Introduction

Words that Count

You may or may not have heard, but over the past two decades a secret and dangerous movement has been growing in humanities departments around the world. Sapping all of the conventional funding out of traditional humanistic pursuits, the so-called 'digital humanities' (or 'DH' to those in the know) brings a grim entrepreneurialism and technocratic mindset to English, history, classics, archaeology—and any other disciplinary space on which it can lay its hands. Seemingly charged with perverting the humanistic foundations of critical thinking and replacing them with techno-solutionist mindsets, the digital humanities are growing and thriving beneath our noses and many seem not even to have noticed the danger. As Roald Dahl wrote of his 'Great Automatic Grammatizator', we will need 'strength, Oh Lord', to resist the machine and the lure of such capital and technology. Dahl's narrator requests the courage to remain pure to art for art's sake, to resist the pull of technology. Contemporary literary critics also need the strength not to surrender to the promise of abundant riches in the digital domain. Give us the strength, Dahl posits, in spurning these new digital forms, 'to let our children starve'.[1]

I jest somewhat. But the study of literature with the aid of computers is undoubtedly controversial. Critics have derided digital methods in literary studies for being: useless (they tell us nothing that we did not already know); trivial (counting the word 'whale' in *Moby-Dick* can tell us only one thing: how often the word 'whale' is used in

[1] Roald Dahl, 'The Great Automatic Grammatizator', in *Someone Like You* (Harmondsworth: Penguin, 1986), pp. 190–209 (p. 209).

Moby-Dick); neoliberal (producing software is the Silicon Valley model of scholarship); and just plain wrong.[2] Proponents, by contrast, have pronounced forcefully on the possibilities for broad-scale literary history beyond the limitations on reading made by the finite human lifespan; on how we can better understand genre and form through visualization and spatialization; and even on the fresh perspectives such methods might bring for rethinking core theoretical assumptions about literature itself.[3] The digital humanities are certainly provocative and divisive.

However, one of the first misconceptions that requires a response lies in the equation of 'the digital humanities' with digital literary

[2] Timothy Brennan, 'The Digital-Humanities Bust', *The Chronicle of Higher Education*, 15 October 2017 <http://www.chronicle.com/article/The-Digital-Humanities-Bust/241424> [accessed 2 November 2017]; Daniel Allington, Sarah Brouillette, and David Golumbia, 'Neoliberal Tools (and Archives): A Political History of Digital Humanities', *Los Angeles Review of Books*, 2016 <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/> [accessed 29 May 2016]; Nan Z. Da, 'The Computational Case against Computational Literary Studies', *Critical Inquiry*, 45.3 (2019), 601–39 <https://doi.org/10.1086/702594>.

[3] For just a selection, see Lisa Samuels and Jerome J. McGann, 'Deformance and Interpretation', *New Literary History*, 30.1 (1999), 25–56 <https://doi.org/10.1353/nlh.1999.0010>; Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (London: Verso, 2007); Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism*, Topics in the Digital Humanities (Urbana, IL: University of Illinois Press, 2011); Franco Moretti, *Distant Reading* (London: Verso, 2013); Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History*, Topics in the Digital Humanities (Urbana, IL: University of Illinois Press, 2013); Tanya E. Clement, 'Text Analysis, Data Mining, and Visualizations in Literary Scholarship', in *Literary Studies in the Digital Age: An Evolving Anthology*, 2013 <https://dlsanthology.mla.hcommons.org/text-analysis-data-mining-and-visualizations-in-literary-scholarship/> [accessed 6 September 2017]; Ray Siemens and Susan Schreibman, eds., *A Companion to Digital Literary Studies*, Blackwell Companions to Literature and Culture (New York: Wiley-Blackwell, 2013); Melissa M. Terras, Julianne Nyhan, and Edward Vanhoutte, eds., *Defining Digital Humanities: A Reader* (Farnham: Ashgate Publishing, 2013); David M. Berry and Anders Fagerjord, *Digital Humanities: Knowledge and Critique in a Digital Age* (Cambridge: Polity Press, 2017); Andrew Piper, *Enumerations: Data and Literary Study* (Chicago, IL: University of Chicago Press, 2018); Martin Paul Eve, *Close Reading With Computers: Textual Scholarship, Computational Formalism, and David Mitchell's* Cloud Atlas (Stanford, CA: Stanford University Press, 2019); Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago, IL: University of Chicago Press, 2019). I will turn more thoroughly to examine these critiques later in this chapter.

studies. For digital approaches to the study of literature are not the same as '*the* digital humanities'. Indeed, as Eric Weiskott eloquently puts it, 'digital technology doesn't transform knowledge in one single way, programmatically, any more than print technology did'.[4] There is, then, really no such thing as a singular 'digital humanities'. As these technologies do not work systematically within any single epistemology in any single unified way on any single set of scholarly objects, it is essential to note that they also work across and within different disciplinary spaces. Historians, archaeologists, classicists, media scholars, ethnographers, theologians, and anthropologists are as likely to call themselves digital humanists as are the (in) famous advocates of distant reading in literary studies.[5] Those who speak of the digital humanities, in the singular, can all too often erase the specificity of disciplinary work outside of their own field. Those in literary studies can be among the worst culprits for this offence.

This is a book, then, that addresses specifically the questions in literary studies that computational methods and technological analyses may answer. I aim to deliver an introduction and overview of developing intersections between digital methods and literary studies to serve as a starting point for those who wish to learn more about the possibilities and the limitations of oft-touted digital humanities in the literary space. The volume intends to engage with the proponents of digital humanities and its detractors alike, aiming to offer a fair and balanced perspective on this controversial topic. This book fuses an introductory background approach and survey with original literary research. It should, therefore, be able to straddle the divide between seasoned digital experts and interested newcomers. That said, by way of a positional disclaimer: I am enthusiastic about the possibilities of

[4] Eric Weiskott, 'There Is No Such Thing as "The Digital Humanities"', *The Chronicle of Higher Education*, 1 November 2017 <https://www.chronicle.com/article/There-Is-No-Such-Thing-as/241633> [accessed 17 November 2018].

[5] For more on this and a range of examples, see Sarah E. Bond, Hoyt Long, and Ted Underwood, '"Digital" Is Not the Opposite of "Humanities"', *The Chronicle of Higher Education*, 1 November 2017 <http://www.chronicle.com/article/Digital-Is-Not-the/241634> [accessed 2 November 2017].

digital methods for literary studies even while recognizing the anxieties around their development.[6]

Many of the fears about digital humanities also stress the term *distant*—as in so-called 'distant reading', the quantitative study of literary texts—with disdain. They worry that the use of computers will take us further away from the joy of reading. As *my* provocation in this book, I instead argue that digital methods can bring us *closer* to literary texts; to give us a new viewport through which to observe their narratives. I even go so far here as to extend this to analogue counterparts of 'digital' approaches, such as tabulation and mapping. *Activities* associated with building databases and digital artefacts, even when conducted non-digitally, can be a way newly to engage with literary works. I attempt to demonstrate this argument through the novel case studies that appear in this book but also with reference to the extant work of others.

One of the most interesting things about literary studies, though, is that although it is now a core humanities subject in the Anglophone Global North, it is not, in fact, actually that old. A disruptive discipline that achieved ascendency to a central place in universities worldwide in just a century and a half, not coincidentally during the rise of the British Empire, 'English language and literature' was founded only in 1828 at University College London.[7] (Although, notably, Birkbeck, UCL's older sibling university, taught literary studies in a higher education context as early as 1823. Many Scottish institutions also had literary texts on their curricula before this date.) Over time English has undergone many mutations and methodological U-turns. Moreover, despite protests from revisionist historians of our discipline,

[6] This enthusiasm may stem from my background as a computer programmer. Conversely, the anxiety arises from my position within literary studies. Some might claim that my advocacy reflects a desire to bring a rare domain knowledge of quantification to the field. Am I merely bringing expertise that I have, but many others do not, to change the older field of literary studies for the worse? Although I would also note that I have conducted much non-digital literary critical work.

[7] Ted Underwood, *Why Literary Periods Mattered: Historical Contrast and the Prestige of English Studies* (Stanford, CA: Stanford University Press, 2013), p. 81; see also Franklin E. Court, *Institutionalizing English Literature: Culture and Politics of Literary Study,* 1750–1900 (Stanford, CA: Stanford University Press, 1992); Gerald Graff, *Professing Literature: An Institutional History* (Chicago, IL: University of Chicago Press, 1989).

there have also long been quantitative practitioners within the space of literary studies.

For instance, Vernon Lee, the famed Victorian and Edwardian-era aesthetician, called for a quantitative analysis of literature—a 'statistical experiment'—in her *The Handling of Words* (1923) after a debate with Emil Reich.[8] The extension of literary mathematics into computational approaches also occurs far earlier in our disciplinary history than many credit. Dartmouth College, for one, offered a module to students entitled 'Literary Analysis by Computer' in 1969.[9] Certainly, a quantitative strain of literary studies—and its extension into computational modes—has been present for quite some time.

Admittedly, the digital era of mass access to computation and the internet—not even mentioning digitized texts—has accelerated the presence of this quantifying urge and brought with it a host of new possibilities but also challenges for literary studies. Among the drivers has been the proliferation of electronic literature, electronics within literature, and their study. For instance, Jessica Pressman notably examined how many contemporary e-literatures—that is, texts born and published digitally to take advantage of electronic affordances—re-work modernist texts to yield 'immanent critiques of their technocultural context'.[10] Further, other well-known scholars such as N. Katherine Hayles have joined the analysis of how contemporary print novels function as texts that emulate or anticipate the possibilities for digital literature, exemplified in Mark Z. Danielewski's *House of Leaves* (2000).[11] Zara Dinnen has also recently shown how digital technologies have become 'banal' in contemporary fiction, rendering us 'unaware of the ways we are co-constituted as subjects with

[8] Nicholas Dames, *The Physiology of the Novel: Reading, Neural Science, and the Form of Victorian Fiction* (Oxford: Oxford University Press, 2007), p. 188.

[9] Annette Vee, '"Literary Analysis by Computer" Offered at Dartmouth, Winter 1969, Working with Paradise Lost. #1960sComputing', *@anetv*, 2017 <https://twitter.com/anetv/status/919219418189660160/photo/1> [accessed 18 October 2017].

[10] Jessica Pressman, *Digital Modernism: Making It New in New Media*, Modernist Literature & Culture, 21 (New York: Oxford University Press, 2014), p. 156.

[11] See N. Katherine Hayles, *Writing Machines*, Mediawork Pamphlet (Cambridge, MA: MIT Press, 2002); Jessica Pressman, 'House of Leaves: Reading the Networked Novel', *Studies in American Fiction*, 34.1 (2006), 107–28 <https://doi.org/10.1353/saf.2006.0015>.

media'.[12] These 'distributed media systems' approaches to literature featuring the electronic, or literature that is born and read within the electronic environment, are of ever-increasing prominence.[13]

That said, some commentators have insisted that digital humanities are not 'to be understood as the study of digital artifacts, new media, or contemporary culture in place of physical artifacts, old media, or historical culture'.[14] I cannot agree entirely with this assessment. In this book, I do hone in on how digital methods—the 'methods of the medium' in Richard Rogers's phrasing—can be applied to literature, whether digital or print.[15] However, I also focus on how digital media condition the possibilities of those literatures. In other words, at various points in this book I read digital artefacts or works that contain digital elements, using more conventional literary critical methods. However, in conjunction with this, I will, of course, turn to how digital tools can bear on those literary artefacts. By necessity, this nonetheless involves some boundaries of exclusion. We *all* use digital technologies in our study of literature already: the ubiquitous Microsoft Word, for instance. Using such software can barely be said to make one a digital literary scholar, though.[16]

What Questions?

What could we include under such rhetoric of 'digital methods' and what types of question might such methods answer? There is, undeniably, a particular type of 'decompositional' thinking that is necessary to use digital approaches.[17] That is to say that computational methods require problems that can be broken into smaller solvable units of

[12] Zara Dinnen, *The Digital Banal: New Media and American Literature and Culture* (New York: Columbia University Press, 2018), p. 1.

[13] N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis* (Chicago, IL: University of Chicago Press, 2012), p. 212.

[14] Anne Burdick et al., *Digital Humanities* (Cambridge, MA: MIT Press, 2012), p. 122.

[15] Richard Rogers, *Digital Methods* (Cambridge, MA: MIT Press, 2015), p. 1.

[16] That said, I do not intend to wade into the quagmire of defining the digital humanities, which has been addressed at great length in publications such as Terras et al.

[17] I borrow this terminology from D. L. Parnas, 'On the Criteria To Be Used in Decomposing Systems into Modules', *Communications of the ACM*, 15.12 (1972), 6; and David West, *Object Thinking* (Redmond, WA: Microsoft Press, 2004).

addressable and empirically verifiable hypotheses, to which some literary interpretative work is not suited.

An example may serve well to illustrate this decompositional—or, computational—thinking. Consider the genre of 'writers' advice'. This form seems almost as old as writing itself, with Plato advising in *Phaedrus* against the very act of writing: 'the man who thinks that he has left behind him a science in writing [. . .] in the belief that anything clear or certain will come from what is written down, would be full of simplicity'.[18] Writers have, indeed, always sought to advise other writers. My suspicion has long been, when encountering such guidance, that authors who dispense it might be hypocritical, that writers do not do as they say.

However, we could go further in breaking down (decomposing) this sample problem of 'writers' advice' into addressable components, as I have been doing in recent work with Erik Ketzan. Indeed, we cannot easily appraise some types of advice. If a writer advises us that the key to excellent writing is to 'write every day', or issues similar diktats, we must take his or her word for it. However, sometimes writers (ill-advisedly) dispense advice that is more susceptible to empirical analysis. The bestselling horror writer, Stephen King, is one such example. In his 2000 memoir, *On Writing*, King tells the reader that 'the adverb is not your friend'.[19] Specifically, King seems to exclude temporal adverbs and adverbial phrases.[20] In this instance, a set of addressable or decomposed problems for computational analysis might be: how frequently does Stephen King use adverbs and does this change throughout his career as his writing matures?[21] In general terms, with many caveats that I will not address here, but as shown in Figure I.1, the answer to this question is: yes. What the critic then goes on to make of this finding remains a matter of interpretation. Undeniably, though, this method allows us to see something about a text that before was unknown.[22]

[18] Plato, *Phaedrus*, trans. Christopher Rowe (London: Penguin, 2005), p. 63.

[19] Stephen King, *On Writing: A Memoir of the Craft* (London: Hodder, 2012), p. 138.

[20] King, *On Writing*, p. 140.

[21] Also addressed in Ben Blatt, *Nabokov's Favorite Word Is Mauve: What the Numbers Reveal About the Classics, Bestsellers, and Our Own Writing* (New York: Simon & Schuster, 2017).

[22] This result is extracted from work in progress that I am undertaking with Erik Ketzan.

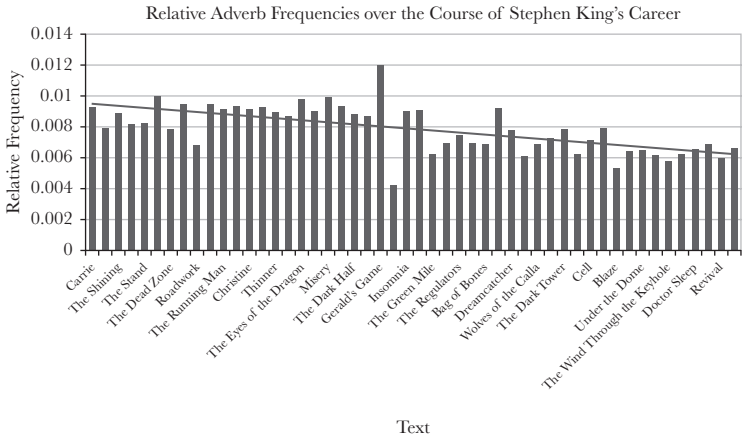

Figure I.1 Relative adverb frequencies over the course of Stephen King's career, excluding temporal adverbs and adverbial phrases using the Stanford PoS tagger english-left3words-distsim model with approximate 3 per cent margin of error. This graph was co-produced with Erik Ketzan.

Did I *need* a computer to produce the result in Figure I.1? I could have sequentially read the novels of Stephen King, marking up the appropriate adverbs and keeping a tally, before plotting these. However, this would probably have taken several months, if not years, of tedious and repetitive reading labour, merely to answer a fundamental empirical question. The computational approach deployed here was not a difference of *type* but a difference of *scale*, *degree*, and *speed*. It is around these matters of repetition, scale, and speed/time that many digital methods in literary studies orbit.[23]

Indeed, the specific trade-off made in the forms of so-called 'distant reading'—computational methods of examining texts—is between resolution and time. There is a 3 per cent margin of error in my process for tagging parts-of-speech in the above experiment on Stephen King's novels. That is to say that for every 100 words processed, approximately three will be misclassified. In the case of *'Salem's Lot*,

[23] For more on this, see Jay Jin, 'Problems of Scale in "Close" and "Distant" Reading', *Philological Quarterly*, 96.1 (2017), 105–29.

to pick on just one example, this means approximately 353 adverbs in my list might not be adverbs. It also means that the system missed some other adverbs that were likewise mislabelled. I lost the resolution and precision that come from actual reading at the expense of having several months more time for other activities and still roughly defining how far off my results might be. That said, there is no guarantee that if I undertook a manual reading exercise and attempted to tag adverbs that I might not also make a comparable number of errors, missing some and incorrectly ascribing others.

In this particular instance of Stephen King's advice, it might be the case that the error margin is too high to draw a sound conclusion. Perhaps the only answer is to read the works 'properly' (although, as above, any such repetitious cataloguing task is also prone to human error). However, many of the problems of scale dealt with by digital literary studies cannot be solved by traditional reading practices. Say, for instance, that one wished to comment on a single year's worth of contemporary fiction and the broad trends within it. But I do not mean award-winning fiction from a single year—I mean *all* fiction. Take the year 2015, for this example. How much would you have to read to be able to say, with absolute certainty, that your statements were accurate across all published fiction in the English language in that year? According to Bowker, almost 220,000 novels were published in English in 2015. Estimating a human lifespan to be approximately 71 years using the World Health Organization's figures, one would have to read ten novels per day, every day from age ten onwards, just to have read all English fiction published in 2015.[24]

Computational methods for the study of literature are not, then, simply an outgrowth of technical capability. Instead, they also respond to specific critiques of canonicity. In a world where it is impossible to read even all the fiction in English published in a single year, the canons to which we devote our time are necessarily limited, but therefore are also biased. We usually delegate to literary prizes and to the internal selection procedures of major publishing houses to filter the

[24] See Erik Fredner, 'How Many Novels Have Been Published in English? (An Attempt)', *Stanford Literary Lab*, 2017 <https://litlab.stanford.edu/how-many-novels-have-been-published-in-english-an-attempt/>; and Eve, *Close Reading With Computers*, introduction.

titles to which we pay attention. This is not a sound basis for examining all literature from any period, though. Under such constraints, 'not reading', as Lisa Marie Rhody puts it, has become 'the dirty open secret of all literary critics'.[25] Between the poles of detailed attention to a limited canon and the void of being unable to read everything sit the digital methods to which the subsequent chapters of this book are dedicated.

*

While the above may have painted a rosy picture of how digital methods might help us with literary empiricism at scale, there are many challenges for our discipline due to digital humanities work. One relates to tool development. Let us say that, instead of King's pronouncements on adverbs above, I had instead taken his advice on similes as the target of my investigation.[26] Here is a reasonable, decomposed question on this subject: 'how frequently does Stephen King use similes compared to a similar corpus of American writers?' However, it turns out that the computational detection and study of simile is a complicated problem with low accuracy rates.[27] Developing the tools that would allow this to work at any scale would take years of software development in cooperation with computer scientists and linguists. Although it might have more general-purpose applications, this development process would likely take longer than reading the material manually; an example of the type of time trade-off that must be considered in any software development.

Another good decomposed question that we might imagine we could answer with computational approaches springs to mind: 'do novels generate similar patterns of affective responses over their plot

[25] Lisa Marie Rhody, 'Beyond Darwinian Distance: Situating Distant Reading in a Feminist *Ut Pictura Poesis* Tradition', *PMLA*, 132.3 (2017), 659–67 (p. 659).

[26] King, *On Writing*, pp. 208–9.

[27] Vlad Niculae and Victoria Yaneva, 'Computational Considerations of Comparisons and Similes', in 51*st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop* (Sofia, Bulgaria: Association for Computational Linguistics, 2013), pp. 89–95 <http://www.aclweb.org/anthology/P13-3013> [accessed 16 November 2018]; Suzanne Patience Mpouli Njanga Seh, 'Automatic Annotation of Similes in Literary Texts' (unpublished Ph.D., Université Pierre et Marie Curie—Paris VI, 2016).

arcs?' That is, do novels share common plot patterns that create similar emotional ebbs and flows in readers? This is precisely one of the questions that Matt Jockers has been attempting to answer using his *syuzhet* software that tries to map sentiment within literary texts. Again, though, this turns out to be a difficult computational task and one that most literary studies scholars could not even begin to work on.[28] In this respect, there is a labour and domain-knowledge problem for the adoption of digital methods within literary studies.

Yet another problematic element for computational methods in literary studies is the legal availability of the texts themselves.[29] In order to perform computation upon a text, as though it were data, one needs a digital copy of the literary work. This may seem to be a trivial matter in the era of the Amazon Kindle. However, the version needed for most digital methods is a plain-text edition, unencumbered by digital rights management technologies (DRM). In the USA and the UK, stripping the DRM off a protected file is a criminal, not 'just' a civil, offence. This means that one cannot be granted permission to remove the DRM from a digital file, even by the rightsholder, regardless of whether it is technically easy to do so. While many scholars working in digital literary studies seem to ignore this legal situation for the sake of convenience and do not remark upon the sources for their work, this is a difficult ethical and legal position. That said, there are two mitigating factors. First, much digital humanities work takes place upon historical literary material out of copyright (although this still

[28] Matthew L. Jockers, 'A Novel Method for Detecting Plot', 2014 <http://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/>; Matthew L. Jockers, 'Requiem for a Low Pass Filter', 2015 <http://www.matthewjockers.net/2015/04/06/epilogue/>; Annie Swafford, 'Why Syuzhet Doesn't Work and How We Know', *Anglophile in Academia: Annie Swafford's Blog*, 2015 <https://annieswafford.wordpress.com/2015/03/30/why-syuzhet-doesnt-work-and-how-we-know/> [accessed 17 November 2018]; Annie Swafford, 'Continuing the Syuzhet Discussion', *Anglophile in Academia: Annie Swafford's Blog*, 2015 <https://annieswafford.wordpress.com/2015/03/07/continuingsyuzhet/>; Benjamin M. Schmidt, 'Do Digital Humanists Need to Understand Algorithms?', in *Debates in the Digital Humanities* 2016, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis, MN: University of Minnesota Press, 2016), pp. 546–55 <http://dhdebates.gc.cuny.edu/debates/text/99>; Matthew L. Jockers, 'Resurrecting a Low Pass Filter (Well, Kind Of)', 2017 <http://www.matthewjockers.net/2017/01/12/resurrecting/>.

[29] Again, I am grateful to Erik Ketzan for first drawing this to my attention.

requires access to a version unencumbered by DRM, which cannot be presumed). Second, the HathiTrust Research Center—a vast archive of 16.7 million items—has extended access to material that is still in copyright for non-consumptive research.[30] This is a praiseworthy and momentous shift that uses a legal defence under US copyright law and vastly expands access to material that would otherwise be unavailable.

Finally, linked to the genesis of electronic literatures, digital literary studies frequently run aground on the rocky shores of evaluation. Datasets, graphs, interactive timelines, software, and other digital artefacts are often not recognized as valid scholarly outputs within the humanities disciplines. A type of 'design practice' sits at the core of this kind of work, which does not necessarily look like work in literary studies.[31] This essentially leads to a situation in which those who perform digital work are asked to coerce their scholarship into existing, recognized media forms for assessment, hiring, promotion, and tenure.[32] Even the fact that citation styles usually require reference to a *page number* encodes an *assumed media form* within a resource locator. Print media remain firmly enthroned at the heart of such citation practices. To counter this, learned societies have formulated sets of evaluation principles for digital scholarship, although uptake remains slow.[33] At the core of this challenge for digital literary studies, though,

[30] HathiTrust Digital Library, 'HathiTrust Research Center Extends Non-Consumptive Research Tools to Copyrighted Materials: Expanding Research through Fair Use', *Perspectives from HathiTrust*, 2018 <https://www.hathitrust.org/blogs/perspectives-from-hathitrust/hathitrust-research-center-extends-non-consumptive-research-tools> [accessed 17 November 2018].

[31] See Burdick et al.

[32] Sydni Dunn, 'Digital Humanists: If You Want Tenure, Do Double the Work', *Vitae*, 2014 <https://chronicleVitae.com/news/249-digital-humanists-if-you-want-tenure-do-double-the-work> [accessed 21 March 2017].

[33] Bethany Nowviskie, 'Where Credit Is Due: Preconditions for the Evaluation of Collaborative Digital Scholarship', *Profession*, 2011.1 (2011), 169–81 <https://doi.org/10.1632/prof.2011.2011.1.169>; American Historical Association, 'Guidelines for the Professional Evaluation of Digital Scholarship by Historians', *American Historical Association*, 2015 <https://www.historians.org/teaching-and-learning/digital-history-resources/evaluation-of-digital-scholarship-in-history/guidelines-for-the-professional-evaluation-of-digital-scholarship-by-historians> [accessed 24 March 2017]; Hamid R. Jamali, David Nicholas, and Eti Herman, 'Scholarly Reputation in the Digital Age and the Role of Emerging Platforms and Mechanisms', *Research Evaluation*, 25.1 (2016),

lies a far more profound anxiety; an act of soul searching by our discipline. This inward-looking stance asks: is digital literary studies *really* literary studies? Should literary studies scholars produce data, code, and graphs…and should we reward them for doing so? Finally, it seems to ask: is this new model a *threat* to our discipline and its evolved state of practice?

The Digital Humanities and Its Discontents

How can we understand a double logic in which digital literary studies work is at once so powerful as to crowd out the traditional humanists, threatening the discipline with total takeover, while at the same time so poorly understood as to need supplementation by traditional publication? How can conducting digital labour in the humanities be seen by some as a sure-fire path to tenure and funding but, simultaneously, a 'risky thing', as Kathleen Fitzpatrick and Mark Sample put it?[34]

Indeed, as I have already implied, far from everyone is happy with the rise of the digital humanities or digital literary studies. The concurrent ascendancy of digital technologies alongside the political rationality known as neoliberalism has made many deeply suspicious of a digital agenda. Daniel Allington, Sarah Brouillette, and David Golumbia even go so far as to claim that:

> Neoliberal policies and institutions value academic work that produces findings immediately usable by industry and that produces graduates trained for the current requirements of the commercial workplace. […] By providing a model for humanities teaching and

37–49 <https://doi.org/10.1093/reseval/rvv032>; but see also Samuel Moore et al., 'Excellence R Us: University Research and the Fetishisation of Excellence', *Palgrave Communications*, 3 (2017) <https://doi.org/10.1057/palcomms.2016.105>.

[34] Kathleen Fitzpatrick, 'Do "the Risky Thing" in Digital Humanities', *The Chronicle of Higher Education*, 2011 <http://www.chronicle.com/article/Do-the-Risky-Thing-in/129132/> [accessed 21 March 2017]; Mark Sample, 'Tenure as a Risk-Taking Venture', *Journal of Digital Humanities*, 1.4 (2012) <http://journalofdigitalhumanities.org/1-4/tenure-as-a-risk-taking-venture-by-mark-sample/> [accessed 24 March 2017]; parts of this section appeared previously in Martin Paul Eve, 'Violins in the Subway: Scarcity Correlations, Evaluative Cultures, and Disciplinary Authority in the Digital Humanities', in *Digital Technology and the Practices of Humanities Research*, ed. Jennifer Edmonds (Cambridge: Open Book Publishers, 2019).

> research that appears to overcome these perceived limitations, Digital Humanities has played a leading role in the corporatist restructuring of the humanities. [. . .] What Digital Humanities is *not* about, despite its explicit claims, is the use of digital or quantitative methodologies to answer research questions in the humanities. It is, instead, about the promotion of project-based learning and lab-based research over reading and writing, the rebranding of insecure campus employment as an empowering "alt-ac" career choice, and the redefinition of technical expertise as a form (indeed, the superior form) of humanist knowledge.[35]

Neoliberalism is probably best defined as a mode of political economy that emerged from the 1980s onwards in which politics is disenchanted by economics as the dominant societal logic.[36] Under such logic, economics must form the basis for all state decisions. The state itself merely ensures that the conditions for market exchange are enforced, even while the state itself must work on a market logic—the state under the supervision of the market, as Michel Foucault had it.[37]

Within such a definition, one can begin to see how the digital humanities might appear neoliberal. If neoliberalism is the disenchantment of politics by economics, then digital methods for studying literature appear as the disenchantment of literature by computers. Further, in Allington, Brouillette, and Golumbia's view, the digital humanities' labour structures are the worst aspect. In the privileging of technocratic knowledge, supposedly over and above humanistic epistemologies, they see a replication of wider societal patterns of precarity within the digital humanities. (Although I note that material textual scholarship, for instance, has long had a technocratic interest in, say, the manufacturing processes of books.) It is not my aim here to refute systematically the arguments of Allington, Brouillette, and

[35] Allington et al.

[36] William Davies, *The Limits of Neoliberalism: Authority, Sovereignty and the Logic of Competition* (Thousand Oaks, CA: Sage, 2014). See also Wendy Brown, *Undoing the Demos: Neoliberalism's Stealth Revolution* (New York: Zone Books, 2015); Wendy Brown, *In the Ruins of Neoliberalism: The Rise of Antidemocratic Politics in the West*, The Wellek Library Lectures (New York: Columbia University Press, 2019).

[37] Michel Foucault, *The Birth of Biopolitics: Lectures at the Collège de France*, 1978–79 (Basingstoke: Palgrave Macmillan, 2008), p. 116.

Golumbia's piece, which moves between specific attacks on the University of Virginia, the Andrew W. Mellon Foundation, and broader statements on labour and neoliberalism, but I will only note that I do find it strange to believe that digital literary studies will produce 'findings immediately usable by industry'. It seems indeed optimistic to think that broad-scale findings about the history of literary genre or gender representation, say, using computational methods, will show themselves to be instantly 'monetizable', to use a current buzz-word.[38]

One of the other criticisms levelled at digital literary studies in this same piece—but echoed elsewhere—is that digital approaches involve 'the displacement of politically progressive humanities scholarship and activism in favor of the manufacture of digital tools and archives'. That is, the claim here is that digital literary studies are apolitical formalism at best and, at worst, immoral in detracting resources from now-conventional modes of political critique in the discipline. One of the most obvious, although distressing, retorts to such a statement is to note that the political effects of literary criticism are often overstated. Certainly, Aime Cesaire, Frantz Fanon, Judith Butler, and many others in the postcolonial and gender studies fields can be said to have had a lasting political legacy. Yet although I do not agree wholeheartedly with her polemical injunction to abandon critique, Rita Felski has a point when she writes that, overall, critique and suspicion in literary studies are 'less heroic and more humdrum and routinized than we might think'.[39] The routinization and normalization of critique in literary studies may have dimmed its power.

That said, it is hard to overstate the influence that critical theory has had upon English departments around the world. This level of disruption to previously formalist departments engendered by literary theoretical paradigms is aptly illustrated in the 'MacCabe Affair' in the UK. In this case, Colin MacCabe was denied tenure at Cambridge University for his support of theoretical approaches—a news story

[38] See, for just such an article, Ted Underwood, 'The Life Cycles of Genres', *Journal of Cultural Analytics*, 1.1 (2016) <https://doi.org/10.22148/16.005>.

[39] Rita Felski, *The Limits of Critique* (Chicago, IL: University of Chicago Press, 2015), p. 47.

that, unbelievably, made the front page of the *Guardian* newspaper.[40] Some commentators fear that DH heralds a return to some prior apolitical, formalist stance for these disciplines.

It is also worth noting that this critique of the digital's apoliticality can apply to any other work of formalist-aesthetic literary criticism. Literary criticism has long straddled aesthetic and thematic approaches using political readings. Certainly, the empirical evidence furnished by digital approaches is usually formalist. However, it is what one does with that evidence that matters.[41] As Lisa Gitelman deftly phrases it, following Geoffrey C. Bowker, 'raw data is an oxymoron'.[42] The political import, or otherwise, of digital work rests upon the use one makes of the words on the page, whether filtered through a computer or whether one reads them by eye. As a final note on this, if the critique is that it is the *time* spent on building tools that is here apolitical (or even immoral), then one might say exactly the same of any kind of reading/thinking/note-taking or processual methodology for the study of aesthetics. All types of literary critical work require a level of background labour that contributes towards the endpoint of an argument. When it is digital labour, though, there seems to be an additional level of criticism.

In recent years, however, there has been an explosive growth in the volume of scholarship that connects digital humanities and ethics. For instance, among the most important of these recent works is Ruha Benjamin's *Race After Technology: Abolitionist Tools for the New Jim Code*. In this book, Benjamin argues that racial prejudice is repeatedly inscribed within algorithms under the cloak of objectivity, a phenomenon she calls 'the New Jim Code', riffing on the informal name for the USA's system of segregation.[43] Likewise of significance is

[40] Francis Mulhern, 'The Cambridge Affair', *Marxism Today*, March 1981, pp. 27–8; Marcus Morgan and Patrick Baert, *Conflict in the Academy: A Study in the Sociology of Intellectuals* (London: Palgrave Macmillan, 2015).

[41] See, for instance, Richard Jean So, *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction* (New York: Columbia University Press, 2020) for an example of the use of data-driven approaches for an ethical end.

[42] Lisa Gitelman, ed., *'Raw Data' Is an Oxymoron*, Infrastructures Series (Cambridge, MA: MIT Press, 2013), p. 1.

[43] Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity Press, 2019).

Charlton D. McIlwain's *Black Software: The Internet and Racial Justice, from the AfroNet to Black Lives Matter*, which charts the story of a vanguard that 'demonstrates how black people have taken technology not originally designed with our concerns in mind' while, at the same time, showing 'how computing technology was built and developed to keep black America docile and in its place'.[44] Further, Catherine D'Ignazio and Lauren F. Klein's *Data Feminism* is a damning indictment of how gender inequality is inscribed in the cultures of big data that permeate our societies. More than simply diagnosing the problem, though, D'Ignazio and Klein offer a powerful critical framework to redress this imbalance for those working to examine how data are used computationally.[45]

There are further works at the intersection of the digital humanities and ethics that bear closer scrutiny. The first is Roopika Risam's *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*; a work that fuses two relevant strands of inquiry. The first is the well-known paradigm of postcolonial studies, in which it is shown that 'the foundations of literary studies and historiography—whether Anglophone, Francophone, Hispanophone, or Lusaphone—are inextricably linked to the rise of European colonialism'.[46] The second is the subject of this book: the digital humanities. The new field of 'postcolonial digital humanities' that Risam posits explores the relationship of digital practice 'to the intersections of race, gender, class, nation, sexuality, ability, and other axes of identity and oppression'. It is a field that 'attends to the politics and theory subtending the creation of scholarship to clear space for new modes of thinking that foreground the particular over the universal and the local over the global in the production of the digital cultural record'.[47]

Perhaps one of the most astute observations of Risam's book lies in her analogy between programming and literature as sharing a communal effort at 'world making'. This may seem far-fetched, but many

[44] Charlton D. McIlwain, *Black Software: The Internet and Racial Justice, from the AfroNet to Black Lives Matter* (New York: Oxford University Press, 2020), p. 7.

[45] Catherine D'Ignazio and Lauren F. Klein, *Data Feminism*, Strong Ideas Series (Cambridge, MA: MIT Press, 2020).

[46] Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Evanston, IL: Northwestern University Press, 2018), p. 25.

[47] Risam, p. 30.

books on the study of programming, such as David West's *Object Thinking*, published by Microsoft, stress that object-oriented programming (OOP) is, at the very least, a form of world *modelling*.[48] For Risam, following Matthew Kirschenbaum, the point is that if 'the coder becomes the world maker, charged with defining the rules and characteristics of the world', then there are both dangers and opportunities in digital approaches. The dangers are that this 'apt description of programming evinces the colonial dynamics of knowledge production' and can end up 'reproducing the hegemonies of the "real" world'. The opportunity that Risam poses is a set of digital worlds that do not fall prey to this 'risk of rehearsal'. Could they, she asks, 'be ones that imagine new forms of resistance through digital knowledge production?'[49]

Risam's work is also very good at undoing the early utopian histories of the internet and cyberculture. As she notes, the initial optimism of scholars such as Frank Biocca, Larry McCaffery, and Michael Benedikt was misplaced. In Risam's words, they saw the internet 'as a space of freedom and creation that exists outside of the iniquities of lived experience'.[50] Yet, as subsequent new media scholars such as Wendy Chun, Anna Everett, and Lisa Nakamura have identified, this democratic space is far less representatively peopled than we might like. The 'putatively democratic space of the internet' has led to the false notion that 'the internet is disembodied and shielded from social inequalities'; a patently untrue assertion that plays out in the replicated racism of, say, artificial intelligence and facial recognition.[51]

[48] West. See also Matthew Kirschenbaum, 'Hello Worlds', *The Chronicle of Higher Education*, 23 January 2009 <https://www.chronicle.com/article/Hello-Worlds/5476> [accessed 13 April 2020].

[49] Risam, pp. 33–4.

[50] Risam, p. 36 points to Larry McCaffery, 'Introduction: The Desert of the Real', in *Storming the Reality Studio: A Casebook of Cyberpunk and Postmodern Science Fiction*, ed. Larry McCaffery (Durham, NC: Duke University Press, 1991), pp. 1–16; Frank Biocca, 'Communication Within Virtual Reality: Creating a Space for Research', *Journal of Communication*, 42.4 (1992), 5–22 <https://doi.org/10.1111/j.1460-2466.1992.tb00810.x>; Michael Benedikt, ed., *Cyberspace: First Steps* (Cambridge, MA: MIT Press, 1994).

[51] Risam, p. 36.

'Where', asked Alan Liu in 2012, 'Is Cultural Criticism in the Digital Humanities?'[52] His own critique therein was that 'the digital humanities are noticeably missing in action on the cultural-critical scene', neglecting the reflexive inflection seen in other adjacent fields, such as new media studies. Yet, I would like to venture, if this movement has been slow in coming, projects such as Risam's extend the digital humanities movement outwards into valuable areas of critical discourse—and demonstrate that there have been figures thinking through this area for some time.

A second area where we see an increase in ethical intersections is in digital cultural history. Although not strictly within the purely literary realm, this is also a massively expanding field. In particular, recent work by Marie Hicks has turned to how women formed the core of early computer operators and workers but were erased from these roles as an official computing 'industry' emerged. This, of course, has profound implications for how computing has spread across the globe in general and carries ramifications, I would argue, for how we consider the adoption of digital technologies in the literary studies space.

By way of background, it is worth noting—as does Hicks—that the term 'computer' originally referred to a person. Specifically, it denoted a woman who was employed to undertake calculations. For, 'in the 1940s, computer operation and programming was viewed as women's work—but by the 1960s, as computing gained prominence and influence, men displaced the thousands of women who had been pioneers in a feminized field of endeavor, and the field acquired a distinctly masculine image'.[53]

Hicks's study is perhaps most valuable for the fact that, while it is a study of a technological area of development, its prime object of focus rests on the social conditions that surround the development of computation. That is to say that it is not the development of technology that interests Hicks, but rather how the field replicated social privilege

[52] Alan Liu, 'Where Is Cultural Criticism in the Digital Humanities?', in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis, MN: University of Minnesota Press, 2012), pp. 490–509 <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/896742e7-5218-42c5-89b0-0c3c75682a2f> [accessed 14 April 2020].

[53] Marie Hicks, *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing* (Cambridge, MA: MIT Press, 2018), p. 1.

despite early engagement by women. Indeed, Hicks discerns a regular phase of feminization early in the development of many new technologies: 'a familiar historical pattern seen in everything from textile manufacturing to typewriting'.[54] Yet this did not happen in computing's switch to electro-mechanical components, primarily because education systems privileged male access to computers as playthings of the future and assumed there was a natural interest among boys in computing and its attendant technologies.

Hicks also points, though, to specific national contexts as contributing to the gendered inequality of the computing industry. The British case presents an instance of 'a top-down government initiative to computerize' that came with attendant 'explicit structural discrimination' in, say, the gendered relative pay structures of the British civil service.[55] The modernization of technologies does not—perhaps self-evidently—come with concomitant social advances.

Indirectly, though, sexuality also feeds into the gendering of labour in the British context. The assumption that underpinned much of the British hierarchy of labour value was that a male breadwinner would have to earn enough to support a nuclear family. By contrast, women who worked were assumed not to have the same wage 'requirements' and hence the entire remuneration system was structured to pay women less than their male counterparts. The assumption, in other words, was that all women were heterosexual and would be married—and thereby provided for. In this way, it is impossible to separate assumed sexuality from gender roles in this instance.

Hicks is, of course, far from the only person to have studied the gendered status of labour in the computational environment. Works by Jennifer Light, Jean Jennings Bartik, Nathan Ensmenger, and Janet Abbate, among others, form the background context against which this most recent study is set.[56] We might further consider Margot Lee

[54] Hicks, p. 2. [55] Hicks, p. 3.

[56] Jennifer S. Light, 'When Computers Were Women', *Technology and Culture*, 40.3 (1999), 455–83; Jean Bartik, *Pioneer Programmer: Jean Jennings Bartik and the Computer That Changed the World*, ed. Jon T. Rickman and Kim D. Todd (Kirksville, MO: Truman State University Press, 2013); Nathan Ensmenger, *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*, History of Computing (Cambridge, MA: MIT Press, 2010); Janet Abbate, *Recoding Gender: Women's Changing Participation in Computing*, History of Computing (Cambridge, MA: MIT Press, 2012).

Shetterly, the author of *Hidden Figures: The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race*, now a major Hollywood motion picture.[57]

There is, ultimately though, a nice payback in Hicks's work. The narrative that she charts is about the downfall of the British computing industry, at least in part because of its gendered pay and labour policies. She also gives us pause for thought about the crisis nature of the emergence of computing's gendered labour in the Second World War and the codebreaking facilities at Bletchley Park. As I write in 2021, the world grapples with the COVID-19 coronavirus pandemic—a moment of crisis comparable in its disruption to the two world wars of the previous century. I would be willing to wager that a programme of technological rebuilding of the economy will be key to many government strategies in the aftermath of this catastrophe. Whether we learn the lessons of Hicks's study in replicating socially unequal labour structures in the digital space remains to be seen.

To return to the intersection of these first two conjoined critiques of DH, though—that the digital humanities are neoliberal and that the digital humanities are apolitical—sits a third: that digital literary studies are useless. Timothy Brennan most pointedly articulated this in his *Chronicle of Higher Education* article, 'The Digital Humanities Bust', but it is a common refrain with which almost anyone who has done digital humanities work will be familiar. For Brennan, in digital literary studies, 'some of their interpretations derive from what they knew "in advance" '. This means, in his view, that 'the findings do not need the data and, as a result, are somewhat pointless'.[58]

There are, though, four distinct defences of digital practices that can be raised against such a critique. The first is that digital practices require validation at the micro level in order to scale. I return to this in the final chapter of this book but suffice to say that if you are developing a piece of software that tests certain properties of literary texts, one needs known conclusions with which to begin. Otherwise, you cannot test that the software works as expected before using it on texts

[57] Margot Lee Shetterly, *Hidden Figures: The American Dream and the Untold Story of the Black Women Mathematicians Who Helped Win the Space Race* (New York: William Morrow and Company, 2016).

[58] Brennan.

that one has not read. Validating that computer models work to the same conclusions as human readers is the only way of showing that a software model might be useful at scales beyond the human. The second response is that this critique can be levelled at most pieces of literary criticism that perform interpretation. The words were on the page in front of us beforehand, after all. Literary critical knowledge is a type of unearthing of latent content that was always obvious, *after the fact*. This is why literary criticism can hold such power: because it retroactively makes something hidden seem obvious. The third response to this argument is that there is a curious utilitarianism in its premise. Since when did literary studies need to have a 'point'? Since when did we demand of literary critique that it be useful? Indeed, this type of utilitarian insistence that digital literary studies deliver something purposeful, useful, and pointed sits in curious tension with the assertion that the digital humanities are neoliberal. How can digital literary studies win?[59] Produce a useful outcome and one is branded utilitarian and neoliberal. Conduct pointless work and one is told that one is not useful enough.

The fourth, final, and strongest rebuttal to the argument that DH can tell us nothing new is that such an assertion is often simply not true. Digital methods can unearth fresh evidence that can overturn critical consensus. It is in this area of contestation—covered in the final chapter of this book—that digital literary studies often best succeeds; these moments where computational discourse interacts with commonly held literary critical precepts and blows them apart.

There is a final and powerful critique of 'computational literary studies' mounted in 2019 by Nan Z. Da that also cannot be ignored: that many of the findings from quantitative, statistical, and digital approaches to the study of literature are, in fact, wrong and inaccurate.[60] Da spent over two years tracking down data from papers that used digital methods to study literature in order to show that the findings are unreplicable and in some cases drastically misinterpreted.

While the fallout from Da's *Critical Inquiry* article will continue, there are a few points worth noting. The first is that Da holds computational literary studies to a higher standard than conventional literary approaches. As Alan Liu put it, quantitative statements about art are

[59] I am grateful to Ted Underwood for this point.

[60] Da.

made all the time in literary criticism, such as 'Wordsworth uses "joy" a lot in important poems like "Tintern Abbey"' and that 'evidence of that sort underlies much of literary studies, going back to close reading'. In other words, for Liu here, Da's criticism of statistical problems in quantitative literary studies is unequally distributed. It picks on the digital, when digital approaches are attempting 'to make it, if not right, [then] better'.[61] This is not, of course, to say that we should not criticize inaccuracies in digital/computational literary studies. It remains important to do so and Da has done a service by pointing to some errors in the secondary literature (although some targets of her critique argue that she has misread them).

The second curious point is that Da's article is extremely critical of the funding that digital approaches reportedly receive. This critique contains elements of a now-common anti-DH polemic: the digital humanities are vastly well funded compared to other areas of the humanities. Yet, the 2019 appropriation of $155,000,000 by the USA's National Endowment for the Humanities contained just 2.97 per cent ($4,600,000) dedicated to digital humanities (and this is not even specifically digital literary studies). Further, this money isn't used to pay for software/infrastructure, as the piece implicitly claims (Da notes that most of the software is free/open source), but for the labour of researchers and developers. Perhaps there is a fair comment to be made on DH's allocation of funding (though it is hardly as large as others make out). But it is disconcerting to see people cheerleading for less money to be put into the study of humanistic objects of inquiry. Perhaps it is not a call for less money to be put into it in general, though, but rather for a reallocation away from digital approaches, as though such funding were a zero-sum game. This, though, plays the very competitive game that we criticize elsewhere, pitting should-be allies against one another, rather than working in concert to ensure a better future for all the humanities.

Finally, there are huge infrastructural implications to Da's piece. In other disciplines, these are already being broached via the rhetorics of

[61] Alan Liu, 'E.g. (Generic Example): "Wordsworth Uses 'joy' a Lot in Important Poems like 'Tintern Abbey'." Evidence of That Sort Underlies Much of Literary Studies, Going Back to Close Reading. Let's Compare the Statistical Validity of _that_ to DH's Attempt to Make It, If Not Right, Better', *@alanyliu*, 2019 <https://twitter.com/alanyliu/status/1106109232661725185> [accessed 17 March 2019].

the reproducibility and replication crises. As Alan Thomas at the University of Chicago Press asked: 'how realistic for authors and publishers' are Da's recommendations of full datawork and replicable software?[62] In the present moment, this is possible. We can lodge most of these artefacts in various preservation-backed repositories with stable identifiers. The question is actually: for how long do we want to be able to replicate a finding? This is a question of usage as opposed to one just of preservation. Certainly, we can make bits and bytes available for a very long time indeed. But how are they interpreted? Usage half-lives of work in the humanities disciplines are long and one might wish to validate some work undertaken, say, six years ago. What guarantee do I have that software written six years ago will still run on the newest operating system? Can you open the files created by your word processor—perhaps the most common tool—from fifteen years ago?

The other challenge is that the term 'data' actually means 'stuff'. Data can range from a tiny CSV representation of a spreadsheet up to terabytes of information. To say to publishers and archivists 'please can I deposit my "data"?', when the spectrum for what that may contain is so wide, is a problem. This is because there is an economic scarcity underlying all digital preservation systems, as the prominent digital preservation expert David S. H. Rosenthal has argued for years. Part of this scarcity consists of pre-selection to militate against all resources being consumed by, say, a single project. Yet blanket calls for all data and software to be available over decadal-plus timespans for replication and repeatability will only be viable while digital literary studies remains a niche, small area. When these data formats and structures are bespoke and customized for specific projects, the problem is even larger. There is an almost directly proportionate relationship between the bespokenness of a digital artefact and the difficulty of preserving it. These are some of the looming challenges for digital literary studies.

[62] Alan Thomas, 'Here Are the First of Nan Z. Da's Suggested Guidelines for Peer Review of Computational Literary Studies, from Her Critique of the Field in @CriticalInquiry. How Realistic for Authors and Publishers?', *@alnthomas*, 2019 <https://twitter.com/alnthomas/status/1106616795534934016> [accessed 17 March 2019].

What Has Digital Literary Studies Ever Done for Us?

Despite the naysayers and the challenges, the remainder of this book is dedicated to an exploration of the contributions that digital literary studies have made, continue to make, and look set to make in future. The format for the work is a fusion of original examples—as per the above section on Stephen King—and surveys of innovative work in the field.

The rest of this book proceeds along four different lines that correspond to chapters: authors and writing; space and visualization; place and maps; and distance and history. By way of cartography, I here outline the conceptual route that the rest of this volume will take.

On authors and writing, two central questions posed by literary theory over the past half-century have been: 'what is a literary text?' and 'what is an author?' Indeed, the university discipline of literary studies has never truly known its precise object of study, which is partially why so many diverse practices of scholarship are lodged within English departments. What might it mean for a text, then, to be particularly 'literary'? Do we know? Is there any discernible aspect within language itself that denotes a work as literary? There are ways that we can begin to address these questions through digital approaches.

In this first chapter, I introduce a range of approaches to the measurement and digital quantification of literary style: stylometry or digital stylistics. This begins with a history of stylometric thinking, ranging from approximately 1851 through to contemporary multi-dimensional fingerprinting techniques, such as Burrows's delta method. I then progress to discuss close vs. large-scale literary reading and the problematic terminology of 'distant reading' (namely, that one can use computational techniques to read closely, despite this also being a type of 'distant' reading).

In the second chapter, I turn to space and visualization. For the common link between the section titles of Franco Moretti's well-known book, *Graphs, Maps, and Trees* (2007), is the visuality of his abstract models for literary history. Indeed, graphs, maps, and trees are all structures by which we can downmix complex, multi-dimensional aspects of literature into approximate two-dimensional (or sometimes three-dimensional) space. Much like conventional literary criticism, visualization yields to us new ways to conceive of narrative,

reorienting texts through fresh optics and augmenting understanding. Visualization is a form of deformance and interpretation, as Lisa Samuels and Jerome McGann would have it.[63]

The third area of exploration in this book—and its third chapter—pertains to place and maps. The 'spatial turn' in the humanities—exemplified in the work of scholars as far apart as Jo Guldi and Robert Tally Jr—draws our attention to how literary texts structure their senses of place. From J. R. R. Tolkien to W. G. Sebald via the Hundred Acre Wood, literary works have often included maps within their pages. Yet such *topoi* sit distinct and apart from the extra-textual world, even when such places are represented therein. Digital approaches to geographic information systems (GIS) have been among the most commonly deployed technologies to think 'around' these issues of space and place. Whether it be in visualizing the multiple pathways taken by Woolf's characters in *Mrs Dalloway* (1925) or mapping the Lake District of the Romantic poets, attention to literary geography has been extensive in the digital world. It is to these themes that this chapter addresses itself.

The final chapter in this book thinks through notions of distance and history. As above, for many years now, more contemporary fiction has been published every year than a person can read in a lifetime. The implications for literary history here are enormous. Field mastery by a single individual is impossible and the systematizing dreams of the early Russian formalists seem far out of reach. One of the ways in which statistical reading has been billed as useful, though, is in overcoming these human limitations. If we cannot read enough ourselves, perhaps, it is posited, we might delegate this work to the machines. In conclusion, I end this book with a very brief summary of where digital methods might lead us and what their continued presence means for literary studies today.

Perhaps there is one area of work in digital literary studies that, in this book, gets somewhat less of a look-in than it might merit: the production of digital textual editions using the Textual Encoding Initiative's TEI standard. This XML format is widely used to represent digital texts—for instance, in textual editing. At the same time, the TEI consortium has already extensively documented the range of

[63] Samuels and McGann.

projects that use this standard: from Inscriptions of Roman Tripolitania to the Darwin Correspondence Project and beyond. In a way, although I do not give much space herein to TEI, this is because a whole book could be (and has been) dedicated to this standard and, still, it would be insufficient to cover all of its ground.[64] Suffice it to say that digital textual editing brings an intense textual focus in the same way as conventional editing; its practices are the very opposite of 'distancing'.

This book cannot do everything. It is impossible not to omit a great deal of valuable work from a survey when writing within the confines of a shorter book and I am certain that many readers will query the selections I have made. I aim nonetheless to give an overview of the scene of contemporary digital literary studies, gesturing towards broad areas for investigation, even while I must inevitably elide many specifics.

[64] For more, I recommend the forthcoming Christopher Ohge, *Inventions of the Text: Editing, Computing, and Publishing Digital Exhibitions of Experience* (Cambridge: Cambridge University Press, 2022).