
GLAM [Galleries, Libraries, Archives and Museums]	More focused digitization projects undertaken by cultural heritage organizations, often focusing on a particular genre or time period. Due to this level of discrimination, the overall quality of the finished product is higher.[4]	New York Public Library Menus Early English Books Online
Folk	The phenomenon of users digitizing their own material for personal use or sharing within a network, often because it is not available otherwise. Digitized material may be in the public domain or still protected by copyright. The latter is commonly known as a shadow library.	Project Gutenberg Lib.ru
Trade	Conducted by publishers or booksellers for the primary purpose of selling individual titles as a commercial ebook rather than a database of titles (such as EEBO or ECCO)	Rosetta Books[5] Amazon's early digitization efforts[6]

3. Thylstrup 2019, 18.
4. Coyle 2006, 642.
5. Adams 2002, 29–46
6. Rowberry 2022, 7–8.

The current project focuses on the intersection between the latter two forms of digitization as underdeveloped topics in discussion around the transition from print to digital publishing especially at the junction between samizdat distribution and the arrival of official digitized material for sale. My choice to decouple shadow libraries from mass digitization acknowledges the difference between large-scale projects posting scanned or pirated PDFs and the more bespoke work of smaller operations that nonetheless circulate material protected by copyright (Thylstrup 2019, 79–100).

In this article, I analyze the effects of digitizing Nabokov's *Pale Fire* as a case study of how trade and folk digitization editorialize the conversion process alongside the more well-studied phenomenon of introducing additional typographical and material errors in the various stages of scanning, character recognition, and correction. In this context, the term 'editorialize' encompasses human's intentional choices in designing a digital publication from the use of automated processes such as spell check through to how they choose to format and present the published text. Nabokov's 1962 novel is an exceptional case study because its complex non-linear structure offered a blank canvas for digitizers to extend when considering how computers offer new affordances for reading complex materials (Rowberry 2012; Ferguson 2013, 101–16). These kinds of decisions demonstrate the importance of curatorial or editorial aspects of digitization on a micro level that are more prominent in trade and folk digitizations than mass digitization projects.

Together with the material transformations of converting analog text to digital data legible by humans on-screen, we can see digitization as 'deformance'. In their article outlining the concept of deformance, Samuels and McGann state:

> A deformative procedure puts the reader in a highly idiosyncratic relation to the work. This consequence could scarcely be avoided, since deformance sends both reader and work through the textual looking glass. On the other side customary rules are not completely short-circuited, but they are held in abeyance, to be chosen among (there are many systems of rules), to be followed or not as one decides. Deformative moves reinvestigate the terms in which critical commentary will be undertaken.
>
> (Samuels and McGann 1999, 36)

It is possible to use this method as a basis for criticism as the "imaginative work has an elective affinity with performance: it is organized as rhetoric and *poiesis* rather than an exposition and information-transmission. Because this is so, it always lies open to deformative moves" (Ibid.). Samuels and McGann note four basic categories of deformance: reordering the text's structure; isolating parts of speech, alliteration, punctuation; altering typography or word choice; and adding to the original (Ibid., 36–37). The process of digitization is fertile territory for exploring the role of deformance in the publishing process. As Whitney Trettien argues, "OCR reprints mutate at the level of text — that is, text as manipulable content" (Trettien 2013). Part of this conversion process may involve explicitly deformative acts including changing word choices, especially when paired with automation that has not been manually corrected afterwards. As I argue in this article, deformance is a useful critical framework for understanding the human and machinic distortions of the digitization process.

The quality of transcription, whether undertaken manually or automatically, is vital to the process of creating useable digital texts. With the rise of Google Books and other high-yield, middling-quality digitization projects, "there is a risk therefore that focusing on the ability of early adopters to use large corpora with a low-quality threshold could encourage a policy of digitizing quickly at the expense of quality" (Gooding 2013, 425–31). In the rush to capitalize on the success of the Kindle, similar shortcomings have emerged again in trade digitization. Ryan Cordell argues that "dirty OCR illuminates the priorities, infrastructure, and economics of the academy in the late 20th and early 21st centuries" (Cordell 2017, 190). Since dirty OCR is an established part of folk and trade digitization, it can be productive to assess how it reflects on the more commercial aspects of print digitization.

Even the best OCR can require extensive correction in order to remain consistent with the print version. In ideal conditions, the most effective OCR software records a 99.98% success rate, or one error per 5000 characters, with text produced after 1950 (Tanner, Muñoz, and Ros 2009). Nabokov's English language works and translations were mostly published after the 1950s, but this still comes with the need for a detailed search for mistakes, as *Pnin,* one of Nabokov's shorter novels still contains almost 300,000 characters. In optimal conditions, this would lead to 60 incorrectly processed characters and given Nabokov's penchant for using multiple languages in a single novel, it is likely that OCR will not pick up some of the more irregular character patterns found in Russian or French

while browsing an English text.[7] For example, in the Penguin Kindle edition of *Pale Fire*, "était" has been replaced by "éait", an artefact of the digitization process rather than a transformission traceable back to its print counterpart.[8]

Nabokov's concerns for the reduced fidelity of his writing through the various stages of the publication process are not unfounded. The more manual or automated interventions in the process of publication, the probability of unintentional errors increases. Randall McLeod terms this phenomenon 'transformission' (Clod 1991, 241–81). As digital media is built upon a culture of copying, transformission is a common occurrence, especially at the point of converting text from one format to another, such as transferring the contents of a Word document into InDesign. A single Word manuscript might form the basis of print editions, EPUB, Kindle file, PDF, and potentially other formats, all of which require additional conversion processes.

Critical digitization studies, spurred by publications such as Nanna Bonde Thylstrup's *The Politics of Mass Digitization*, offer a complementary framework for understanding these kinds of transformissions (Thylstrup 2019). This can range from the macrostudy of an entire mass digitization project through to exploring a single text from multiple perspectives.[9] Within the context of ebooks, Alan Galey's analysis of the material differences between the Kobo and print edition of Johanna Skibsrub's *The Sentamentalist* demonstrates transformissions can occur even when an ebook and print edition start from the same born-digital files. For example, Galey notes that the Kobo edition still closes "with a note on an absent typeface" that has been replaced by Georgia, "a screen font developed by Matthew Carter for the Microsoft Corporation [that] completely reverses the typographic politics of the print edition" (Galey 2012, 228–29).

Bonnie Mak demonstrates that Early English Books Online's choice to digitize the microfilm surrogates of rare books rather than the original objects can exaggerate features present in the microfilm medium, which might not have been visible in the original conversion, rather than represent the original print (Mak 2014, 1515–26). Ryan Cordell argues for a deeper understanding of how the automatic process of digitization shapes digital editions. Cordell proposes to view Optical Character Recognition (OCR)

7. See Peng et al. 2013, 1–8, for an overview of the problems in multilingual OCR.
8. Nabokov 2011, loc. 3235.
9. See Gerlach and Font-Clos 2020, 126; see also Trettien 2013.

as "a species of compositor: prone to transcription errors, certainly, but nonetheless resetting the type of its proof texts into .txt or .xml files rather than a galley" (Cordell 2017, 196). This recasts the performance of digitization as dynamic and historically contextualized rather than declaring poor quality scanning as an anomaly of the digital era.

Nabokov and the Book Trade

Vladimir Nabokov provides a useful case study of an author embedded in what Rebecca Rosen has termed "the missing twentieth century", an era of publishing spanning the 1920s to 2000s that has reduced visibility in contemporary publishing as copyright legislation prevents reprints or digital distribution (Rosen 2012). Nabokov's career covers the missing twentieth century from the 1916 publication of *Stikhi* (*Poems*) to his last novel, *Look at the Harlequins!* in 1979, which arrived prior to the mainstream use of digital workflows in the publishing industry and therefore any digital copies would necessitate digitization. It should, however, be acknowledged that since the publication of *Lolita* in 1955, Nabokov has enjoyed a privileged position within the literary canon and cannot therefore be considered representative as a twentieth century author prior to the digitization discussed in this article, although occasionally some of his novels, including *Pale Fire*, have fallen out of print in the United Kingdom.

Other than *Pale Fire*, his work is relatively unexceptional materially and typographically, as he favored linguistic experimentation.[10] The materiality of Nabokov's work has only recently become of interest to scholars, as a result of the controversial posthumous publication of *The Original of Laura* as a perforated facsimile of his work-in-progress index cards, and more recently, the digital publication of his back catalog as ebooks. Nonetheless, his high profile and the unique nature of *Pale Fire* within his bibliography offer an opportunity to extend current debates in critical digitization studies.

The influence of outside agents is clear in Nabokov's fraught relationship with Maurice Girodias, the original publisher of *Lolita*. The book's

10. This is not to say that Nabokov has not been appropriated for digital humanities studies. There is a long history of digital humanities work on Nabokov despite his works remaining under copyright starting with Ted Nelson's hypertext prototype featuring *Pale Fire* and several innovative works based upon Nabokov's use of language. See also Parrish 2008 and Grant 2012.

association with Girodias's Olympus Press threatened to condemn the novel to the label of pornography alongside the publisher's backlist. The relationship between Girodias and Nabokov was not entirely antagonistic, however, since Girodias, a native French speaker, suggested that Nabokov remove superfluous French from the manuscript, an attempt at hypercorrection to avoid detection as the author of the planned anonymous publication (Nabokov 1990, 166). Dmitri Nabokov's footnote to his father's short story, "Sounds", highlights a similar non-authorial incursion: "The sharp-eyed bilingual reader familiar with the original Russian may note the substitution of 'hairless' for 'irresolute'. The two words resembled each other in Russian, and 'irresolute' was almost certainly the result of a copyist's slip" (Nabokov 1997, 19).

Nabokov responded to his new-found international fame after *Lolita* by critiquing the Anglo-European book trade in all his subsequent novels. Most of his protagonists are authors or agents of the book trade, which reached its pinnacle in his last published novel, *Look at the Harlequins!* in which Nabokov burlesques his own bibliography (Nabokov *1996,* 565–747). Nabokov's most poignant use of the book trade, however, comes earlier in his career, while he was writing *Lolita*, *Pale Fire,* and *Ada*, the three novels in which he most intensely scrutinizes the agents of print outside of the author. In these novels, Nabokov refers to the processes of composition and reception on nearly every page. Publishing professionals' appearances highlights the novels' 'unfinished' quality, they are presented to the reader without final copyediting, in Humbert Humbert's note for the printer of *Lolita:* "Lolita, Lolita, Lolita, Lolita, Lolita, Lolita, Lolita, Lolita, Lolita. Repeat till the page is full, printer", which is left ignored (Nabokov 2000, 109). Ada's fictional editors Ada and Van Veen also present their final text with marginal comments to each other intact, such as "Hue or who? Awkward. Reword! (marginal note in Ada Veen's late hand)" embedded directly into the narrative (Nabokov 1990, 12). The theme of the book trade is more central to *Pale Fire* than the two other books. Unlike *Lolita* and *Ada's* embedded theme of the book trade which must be teased out through close rereading, *Pale Fire* revolves around a literary heist and the politics of publishing and creating a scholarly edition of a posthumous text.

It is worth briefly recapitulating the complex plot of *Pale Fire*. The American poet John Shade has been killed just before finishing his greatest poem, "Pale Fire", an autobiographical treatise on mortality, eschatology, and his daughter's suicide. Sybil Shade, the poet's widow, has entrusted her husband's unfinished work to Shade's friend, Charles

Kinbote, for publication with his scholarly exegesis to immediately boost the reputation of the poem. The novel is structured as a scholarly edition: John Shade's poem is sandwiched by Kinbote's foreword, commentary, and an index. The reader of Kinbote's commentary soon realizes he is an inattentive commentator who ignores standard scholarly exposition to instead focus on Shade's hidden theme: Zembla and its exiled king, Charles Xavier. Kinbote's reckless editing reaches its pinnacle when he asserts, without evidence, that Shade's intentions were to repeat the first line at the end of the poem, foreshadowing the poet's murder. As the commentary unravels, Kinbote gleefully drops hints about his big secret, that he is in fact the exiled king of Zembla, who is indirectly responsible for Shade's death as the bumbling assassin, Gradus, hits Shade in the crossfire of gun shots meant to kill Zembla's former king. It is more likely, however, that the murderer is the escaped convict, Jack Grey, who mistakenly shoots John Shade, presuming he is Grey's incarcerator and Shade's neighbour: Judge Goldsworth. The spectre of a mad Professor Botkin problematizes the notion of Zembla, as a possible true identity of the narrator, who has dreamed up Zembla while spying on Shade and obtaining the manuscript through illegitimate means.

While *Pale Fire*'s composition and publication history was simpler than Nabokov's struggle to bring *Lolita* to market,[11] its afterlife matches, if not outpaces, his most famous text. Putnam's text from 1962 was contracted to Weidenfeld and Nicolson for hardback rights who subsequently licensed the text in the UK to Penguin and Random House. In the United States, in 1989 Vintage included Nabokov's corrections to the original Putnam publication and Brian Boyd's editorial work for the Library of America editions has created the most reliable version of *Pale Fire* so far. There have been over fifty editions in various formats of the novel. Figure 1 (below) summarizes the transmission history of the major Anglophone editions of the text, noting where various trade and folk digitization projects sourced their material.

This stemmatization reveals a divergence in the development of digital editions. One peculiar example is Dmitri Nabokov's corrections to the Penguin edition of *Pale Fire* when renegotiating the contracts in May 1993. Miranda McAllister later suggested in an internal memorandum that Dmitri's corrections meant they did not have to offset the more up-to-date Vintage edition.[12] Dmitri notes five of his father's twelve corrections to

11. See Boyd 1991, chapters 10–16.

12. Miranda McAllister to Mark Handsley, personal correspondence, 18 November 1995, DM2083/13/1. Penguin Archives, University of Bristol.

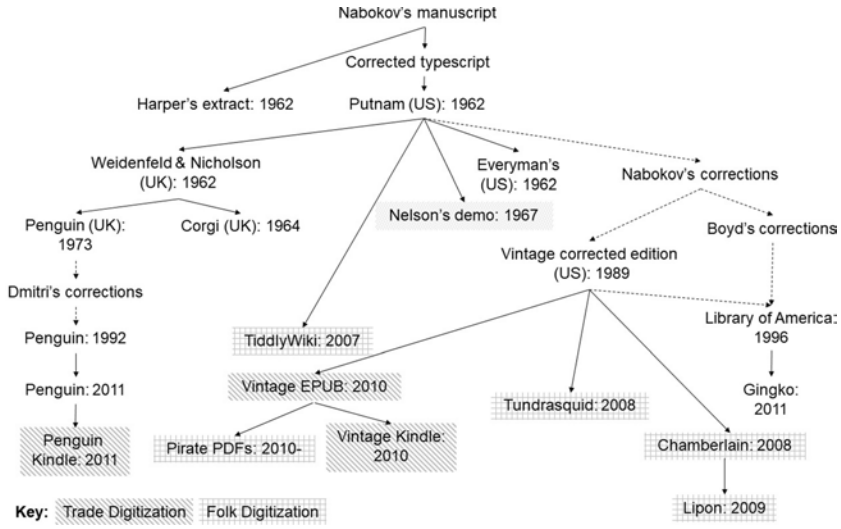


Figure 1. Stemmatization of editions of *Pale Fire*.

the 1962 version of the novel previously integrated into the 1989 Vintage as well as two further corrections, "Catskin" should be changed to "Catkin"[13] — incidentally a remnant from the 1962 version that was not noted as a correction by Nabokov — and a correction of the capitalization of "for example",[14] an artefact from earlier publications that remained uncorrected in 2000 and 2010 editions.[15] The former was the effect of an overzealous proofreader, while the latter has remained a constant artefact that has often gone unnoticed in editing *Pale Fire*, as it is not included in either Nabokov's or Boyd's list of corrections. Print versions prioritize the consistency of the poem as the centrepiece of the novel compared to the many variants that occur round the text. The transmission history of the novel, as with many other texts, is further complicated by additional reprints from the offsetting of the first British edition to the cheap Corgi paperback reprint which required resetting, leading to the potential for further typographic errors.

13. Nabokov 1995, Commentary to line 171, 547.
14. Ibid., Commentary to line 627, 609.
15. Dmitri Nabokov, corrections to Penguin *Pale Fire*, May 1993, DM2083/13/1. Penguin Archives, University of Bristol.

The most popular editions in both the United Kingdom (Penguin) and United States (Vintage) have been reprinted over ten times. Textual variation has inevitably crept in with these various editions and their association with ebook versions in Britain and the United States respectively has further complicated this situation. The Penguin Kindle ebook is of particular interest, as early editions did not feature an index. The print version it is derived from, the 2011 eleventh reprint, features an index, as does the iBooks version, which is based upon the same print copy. It is unclear why the early Kindle version, one of the most popular digital formats, would not have the index, a vital part of the novel's fictional conceit. Without an index, readers do not experience the full text through the additional links in the index as well as the mysterious last entry: "*Zembla*, a distant northern land", which appears without page numbers. On the manuscript index card, Nabokov writes a circled instruction to the typesetter: "nothing more, no ref" (Quoted in Roth 2015, 24). Further to this, Nabokov had erased a comment that the "index [is] not completed by the late Dr. ^Vseslav^ Botkin" which Roth notes would have removed much of the mystery of the text (Roth 2015, 25). It is clearly important to include the index as the dénouement of the text. Penguin has since updated the Kindle edition to include an index but it is prefaced by a generic disclaimer about ebook indexes:

> *The page references in this index correspond to the printed edition from which this ebook was created. To find a specific word or phrase from the index, please use the search feature of your ebook reader.*[16]

The *Pale Fire* index does not use page references, however, as the following paragraph retains the original text's explanation that the index does not reference pages but rather the line of commentary, which remain consistently numbered through the print and digital editions. The insertion of this generic disclaimer, that can be found across various Penguin titles with indexes, reveals the tensions between personalising each ebook title and automatically inserting medium-specific information regardless of the context.

Nabokov's Transformissions

Digitization places a greater focus on the interpretive complexities of transformission, as users of digital repositories are made aware of the

16. Nabokov 2011, loc. 4228.

differences between print and its digital surrogates. It is no coincidence that textual materiality is enjoying a surge in popularity with the rise of digital humanities: digitization projects constantly remind us of the source text's materiality through its perceived absence in digital form. Digitization provides the material interface between a print artefact and a digital object and must be taken seriously in understanding a text's material transmission.

Practices of trade digitization are often obscured as part of commercially sensitive processes. For example, Amazon has not revealed the exact processes involved to ensure the availability of a large back catalog of 88,000 ebooks for the Kindle's launch in 2007. The most accurate records of their digitization practices can be seen in patent filings including "Method and system for automatically classifying page images" and "Creating an electronic book using video-based input".[17] These patents demonstrate Amazon have developed advanced techniques for speeding up the process of OCR through composites of video frames rather than taking individual images, as well as their reliance on human intervention when automation fails. It is unclear, however, how much digitization work Amazon has conducted since the release of the Kindle 1. These patents offer a rare peak into the development of trade digitization which is substantially different to the more widely discussed mass digitization. Extracting text becomes more important while there is less of a focus on provenance, for example.

There has been a shift in focus with digitization projects. Amy Earhart notes that the late 2000s marked "a rapid decrease of new digital recovery projects. The proliferation of early activist projects by scholars interested in recovery [through digitization] dropped precipitately as literary scholarship shifted its focus away from canonicity" (Earhart 2015, 76). In a footnote to this comment, Earhart argues: "The advent of Google Books would have an impact on digitization projects [. . .] Texts that libraries do not purchase are not digitized, which leads to gaps in coverage, particularly gaps in coverage of unrepresented authors" (Earhart 2015, 142). Similarly, the agendas behind trade digitization shape the future potential for scholarly digitization projects, as funding bodies are unlikely

17. Bradley Jeffery Behm and Brent Eric Wood, Method and system for automatically classifying page images, United States Patent no. 8,306,326, issued 6 November 2012. Vasant Manohar, Sridhar Godavarthy, and Viswanath Sankaranarayanan, Creating an electronic book using video-based input, United States Patent no. 9,191,554, filed 14 November 2012, and issued 17 November 2015.

to support applications to create higher fidelity copies of works that have already been digitized.

Publishers or online booksellers are responsible for trade digitization, although this can be outsourced to third parties who specialize in the rapid digitization of texts. While publishers who digitize their backlists provide clear value for their businesses, the phenomenon of booksellers digitizing titles is more puzzling. Retailers who wished to create a critical mass of titles for their ebook platforms could not rely on publishers but had to digitize texts with the expectation that sales of the title will be sufficient to make a profit. Amazon's early digitization project for the Kindle required Amazon to retain exclusive rights for ebook sales of titles digitized by the company. Fortunately, given the scale and resources of Amazon compared to small publishers, this form of digitization was only prevalent in the early development of the Kindle (2004–2008) and since the value of the ebook market has been established, independent publishers have begun to digitize their backlists, although this can still be costly and require external partnerships such as MIT Press's collaboration with Arcadia and the Internet Archive (MIT 2017).

Given the commercial objectives of these scans, there is not the same duty of care as with other digitization projects. As a result, despite the general improvements in scanning and correcting texts, a trade digitized edition may feature errors as the result of computational processes that have further destabilized the presumed authority of print. The goal of trade digitization is to produce a readable copy of a text at the lowest cost rather than a high-quality edition which is often required for further analysis. It is now more difficult for an author, or their estate, to preserve carefully edited and compiled versions of texts as copies proliferate through authorized and illicit editions in cheap paperback and digital forms. The texts that need to be digitized from analog formats are at highest risk as the chances for transcription errors are multiplied through the conversion process.

In the absence of an authoritative digital edition of a text, variants can proliferate across reading platforms as publishers and readers seek to digitize their own version of the text. This is no different for Nabokov's works. Through both intention and historical accident, Nabokov's pre-publication materials survive only as fragments, so his extra-textual and post-publication revisions offer an opportunity for genetic analysis from the point of publication to the latest editions. The current project emerges from this tension between the impossibility of controlling one's own reception and Nabokov's constant concern for the stability of his literary output. While the Nabokov Estate has been reasonably successful

in maintaining a modicum of control over print editions of his works, the emergence of digital editions and ebooks introduces new and unexpected variations (Leving and White 2013). Accurate digital transcription is expensive, with no guarantee that large numbers of readers will purchase the digital text. This is exacerbated by the need to constantly update ebooks when errors are reported to avoid the title being removed from the bookstore (Elsey 2016).

Post-1922 digitization projects are often disadvantaged by a lack of authoritative data about the process of a title's transmission history. The digitization of early modern publications started with cataloguing efforts of the Short Title Catalogue that led to Early English Books Online. Progress was slow but methodical, ensuring the digitizers' workflow was robust and effective. Conversely, this detailed work of collation, cataloguing, and mapping copy variation has yet to begin for many post-war authors although it can bring up unexpected discoveries such as discrepancies between the US and British publication of David Mitchell's *Cloud Atlas* (Eve 2016, 1). The act of digitization for texts published before the widespread adoption of digital publishing workflows introduces further variants through the various mechanical and manual processes involved in creating a finished publication.

Editing a Digital *Pale Fire*: Beyond plain text

A digital edition of *Pale Fire* highlights some of the editorial challenges present in conversion that are often ignored at the expense of typographic errors: How do you render material and typographic features in digital form? The origins of a hypertext edition of *Pale Fire* precede most other digitizations of Nabokov's text and could have produced a moment equal to Douglas Engelbart's "Mother of All Demonstrations" (Rowberry 2011, 319–24). Ted Nelson was working on a prototype hypertext system, Hypertext Editing System (HES) at Brown University with Andreas Van Dam and additional support from IBM in the late 1960s. Nelson picked *Pale Fire* alongside a technical manual and Vannaver Bush's "As We May Think" to demonstrate the potential abilities of HES. Ted Nelson defined *Pale Fire* as "a real hypertext", and suggested "many authors have used this form for different purposes, none more strikingly than Nabokov in Pale Fire" (Nelson 1968, 2, 3). After Nelson's first efforts, many other readers proposed that a digital edition should be created, but the majority of these proposals were never complete or published but they included a variety of

platforms including Jimmy Gutermann's Storyspace version and Charles Cave's Perl adaptation.[18] Jerry Friedman and Matthew Roth conceived of a wiki *Pale Fire* concordance to allow users to plot core themes throughout the narrative (Roth 2006). The wide uptake of the Web has led to further digital editions, often made public. Rather than worrying about sharing for copyright reasons, the Web coupled with DMCA takedown culture created an environment where users felt safe to post their own full-text versions of *Pale Fire*.

The problem of a digital edition of *Pale Fire* has been tackled by many, but it is worth considering in greater detail the roadblocks and problems they have faced when conceptualizing their projects. Most significantly, international copyright law still protects *Pale Fire*, since it was published just 60 years ago. This has not stopped some of the proposals suggesting that the Nabokov Estate would be welcome to contact them to discuss sharing the document more publicly — an unlikely proposition. That is not to say that the Nabokov Estate would reject what it considers to be a worthy proposal, since they granted Ted Nelson permission to use *Pale Fire* and Brian Boyd's *Ada Online* offers a free copy of Nabokov's longest novel. It is no surprise given Dmitri Nabokov's constant fight with post-USSR collapse pirates that the earliest trace of a web-based *Pale Fire* emanates from a Russian site.[19] Following the release of the Russian web-based hypertext, a spate of web editions created for personal use have been shared, in part because there was not an authorised ebook edition until 2012.

Since there are no scholarly digital editions of Nabokov's works, digital copies are the result of either trade or folk digitization. Although trade ebooks are published, and therefore legitimized, by legacy publishing houses, they are not straightforward facsimiles or surrogates for the print version. For example, Alan Galey notes Kindle editions initially open at the beginning of the first chapter or introduction, ignoring any front matter thus distorting the reader's perception of the text (Galey 2012, 216). Nabokov's works have not survived this conversion intact, as the index of Penguin's edition demonstrates.

The form of the page in the print edition is equally important in making connections, as occasionally links can be made on the same page or opened recto and verso. In the 2000 Penguin edition, for instance, the notes to line 376 and 347–48 both refer to T.S. Eliot in close proximity,

18. See Guterman 1994; see also Cave 2001.
19. Nabokov 1990, x.

creating an unintentional link through what Mark Bernstein terms "spatial hypertext" (Bernstein 2011, 103–12) or connections made through visual structural clues, although Kinbote's commentary to line 376 does make the connection implicitly by stating "I believe I can guess (in my bookless mountain cave) what poem is meant; but without looking it up I would not wish to name its author."[20] As the 2000 Penguin edition was the first copy of the text I encountered, I had assumed that this was an intentional use of the spatial arrangement of text on the page in order to undermine Kinbote's editorial ability, but the larger typography in the Putnam's edition pushes the second gloss over the page demonstrating it was instead unintentional. The three texts with the greatest authorial and editorial interventions — the Putnam (and off-prints), Vintage and Library of America editions — do not feature the resonance as the reader must turn the page in order to read the second note. Of the other editions examined, the cheap reprints all feature the two annotations on the same page or spread. In the digital editions, all but the ebooks chunk up the nodes into separate pages, so the connection is harder to make and the ebooks rely on the reader having the settings in a customized, and unusual, typographic setup to see similar results. This example demonstrates the importance of the edition one encounters in interpretation and the user's own experience with the text. It is likely the reader would meet this juxtaposition anyway with a linear reading as there are no cross-references blocking the way other than another note on John Shade's rhyming patterns.[21] The two-dimensional page allows the reader to make these serendipitous connections that may not be possible if flattened out by the screen.

Towards a taxonomy of trade digitization errors

The process of digitization produces a wealth of errors that can be the result of automatic or manual interventions, which often produce radically different outputs. These can be intentional: for example, an editorial style guide might require a change in the formatting that leads to an error when the ebook is compiled. This is often less discussed than the unintentional changes that are not caught in the production process. In this section, I outline five of these categories that I encountered in the various editions of *Pale Fire* discussed in this article: OCR, manual transcription, spell check,

20. Nabokov 1995, 578.

21. Nabokov 1995, Commentary to lines 367–70, 577.

Knuthian, and mark-up errors. This typology is not comprehensive but demonstrates that we should consider more factors than just OCR. These errors reveal the challenges for creating an accessible, readable digital copy of a text originally produced for print.

OCR Errors

Unfortunately, the process of digitization is not as straightforward as taking photographs of the book and automatically generating a perfect digital copy of the text (Drucker 2009, 150–51). OCR is the process of transforming an image of text into a machine readable, and searchable, copy but its success rate may be low with errors going uncorrected unless a human checks the machine converted version against the original. The OCR algorithm must first identify the constituent characters of a word (for example, separating 'dear' into four discrete symbols: 'd', 'e', 'a', and 'r') and then find the closest match for each character in context through training data. The detection of individual characters is an important first step as the 'd' glyph could easily be misinterpreted as 'cl'. This is a nontrivial error: for example, 'clear' and 'dear' may both be used as adjectival modifiers, and 'a dear day' is instead rendered 'a clear day', turning an expensive day out into a clear sky. This transcription error must be corrected manually. While sophisticated OCR algorithms can note the likelihood of a character being poorly transcribed, in this case it is likely to go undetected.

This is a common problem for authorized publishers too since the original text was set mechanically, and one can see errors creeping into the transcription process for the ebook versions of the novel. Richard Davis's TiddlyWiki adaption, the oldest English-language digital version aptly derives from the original Weidenfeld edition as the page breaks are identical in both texts despite TiddlyWiki not requiring any. Davis leaves further traces of the automation of his production as some of the faulty OCR has remained in the finished product such as "tile window", "lor tile", "experincced it", "tile skill", and "111y fall". The OCR evidently has problems with distinguishing the difference between "the" and "tile" and "My" and "111y" in the small type of the Weidenfeld edition. Whitney Trettien argues such marks exist as the zombie remnants of print culture as the characters that undergo digitization must initially be treated as images, rather than manipulable digital objects (Trettien 2013). This digital textuality shapes the edition and requires human intervention that is often lacking at the level of the folk edition. The

algorithms running behind the OCR can handle line breaks, but when it comes to italicized text or the differences between "M", "W" and related to symbols, the mechanisms begin to break down. This needs careful human post-processing as these errors may not necessarily be noted by automated procedures.

Manual Transcription Errors

The other folk projects do not feature such explicit OCR errors, but manual transcription can produce more human errors. We can see the marks of manual transcription through such errors as "andti-Karlist", which emerge from the typist's muscle memory of the *3-gram* "and" that is less likely to occur with OCR errors (Davis 2007). We can also see other traces of the copying process, and none more clear than the Lipon edition. Both Chamberlain and Tundrasquid use the Vintage edition, but Lipon does not declare which edition he used to create his version. A collation of Lipon's edition compared to Chamberlain's reveals the presence of many typographical errors that have been replicated by the two that indicate that Lipon has copied Chamberlain's text and only corrected certain errors. Lipon has carefully corrected the errors in the poem, as his digital edition is an attempt to showcase his poetic reading, but the commentary does not receive the same attention. Over 350 of the errors unique to Chamberlain also appear in Lipon's version and only 10% of errors which originate from Chamberlain's edition have been corrected in Lipon's update. Moreover, a closer examination of the presence of an absurd variant in both editions is the result of copying rather than serendipity. In the opening paragraph to Kinbote's commentary, he states "We can visualize John Shade . . . [as] he picks up from the turf that compact ovoid body"[22] while this has been transformed to "he picks up from the turd" in both the Chamberlain and Lipon. These marks suggest that Lipon has reused Chamberlain's cleaned version of the text rather than attempting the intensive process of digitization himself. This reveals the importance of proofreading on the reading experience. A similar phenomenon can be noted in the pirated PDFs, which originate from the 2010 Vintage eBook as the metadata remains on both the copyright page and the file information. If the digitizer starts with a text other than the original Putnam edition in consultation of Nabokov's corrections, the edition is likely to proliferate pre-existing errors.

22. Nabokov 1995, Commentary to lines 1–4, 486.

Spell/Grammar Check Errors

The probabilistic techniques that determine the quality of OCR are mirrored by the typical reliance on spell checking software in the process of folk and trade digitization. This post-processing also mediates digitization projects which may have relied on manual transcription. Spell check software is one example of the challenges facing Natural Language Processing (NLP).[23] These projects often explore the complexity of these statistical methods, but it is sufficient here to instead focus on Peter Norvig's lightweight Python 3 implementation of a Bayesian spell check algorithm that replicates the basic functionality of Google's autocomplete mechanism (Norvig 2016). Norvig's choice of variables — 'deletes', 'transposes', 'replaces', 'inserts' — serendipitously evokes Samuels and McGann's list of potential deformative acts, albeit in this case to attempt to improve the quality of the original. The algorithm works through querying how closely a chosen word comes to matching the same letter patterns in a pre-defined dictionary. Words are then ranked according to their closeness and popularity based upon a word list populated by corpora from Wikipedia, Project Gutenberg, and the British National Corpus. Norvig claims the results are accurate 80–90% of the time despite its small file size. These probabilistic spell check methods can have a significant impact on digitized works that depend on them. For example, the aforementioned "turd"/"turf" mistake in the Chamberlain/Lipon versions likely stems from an unfortunate instance of automated spell check in a *Pale Fire* digitization (Chamberlain 2007).

The Tundrasquid version reveals the most information about its composition history as the metadata in the HTML indicates that the project began its life in Microsoft Frontpage 4.0, and this is confirmed by a couple of the incidentals that read as though they have been run through Microsoft Office spellcheck such as "student who" being transformed into "student that". The other HTML versions do not document their provenance so explicitly but can be compared on the length of the pages themselves. Tundrasquid and Lipon separate their nodes of the commentary into separate pages, while Chamberlain has included the section as a single page. Chamberlain's layout is more amenable to serendipitous discovery as the reader can navigate the page, or even search for terms, and discover new connections, while having the separate chunks gets rid of some of this ability.

23. See Hodge and Austin 2003, 1073–81 and Choundhury et al. 2007, 81–88 for overviews of complex methods of spell checking.

Spelling and grammar check errors indicate another important axis in understanding digitization for human consumption: the style guide. Publishers use either in-house or industry standard style guides to ensure their products read as if part of their brand. These guidelines are malleable and may be subject to change, especially in digital form. Teresa Elsey documents the challenges of the constantly shifting standards which might include editorial updates (for example, changing Web to web), updating incorrect or out-of-date information, or altering the formatting to account for the latest design conventions (Elsey 2016). This creates a tension between maintaining the perceived fidelity of the print edition or creating a digital publication that adapts to contemporary standards.

While the style guide is a larger part of trade publishing, digital tools such as the word processor can accelerate automation in adopting these changes where necessary such as in the overly judicious use of find and replace to ensure a text is more compliant to the publisher's style guide. For example, adopting American English from a document produced in British English through automated means may use a 'greedy' find and replace tool that changes every instance of 'our' to 'or' even if it is not the suffix of a word such as 'color'.[24] While there are no documented instances of this phenomenon within my corpus, it is related to the tools embedded within the conversion process beyond producing a plain text or facsimile file.

Knuthian Errors

Davis's TiddlyWiki adaption reveals another error type in an attempt to turn the original text into the fictional index cards described in Kinbote's commentary. Davis's compromise is to use premade software to achieve the desired effect, which necessarily restricts some of the possibilities. Nonetheless, Davis runs into problems when he attempts to reconcile the structure of Shade's index cards in the same format as Kinbote's commentary. The materiality of the index card suits the format of the short node, but Kinbote's Foreword is chunked up according to the pages of the 1962 Putnam edition which include errant line breaks, a phenomenon which did not need to be accurately replicated. Similarly, many of the incidentals from both print and digital versions revolve on the justified line breaks of the original text, which introduced variants such as "mid-summer".[25]

24. Regular expressions (or 'regex' for short) offer a solution to this problem, but their technical barriers to entry are much higher and their use requires a good understanding of any edge cases within a given document.

25. Nabokov 1962, 19.

I call these line-break transcription errors "Knuthian", a reference to Donald Knuth's extensive work on the 'hyphenation' algorithm for his TeX markup language that would automatically include these line breaks for print-ready publications.

The error has only been corrected in the few places that have examined the original manuscript or corrections such as the Vintage and Library of America editions. Where the manuscript is consistent with the published version, it is clear that many of the ambiguous hyphenated line breaks were printing artefacts rather than Nabokov's intentions, who tended to be lax with his use of hyphenation. The most extreme uptake of this is the original Penguin Kindle edition, which has been directly transferred from the print version, leaving line dashes in place that were primarily included for the page. These are marks that remind the reader that they are engaging with a text produced before mass digitization and it is easy to keep these traces of print culture in later editions simply because they appear to be correct.

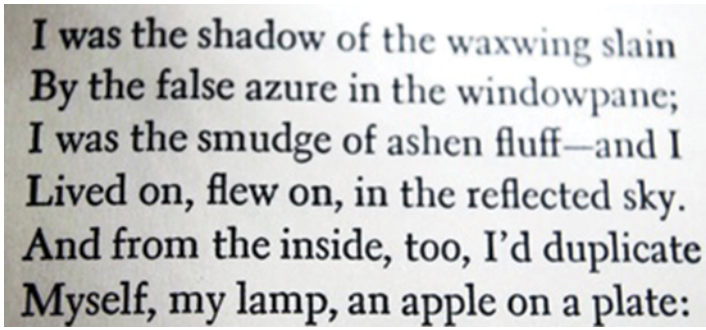
Mark-up Errors

Figure 2 shows the potential for errors in the presentation of the text as an ebook. In the print edition, the poem is set in standard Roman script, but both the Google Play Store and Kindle editions introduce extra italics and links. There is a further distinction between the two digital editions, as the first line of the Kindle edition does not feature italics due to differences in how the two format's compilers render the same source material. [26]

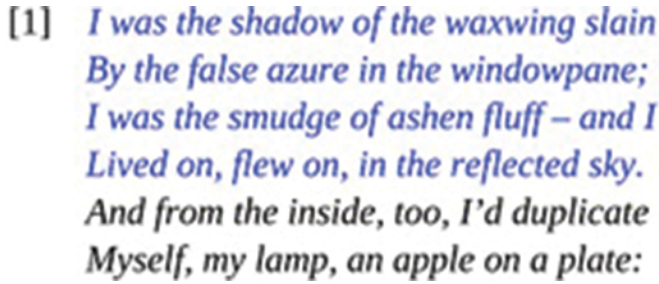
An examination of the Kindle edition's formatting reveals the source of the error is the result of both inattentive and overly fastidious post-digitization editing.[27] The problem emerges from overzealous link tagging rather than intentionally incorrect formatting. Every link in *Pale Fire* has been rendered bi-directional in this ebook. This normally does not cause a problem, but since the first and last lines of the poem are the same, the first line features links to two separate locations in the commentary concurrently. These links both appear in the HTML, despite a lack of support for multi-directional linking within the specification that produces the visual and technical errors. The original EPUB file maintains the two links, but once interpreted by software, ignores the second links completely. Conversely, the Kindle edition was validated by KindleGen

26. Nabokov 2011.

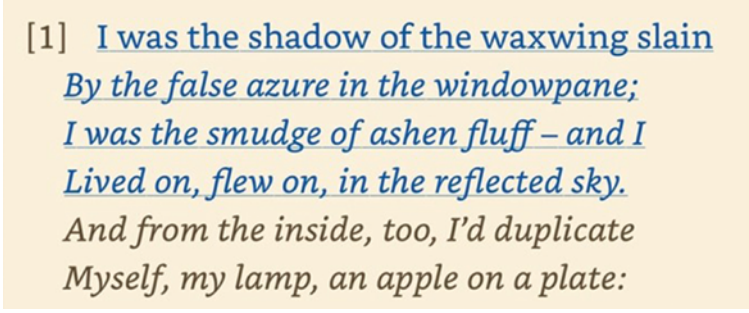
27. As interpreted by KindleStrip, which reinterprets Kindle ebook file formats into their original EPUB form.


I was the shadow of the waxwing slain
By the false azure in the windowpane;
I was the smudge of ashen fluff—and I
Lived on, flew on, in the reflected sky.
And from the inside, too, I'd duplicate
Myself, my lamp, an apple on a plate:

2a


[1] *I was the shadow of the waxwing slain*
By the false azure in the windowpane;
I was the smudge of ashen fluff – and I
Lived on, flew on, in the reflected sky.
And from the inside, too, I'd duplicate
Myself, my lamp, an apple on a plate:

2b


[1] I was the shadow of the waxwing slain
By the false azure in the windowpane;
I was the smudge of ashen fluff – and I
Lived on, flew on, in the reflected sky.
And from the inside, too, I'd duplicate
Myself, my lamp, an apple on a plate:

2c

Figure 2. The transformission of the opening six lines of Vladimir Nabokov's *Pale Fire* from the original publication (Putnam 1962) to the Google Play Books (Nabokov 2011) and Kindle Penguin Editions.

which is a compulsory step to upload any material to the Kindle Store.[28] The parser noted the error and attempted to fix it. Through this fix, the shorter link has simply been placed outside of the first, and since the tags are embedded within the links, they were also accidentally removed from the first line of the Kindle edition. This is an example of a *format-specific parser* error whereby an automated final check of an ebook or website can flag issues and offer automatic solutions that alter the content if not carefully checked.

Conclusion

The digitization-based transformissions of *Pale Fire* I have discussed reveal how analysis of trade digitization must consider editorial and platform issues rather than the narrower focus on OCR. A digitized text that extends beyond a photographic facsimile involves interpretive decisions about formatting and other material elements. This formatting will most commonly appear in the form of a mark-up language which can be either presentational (font, textual structure, media, page layout) or semantic (document structure, metadata, revision history). While scholarly digitization often focuses on semantics that can be reassembled into various publication types, trade digitization instead emphasizes the presentation of a text for remediation as a PDF or ebook. The latter approach can often ignore what the text means and requires close attention to each individual format, which may be helpful for digital indexing and other tools that are often lauded as benefits of reading on-screen. Indexing mechanisms such as the Kindle's X-Ray are often superficial in their use of semantics and do not recreate the richness of a print index.[29] Rather than bemoaning the poor quality of many of these commercially produced ebooks, it is instead worth considering the processes that led to these conditions.

University College London

28. It is impossible to create a native Kindle file, as Amazon have not released a complete toolset for creating them. Instead, publishers can create something approximating the Kindle file and then run it through the latest convertor.
29. Bothma and Prinsloo have demonstrated the consequences of on-the-fly semantic evaluation of content for the Kindle's built in dictionary lookup function and shown why overlooking parts of speech markup and other features familiar to many academic digitization projects but sadly lacking from trade digitization can result in a poor quality publication: see Bothma and Prinsloo 2013, 165–98.

Works Cited

Adams, Caryn J. 2002. "Random House v. Rosetta Books". *Berkeley Technology Law Journal,* 17.1: 29–46.

Behm, Bradley Jeffery, and Brent Eric Wood. 2012. Method and system for automatically classifying page images. United States Patent no. 8,306,326, issued 6 November.

Bernstein, Mark. 2011. "Can We Talk about Spatial Hypertext?" In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia*, edited by P. M. E. De Bra and K. Gronbak, 103–12. New York: ACM.

Bodó, Balázs. 2018. "Library Genesis in Numbers: Mapping the Underground Flow of Knowledge". In *Shadow Libraries: Access to Knowledge in Higher Education*, edited by Joe Karaganis, 53–78. Cambridge: MIT Press.

Bothma, Theo J.D., and D.J. Prinsloo. 2013. "Automated Dictionary Consultation for Text Reception: A Critical Evaluation of Lexicographic Guidance in Linked Kindle e-Dictionaries". *Lexicographica,* 29.1: 165–98.

Boyd, Brian. 1991. *Vladimir Nabokov: The American Years*. Princeton: Princeton University Press.

Cave, Charles. 2001. "Creating a hypertext version of Pale Fire". Posting to NABOKV-L. 8 May 2001.

Chamberlain, Shannon. 2007. "Pale Fire Project". http://www.shannonrchamberlain.com/palefiremain.html.

Choundhury, Monojit, Markose Thomas, Animesh Mukherjee, Anupam Basu, and Niloy Ganguly. 2007. "How Difficult Is It to Develop a Perfect Spell-Checker? A Cross-Linguistic Analysis through Complex Network Approach". In *Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing,* edited by Chris Biemann, Irina Matveeva, Rada Mihalcea, and Dragomir Radev, 81–88. New Jersey: Association for Computational Linguistics.

Clod, Random. 1991. "Information on Information". *TEXT: Transactions of the Society for Textual Scholarship,* 5: 241–81.

Cordell, Ryan. 2017. "'Q i-Jtb the Raven': Taking Dirty OCR Seriously". *Book History,* 20: 188–225.

Coyle, Karen. 2006. "Mass Digitization of Books". *The Journal of Academic Librarianship,* 32.6: 641–45. https://doi.org/10.1016/j.acalib.2006.08.002.

Davis, Richard. 2007. "Using TiddlyWiki as a Hypertext Engine for *Pale Fire*". *PALE FIRE by Vladimir Nabokov — An Experimental Hypertext Edition* (blog), 18 January. http://palefire.tiddlyspot.com/.

Drucker, Johanna. 2009. *Speclab: Digital Aesthetics and Projects in Speculative Computing*. Chicago: The University of Chicago Press.

Earhart, Amy E. 2015. *Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies*. Ann Arbor: University of Michigan Press.

Elsey, Teresa. 2016. "When Nothing Ever Goes Out of Print: Maintaining Backlist Ebooks". *Medium* (blog), 6 April. https://medium.com/@teresaelsey/when-nothing-ever-goes-out-of-print-maintaining-backlist-ebooks-fcd63e680667.

Eve, Martin Paul. 2016. "'You Have to Keep Track of Your Changes': The Version Variants and Publishing History of David Mitchell's Cloud Atlas". *Open Library of Humanities,* 2.2 (10 August): 1. https://doi.org/10.16995/olh.82.

Ferguson, Andrew. 2013. "Mirror World, Minus World: Glitching Nabokov's *Pale Fire*". *Textual Cultures,* 8.1: 101–16.

Galey, Alan. 2012. "The Enkindling Reciter: E-Books in the Bibliographical Imagination". *Book History,* 15.1: 210–47.

Gerlach, Martin, and Francesc Font-Clos. 2020. "A Standardized Project Gutenberg Corpus for Statistical Analysis of Natural Language and Quantitative Linguistics". *Entropy,* 22.1: 126. https://doi.org/10.3390/e22010126.

Gooding, Paul. 2013. "Mass Digitization and the Garbage Dump: The Conflicting Needs of Quantitative and Qualitative Methods". *Literary and Linguistic Computing,* 28.3: 425–31.

Grant, Angelynn. 2012. "MIT Design Project: Navokov's Signs and Symbols". http://angelynngrant.com/nabokov/.

Guterman, Jimmy. 1994. "Nabokov hypertext." Posting to Alt.hypertext. 28 March 1994.

Hodge, V. J., and J. Austin. 2003. "A Comparison of Standard Spell Checking Algorithms and a Novel Binary Neural Approach". *IEEE Transactions on Knowledge and Data Engineering,* 15.5: 1073–81.

Leving, Yuri, and Friedrick White. 2013. *Marketing Literature and Posthumous Legacies: The Symbolic Capital of Leonid Andreev and Vladimir Nabokov.* Lanham: Lexington Books.

Mak, Bonnie. 2014. "Archaeology of a Digitization". *Journal of the Association for Information Science and Technology,* 65.8: 1515–26.

Manohar, Vasant, Sridhar Godavarthy, and Viswanath Sankaranarayanan. 2015. Creating an electronic book using video-based input. United States Patent no. 9,191,554, filed 14 November 2012, and issued 17 November 2015.

MIT Press. 2017. "MIT Press Announces a New Collaboration". Boston: MIT Press. https://mitpress.mit.edu/blog/mit-press-announces-new-collaboration.

Moore, Samuel A. 2020. "Revisiting 'the 1990s Debutante': Scholar-Led Publishing and the Prehistory of the Open Access Movement". *Journal of the Association for Information Science and Technology,* 71.7: 856–66.

Nabokov, Vladimir. 1962. *Pale Fire*. New York: G. P. Putnam's Sons.

———. 1990. *Vladimir Nabokov: Selected Letters 1940–1977.* Edited by Dmitri Nabokov and Matthew J Bruccoli. London: Weidenfeld and Nicolson.

———. 1997. *Collected Stories*. London: Penguin.

———. 1995. "Pale Fire". In *Novels 1955–1962: Lolita, Pnin, Pale Fire, Lolita: A Screenplay*, edited by Brian Boyd, 437–668. New York: Library of America.

———. 1996. "Look at the Harlequins!". In *Nabokov: Novels 1969–1974,* 565–747. New York: Library of America.

———. 2000. *The Annotated Lolita*. edited by Alfred Appel Jr. London: Penguin.

———. 2011. *Pale Fire [Kindle Edition]*. London: Penguin..

Nelson, Theodor. 1968. "IBM Demo: Oral script and stage directions". Brown University. https://archive.org/details/ibd-1967.

———. 1987. *Dream Machines*. Redmond: Tempus.

Norvig, Peter. 2016. "How to Write a Spelling Corrector". http://norvig.com/spell-correct.html.

Parrish, Allison. 2008. "Nabokov's Synaesthesia: A Text Transformer". http://itp.nyu.edu/~ap1607/a2z/week01/.

Peng, Xujun, Huaigu Cao, Srirangaraj Setlur, Venu Govindaraju, and Prem Natarajan. 2013. "Multilingual OCR Research and Applications: An Overview". In *Proceedings of the 4th International Workshop on Multilingual OCR*, 1–8. New York: ACM.

Rosen, Rebecca. 2012. "The Missing 20th Century: How Copyright Protection Makes Book Vanish". *The Atlantic*, 30 March. http://www.theatlantic.com/technology/archive/2012/03/the-missing-20th-century-how-copyright-makes-books-vanish/255282/.

Roth, Matthew. 2015. "The Composition of *Pale Fire*". *Nabokov Online Journal,* 9: 24. https://www.nabokovonline.com/uploads/2/3/7/7/23779748/2_noj_9_2015_roth_composition_of_pale_fire.pdf.

Rowberry, Simon Peter. 2012. "'His and My Reader': Rereading *Pale Fire* Hypertextually". *Nabokov Online Journal,* V–VI.

———. 2022. *Four Shades of Gray: The Amazon Kindle Platform*. Cambridge: MIT Press.

Samuels, Lisa, and Jerome J. McGann. 1999. "Deformance and Interpretation". *New Literary History,* 30.1: 25–56.

Schonfeld, Roger C. 2003. *JSTOR: A History*. Princeton and Oxford: Princeton University Press.

Tanner, Simon, Trevor Muñoz, and Pich Hemy Ros. 2009. "Measuring Mass Text Digitization Quality and Usefulness: Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century Online Newspaper Archive". *D-Lib Magazine,* 15.7/8. https://doi.org/10.1045/july2009-munoz.

Thylstrup, Nanna Bonde. 2019. *The Politics of Mass Digitization*. Cambridge: MIT Press.

Trettien, Whitney Anne. 2013. "A Deep History of Electronic Textuality: The Case of English Reprints Jhon Milton *Areopagitica*". *DHQ: Digital Humanities Quarterly*, 7.1. http://www.digitalhumanities.org/dhq/vol/7/1/000150/000150.html.