



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/233834/>

Version: Accepted Version

Article:

Middleton, E. and Taylor, J.E. (2025) Introduction: Where Next for Digital Victorian Studies? *Journal of Victorian Culture*, 30 (4). vcaf015. pp. 289-291. ISSN: 1355-5502

<https://doi.org/10.1093/jvcult/vcaf015>

© 2025 Leeds Trinity University. This is an author produced version of an article published in *Journal of Victorian Culture*. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Introduction: Where Next for Digital Victorian Studies?

Abstract: This introduction to the Digital Forum considers some of the issues facing researchers in the digital space today, including questions of scale raised as much by the creation of small datasets as by the problems and possibilities of Large Language Models. It introduces contributions to the Digital Forum by Mia Ridge, Zoe Alker, Thomas Smits and Roopika Risam, which explore issues around digital preservation in the heritage sector, digital literacy and digital history, Computer Vision for the analysis of images in periodicals, and the role of recovery projects in the Digital Humanities landscape.

Keywords: Digital Forum, digitization, distant reading, digital history, databases, intersectionality, Computer Vision, historical recovery

Since its inception in 2009, the *Journal of Victorian Culture*'s Digital Forum has sought to 'make visible the many ways that digital scholarship underpins scholarly practices in nineteenth-century studies'.¹ In that time, as the Digital Fora have collectively exemplified, digital resources have changed what and how we teach, research, and engage with the Victorian period. In the Digital Forum's lifetime, massive online collections and smaller collection-based digitisation projects have transformed our access to archives; the development of easy-to-use tools, such as Google Ngrams or ArcGIS's StoryMaps, have redefined who can 'do' digital Victorian studies; and the rapid development of machine learning and artificial intelligence promises an even swifter, even more radical, metamorphosis in the scope of what is achievable for the mid-twenty-first-century Victorianist.

Yet, what is perhaps most remarkable is the strange persistence of key questions about the practice, practicalities and ethics of digital approaches to Victorian studies. Are we, for instance, still 'readers' of Victorian texts, or have we become 'users' of Victorian-derived resources?² How do we select what material is digitized – and how do we continue to justify what, or who, is 'lost' in the process?³ Can we still understand a nineteenth-century media ecology when our own – including the ways we use digital technologies to position historic media – is so very different?⁴ And, throughout these issues, a core concern percolates: how do we make sure that we – as academics, students, and heritage professionals – continue to respond rigorously and critically to nineteenth-century sources, when the speed and scale of our methods each encourage forms of distance or – at worst – disengagement?⁵

These questions are perhaps even more urgent now than they have been even over the last decade or so. While it is not revolutionary to talk about scale in digital humanities research, we are on the cusp of a revolution of scale, as the essays in this Forum collectively demonstrate. The field of DH research is characterized by debate about the ethics and approaches demanded by working at scale, and its foundational critics and projects have coalesced around Victorian literature: in defining the term 'distant reading' in his 2000 *Conjectures on World Literature*, Franco Moretti called distance 'a condition of knowledge', arguing that 'If we want to understand the system in its entirety, we must accept losing something'.⁶ Subsequent critics have probed that 'loss' in intersectional terms: notably Lauren F. Klein has argued that 'it's not a *coincidence* that distant reading does not deal well with gender, or with sexuality, or with race', but that distant reading as practised occludes these kinds of intersectionalities. She asserts that we need to reexamine the very creation of our datasets and models to consider what is hidden and otherwise left out.⁷ This need has become all the more pressing as Large Language Models have become prominent in the digital space in all areas of work.

¹ James Mussell, 'Digital Spaces: Introduction', *Journal of Victorian Culture* 18.2 (2013), p.263.

² James Mussell, 'Readers and Users', *Journal of Victorian Culture* 15.1 (2010), p.137.

³ James Mussell, 'The Postcolonial Archive', *Journal of Victorian Culture* 19.3 (2014), p.383; Zoe Alker and Christopher Donaldson, 'Digital Heritage', *Journal of Victorian Culture* 23.2 (2018), p.220.

⁴ James Mussell, 'Nineteenth-Century Newspapers in the Digital Age', *Journal of Victorian Culture* 17.2 (2012), p.221.

⁵ Zoe Alker and Christopher Donaldson, 'Workflow', *Journal of Victorian Culture* 22.2 (2017), p.222; Christopher Donaldson and Zoe Alker, 'Digital Literacy', *Journal of Victorian Culture* 24.3 (2019), pp.329-30.

⁶ Franco Moretti, 'Conjectures on World Literature' *New Left Review* 1 (2000): pp.57-58.

⁷ Lauren F. Klein, 'Distant Reading After Moretti' (2018) <<https://lklein.com/digital-humanities/distant-reading-after-moretti/ing-after-Moretti-Lauren-F.Klein>> [accessed 4 February 2025]; see also Laura C. Mandell, 'Gendering Digital Literary History: What Counts for Digital

The term ‘distant horizons’, from the title of Ted Underwood’s 2019 book, strikes us as a more sympathetic nomenclature, gesturing to the phenomenal scale of data LLMs now work with but also to the future potential offered by this radically transforming world. His focus on bringing the pattern recognition and visualization potential of working with large datasets into careful dialogue with more ‘conventional’ literary and historical methods is reinforced and reflected in the pieces in this issue. This also speaks to a shift towards a ‘critical digital humanities’,⁸ which seeks to draw attention to the biases encoded into tools, underlying collections, and animating our search strings or research questions. As Mia Ridge puts it in her article on large-scale digital collections, the future ‘needs historians’.⁹

Ridge’s article ‘Thinking Big? Towards the Future of Large-Scale Digital Collections’ looks back on the recent history of digitization and digital collections at the British Library. She reflects on the implications of the British Library’s recent major cyber incident, and the ways in which it rendered the costs and complexities of digital infrastructure very visible to the wider public. The piece pulls out lessons from the Living with Machines project, and how digital research can feed back into the collections on which it is based. Her contribution underscores the need to document digitization in order to understand the peculiarities of the data: while she offers ways in which less technical researchers can work with the collections playfully and with nuance, she argues that a kind of digital literacy that builds on and extends the traditional historian’s skillset is necessary to unlock the potential of large collections and handle their peculiarities sensitively.

In ‘Big Data, Digital History and the Development of Digital Literacy in the Era of Artificial Intelligence’, Zoe Alker similarly grapples with concerns about uncritical dependence on digital sources. Now is the time, she argues, to ‘renew existing forms of historical data criticism’ in response to the increasing ubiquity of AI tools such as ChatGPT. Alker describes how she has imbricated data criticism in her research on a series of databases and web-resources, particularly *Data Mining Convict Tattoos* and *Skin and Bone*, online archives developed through combinations of rules-based with critically reflexive machine learning. These are not just source repositories, she argues, but indicators for how we might develop ‘new forms of historical source criticism’ that are vital for critical digital and Victorian studies.

Thomas Smits is also concerned with the critical application of machine learning. ‘Machine(s) Learning about Victorian Visual Culture’ demonstrates how far Computer Vision (CV) technologies have advanced in the last decade, and uses the project ‘Multimodal AI, Image Analysis, and the Illustrated Periodical Press’ as an exemplar to suggest how new types of multimodal machine learning can be applied ‘meaningfully’ to study the millions of images from published (and digitized) Victorian texts. Smits outlines how nineteenth-century images have been uniquely instrumental in advancing CV approaches, and suggests how combining data from texts and images to train new models can transform the ways we work alongside AI in our digital archives. For Smits, like Alker, developing methods for critically and thoughtfully integrating AI into our approaches will be a keystone of digital Victorian Studies in the coming years; he argues that we need to look back to the

Humanities’, in *A New Companion to Digital Humanities* ed. by Susan Schreibman, Ray Siemens, and John Unsworth (Wiley Blackwell, 2016): pp.511–523.

⁸ See, for example, the special issue of *Digital Humanities Quarterly* on ‘critical code studies’ (*Digital Humanities Quarterly* 17.2 [2023]

<<https://www.digitalhumanities.org/dhq/vol/17/2/index.html>>).

⁹ Ted Underwood, *Distant Horizons* (University of Chicago Press, 2019).

Victorian period's 'visual experience' in order to construct the 'visual patterns' through which AI-driven scholarship primarily generates its knowledge.

In 'Race, Data, and Representation', Roopika Risam reflects on how digital humanities can recover stories left out of the (digitized and digital) cultural record, and considers where representation and recovery fit into questions of scale. Drawing on several major projects, including *Reanimate*, which recovers writing by women overlooked in media history, she argues not only that 'our data itself is inextricably linked to history, race, colonialism, gender, and other axes of identity and oppression that confer or withhold relative amounts of privilege for the literatures, histories, and cultures we study', but that the methods, approaches and workflows further this inequality. Risam shows that we need to create more diverse datasets before we can address questions about representation at scale, and that methods need to take into account, for example, the 'particularities of Black data'. *The Pan-Africanism Data Project* aims to address some of these issues, and Risam explores the possibilities and challenges thrown up by the project. The piece offers a hopeful perspective on the future of DH, and how it might advance efforts to improve representation, not only in collections but also in tools and resulting analysis.

The essays in this forum thus explore the exciting possibilities of Victorianist research at scale, while outlining the need to take into account biases, peculiarities of specific datasets, the need to create and modify both bespoke datasets and digital approaches, and the ways in which digital research calls for a humanist skillset. As the articles here attest, Victorianists, with their expertise in critically assessing complex multi-media ecologies, are particularly well-placed to experiment with addressing these challenges.