

# Digital Humanities

Simon Burrows and Michael Falk

*This is a draft of a chapter/article that has been accepted for publication by Oxford University Press in the forthcoming book Oxford Encyclopedia of Literary Theory edited by John Frow due for publication in 2020.*

## Summary

This article considers offers a definition, overview and assessment of the current state of Digital Humanities particularly with regard to its actual and potential contribution to literary studies. It outlines the history of Humanities Computing and Digital Humanities, its evolution as a discipline including its institutional development and outstanding challenges it faces, and considers some of the most cogent critiques it has faced, particularly from North American based literary scholars, some of whom have suggested it represents a threat to centuries old traditions of humanistic inquiry and particularly to literary scholarship based on the tradition of close reading. The article shows instead that Digital Humanities approaches gainfully employed offer powerful new means of illuminating both context and content of texts, to assist with both close and distant readings, offering a supplement rather than a replacement for traditional means of literary inquiry. The digital techniques it discusses include stylometry, topic modelling, literary mapping, historical bibliometrics, corpus linguistic techniques, and sequence alignment, as well as some of the contributions that they have made. Further, the article explains how many key aspirations of Digital Humanities scholarship, including interoperability and linked open data, have yet to be realised, and considers some of the projects that are currently making this possible and challenges that they face. The article concludes on a slightly cautionary note: what are the implications of the Digital Humanities for literary study? It is too early to tell.

## Keywords

Digital Humanities; Close Reading; Distant Reading; Historical Bibliometrics; Stylometry; Corpus analysis; Eighteenth-Century Studies; Literary History; Literary Mapping

## Defining Digital Humanities: Who's in the tent?

Defined succinctly, "Digital Humanities" involves the application of computational techniques to traditional humanities problems, both as a scholarly practice and as the study thereof. "Digital Humanities" thus describes both a technology-empowered methodological approach (or approaches) and a self-reflective critical component. Nevertheless, the precise meaning of the term has proven unstable and subject to debate since it supplanted the earlier

label, “Humanities Computing”, shortly after the millennium.<sup>i</sup> This instability is partly the result of technological shifts as visualisation, immersive virtual reality, social media, 3-D modelling, online gaming and machine learning have opened new vistas for researching, exploring, curating, presenting and understanding objects, the archive, and the human condition. However, much debate and redefinition has occurred as a conscious academic strategy, as digital humanists have sought to emphasise that digital scholarship in the humanities now involves far more than opening up archives and texts to computational analysis through digitization.<sup>ii</sup>

It remains an open question whether Digital Humanities should be considered as a discipline, a research field or a movement. Some advocate a capacious, “Big Tent” definition of Digital Humanities, incorporating a “shared core of like-minded scholars who explore digital frontiers to undertake work in the humanities”.<sup>iii</sup> This is a widely shared ideal, but it faces two key problems. First, nearly all Humanities scholars today rely on digital technology, but “if everyone is a Digital Humanist, then no one is really a Digital Humanist”.<sup>iv</sup> This has led some to predict that the term “Digital Humanities” will eventually “wither away” like Marx’s socialist state.<sup>v</sup> The second problem is that the Big Tent might not be as Big as it seems. Scholars of literature and linguistics – with valuable contributions from cognate areas such as biblical studies, classics and medieval studies – have dominated Digital Humanities since its inception, probably because computer analysis of text has historically been simpler than analysis of sound or image.<sup>vi</sup> This dominance across much of the field’s history has been confirmed by recent statistical analysis, which shows that between 1966 and 2004, 64% of the articles in the field’s two most influential journals were devoted to the study of text.<sup>vii</sup> However, in subsequent years the expansion of the field has embraced – or some might prefer to say brought it into conversation with – disciplines as diverse as art history, design and architecture through to social media analysis, social robotics, and many areas traditionally associated with the social sciences. Whether Digital Humanities is a Big Tent or a narrow trench, however, there is no doubt that it has radical implications for the scale and conduct of much humanities research, for modes of inquiry and analysis, and for the types and sophistication of questions that scholars can meaningfully ask of traditional sources. This chapter shows how several core questions of literary theory—meaning, interpretation, textuality—have been affected by Digital Humanities.

### **The Distant Reading Debate: Can Computers Read?**

In the first two decades of the millenium, the role and significance of Digital Humanities has become one of the most hotly contested and controversial areas of humanities study. In literary studies, the controversy began with a single sentence. In the year 2000, Franco Moretti observed that literary scholars only ever consider a tiny fraction of the world’s books. If they were ever to transcend the limits individual scholarship, he suggested, then they would need “a little pact with the devil: we know how to read texts, now let’s learn how *not* to read them.”<sup>viii</sup> Moretti promoted a new form of “distant reading”, in which scholars would use

digital methods to analyse “units much smaller or much larger than the text: devices, themes, tropes—or genres and systems”.<sup>ix</sup> Though Moretti later recanted his *pact with the devil*—“it was meant as a joke”—it has remained a totemic statement of a new concept of textual interpretation.<sup>x</sup>

Since the early 2010s, a series of scholars have begun to answer Moretti’s call. Jodie Archer, Matthew Jockers, Ted Underwood and Andrew Piper have all published scholarly monographs exploring trends that cut across the boundaries of text, genre and period,<sup>xi</sup> while Moretti and his colleagues at the Stanford Literary Lab have published a series of often field-defining pamphlets.<sup>xii</sup> These scholars not only claim to have made novel empirical discoveries; they claim to reformulate fundamental concepts of literary theory. Underwood, for instance, claims that digital methods fatally compromise the concept of period, arguing that the “largest patterns organizing literary history” do not fall neatly into the boxes scholars draw around particular times and places.<sup>xiii</sup> Two decades after Moretti’s pronouncement, distant reading has arrived in literary studies.

Moretti’s vision of distant reading has inspired three kinds of critique. First, there are those who see it as an assault on a liberal tradition of humanistic inquiry. Digital Humanities is a Trojan horse concealing a neoliberal agenda: it hoovers up research funding, devalues the free intellect of the scholar, and promotes the idea that literary studies should be applied and factual rather than critical and interpretative.<sup>xiv</sup> David Golumbia suggests that Digital Humanities could bring about the “Death of a Discipline”, as “professionals who are not humanists” (e.g. librarians and computer scientists) become “engaged in setting standards for professional humanists”.<sup>xv</sup> Though arguments like these have failed to deter distant readers, they have some relevance to the institutionalisation of Digital Humanities (see “The Digital Divide”).

The second critique is of more recent origin. In 2019, Nan Da published “A Computational Case Against Computational Literary Studies”, a thrilling essay that turns the techniques of distant reading against it.<sup>xvi</sup> She replicates a number of digital studies, showing that minor adjustments to certain parameters can completely change the results. The fundamental problem, she argues, is that computational techniques are necessarily reductive, producing simplified models of text based on the arrangement of words. While such reduction may be useful in contexts like legal document discovery, “there is no rationale for such reductionism” in literature—in fact, the discipline is about *reducing* reductionism”.<sup>xvii</sup> Her critics point to her narrow conception of “Computational Literary Studies”, her selective citation of the literature, and her insistence that statistics are only useful if they provide clear causal explanations.<sup>xviii</sup> Whether Da is right or wrong, it is undeniable that she has almost single-handedly put the debate on a new technical footing.

The third critique is the oldest and most interesting from a theoretical perspective, because it raises fundamental questions of meaning and interpretation. In everyday language, people say that a computer “reads” a file or a disk, but as Johanna Drucker observes, “computers do not interpret; they simply find patterns”.<sup>xix</sup> Computers have no conception of external reality, and therefore cannot comprehend how language is used by humans in concrete situations. Some argue that close reading is therefore unthreatened by distant reading.<sup>xx</sup> Others argue that digital methods “deform” or “transform” texts, enabling humans to discover or create new meanings that were not there before.<sup>xxi</sup> Still others argue that information theory provides a basis for linking patterns and meaning.<sup>xxii</sup> Moretti and his school ground their approach in Russian formalism and the philological tradition of Leo Spitzer and Erich Auerbach. Drucker herself argues that “modelling” can bridge the gap between the computer’s pattern and human meaning. Needless to say, the question of pattern and meaning is one of the most fruitful areas of theoretical inquiry in Digital Humanities today.

While the distant reading debate has brought Digital Humanities to the heart of literary theory, it has tended to obscure an older and possibly richer tradition of digital text analysis. At the heart of Moretti’s vision is a desire to see past individual texts and authors to reveal the “*champs littéraire*” [literary field], the field of cultural, social and economic power that shapes literary production.<sup>xxiii</sup> Both proponents and opponents of distant reading alike tend to argue that digital methods are necessarily “crude” or “brute”, and therefore inapt for more fine-grained analysis.<sup>xxiv</sup> An earlier tradition argued precisely the opposite: digital analysis can be extremely subtle, and is especially apt for studying the particularity of texts and the individuality of authors. In the 1970s, Robert Cluett rigorously studied the prose of authors such as Philip Sidney and Ernest Hemmingway, seeking to identify the precise linguistic features that distinguished their characteristic styles.<sup>xxv</sup> The following decade, John Burrows carefully sifted the language of Jane Austen’s novels, showing how the individuality of her characters could be seen in the patterns of high-frequency words like “of” and “the”.<sup>xxvi</sup> Burrows strongly defended the concept of “idiolect”: since the computer could so easily distinguish the different languages of individual characters and authors, he said, the anti-individualist philosophy of language espoused by Roland Barthes, Jacques Derrida and Michel Foucault was empirically baseless.<sup>xxvii</sup> Perhaps this unorthodox theory contributed to the relative obscurity of Cluett and Burrows’s approach in the new millennium, though their work is fundamental to modern stylometry, and there seems to have been a revival of small-scale digital reading today.<sup>xxviii</sup>

### **Modelling Text: Broader Currents**

A second major strand of Digital Humanities has reshaped fundamental aspects of literary scholarship without rousing the controversy of distant reading. Distant reading is underpinned by the idea of computers as information processors, which read in text and reveal the underlying patterns. This idea of computers is false, argues Willard McCarty, one of the most prominent theorists in this second strand. Computers are in fact “modelling machines, not

knowledge jukeboxes”.<sup>.xxix</sup> Whenever a computer is used to study a text or other artifact, the text or artifact must be reduced to “computational form”, but since no computer model can capture all the ripples and complexities of reality, there is always some residual complexity that the model cannot explain.<sup>.xxx</sup> This residual complexity or gap drives a creative process of interpretation. On the one hand, modelling knowledge forces scholars to make their concepts and assumptions explicit. On the other hand, modelling the text compels them to confront its difficult, “computationally unknown” aspects.<sup>.xxxi</sup> As they become more aware of the gap between their concepts and reality, they are driven to improve their models and begin a playful process of testing and manipulating ideas in digital space.<sup>.xxxii</sup>

This vision of interactive modelling has transformed scholarly editing and book history, and has accordingly transformed how nearly all literary scholars encounter literary objects.<sup>.xxxiii</sup> Scholarly editors have always been aware of the instability of literary texts, as different versions of a text proliferate, and editors combine texts into holistic *oeuvres* representing the writer’s total vision. Digital editions allow editors to model this instability more thoroughly than ever before. As Jerome J. McGann explains, it is not simply that digital editions can “store vastly greater quantities of documentary materials”, and can “organize, access, and analyze” them more quickly than paper-based editions; hyperlinked digital editions also lead to a new kind of “decentered” textuality, in which no part or version of a text is prioritised over another.<sup>.xxxiv</sup> Likewise, stylometry has transformed how scholars understand the problem of authorship attribution. Under the influence of John Burrows, Patrick Juola and the Polish School of Stylometry, scholars have learnt to build statistical models of individual style, and distinguish authorship as a signal in the noisy flow of text. In book history and periodical studies, databases have destabilised the very concept of the “book”, “article” or “issue”, and scholars have had to develop new ways to model the production and consumption of text, as outlined in the case study below. What unites all these enterprises is the effort to create adequate computer models of texts, authors, or books, and the resultant need to redefine the very concepts scholars were trying to model in the first place.

Surveying this situation, Katherine Bode argues that the transformation of textual scholarship renders the entire distant reading debate otiose. Moretti’s vision of “*not reading*” texts is an empty dream, because all texts have already been interpreted according to whatever model was used to digitise or edit them in the first place; the dream of Moretti’s critics, that the close reader can exercise subjective readerly freedom, is empty for the same reason.<sup>.xxxv</sup> It must be said that more sophisticated distant readers grasp Bode and McCarty’s point, and have recast their approach as a kind of modelling.<sup>.xxxvi</sup> Meanwhile there has been an explosion of interest in new modelling techniques, including network analysis, literary mapping, and exciting experiments in gaming, virtual reality and augmented reality.<sup>.xxxvii</sup> In the Global South, pioneering scholars are pursuing various kinds of decolonial “world making”.<sup>.xxxviii</sup> In these ways, McCarty’s vision of interactive modelling continues to transform the way literature is studied and experienced.

## **The Emergence and Institutionalisation of Digital Humanities**

In order to bring about this new world of decentred texts and distant readers, scholars in Digital Humanities have erected a large scholarly infrastructure. Scholarly journals, academic programmes, job vacancies and research centres have proliferated, particularly since 2010,<sup>xxxix</sup> along with national academic organisations, affiliated to a global body, the Alliance of Digital Humanities Organisations (ADHO), which organises an annual international conference. Annual training camps have also emerged, staffed by academic enthusiasts who donate their time freely, including the Digital Humanities Summer Institute, the Oxford Digital Humanities Summer School, and Digital Humanities Down Under. This open team culture reflects the collaborative and interdisciplinary nature of Digital Humanities work and the commitment and openness of many practitioners to new modes of scholarly collaboration, public engagement and open publishing practices.

University, national and international infrastructures continue to evolve to promote, host and sustain Digital Humanities work,<sup>xi</sup> and to accommodate academic practices, outputs collaborations and careers that defy traditional metrics for academic evaluation, accreditation, publication and sustainability.<sup>xi</sup> These efforts have had mixed results, and scholars and administrators continue to disagree on how to establish effective research centres. Labs, social and creative spaces, software, hardware and access to supercomputers and digital storage may all play a part. However, many scholars concur with Laurent Dousset, a prominent French scholar who advised the French government that the primary infrastructure lies in people.<sup>xlii</sup> University administrations and funding bodies have been slow to absorb this lesson, which has led to well-known problems: broken and unstable teams; lack of career progression; loss of key institutional or project knowledge; projects delayed or scuttled; and outputs going off-line prematurely. By contrast, successful centres such as Stanford University's Centre for Spatial and Textual Analysis (CESTA) or Sheffield's Humanities Research Institute (HRI) have invested heavily in people, realising that the future development of Digital Humanities depends upon secure jobs and careers for both academic and technical staff.<sup>xliii</sup> Even in these centres, however, the situation can seem precarious.<sup>xliv</sup> Every project – and accompanying years of work – seems to topple perpetually on the verge of a precipice, menaced by the possibility of an undiscovered bug, the departure of a key person with irreplaceable knowledge, or the inability to secure funding to maintain an online resource. Barriers, even in privileged institutions, remain formidable.

Nevertheless, computationally-based research in the Humanities has a long pedigree: “Humanities Computing” can trace its origins back to the 1940s and 1950s. Its earliest pioneers include Professor Josephine Miles, who with an all-female team of students and punch-card operators between 1951 and 1956 produced a “Concordance to the Poetical Works of John Dryden”,<sup>xlvi</sup> and the Italian Jesuit priest Roberto Busa, who in 1946 began compiling a concordance of the nine million words in the sprawling work of Saint Thomas

Aquinas, eventually with the support of IBM.<sup>xlvii</sup> By the mid-1960s there were enough practitioners to support a journal, *Humanities and Computing*, and the first specialised academic associations were founded in the 1970s.<sup>xlviii</sup> Although the early history of Humanities Computing/ Digital Humanities has often been written from an Anglo-Saxon perspective, significant work was conducted outside Britain and America.<sup>xlix</sup>

Much of this early work was revolutionary, even if its impact took decades to register. In 1965-70, for example, François Furet and his collaborators published *Livre et Société dans la France du XVIII<sup>e</sup> Siècle*.<sup>i</sup> Using computational analysis of French bureaucratic and book trade records, Furet and his team offered foundational insights into reading prior to the 1789 revolution, and this work is only now being surpassed.<sup>ii</sup> A few years later, another French study pioneered the use of descriptive markers to assess the content of the pre-revolutionary newspaper press.<sup>iii</sup> The revolutionary research possibilities of large-scale digitization of extensive runs of newspapers using optical character recognition (OCR) and powerful bespoke search and analytic tools were only realised two decades later.<sup>iiii</sup>

These developments were accompanied from the mid-1990s by others which further empowered Digital Humanities research. These included the mass digitization of archives, objects and printed texts; the advent and uptake of the internet; and the Text Encoding Initiative (TEI), a scholarly initiative which created machine-readable text encoding “guidelines for the creation and management in digital form of every type of data created and used by researchers in the humanities”.<sup>liv</sup> With the publication of the first TEI guidelines in 1994, the humanities community had for the first time common digital standards and mark-up to facilitate research, teaching, data curation and preservation, and, in due course, interoperability.<sup>lv</sup>

### **The Digital Divide: Money, Power, Empire**

We have seen already that Digital Humanities has inspired considerable theoretical and methodological debate. Its institutionalisation has raised further criticisms. Critics worry about the creeping encroachments of neoliberalism and neocolonialism on academe, especially given the perceived technical and financial barriers of entry into the field.<sup>lxviii</sup> Teams of researchers and technologists can come with eye-watering price tags. The same can be said of indispensable research databases published by Gale-Cengage, ProQuest and Adam Matthew Digital. For some, these costs represent an unwelcome corporate invasion of the traditional research space, particularly when accompanied by attempts to monetise humanities research.<sup>lxix</sup> As winning funding has become an increasingly important activity for scholars, the expense of Digital Humanities research has perversely become one of its most prized aspects. Large ERC grants for projects such as Radboud University’s MEDATE project can reach 2M euros, whilst projects such as Oxford’s “Cultures of Knowledge”, Sheffield University’s “The Old Bailey Online” and its successors, or Western Sydney

University's "French Book Trade in Enlightenment Europe" project have often received repeat grants in six or seven figures. Major players have therefore emerged, surrounded by a cadre of precariously employed early-career scholars who often find themselves moving internationally as digital projects start and finish.<sup>lxx</sup> Yet these projects are collectively dwarfed in scale by the most ambitious project to date, the ECR-backed Time Machine project, which requires funding on a breathtaking scale. It involves, at the time of writing, over 400 partner organisations in 34 countries and has received, one million euros of ERC preliminary funding for its strategic planning phase alone.<sup>lxxi</sup> For the flagship Venice Time Machine initiative alone, there are plans to digitize, analyse and make accessible the entire Venetian state archives, which occupy 80 kilometres of archival storage.<sup>lxxii</sup>

The fear of a growing "digital divide" between wealthier and poorer researchers, institutions, libraries and countries is thus not without foundation. Nevertheless, scholars in the Global South have found ways to embrace Digital Humanities, as evidenced by pioneering efforts such as India's DHARTI, the Network for Digital Humanities in Africa, and the global movement for Indigenous Data Sovereignty. Such scholars have also pioneered approaches to digital scholarship that harness the power of ordinary computers, eschewing centralised computing clusters and subscription databases. The movement is known as "*jugaad* in India, *gambiarra* in Brazil, *rebusque* in Columbia, *jua kali* in Kenya, and *zizhu chuangxin* in China", or as "minimal computing" in the Anglophone world.<sup>lxxiii</sup> This is a vital and creative movement, though of course *jugaad* ["making do"] would not be necessary if the digital divide were not a reality.

Given the huge expense and skewed allocation of resources, critics ask: has Digital Humanities been worth the expense? As noted above, some critics argue that Digital Humanities is an uncritical enterprise, with a covert neoliberal agenda and an overt tendency to prioritise technicalities over critique. Certainly some Digital Humanities work necessarily engages with technical problems at the expense of humanistic ones, but much Digital Humanities research remains highly politically engaged. Cases in point include the "Slave Voyages" project's attempts to database and visualise three centuries of monstrous transatlantic human trafficking or the "Colonial Frontier Massacres" project's mapping of the true extent of settler violence against indigenous populations in Central and Eastern Australia between 1788 and 1930. Indeed, such digital projects have a capacity to engage and mobilise much broader publics interested in history, genealogy, demography or racial politics than much traditional research, particularly when they offer online resources with intuitive interactive tools.<sup>lxxv</sup> Nor was the first attempt at a distant-reading of eighteenth-century British erotica any less politically-engaged than the scholarship that preceded it. Indeed, its feminist conclusions might appear more authoritative for resting on the comprehensive – if not quite exhaustive – evidential base provided by datamining Gale-Cengage's magnificent Eighteenth-Century Collections Online (ECCO) digital resource.<sup>lxxvi</sup> Furthermore, digitally-empowered techniques such as literary mapping – which uses digital techniques to explore the temporal-spatial dimensions of fictional settings to illuminate where events happened and



the geospatial parameters for action within a novel – offer literary scholars (among others) previously unimagined means to develop deeper understandings of setting, plot, action and even mood and emotional associations.<sup>lxxvii</sup> They also offer new means for engaging audiences, particularly once a VR dimension is added.<sup>lxxviii</sup>

### **Digital Humanities in Practice: A Case Study**

Digital Humanities now offers such a dazzling arsenal of tools, techniques and possibilities to researchers, that a comprehensive catalogue or typology is beyond the scope of a single article. This section considers a range of the most prominent techniques in literary studies, and highlight some of the practical issues encountered by scholars who use them. Many of the most innovative projects and digital resources bring together more than one technique or approach. Indeed, the next “big thing” in Digital Humanities is likely to be the development of tools to exploit multiple datasets using linked open data techniques, though the full promise of the so-called “semantic web” has yet to be realised in practice.<sup>lxxix</sup> In the meantime, the main practical effect of Digital Humanities has been to change the way scholars encounter the archive. At the touch of a button, a student can search an archive as easily as a professor, and could potentially retrieve in seconds material that would not so long ago have taken a lifetime of work to discover.<sup>lxxx</sup> Though the promise of such encounters is great, scholarly expertise is still required to analyse interpret and explain the significance of the information gathered, and in practice digital scholarship encounters numerous unforeseen barriers. To demonstrate this practical dimension, this chapter now considers several examples that relate to a particular historical and literary period, the eighteenth-century European enlightenment. This is an area of study where Digital Humanities has had a particularly large impact.<sup>lxxxi</sup>

Much effort and expense in the early stages of the Digital Humanities revolution was expended on creating new digital editions and large digital text corpora. In both cases, the fundamental problem was the same: digitizing the text could be done relatively simply, but editing and organising it was labour-intensive, and required the development of complex new models of the “decentred” textuality (see “Modelling Text”). This mismatch has plagued many important digitization efforts, including Google Books and ECCO – which used optical character recognition (OCR) to transcribe thousands of historic books. These made huge amounts of text available, but the transcriptions were poor and in the case of Google Books the bibliographic data was patchy and impeded systematic discoverability. ECCO provides significantly better bibliographic data, relying on library MARC records to annotate each file. But even this data, gathered and curated by generations of librarians to varying standards, is inadequate for many scholarly purposes. A scholar trying to estimate how many books were published in different towns, for example, would be unable to discover whether a book published in “Richmond” was published in Surrey (UK), Yorkshire (UK), Virginia (US), or Jamaica. Discerning which was intended in each individual case would be a daunting task in a dataset of 220,000 volumes. Similarly, the database records 91,875 distinct publishers, but

since every inconsistently placed apostrophe or name variant generates a distinctive “result”, the true number of publishers is far lower. To realise the full potential of the data, such ambiguities need to be fully resolved by painstaking textual scholarship. This daunting data-cleaning task is being undertaken by the Digital Humanities group at the University of Helsinki under the direction of Mikko Tolonen.

Luckily, along with mass-digitisation projects like ECCO, there have been a range of projects collating key metadata. The research required to clean and edit ECCO is greatly aided by the Consortium of European Research Libraries (CERL) Thesaurus and its feeder databases (eg. Data@bnf and the British Book Trade Index (BBTI)), though expertise and resources are required to make use of these highly technical resources. As discussed below, when accurate, authoritative metadata is combined with cleaned up text, the possibilities for literary scholarship are immense. And as semantic web technologies become more widespread, it should soon be possible to automate and speed up large parts of the process.

Digital reading and modelling methods play a double role in this process of cleaning and editing. On the one hand, until the data itself is properly cleaned and organised, digital methods are of little use for literary interpretation. On the other hand, certain methods such as stylometry can be of great use during and beyond the cleaning phase. Stylometry has helped to confirm the long-suspected collaboration of Christopher Marlowe in the writing of Shakespeare’s plays, for example, and to identify the authors of scandalous political *libelle* pamphlets attacking Marie-Antoinette in the 1780s.<sup>lxxxiii</sup> In this way, it can help to provide the metadata on which large-scale digital studies rely. As data cleaning proceeds, more reading and modelling techniques become useful. One popular technique is *topic modelling*, is used to identify clusters of co-occurring words, or “topics”, in order to suggest what a document is about. It can thus be used to identify and interrogate recurring themes within a large set of texts.<sup>lxxxv</sup> Other popular interpretative techniques include *sentiment analysis*, which predicts the emotional charge of a text based on its vocabulary and syntax, *collocation analysis*, which finds pairs, triplets or larger sets of words that tend to co-occur with one another, and *word vectors*, where words are represented as points in a high-dimensional space that model their semantic relationships with one another.

Whilst these mathematically sophisticated techniques attract attention, it is often mathematically simpler methods that have proven most useful in practice, even when confronting ‘big data’. In one study, Clovis Gladstone and Charles Cooney used string-matching sequence alignment techniques – originally developed by computer scientists for applications such as spelling correction to identify literary “common-places” – frequently repeated phrases – in the ECCO database. They identified the *Bible* as the origin of 58.5% of all those whose origins could be determined, a surprising finding for a century associated with elite religious scepticism and growing secularisation.<sup>lxxxvi</sup> Their results bring traditional theories of secularisation into question, hinting at a culture still immersed, to an unsuspected

degree, in religious imagery. Thus, string-matching technique initially developed in bioinformatics, offers support to the hypothesis that the enlightenment rational critical tradition had largely religious origins, stemming from the habit of finding disputational evidence in scriptural texts.

Techniques like topic modelling and string matching analyse the text. But as scholars like Bode are so right to point out (see above, “Modelling Text”), the metadata painstakingly collected by editors and book historians often harbours equally important insights. In book history, the use of such data is often referred to as *bibliometry*. Recent industrial-scale bibliometry conducted for the “Mapping Print, Charting Enlightenment” database project has suggested that religious texts were far more prevalent in pre-revolutionary France than previously thought. Previous scholars relied on sources that under-reported such books; only large scale digital bibliometric analysis of supply-side sources produced by the publishing industry could capture the production of religious writing on a grand scale.<sup>lxxxvii</sup> Likewise, the MEDiate project’s study of private library catalogues and book ownership is uncovering the kinds of books that “mediated” between religious and secular worldviews, demonstrating the importance of overlooked authors like Madame Le Prince de Beaumont and Stephanie de Genlis.<sup>lxxxviii</sup> Thus the textual methods used by Cooney and Gladstone can be combined with bibliometry to provide a multifaceted view of enlightenment secularisation.

Our emphasis on the European enlightenment is not coincidental. As Digital Humanities emerged, scholars of eighteenth-century Britain and France in particular had the good fortune to enjoy a particularly privileged position.<sup>lxxxix</sup> The materials available to them were both immense and finite, making it possible to build large and comprehensive databases. Further, eighteenth-century scholars very early benefitted from a number of highly significant digital products. ECCO and the Burney collection of British newspapers gave them digital access to the lion’s share of all eighteenth-century printed products in English. “The Old Bailey Online” digitized three centuries of London’s criminal court records, making them publically accessible online with excellent research and analytic tools.<sup>xc</sup> Historians of France also had ready access via the “Electronic Enlightenment” database to a corpus of 70,000 letters written to and from such luminaries as Voltaire and Rousseau. Its superb metadata has empowered the work of Stanford’s outstanding “Mapping the Republic of Letters” project and the rich prosopographical insights which have stemmed from it.<sup>xcii</sup> These tools have been supplemented by the riches assembled on the French Bibliothèque nationales’ Gallica, a national digitization project surpassing all others, and extensive historical bibliometric initiatives such as “Mapping Print, Charting Enlightenment” and MEDiate.<sup>xciii</sup> Thus digital humanists have been able to use the eighteenth-century as a laboratory for experimentation, and eighteenth-century scholars have been at the forefront of attempts to realise the potential of linked data. In 2016 an annual international symposium and scholarly network, Digitizing Enlightenment, was established to further these efforts.<sup>xciv</sup>

One attempt to realise such ambitions is the “Libraries, Reading Communities and Cultural Formation in the Eighteenth Century Atlantic World”, an international project based at Liverpool University (UK).<sup>xcv</sup> It intends to create a database of every extant eighteenth-century subscription library catalogue in the English speaking world (covering about 80 institutions) together with borrower records and, primarily through ECCO, the actual texts held. In this way, the project will enable a fundamentally new kind of literary history, in which corpus analysis is weighted according to readership. Rather than treating texts as fixed units published at a particular time, the project will treat them as items that flow and persist across time and space, overcoming Bode’s objections to bibliographically naïve forms of distant reading. This marriage of corpus linguistics, historical bibliometrics, discourse analysis and big data analytics offers a powerful new means for exploring and conceptualising the significance of literary texts and non-fiction works and their relationship to processes of social, cultural and political change.

If the digital revolution has brought new ways of exploring the production, dissemination and content of texts, it has perhaps even greater potential for recovering and understanding reader response. Reader response, despite its interest to literary and book historians and the best efforts of educationalists and psychologists studying living subjects, remains relatively little understood. In past historical contexts it is particularly difficult to uncover how texts were read, since the act of reading generally leaves little tangible trace. Scholars have generally been forced to rely on printed reviews by high-brow literary reviewers or the equally atypical reading journals or common place books of diarists like Samuel Pepys or the Sheffield apprentice Joseph Hunter.<sup>xcvi</sup> The pioneering book historian Robert Darnton famously wrote a seminal article on reader responses to Rousseau based on just two readers.<sup>xcvii</sup>

Pioneering attempts to overcome this problem have been made by the Reading Experience Databases (REDs) that exist or are planned for several English speaking countries (Britain, Canada, Australia, New Zealand), the Netherlands, and tentatively Finland. These projects, based largely on volunteer crowd-sourced labour, gather data on documented reader experience wherever encountered, in some cases (e.g. Britain) over time spans as long as 500 years, and in others at fixed moments (the New Zealand RED, for example, focuses on reading during World War I).<sup>xcviii</sup> This approach is extremely labour-intensive but yields rich information. The online data-entry form gathers data on time of day, location, social context and postures in which acts of reading occurred, as well as on individual readers and their critical responses to texts. REDs can be hard to maintain, and are often skewed towards the experience of the most voracious readers. Even the largest RED, the British, is limited: 30,000 entries covering 500 years of reading equates to an average of just 60 documented acts per year.<sup>xcix</sup> Nonetheless such databases put the study of reading on a fundamentally new footing.

The RED approach to reader experience is not, however, the only possible one. With the proliferation of digitized sources, improved OCR and improvements in machine learning and sentiment analysis, harvesting dispersed traces of reading is increasingly feasible. The sources for such work include published book reviews; private correspondence; manuscript newsletters; reading journals; mentions of reading in fictional and non-fiction works; footnote citations and commonplaces which cross reference texts; readers' marginal annotations; school exercise books and university essays; publishers correspondence; and censors' reports.<sup>c</sup> There are, additionally, reports of police spies and agents; court, police and inquisitorial records; and mass observation diaries. Scholars regularly mine sites such as Goodreads and LibraryThing for data about reading practices today, as well as studying probably the most intense and digital form of reader response: fan fiction.<sup>ci</sup>

The discussions of infrastructure, funding and ambitious mega-projects in this chapter should not be taken to imply that the significance, utility and quality of Digital Humanities projects is best measured by scale. As the *jugaad* movement demonstrates, technical barriers to entry are not necessarily very great. Effective and impressive projects have been conducted using technology no more complex than a spreadsheet. Recent work of Cheryl Knott is a case in point: she studied six tiny colonial libraries' holdings, which pointed to significant differences in reading cultures either side of the Potomac.<sup>cii</sup> A recent project overseen by Gary Kates was similarly modest in method, though great in aspiration. Kates' students collated Worldcat data on almost 5,000 eighteenth-century editions of 171 leading titles which historians have associated with the Enlightenment, across all major European languages, to assess which books were most frequently reprinted across the century. The most popular proved to be political works, particularly novels, several of which were reprinted across the entire century. Montesquieu, Rousseau and Marmontel were unsurprisingly among the most reprinted authors, but the two most reprinted books were Madame de Graffigny's *Lettres peruviennes* (282 editions across the century) and Fénelon's *Télémaque*, an apology for tempered monarchy, which first appeared in 1699 and ran to an astonishing 445 editions.<sup>civ</sup> This study, whose primary research tool was the search bar on WorldCat, seems to have turned back the clock on enlightenment historiography by three generations, suggesting that constitutional monarchy was the period's most popular political theme.<sup>cv</sup>

Whilst most projects discussed here gathered their own data, existing textual databases and other digital resources offer near infinite research possibilities for Digital Humanities researchers and students, and many major digitized research collections are now coming with research tools attached, such as Gale-Cengage's Digital Scholar Lab. Most digital humanists, however, use free software packages such as Voyant Tools, Gephi and the *stylo* package of Maciej Eder and Jan Rybicki (see "Links to Digital Materials"). Increasingly, humanists are also learning to code, particularly in R, Python, and JavaScript. They release their code on Github under permissive licences, allowing re-use, learning, and scholarly access to the bases of research. Online forums such as StackExchange provide copious free advice to beginning

programmers. As Digital Humanities has matured, practitioners have begun to produce comprehensive textbooks describing basic methods and best practices (see “Further Reading”). Yet it in the Digital Humanities research space, agreed methodologies and best practice are not enough. As James E Dobson has argued, computational tools themselves need to be continually critiqued and since the humanities, unlike the sciences, lack a defined knowledge frontier, both methods and findings subjected to constant re-examination and reinterpretation. Envisaged thus, the Digital Humanities have the potential to continue in the best traditions of critical humanistic enquiry.<sup>.cvi</sup>

## Conclusion

Nothing in the discussion presented here suggests an intent, desire or ability for Digital Humanities to replace traditional literary study. Rather, this article has sought to explore the power of computationally-enhanced humanities to accommodate the subjective insights of close reading of texts within wider appreciations of the cultural, social, and literary contexts. This involves applying new methods to interpret, understand and experience texts, by revealing patterns within, between or beyond individual texts which are not evident or obvious to unaided human perception. It involves the development of new models to make sense of these patterns, and determine their connection to the world beyond the text. The tools, digital archives, datasets, ontologies and techniques used to read and model literature are still at a relatively early stage, as are scholars’ abilities to interpret the complex visualisations and numerical outputs needed to understand them. These new technologies have changed how scholars encounter literature and provoked considerable theoretical debate. Digital Humanities remains a contested term describing a field in flux. If asked to assess its impact it would be wise to concur with Chinese premier Zhou Enlai’s apocryphal response when asked to assess the impact of the yet greater revolution that struck France in 1789. “It is too early to say”.<sup>.cvii</sup>

## Guide to the Literature

The literature on Digital Humanities is voluminous. DARIAH-DE, Germany’s peak Digital Humanities body, maintains a [useful biography](#) online. The “Further Reading” section lists a number of useful general texts, case studies and textbooks. In contrast this section focuses on major works of theory and criticism that have driven debate in digital literary studies.

In the field of computational literary criticism, key early works include Robert Cluett’s *Prose Style and Critical Reading* (1976) and John Burrows’ *Computation into Criticism* (1987).<sup>.cviii</sup> Cluett and Burrows harnessed the power of early research computers to study subtle stylistic fluctuations in classic texts, and formulated a strong theory of language as an index of individual personality to justify their methods. Recently several scholars have revived this early vision of precise and subtle computational literary criticism. Stephen Ramsay’s *Reading Machines* (2011) promotes an anarchic form of digital reading rooted in the philosophy of

Paul Feyerabend; Monika Bednarek's *Languange and Television Series* (2018) combines corpus linguistics with screenwriter interviews to show how race, class and regional identity are represented in contemporary teleplays; and Martin Paul Eve's *Close Reading with Computers* (2019) promotes a vision of "narrow deep" reading with computational tools.<sup>cix</sup>

As mentioned above ("The Distant Reading Debate"), in recent decades the dominant force in computational literary criticism has been towards "distant reading", the statistical analysis of large corpora of texts to analyse long-term trends. This project is most often associated with Franco Moretti, whose early papers on the subject are collected in *Distant Reading* (2013).<sup>cx</sup> Recent works in a similar vein include Matthew Jockers's *Macroanalysis* (2013) and *The Bestseller Code* (2016), co-written with Jodie Archer; Andrew Piper's *Enumerations* (2018); and Ted Underwood's *Distant Horizons* (2019).<sup>cxii</sup> All these works contain useful theoretical reflections as well as practical applications of large-scale analysis. No reader of these works can afford to ignore Nan Da's thrilling critique, "The Computational Case against Computational Literary Studies" (2019), and the numerous responses it has provoked.<sup>cxii</sup>

An alternative vision of Digital Humanities sidelines the question of "reading", considering instead how computers allow scholars to build models of literary history. In *Humanities Computing* (2005), Willard McCarty describes a method of "interactive modelling", in which scholars continually grapple with the computer's inability to fully capture reality.<sup>cxiii</sup> Moretti himself once promoted a similar method, and in works such as *Atlas of the European Novel* (1998) and *Graphs, Maps, Trees* (2005), he utilised modelling techniques such as digital mapping and network analysis.<sup>cxiv</sup> In *A World of Fiction* (2018), Katherine Bode offers a profound critique of the way scholarly databases model literary history.<sup>cxv</sup> In this vein, Laura Mandell's *Breaking the Book* (2015) and Alan Liu's *Friending the Past* (2018) consider how digital media change the way scholars encounter the literary past.<sup>cxvi</sup>

This change is evident from the interrelated fields of stylometry, book history and scholarly editing. Stylometry is now a standard tool of authorship attribution for scholarly editors. Readers interested in the underlying theory may begin with articles by John Burrows, Patrick Julioa, Jan Rybicki and Maciej Eder.<sup>cxvii</sup> Readers interested in how Digital Humanities is transforming book history and the concept of the scholarly edition should consult Jerome J. McGann's *Radiant Textuality* (2001), Bode's *World of Fiction* and Paul Eggert's *The Work and the Reader in Literary Studies* (2019).<sup>cxviii</sup> For practical examples of digital book history, or "historical bibliometrics", readers may consult Simon Burrows and Mark Curran twin volume's on *The French Book Trade in Enlightenment Europe* (2018), and Bode's *Reading by Numbers* (2012).<sup>cxix</sup> The fascinating history of digital books lies beyond the scope of this chapter: interested readers should instead consult the articles on "Reading in the Digital Age", "E-Text" and "Hypertext Theory" in this volume.

To conclude this discussion, no survey would be complete without noting Roopika Risam's *New Digital Worlds* (2019).<sup>cxx</sup> Risam's work is not confined to literary studies, but she identifies crucial power imbalances that undercut the ideal of Digital Humanities, and describes the pioneering interventions of scholars and theorists from the Global South.

## **Further Reading**

### ***I. Digital Humanities, general***

Arthur, Paul Longley and Katherine Bode. *Advancing Digital Humanities: Research, Methods, Theories*. Houndsmills, Basingstoke: Palgrave Macmillan, 2014.

Bodenhamer, David J., John Corrigan, and Trevor M. Harris, eds. *The Spatial Humanities*. Bloomington: Indiana University Press, 2010.

Burdick, Anne, et al. *Digital\_Humanities*. Cambridge, Mass.: MIT Press, 2012.

Crompton, Constance, Richard J. Lane and Ray Siemens, eds. *Doing Digital Humanities: Practice, Training, Research*. New York: Routledge, 2016.

Gold, Matthew K., and Lauren F. Klein, eds. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2016.

Hirsch, Brett D, ed. *Digital Humanities Pedagogy: Practices, Principles and Politics*. Open Book Publishers, 2012.

McCarty, Willard. *Humanities Computing*. Basingstoke: Palgrave, 2005.

Risam, Roopika. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis and Pedagogy*. Evanston: Northwestern University Press, 2019.

Terras, Melissa, Julianne Nyhan and Edward Vanhoutte, eds. *Defining Digital Humanities. A Reader*. New York: Routledge, 2013.

Unsworth, John, Susan Schreibmann and Ray Siemens, eds. *A New Companion to the Digital Humanities*. Chichester: John Wiley & Sons, 2015.

### ***II. Digital Humanities, literary studies***

Archer, Jodie, and Matthew Jockers. *The Bestseller Code: Anatomy of the Blockbuster Novel*. London: Penguin, 2016.

Armstrong, Nancy, et al. "Theories and Methodologies: On Franco Moretti's Distant Reading", *PMLA* 132.3 (May 2017): 613-689.



- Bednarek, Monika. *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge Applied Linguistics. Cambridge: Cambridge University Press, 2018.
- Bode, Katherine. *Reading By Numbers: Recalibrating the Literary Field*. London: Anthem, 2012.
- Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press, 2018.
- Burrows, John. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Criticism*. Oxford: Clarendon Press, 1987.
- Cluett, Robert. *Prose Style and Critical Reading*. New York: Teachers College Press, 1976.
- Da, Nan. "The Computational Case against Computational Literary Studies" *Critical Inquiry* 45.3 (2019): 601-639
- Eve, Martin Paul. *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's "Cloud Atlas"*. Open access ebook.. Stanford: Stanford University Press, 2019. <https://doi.org/10.21627/9781503609372>.
- Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. Urbana, Chicago and Springfield: University of Illinois Press, 2013.
- Liu, Alan. *Local Transcendence: Essays on Postmodern Historicism and the Database*. Chicago: Chicago University Press, 2008.
- Liu, Alan. *Friending the Past: The Sense of History in the Digital Age*. Chicago: Chicago University Press, 2018.
- Mandell, Laura. *Breaking the Book: Print Humanities in the Digital Age*. Chichester: Wiley-Blackwell, 2015.
- McGann, Jerome J. *Radiant Textuality: Literature After the World Wide Web*. Houndmills: Palgrave, 2001.
- Moretti, Franco. *Atlas of the European Novel: 1800-1900*. London and New York: Verso, 1999.
- Moretti, Franco. *Distant Reading*. London and New York: Verso, 2013.
- Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. London and New York: Verso, 2005.
- Ramsay, Stephen. *Reading Machines: Towards an Algorithmic Criticism*. Urbana, Chicago and Springfield: University of Illinois Press, 2011.
- Siemens, Ray and Susan Schreibman, eds. *A Companion to Digital Literary Studies*. Chichester: Wiley, 2013.
- Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: Chicago University Press, 2019.

### ***III. Case Studies: Digital Studies of the European Enlightenment***

Burrows, Simon and Glenn Roe, eds. *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*. Liverpool: Oxford Studies in Enlightenment, 2020.

Burrows, Simon. *The French Book Trade in Enlightenment Europe II: Enlightenment Best-Sellers*. London and New York: Bloomsbury, 2018.

Curran, Mark. *The French Book Trade in Enlightenment Europe I: Selling Enlightenment*. London and New York: Bloomsbury, 2018.

De Bolla, Peter. *The Architecture of Concepts: The Historical Formation of Human Rights*. New York: Fordham University Press, 2013.

Edmondson, Chloe, and Dan Edelstein, eds. *Networks of Enlightenment. Digital Approaches to the Republic of Letters*, Oxford University Studies in the Enlightenment. Liverpool: Liverpool University Press, 2019.

Furet, François, et al, eds. *Livre et Société dans la France du XVIII<sup>e</sup> Siècle*. 2 vols. Paris: Mouton, 1965-1970.

Paule Jansen, ed. *L'Année 1778 à travers la presse traitée par ordinateur*. Paris: Presses Universitaires de France, 1982.

### ***IV. Textbooks in Digital Method***

Arnold, Taylor and Lauren Tilton. *Humanities Data in R: Exploring Networks, Geospatial Data, and Text*. Cham: Springer, 2015.

Gregory, Ian. *A Place in History: A Guide to Using GIS in Historical Research*. Oxford: Oxbow Books, 2003. [A [revised edition](#) is available free on Gregory's ResearchGate profile].

Jockers, Matthew. *Text Analysis with R for Students of Literature*. Cham: Springer, 2014.

Patrick Juola. "Authorship Attribution", *Foundations and Trends in Information Retrieval* 1, no. 3 (2006): 233-334.

Newman, Mark. *Networks*. 2<sup>nd</sup> edition. Oxford: Oxford University Press, 2018.

Rockwell, Geoffrey, and Stéfan Sinclair. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: MIT Press, 2016.

### **Links to Digital Materials**

[1947 Archive](#): A pioneering oral history archive, which collects testimonies of Partition from survivors in India, Pakistan and Bangladesh. An example of digital “world making” in the Global South.

[ADHO](#): The Alliance of Digital Humanities Organisations, which runs the yearly *Digital Humanities* conference and co-ordinates the *Digital Scholarship in the Humanities* journal.

[DARIAH-DE bibliography](#): A reasonably comprehensive bibliography that is constantly updated, and can be downloaded in the convenient form of a Zotero library.

[Drama Corpora Project](#): A huge database of German, Russian, Italian and other plays, an excellent example of network analysis, text analysis and Linked Open Data.

[Gephi](#): The most popular program for network analysis. Free and relatively intuitive, it provides support for a range of visualisation and analysis techniques.

[Humanist discussion group](#): This is the leading international forum for the discussion of Digital Humanities, and has been moderated by Willard McCarty since 1987.

[Mapping Emotions in Victorian London](#): An interesting application of geospatial analysis and text analysis.

[Programming Historian](#): Provides free, peer-reviewed, and well-pitched tutorials in numerous digital technologies and techniques, in English, Spanish and French.

[Python](#): Along with R, probably the most popular programming language in Digital Humanities. There are numerous online tutorials.

[QGIS](#): One of many free geographic information systems used for mapping and geospatial analysis.

[Rosetti Archive](#): Jerome J. McGann’s influential digital edition of Dante Gabriel Rosseti, which drove early debates about digital scholarly editing.

[RStudio](#): R is one of the most popular programming languages in Digital Humanities, and RStudio is a free Integrated Programming Environment (IDE) that makes the language easier to use. Many key tools in Digital Humanities are released as “R packages”, such as Jan Rybicki and Maciej Eder’s [stylo](#) package for authorship analysis, Matthew Jockers’s [syuzhet](#) package for sentiment analysis, and David Mimno’s [mallet](#) package for topic modelling.

[Text Encoding Initiative](#): The TEI website contains detailed information about the TEI specifications, and links to training and resources.

[Trove](#): A leading example of a national newspaper and document archive, built on open data and crowdsourcing principles, subsequently emulated around the world.

[Voyant Tools](#): A popular suite of free text analysis tools, built by Geoffrey Rockwell and the late Stéfan Sinclair.

[Zotero](#): The bibliographic software of choice for many digital humanists. The free version is highly functional, and allows for easy use of the DARIAH-DE bibliography.

## Citations

- i On the discipline's emergence and self-presentation, see Melissa Terras, Julianne Nyhan and Edward Vanhoutte, eds, *Defining Digital Humanities. A Reader* (Routledge, 2013); Matthew K. Gold and Lauren F. Klein, eds, *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2016).
- ii The rhetorical shift has been attributed to John Unsworth, Susan Schreibmann and Ray Siemens in *A Companion to the Digital Humanities*, <http://www.digitalhumanities.org/companion/> (2004). The term entered wider circulation around 2010, especially after the *New York Times* (10 November 2010) ran a front-page article featuring a prize-winning student-produced visualisation from Stanford University's "Mapping the Republic of Letters" Digital Humanities project.
- iii Melissa Terras, "Peering Inside the Big Tent", in *Defining Digital Humanities*, eds. Terras, Nyhan and Vanhoutte, 263-70, 267.
- iv *Ibid*, 268.
- v A compelling refutation of this position was offered as early as 1998 by Willard McCarty in "What is Humanities Computing? Toward a Definition of the Field", Centre for Humanities Computing, <http://www.mccarty.org.uk/essays/McCarty.%20What%20is%20humanities%20computing.pdf> (retrieved 23 August 2019).
- vi Matthew Kirschenbaum, "What is Digital Humanities, and What is it doing in English Departments?", in *Defining Digital Humanities*, 201.
- vii Chris Alen Sula and Heather V. Hill, "The Early History of Digital Humanities: An Analysis of *Computers and the Humanities* (1966–2004) and *Literary and Linguistic Computing* (1986–2004)", *Digital Scholarship in the Humanities* 34, supplement 1 (2019): I, 190–206.
- viii Franco Moretti, "Conjectures on World Literature", *New Left Review* 1 (2000): 54–68. The essay is reprinted in *Distant Reading* (London and New York: Verso, 2013), 43-62.
- ix *Ibid*.
- x *Distant Reading*, 44.
- xi Jodie Archer and Matthew Jockers, *The Bestseller Code: Anatomy of the Blockbuster Novel* (London: Penguin, 2016); Matthew Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana, Chicago and Springfield: University of Illinois Press, 2013); Andrew Piper, *Enumerations: Data and Literary Study* (Chicago: Chicago University Press, 2018); Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: Chicago University Press, 2019).
- xii <https://litlab.stanford.edu/pamphlets/>
- xiii Underwood, *Distant Horizons*, x.
- xiv For this critique, made generally in the print media, see Daniel Allington, Sarah Brouillete, and David Golumbia, "Neoliberal Tools (and Archives): A Political History of Digital Humanities", *Los Angeles Review of Books*, 1 May 2016; Adam Kirsch, "Technology is taking over English Departments. The False Promise of Digital Humanities", *The New Republic*, 3 May 2014; Carl Straumsheim, "Digital Humanities Bubble", *Inside Higher Education*, 8 May 2014; Timothy Brennan, "The Digital Humanities Bust", *The Chronicle of Higher Education*, 15 October 2017.
- xv David Golumbia, "Death of a Discipline", *Differences* 25.1 (May 2014): 157.
- xvi Nan Z. Da, "The Computational Case against Computational Literary Studies", *Critical Inquiry* 45.3 (2019): 601–639, doi:[10.1086/702594](https://doi.org/10.1086/702594).
- xvii *Ibid*, 638.
- xviii Katherine Bode, "Computational Literary Studies: Participant Forum Responses, Day 2", *In the Moment* (blog), April 2019, <https://critinq.wordpress.com/2019/04/02/computational-literary-studies-participant-forum-responses-day-2-3/>; Andrew Piper, "Do We Know What We Are Doing?", *Journal of Cultural Analytics*, Debates, January 2020, doi:[10.22148/001c.11826](https://doi.org/10.22148/001c.11826); Ted Underwood, "Critical Response II. The Theoretical Divide Driving Debates about Computation", *Critical Inquiry* 46. 4 (2020): 900–912, doi:[10.1086/709229](https://doi.org/10.1086/709229).
- xix Johanna Drucker, "Why Distant Reading Isn't", *PMLA* 132.3 (2017): 628–635, doi:[10.1632/pmla.2017.132.3.628](https://doi.org/10.1632/pmla.2017.132.3.628), 629.
- xx Barbara Herrnstein Smith, "What Was "Close Reading"? A Century of Method in Literary Studies", *Minnesota Review* 87.1 (2016): 73.
- xxi Jerome J. McGann and Lisa Samuels, "Deformance and Interpretation", in Jerome J. McGann, *Radiant Textuality: Literature After the World Wide Web* (Houndmills: Palgrave, 2001), 105-35; Stephen Ramsay, *Reading Machines: Towards an Algorithmic Criticism* (Urbana, Chicago and Springfield: University of Illinois Press, 2011).
- xxii See "[Information and Meaning](#)", in this volume.

- xxiii See Mark Algee-Hewitt et al, “Canon/Archive: Large-Scale Dynamics in the Literary Field”, *Pamphlets of the Stanford Literary Lab* 11 (2016), 3-5; Pierre Bourdieu, “Le Champs littéraire”, *Actes de la Recherche en Sciences Sociales* 89 (1991): 3-46.
- xxiv Underwood, *Distant Horizons*, xxi; Brennan, “The Digital Humanities Bust”.
- xxv Robert Cluett, *Prose Style and Critical Reading* (New York: Teachers College Press, 1976).
- xxvi J. F. Burrows, *Computation into Criticism: A Study of Jane Austen’s Novels and an Experiment in Criticism* (Oxford: Clarendon Press, 1987).
- xxvii Burrows, *Computation into Criticism*, 93-95. See also John Burrows, “Rho-Grams and Rho-Sets: Significant Links in the Web of Words”, *Digital Scholarship in the Humanities* 33.4 (2018): 725.
- xxviii See for instance Monika Bednarek, *Language and Television Series: A Linguistic Approach to TV Dialogue*, (Cambridge: Cambridge University Press, 2018), doi:[10.1017/9781108559553](https://doi.org/10.1017/9781108559553); Martin Paul Eve, *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell’s “Cloud Atlas”*, Open access ebook (Stanford: Stanford University Press, 2019), doi:[10.21627/9781503609372](https://doi.org/10.21627/9781503609372).
- xxix Willard McCarty, *Humanities Computing* (Houndmills: Palgrave, 2005), 27.
- xxx *Ibid*, 25.
- xxxi *Ibid*, 38.
- xxxii On modelling as an exploratory and experimental tool, see Dennis Yi Tenen, “Towards a Computational Archaeology of Fictional Space”, *New Literary History* 49. 1 (2018): 119-47. On the challenges of modelling in the humanities see Julia Flanders and Fotis Jannidis, “Data Modelling” in *A New Companion to the Digital Humanities*, eds. John Unsworth, Susan Schreibmann and Ray Siemens (Chichester: John Wiley & Sons, 2015), 239-37.
- xxxiii See Alan Liu, *Local Transcendence: Essays on Postmodern Historicism and the Database* (Chicago: Chicago University Press, 2009); Alan Liu, *Friending the Past: The Sense of History in the Digital Age* (Chicago: Chicago University Press, 2019); Laura Mandell, *Breaking the Book: Print Humanities in the Digital Age* (Chichester: Wiley-Blackwell, 2015).
- xxxiv McGann, *Radiant Textuality*, 70; see also Paul Eggert on how digital editing expands the gap between the “archival” and “editorial” impulses: *The Work and the Reader in Literary Studies: Scholarly Editing and Book History* (Cambridge: Cambridge University Press, 2019), chap. 5.
- xxxv Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (University of Michigan Press, 2018), chap. 1.
- xxxvi See, for instance, Underwood, *Distant Horizons*, xv-xvii.
- xxxvii On networks: Vincent Labatut and Xavier Bost, “Extraction and Analysis of Fictional Character Networks: A Survey”, *ACM Computing Surveys* 52.5 (2019): 89:1–89:40, doi:[10.1145/3344548](https://doi.org/10.1145/3344548). On literary mapping: Sara Luchetta, “Exploring the literary map: An analytical review of online literary mapping projects”, *Geography Compass* 11.1 (2017), <https://doi.org/10.1111/gec3.12303>. On gaming, VR and AR, see the *Hyde* project, “which adapts Stevenson’s classic novella, *Strange Case of Dr Jekyll and Mr Hyde*, into a pervasive media game driven by players’ bio-data”: <http://www.react-hub.org.uk/projects/alumni-books-print/hyde/>.
- xxxviii Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis and Pedagogy* (Evanston: Northwestern University Press, 2019), 32-36.
- xxxix As early as 2008, Diane M. Zorich, *A Survey of Digital Humanities Centers in the United States* (Washington: Council on Library and Information Resources, November 2008), 48, listed 32 surveyed organisations.
- xl Large scale national or transnational infrastructure projects include the ERC’s Europeana and DARIAH initiatives, France’s Gallica, or Australia’s Humanities Networked Infrastructure (Huni).
- xli These issues are, for example, raised in the Australian Association for Digital Humanities (AaDH) submission to the Australian Academy for the Humanities “Future Humanities Workforce” consultation at <https://aa-dh.org/join/about/advocacy/future-humanities-workforce-consultation/>, retrieved 28 July 2019.
- xlii Laurent Dousset, “The politics of interoperability in France”, unpublished paper to the Digital Humanities Research Group, Western Sydney University, 3 May 2018.
- xliii See the AaDH submission to the Australian Academy for the Humanities “Future Humanities Workforce” consultation, which calls for “initiatives designed to address concerns around job security, and recognition and career development pathways for the “alt-ac” positions that many ECRs enter (particularly in the digital humanities) as well as pathways for people to flow back and forth between such positions and more traditional academic roles.”
- xliv Dan Edelstein, “Mapping the Republic of Letters: History of a Digital Humanities Project”, in *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*, eds. Simon Burrows and Glenn Roe (Liverpool: Oxford Studies in Enlightenment, forthcoming 2020), ch. 3.
- xlvi Rachel Sagner Buurma and Laura Heffernan, “Search and Replace: Josephine Miles and the Origins of Distant Reading”, *Modernism / Modernity Print+* 3.1 (11 April 2018), at <https://modernismmodernity.org/forums/posts/search-and-replace>, retrieved 28 July 2019.



- xlvi Thomas N. Winter, “Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance”, *The Classical Bulletin* 75.1 (1999), 3-20.
- xlviii ADHO’s website states that the Association for Literary and Linguistic Computing (now known as the European Association for Digital Humanities) was founded in 1973 and the Association for Computers and the Humanities in 1978: <http://adho.org/about> retrieved 30 June 2019.
- xlix See, for example, Susan Hockey, “The History of Humanities Computing” in *A Companion to the Digital Humanities*, eds. Unsworth, Schreibmann and Siemens, 1-19. In her opening paragraphs Hockey mentions Busa, the Trésor de la Langue Française (now ARTFL), and the Institute of Dutch Lexicography at Leiden, but thereafter mostly discusses American, Canadian and British projects, notwithstanding a couple of later examples from Italy and Norway.
- i François Furet, et al, eds. *Livre et Société dans la France du XVIII<sup>e</sup> Siècle*, 2 vols (Paris, 1965-1970).
- ii See the outputs of the French Book Trade in Enlightenment Europe project, notably the later chapters of Simon Burrows, *The French Book Trade in Enlightenment Europe II: Enlightenment Best-Sellers* (London and New York: Bloomsbury, 2018), which draw on many of the sources used by Furet and his collaborators.
- iii Paule Jansen, ed., *L’Année 1778 à travers la presse traitée par ordinateur* (Paris: Presses Universitaires de France, 1982).
- iiii The best and most comprehensive of digitized newspaper collections include the Burney collection, digitized by Gale-Cengage, and the Australian newspapers in Trove, whose OCR has been meticulously corrected by crowd-sourced volunteers.
- liv Lou Burnard, *What is the Text Encoding Initiative? How to add intelligent markup to digital resources* (Open Edition Press, 2014), introduction. Online edition, retrieved from <https://books.openedition.org/oep/679> on 30 June 2019.
- lv See “TEI: Text Encoding Initiative” at <https://tei-c.org/>. Retrieved 30 June 2019.
- lvi Brennan, “The Digital Humanities Bust” is particularly vocal on this point.
- lvii On Digital Humanities and the neoliberal agenda, see Allington, Brouillete and Golumbia, “Neoliberal Tools (and Archives): A Political History of Digital Humanities”.
- lviii AaDH submission to the Australian Academy for the Humanities “Future Humanities Workforce”, cited above, note 6.
- lix <https://www.timemachine.eu/institutions/>; “Time Machine Heralds New Era”, press announcement, 15 March 2019, at [https://tu-dresden.de/tu-dresden/newsportal/news/time-machine-laeutet-neues-zeitalter-ein?set\\_language=en](https://tu-dresden.de/tu-dresden/newsportal/news/time-machine-laeutet-neues-zeitalter-ein?set_language=en), both retrieved 28 July 2019.
- lx “Venice Time Machine” Brochure at <https://vtm.epfl.ch/wp-content/uploads/2019/01/BrochureVTM.pdf> retrieved 28 July 2019.
- lxi Risam, *New Digital Worlds*, 43.
- lxii See <https://www.slavevoyages.org/voyage/database>; <https://c21ch.newcastle.edu.au/colonialmassacres/> both retrieved 28 July 2019.
- lxiii See Jennifer Ann Skipp, “British Eighteenth-Century Erotica: A Reassessment”, PhD thesis, University of Leeds, 2007. Skipp’s contentions about the sexual libertine and misogynist social attitudes underpinning this literature are based on a digitally-empowered empirical analysis of sexual acts depicted in a corpus of almost 850 erotic texts. No previous scholar had identified more than 350.
- lxiv See, for example, Ryan Heuser, Franco Moretti and Erik Steiner, “The Emotions of London”, *Pamphlets of the Stanford Literary Lab* 13 (2016) at <https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf> retrieved 3 September 2020.
- lxv For an overview of this sub-field see Sara Luchetta, “Exploring the literary map”. Major literary mapping projects include *A Literary Atlas of Europe* (<http://www.literaturatlas.eu/en/2012/03/23/ein-literarischer-atlas-europas-poster/>) and *Mapping the Lakes: A Literary GIS* (<http://www.lancaster.ac.uk/mappingthelakes/>), both retrieved on 17 January 2019.
- lxvi For a vision statement for one domain of inquiry – library history – see Simon Burrows, “Locating the Minister’s Looted Books: From Provenance and Library Histories to the Digital Reconstruction of Print Culture”, *Library and Information History* 31.1 (2015): 1-17.
- lxvii To take but one example from the authors’ experience, a student wanting to study the iconography of the French revolution in Britain for her BA dissertation located 3,000 newspaper articles on “trees of liberty” drawn from across the eighteenth century with a single search query using the Burney collection.
- lxviii See also Burrows and Roe, eds., *Digitizing Enlightenment*.
- lxviiii See the two-part treatment of Robert A. J. Matthews and Thomas V.N. Merriam, “Neural Computation in Stylometry: An Application to the Works of Shakespeare and Fletcher”, *Literary and Linguistic Computing* 8.4 (1993) and 9.1 (1994); Simon Burrows, *A King’s Ransom: The Life of Charles Théveneau de Morande, Blackmailer, Scandalmonger and Master-Spy* (London: Continuum, 2010), 152, 246 n. 124.
- lxviiii See for instance Matthew L. Jockers and David Mimno, “Significant Themes in 19th-Century Literature”, *Poetics* 41:6 (2013): 750–769, doi:<http://dx.doi.org/10.1016/j.poetic.2013.08.005>.

- lxxxvi Clovis Gladstone and Charles Cooney, “Opening New Paths for Scholarship: Algorithms to Track Text Reuse in Eighteenth Century Collections Online (ECCO)” in *Digitizing Enlightenment*, chap. 14.
- lxxxvii Simon Burrows, “Forgotten Best-Sellers of Pre-Revolutionary France”, *French History and Civilisation: Papers from the George Rudé Seminar*, vol. 7, 2016 seminar (2017), 51-65.
- lxxxviii Alicia Montoya, “Shifting Perspectives and Moving Targets: From Conceptual Vistas to Bits of Data in the First Year of the MEDiate Project” in *Digitizing Enlightenment*, eds. Burrows and Roe, 195-218; Alicia Montoya, “French and English Women Writers in Dutch Library Catalogues, 1700-1800. Some Methodological Considerations and Preliminary Results”, in *I Have Heard about You”. Foreign Women’s Writing Crossing the Dutch Border: from Sappho to Selma Lagerlöf*, eds. Suzan van Dijk et al. (Hilversum 2004), 182-216.
- lxxxix See Paddy Bullard, “Digital Humanities and Electronic Resources in the Long Eighteenth Century”, *Literature Compass* 10.10 (2013): 748–760, doi:[10.1111/lic3.12085](https://doi.org/10.1111/lic3.12085).
- xc <https://www.oldbaileyonline.org/>; see also its successor projects, <https://www.londonlives.org/> and <https://www.digitalpanopticon.org/>, retrieved 28 July 2019.
- xcii See <https://www.e-enlightenment.com/>; <http://republicofletters.stanford.edu/>, retrieved 28 July 2019.
- xciii <https://gallica.bnf.fr/accueil/en/content/accueil-en?mode=desktop>, retrieved 28 July 2019.
- xciv Meetings of the Digitizing Enlightenment symposium have taken place in Western Sydney University (2016), Nijmegen (2017), Oxford (2018) and Edinburgh (2019). The 2020 symposium is scheduled for Montpellier.
- xcv This project, funded by the British AHRC, is headed by Mark Towsey and has eight investigators and nine “impact partners” based in Britain, Australia and North America.
- xcvi Stephen M. Colclough, “Procuring Books and Consuming Texts: The Reading Experience of a Sheffield Apprentice, 1798”, *Book History* 3 (2000): 21-44.
- xcvii Robert Darnton, “Readers respond to Rousseau: the fabrication of Romantic sensibility” in Robert Darnton, *The Great Cat Massacre and other episodes in French Cultural History* (New York, Basic Books, 1984), 217-56.
- xcviii See <https://www.open.ac.uk/Arts/reading/UK/> retrieved 14 January 2019. On the New Zealand RED initiative see <https://nzredblog.wordpress.com/nzred/>, last visited 28 July 2019.
- xcix Statistics from <http://www.open.ac.uk/Arts/reading/about.php> retrieved 28 July 2019.
- c Nicole Moore, ed., *Censorship and the Limits of the Literary: A Global View* (London: Bloomsbury, 2015); Robert Darnton, *Censors at Work: How States Shaped Literature* (New York: Norton, 2015). See also Nicole Moore’s chapter on “Censorship” in this volume.
- ci For leading scholarship in fan-fiction studies see Karen Hellekson and Kristina Busse, *The Fan Fiction Reader* (Iowa City: University of Iowa Press, 2014).
- cii See Cheryl Knott, “Uncommon Knowledge: Late Eighteenth-Century American Subscription Library Collections” in *Before the Public Library: Reading, Community, and Identity in the Atlantic World, 1650-1850*, eds. Mark Towsey and Kyle Roberts (Leiden: Brill, 2018), 149-73.
- civ The authors are grateful to Professor Kates for access to the beta version of his online database and permission to publish preliminary provisional summary statistics first presented at Digitizing Enlightenment 2 in Nijmegen in 2017.
- cv This finding supports Annalien de Dijn, who rejects a politically radical enlightenment, as espoused by three generations of American enlightenment scholars, notably Peter Gay, Robert Darnton and Jonathan Israel. See Annalien de Dijn, “The Politics of Enlightenment from Peter Gay to Jonathan Israel”, *Historical Journal* 55.3 (2012): 785-805.
- cvi James E. Dobson, *Critical Digital Humanities: The Search for a Methodology* (Champaign, IL: University of Illinois Press, 2019). For an illustrative example of an attempt to propound a “model of reading literary texts that”, in the words of its authors, “synthesizes familiar humanistic approaches with computational ones”, see Hoyt Long and Richard Jean So, “Literary Pattern Recognition: Modernism between Close Reading and Machine Learning”, *Critical Inquiry* 42.2 (2016): 235-67,
- cvi Oxford *Essential Quotations*, ed., Susan Ratcliffe, 4 ed. online edition (2016) citing the *Financial Times* of 10 June 2011, reports both American and Chinese sources confirm that in fact Zhou was referring to 1968 Paris student rising. Retrieved from <https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00018657> on 4 September 2020
- cvi Cluett, *Prose Style and Critical Reading*; Burrows, *Computation into Criticism*.
- cix Ramsay, *Reading Machines*; Bednarek, *Langauge and Television Series*; Eve, *Close Reading with Computers*.
- cx Moretti, *Distant Reading*.
- cxii Jockers, *Macroanalysis*; Archer and Jockers, *The Bestseller Code*; Piper, *Enumerations*; Underwood, *Distant Horizons*.

- cxiii Da “The Computational Case against Computational Literary Studies”. For responses, see in particular the three “Critical Responses” published by Leif Weatherby, Ted Underwood, and Nan Da in *Critical Inquiry* 46 (2020): 891-924.
- cxiiii McCarty, *Humanities Computing*.
- cxiv Franco Moretti, *Atlas of the European Novel: 1800-1900* (London and New York: Verso, 1998); *Graphs, Maps, Trees: Abstract Models for a Literary History* (London and New York: Verso, 2005).
- cxv Bode, *A World of Fiction*.
- cxvi Mandell, *Breaking the Book*; Liu *Friending the Past*.
- cxvii See in particular John Burrows, “‘Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship”, *Literary and Linguistic Computing* 17.1 (2002): 267–287; Patrick Juola, “Authorship Attribution”, *Foundations and Trends in Information Retrieval* 1.3 (2006): 233-334; Jan Rybicki and Maciej Eder, “Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?”, *Literary and Linguistic Computing* 26.3 (2011): 315–321.
- cxviii McGann, *Radiant Textuality*; Bode, *World of Fiction*; Eggert, *The Work and the Reader in Literary Studies*.
- cxix Mark Curran, *The French Book Trade in Enlightenment Europe I: Selling Enlightenment* (London and New York: Bloomsbury, 2018); Burrows, *The French Book Trade in Enlightenment Europe II: Enlightenment Best-Sellers*; Bode, *Reading By Numbers: Recalibrating the Literary Field* (London: Anthem, 2012).
- cxx Risam, *New Digital Worlds*.