

## PART IV

# Conclusions



## CHAPTER 18

# The Common Landscape of Digital History

## Universal Methods, Global Borderlands, *Longue-Durée* History, and Critical Thinking about Approaches and Institutions

Jo Guldi

In old-fashioned social history, the study of the ‘common landscape’ used to serve an index of social and cultural difference.<sup>1</sup> Take two regions, two neighbourhoods or two houses, side by side: their differences illuminate cultural ideas about hierarchy, the reality of divergent incomes, and separate relationships with the material world. Just so, the current volume invites us to conduct a survey of the ‘common landscape’ of digital history in all its variation.

‘Field’ though it might be in name, the domain of history as practised by scholars of different methodological and political orientations, geographical and temporal subjects of study, and institutions around the globe is really more of a patchwork of different fields and sub-fields, connected by an infrastructure of main-travelled roads and divergent footpaths that only precariously serve the whole. Some of these fields are closely guarded by an embattled elite, others plowed by an army of workers, still others remote provinces known only to a handful of toilers. Here and there, social historians and

---

### How to cite this book chapter:

Guldi, J. (2020). The common landscape of digital history: Universal methods, global borderlands, *longue-durée* history, and critical thinking about approaches and institutions. In M. Fridlund, M. Oiva, & P. Paju (Eds.), *Digital histories: Emergent approaches within the new digital history* (pp. 327–346). Helsinki: Helsinki University Press. <https://doi.org/10.33134/HUP-5-18>

business historians have already harvested crops for generations. The bumper harvest of the future promised by new technology frightens some with its scale and lack of care, a question addressed by at least two chapters in this volume. In other places, however, tinkers deploy the new algorithms to cottage-sized gardens of their own liking. Whether they garden with medieval heretics or parliamentary discourse, these cottage-gardeners are increasingly dabbling in some new technology—be it topic models, vectors or other tools found in this volume. Their work is just the garden. The landscape itself is changing as a result of these multiple efforts: it is a changing ecosystem, rebalancing in reaction to human labour, sometimes enclosed or guarded.

The landscape metaphor, to be sure, obscures strong forces of historical change. In Chapter 4 above, Mats Fridlund offers an oceanic image for history and its trends. Forces whence we know not where transform the discipline, moving everything with them. For Fridlund, technology is the wave: in the modern era, technology is the reigning factor. Even scholars who eschew quantitative methods depend upon the word processor, JSTOR articles and newspaper indexes that self-proclaimed non-digital historians depend upon. Using the analytical insights of the history of technology, Fridlund deconstructs the digital, and argues that, like nature, technology is always already with us.

Indeed, there is something elemental in the transformation of scholarship in the modern era. New work with algorithms potentially participates in such a tide, and what is thrilling about it is the sense that any scholar, anywhere in the world, might contribute to its movement. Approaches to the study of history developed by one cadre of researchers working on 19th-century Finland may rapidly translate to studies of 19th-century Britain, of the 20th-century United States, or of medieval China.

Yet, the diversity of different specialisations, periods and interests persists, despite the waves and winds that blow through from time to time. As the rich and diverse studies of this volume have demonstrated, digital history is not so much a field or sub-field in this rich and varied landscape, as a universal approach to history. The practices and methods of digital history are transforming plots here and there across the entire landscape: here a neural net, there a topic model, elsewhere a map or a social network, spanning the entire range of periods, geographies, orientations and institutions, such that no field of historical scholarship, however remote, is today too far away from some garden bed where some scholar has applied a computer-aided practice to their labours.

For proof of this intermingling diversity, one has only to look at the rich set of different algorithms and questions in this volume. Each chapter contributes a different orientation and algorithm, such that the volume as a whole surveys a wide variety of different times, places and orientations. Nonetheless, the neural nets used to study medieval heresy have been used elsewhere to analyse the history of photography, and the topic models used here to analyse German humanist discourse have been used elsewhere to study the history of British parliamentary debates about infrastructure and 20th-century American newspaper coverage of cities.

Surveying digital history thus implies looking over the whole of the rich landscape of spaces that constitute traditional history. All of its periods, geographies and intellectual orientations are being reworked according to collective investments, new approaches, and the questions and problematics opened up thereby.

But is a common set of approaches, shared between knowledge-workers, necessarily the same as a hegemonic project of empire? What is the difference between a 'universalising' set of common practices and a 'universal' history, a single narrative that is supposed to provide a final answer for all times and places? Will the oceanic wave of digital history drown out the lovingly cultivated diversity of the cottage farmer?

In this chapter, I will set out to answer these questions, exploring how the methods of the digital historians exemplified in this volume overlap with the universally applicable discourses of critical theory, how far the two are exclusive and where they are now collaborating together to forge a new set of critical approaches to the study of time. I will use the examples in this volume to test and refine earlier assertions about the trajectory of digital history with regard to the *longue durée*, microhistory, critical theory and politics. I will conclude that the imprint of the *longue durée* is clear; the political implications of practising digital history less so. Along the way towards answering those issues, this chapter will remark on several major features of the common landscape of history as it is now evolving.

This chapter will therefore raise questions about the institutional, national and geographic alignments that make 'doing digital history' possible, reflecting further on some of the themes of institutional investment covered in this book. It will raise questions about the future institutional geography of scholarship, drawing on the implications raised by this volume—one of the most methodologically and forward-thinking volumes on digital history at present—coming from a European nation whose sometime borderland status flags important changes in the geography of scholarship under the digital turn.

Finally, this chapter will look ahead to new trends in scholarship visible in the contributions from this volume, notably: a rise of methodological articles that model the 'bridge' between close and distant readings, the rising importance of scholarship that targets the 'fit' between questions and algorithms, and the increasing importance of international institutional collaboration. As the chapters in this volume demonstrate, digital history is well on its way to establishing theories and methodologies that satisfy these most critical criteria of scholarly investigation; in the years that follow, scholars who pay attention to these themes will have even more to look forward to.

### The Universalism or Common Space of the Digital Humanities

The discussion of methods herein, one might say, offers a truly 'universal' tendency for scholarly exchange, and it is worthwhile pausing to understand what we mean by that. In another volume, the reader who specialises in medieval

history might flip only to the chapter written by the medievalist. But it is a truism that in digital history, and the digital humanities more generally, to pass over the other sections would be an error. The methods developed for one sub-field will be relevant to the next sub-field tomorrow if they aren't already today.

We see in these critical reflections the shape of new standards for historical work as scholars puzzle over the fit between particular algorithms and questions. In Chapter 16, Välimäki and his co-authors (Reima Välimäki, Aleksi Vesanto, Anni Hella, Adam Poznański and Filip Ginter) used neural nets to confirm the authorship of the *Refutatio Errorum*, an anti-heretical treatise from Germany in the 1390s, establishing, in the process, a new standard for author detection. In Chapter 15, Heidi Hakkarainen and Zuhair Iftikhar prove that topic models are well-suited to engaging Koselleck's idea of concept history, linking concepts with temporality. The chapter describes a vast culture of experimentation and discovery, as scholars try out competing algorithms, testing the fit of each method to the scholarly questions independently identified as problematic in each field.

To dub such a capacity for common meeting 'universalism' is to underscore that anyone might play with any of the ideas at stake, even while an enormous pluralism existed of period, subject and political orientation. It would certainly not imply the universal applicability of any one fact or conclusion reached by the research, which is subject, as all historical research is, to the revisions of new discoveries, new archives and new approaches. Perhaps an even better term than 'universalism' for what we are investigating would be the 'common space' of the city that Hannah Arendt identified as a metaphor for the best strivings of both Enlightenment and democracy: that they could be accessed by anyone, that they enlivened and illuminated all lives that touched them.<sup>2</sup>

By employing the image of 'common space' in the city as a metaphor for a certain aspiration in discursive activity, Arendt underscored the question of access: since the European Enlightenment, the modern city was defined by spaces of equal access, spaces that didn't shut people out. We might say the same about digital history: whatever the critique from the outside, the practitioners of digital history have taken pains not to shut out any practitioners whomsoever, and many of them have worked at length to convert digital tools into material for critically inspecting empire, gender and race.<sup>3</sup>

What about the contention that digital history is itself imperial and *universalising* in nature, threatening to draw all history practitioners into a single method, problem of study and macrohistorical overview of the *longue durée*? A notorious example of universalising claims by biologist interlopers relates to a decade ago, and the haunting claim, made in the national media of the United States, that historians would be obviated by the coming of the computer.<sup>4</sup> As digital history is actually practised, we see very little of this. More relevant is a portrait of individual scholars or scholars in small units working together to execute some new perspectival opening onto their sub-field—medieval heresy, the Finnish Parliament, the Finnish borderlands or European humanists.

Not one chapter in this volume gestures towards a universalising macrohistory of the *longue durée* that would eliminate the perspective of workers or feminists beneath the triumph of the nation or empire. Not one chapter in the whole book even claims to dispense with or obviate the field of social history, biography, intellectual, cultural or political history, although most chapters build on aspects of other forms of history in some respect. To paint the common landscape of scholarship, the artist would truly have to recognise a thousand *digital histories*, not one digital history, stretching along the plain, informed by exchanges, building on works already in progress across the land.

Another version of the complaint against digital history as *universalising* and therefore coercive borrows an image from Lawrence Stone, who in the 1970s warned against quantitative enterprises of all kinds as commandeering graduate students into massive, pyramidal projects where their intellectual inquiry is dictated from above and individual initiative is squashed.<sup>5</sup> This, too, seems not to be the case. In the quarters of digital humanities where graduate students are enlisted, they are often featured as first-authors on projects that they were invited to craft with the skill and support of leaders in the field.<sup>6</sup>

Graduate students and early-career faculty in this volume, including Reetta Sippola and Matti La Mela, are particularly distinguished as early adopters and explorers of new technology, who have been willing to extend their training in some sub-fields to adopt new technologies, test a theory or explore a time period. The borrowing—from computer science, statistics, the digital humanities, linguistics or library science—means that these early adopters are fast at work in building up the discursive commons that Arendt so praised.

Whatever we call it, exchange between sub-fields has quite a bit going for it. Discursive commons, technological borrowing and other such scholarly programmes of exchange represent an instinct for common space or the universal that, according to science, has been waning of late in the academy. This tendency to narrow the study of the universal back to readings in one's sub-field, we learn, is not merely a habit of the humanities, but can be quantitatively identified in the social sciences and sciences as well. According to the research of sociologists of knowledge such as James Evans, scholars today generally cite fewer readings directly outside their realm of concern than did scholars a generation ago. It seems to be the case that internet-enabled web catalogues have restricted universal reading habits of borrowing from nearby disciplines, whether for critical theory or for other inspiration. The sheer overwhelming scale of available knowledge left scholars paralysed by the task of keeping up with their nearest cohort.

History seems to indicate that information economies more generally are marked by a pulse of broadening the process of collecting information and refining the information thereby collected. This pulse of information analysis has been studied across decades-long exchanges about early-modern botanical knowledge, as well as on the personal scale of Darwin's notebooks.<sup>7</sup> What we are calling 'universal' moments or moments of 'common spaces' seem to be

moments of expansion, when scholars looked to particular discussions as relevant for scholars who worked across a broad variety of periods and places.

Today, the digital humanities partake of a similar moment of universalism, in which scholars of 1980s online culture, 19th-century novels and Chinese medical texts regularly meet and compare notes, finding algorithms to borrow from one another. As a result, a discussion of methods offers a meeting ground on a broad scale, as well as an opportunity to compare notes across different sub-fields and disciplinary orientations.

There have been other moments of expansion on a theoretical level, where humanities disciplines are reforged through insights from without. Such, for instance, was the impact of the steady importation of continental philosophy and critical theory into the humanities since the 1970s. A Freudian reading of Augustine could be interesting to those contemplating the nature of the biographical subject during the American Revolution precisely because the *method* could be so easily transported and applied to other fields and subjects. By the same token, a Foucauldian reading of colonial India might, in theory, interest readers from the social history of industrialisation.

Indeed, it is possible that methodological moments of unification offer a necessary antidote to the paradigm of modern specialisation of knowledge, with its tendencies to mince fields into sub-fields and further sub-fields, with the concomitant risk of knowing more and more about less and less. The arrival of critical theory in the 1970s meant that the scholar of Virginia Woolf and of Classical Athens could find both a common meeting ground and a common language in terms of inclusiveness, femininity, the knowledge of the state and the construction of the individual.

At the same time, however, the digital humanities are beginning to see a moment of critical inward inspection, of the refinement of processes and pipelines. The 'universal' impulse in the digital humanities is thus giving way to another phase of information analysis—one predicated upon the close inspection and comparison of algorithms, the attention to metadata, and the examination of OCR errors and named-entity detection. As Kimmo Elo points out, this domain of attention is a necessary part of the process of refinement if the garden of earthly methodologies is to bear fruit. The labour of refinement and inspection will almost certainly be a domain of work that requires the labour of historians.

If building interoperable tools and applying them to great questions of history represents the universal access of the city square—liberating with its sense of wide access and possible exchange—the work of refining metadata and inspecting tools is more like the garden plot of digital history, a place that requires focused attention and hard work to produce useful results. The metaphor is complete if we imagine that the garden of tool- and data-refinement produces useful results that can be taken back to the universal exchange of the city square. As Johan Jarlbrink points out, inspecting the results of metadata analysis through simple methods such as the 'tally' stands to help us unpack the 'black box' of digital learning.



A volume of the present kind offers a rare glimpse into how the entire breadth of history is shaped—in its temporal range, from studies in the history of Finnish feminism to medieval heretics to labour politics, and in its methodological range, from the prosecution of new frontiers with existing tools like topic modelling to the close attention to algorithms, metadata and tallies of OCR errors. The scholar who represents each period, place, theme and method has a separate body of texts and different methods, to be sure; indeed, the scholarship in question was in part selected so as to adequately represent the diversity of possible methods, approaches and statistical rigour practised across history as a whole.

In offering a meeting ground of methods and periods, a methodological volume offers an important service to the discipline as a whole. The scholar of medieval heresy may wind up later borrowing the tools from the scholar of humanism or labour politics, even if their data for the moment looks entirely different. Digital practices tend to lend themselves across formats, political interests, historiographical orientations, periods or geographies. Thus, digital tools draw historians back to a certain methodological universalism, even in the face of other kinds of plurality, insofar as they encourage practices of reading-across-boundaries, in the form of conferences or volumes, like this one, that reward the practice of rich learning in new directions.

### **The Pedagogical Role of Discourse**

In order to enjoy knowledge as a commons, one mandate is that the experts of the commons must be motivated with an eagerness to explain, to render accessible the more difficult concepts that they have assembled for the use of others. In the era of critical theory, difficult writers from Marx to Heidegger got interpreters who translated their concepts into ready-to-wear essays: 'Benjamin for Historians', the beginner's guide.<sup>8</sup> Digital humanities can only claim to be a 'commons' accessible to all sub-fields of history insofar as it too has been equipped with multiple translation projects, rendering difficult statistics and algorithms within the grasp of the total novice.

The present volume is a monument to the pedagogical impulse of digital history. The writers assembled here have taken pains to draw down the abstractions of algorithms, statistics and databases into language of period, inquiry and method familiar (or at least tractable) for traditional historians. Each chapter introduces its method, algorithm and dataset in careful detail, presuming little prior acquaintance with historical method. Writers test and explore the possibility of 'false positives' raised by misinterpreting topic models. They ask about the bridge between macroscopic 'overviews' of the material and 'close reading', and how an overview can or should guide the reader back to individual texts or episodes. They seek to open up the 'black box' of digital analysis, to unpack and critique its workings, and to thus devise a new machine for critical analysis of the past. The result is a series of essays that

are pedagogically precise: an instructive model about how to describe a new process for the use of other scholars. These are the tactics that historians of the future should recognise as a new standard for how openly, precisely and clearly we write—for what a truly ‘universal’ or ‘common’ practice of historical reasoning looks like.

The authors of the individual chapters of this volume have taken the editors’ direction for clarity, instruction and critical thinking to heart. In Chapter 9 on feminist history in Finland, the question under investigation is the correct approach to take, and Heidi Kurvinen playfully offers her own experiences of success and failure as an example for other scholars venturing onto this unpredictable frontier of knowledge, where the rules of adequate preparation are not entirely spelled out as yet. In Matti La Mela’s Chapter 11 on the freedom to roam in Finnish parliamentary debates, he walks the reader through a step-by-step recreation of the process, suitable to educating historians with no prior use of the algorithms or techniques at stake. In Chapter 16 by Välimäki and colleagues, the authors critically explore the use of each part of their data (the examples of anti-heretical texts with known authors used to train the neural net, the authors proposed by other scholars) and each algorithm (the Support Vector Machines used to detect authorship, the vector machine used to clean the data). Computational author detection, they urge, should become the new currency of the discipline, amplifying other criteria of authorship detection such as structure, manuscript tradition and argumentation.

The writers in this volume have also shouldered the burden of offering their own interpretation of the new approaches necessary to digital history. From the ‘tensor history’ of Timo Honkela’s foreword to the ‘resource criticism’ of Mats Fridlund’s Chapter 4, their inquiries point to the importance, as history adopts new methodologies, of adequately spelling out the limits of an inquiry, the sources of the data, the silences and limits of each inquiry and the possibilities opened by particular algorithms.

I read the theoretical trajectory of this volume as a powerful demand that each historical encounter explicitly renders obvious its limits, both in terms of sources and in terms of methods. And if that seems like something that historians have already done in carefully describing their paper archives and theoretical baggage, consider this: what if every historical study that leans, in part or in whole, on secondary sources, newspapers, parliamentary debates or other digitised corpora considered not only the limits of the micro-archive, but also the limits of the macro-archive? We would have to choose forms of analysis that allow us to form an overview of the archive as a whole, for the purposes of both *longue-durée* analysis and acknowledging the historical situatedness and inherent bias of each archive. In essence, we are being invited into a new age of historical criticism, one that brings the ‘capital/periphery’ critique of post-colonial studies home in the sense of recognising the limits of data and critique, acknowledging the way in which our source-base and view of history has been shaped by power and limited all along.

## The Universal Borderlands of the Digital Humanities

One token of the universalism of the digital humanities is that the contributions of a group of historians stationed mainly in Finland (rather than Britain or the United States) could dare to claim for itself so vast a title as *Digital Histories*, as if aiming to define the new field. Nor is the volume bound by an exclusive orientation to Finnish history: the subjects of this volume range across Europe. Perhaps because of the 'universal' tendency of digital humanities research outlined above, digital humanities research in Finland is, indeed, as diverse as the historical discipline in Finland. The present volume represents an earnest attempt to define the historiographical range of questions presented by historians working with digital techniques, to cover the range of algorithmic and metadata practices used by our colleagues, and to introduce new scholars to best practices and techniques that merit attention from scholars of any time period or geography.

The digital humanities provide an arena of study that is 'universal' in that it incorporates broad engagement with international currents in intellectual and cultural history, feminist history and the history of science. The topics herein contained range from medieval studies of authorship among Waldensian heretical texts to the popular reception of scientific astronomy in the 18th century. Far from being rigid and inflexible, the chapters here show off an astonishing variety of methods, each the mirror of a historical problem with its own historiographical legacy stretching back over decades. The research projects in question demonstrate something of the expansiveness of interest and time period that could be found within most national traditions.

International sharing of methods and data has been intense over the last decade, among historical practitioners, and this international intensity has raised a number of new international capitals for digital humanities research—among them digital humanities centres and nodes of excellence in Umeå, Uppsala, Venice, the Max Planck Institute Berlin, the Dutch national research infrastructure CLARIAH, the University of Sussex, the Language Bank of Finland, the University of Helsinki and the University of Turku. On a global level, digital humanities researchers in Singapore, Taiwan, China and Latin America are producing significant demonstrations of new methods. The map of digital history practitioners on the avant-garde of methodology is both more international and marked by the presence of younger universities and research institutes than a more traditional map of excellence in the humanities. With all respect to the digital humanities summer institutes that introduce many scholars to the techniques of DH, this is simply not a field of which one gains mastery by a single visit to a great master at Oxford, Cambridge or Harvard: the field is moving too quickly, with nodes of excellence developing in seemingly improbable places.

The conditions for international participation in digital history are set by trajectories that have something to do with the existence of national traditions

of history in all of those places, as well as the rise of an international information economy over the lifetime of the authors of the present volume. As Paju's Chapter 2 explains, historical researchers in Finland have experimented with digital techniques since the 1960s, much as American researchers trace their roots back to experimentations with punch cards and Latin codices in the 1950s. Like universities around the world, Finland has a national tradition of historical research with centres of excellence of teaching and learning. Finnish scholars, building on generations of institutional development, have had the opportunity to theorise important questions about statistical measures and AI, close reading and distant reading, and the role of learned societies in building and maintaining the intellectual infrastructures of today.

As such, the volume represents an event horizon within the global practice of historical scholarship which marks the rise of institutional 'borderlands' that are less well-established than the Ivy League or the ancient universities of Western Europe. Finland's distinctiveness within the digital humanities thus offers a paradigmatic path for other national bodies of researchers who wish to vie for distinction on the frontiers of interdisciplinary knowledge-making. Like many other economies in the developed world, and many in the developing world as well, Finland is heir to the information economy with all of its perquisites. Like many privileged departments in Europe, Canada, North America and Australia, Finland's history departments, libraries and language banks have benefited from an aggressive institutional programme of digitisation and support. These three preconditions—a tradition of historical study, participation in the international information economy and institutional development funded on a national or international level—make it possible for scholars and universities to mark themselves out for distinction within the space of digital history research.

The 'universal' power of the digital humanities has thus established an arena where newer institutions and national traditions of historical research can play, on equal terms, with the oldest and most distinguished universities in the world. The present volume demonstrates how scholars from a European borderland have harnessed this power to demonstrate their engagement with new methods, tools and critiques.

### Digital Directions: The *Longue Durée*, Identity, etc.

Critical readers will want to know whether the digital histories of this volume are closing out or displacing other kinds of inquiry. By implication, they wonder whether departments that choose to invest in digital research are necessarily thereby foreclosing on other kinds of research strategy. The rationale behind fears such as these are located in the real, historical experience of intellectual 'turns' in the academy: one meeting ground sometimes displaces another, and this was true of critical theory. Topics of study inherited by the 20th-century humanities from the 19th century and classical precedents included the

existence of 'genres' in fiction, the search for ideal character in the genre of biography, the ideal of succeeding generations of 'reform' in the study of political regimes and the history of progress through a cascading series of perfections (whether in the form of intellectual history, the history of science or the history of technology).

As scholars came to understand the past with the aid of critical theory, old categories of research (genre, character, reform and progress) each held within them a doctrine that was itself a historical construct. What critical theory did for the liberal conscience of the university was to create a series of substitutions where a former doctrine was broken open in favour of a series of new research questions. Many of those questions were informed by politics (for example, jettisoning the doctrine of empire's beneficence in favour of a series of critical questions). Rather than taking the 19th-century agenda of matters for historical investigation as a given, it was possible to subject each of them to deconstruction and historical analysis. In the process, a new set of research questions emerged: instead of character, there was agency, and with it the question of under what conditions it became possible for an individual or a certain group to exert change over their fate or the course of collective experience. Instead of technology as the progress of inventions from one generation to the next, the history of science and technology was reborn as a series of questions about the ideology of science and technology; their affiliation with empire, masculinity and capital; the institutions that support them; and the illusion of forward progress. Critical theory buried the naïve liberal scholarship that came before it, replacing it with a series of new research questions.

Just as critical theory pushed out the set of uncritical liberal targets of research that came before it, so, it might be expected, will the new goals of digital history displace some of the focus of the scholarly record before them. It is too soon to tell what the subjects of replacement will be; the relationship between digital history and earlier generations of history is still in formation. Moreover, the number of digital history papers that directly counters an extant historical theory is very small, in comparison with those in digital literature, where scholars such as Ted Underwood and Andrew Piper have explicitly taken on some of the mainstream conclusions of the field and shown how digitally produced knowledge overturns received wisdom about, say, the idiosyncrasy of Flaubert.

While digital history remains, as yet, immature, the field as a whole is guided by theories of when, how and whether digital history will call for a revision of lasting tropes in the discipline. In *The history manifesto*, my co-author and I pushed the strongest possible case for a revolution in critical thinking abetted by access to digital tools, and we sought to describe what that might look like: in brief, we conjectured that *longue-durée* timescales of 100 years or more would displace microhistory on the scale of the human life or shorter. We advanced some related claims about the political crises of the present and the new *longue-durée* inquiries that they might provoke (shifting attention to climate change

and economic inequality over identity politics). How have those prophesies held up when it comes to the writing of ‘digital history’ five years later?

One of our guesses was the renewed importance of *longue-durée* perspectives, given the fact that digital research made trivial the repetition of cultural-studies-type analytics on long time scales using scales of material that would be impossible for the traditional reader. Just such an approach is represented, in the current volume, by several uses of newspaper corpora. In Chapter 15, Heidi Hakkarainen and Zuhair Iftikhar prove that, over 21 years, the language of Austrian newspapers demonstrates the growing importance of historicist thought. They chart the succession of disciplines through which ideas of humanism spread, a modelling of ideas like a contagion, a transition from education to the reformation to revolution and later the death penalty. They reveal how newspapers gradually connected the human, the spiritual, democracy and the future, a new language evidenced by the increasing importance of the terms *Zukunft* (future) and *Zeit* (time) to the definition of humanism. In Chapter 11, Matti La Mela uses the Finnish parliamentary debates to investigate the freedom to roam over a century. On an even longer time span, in Chapter 12, Pasi Ihalainen (with the help of Aleksi Sahala) identifies several kinds of discourse about internationalism since the founding of the League of Nations, including internationalism typically opposed to (but sometimes arising from) *nationalism*; variants of *party* and *labour* internationalism; the *spirit* of nationalism linked to ideas and values; and the promise of *peace* and *democracy* arising from internationalism, especially after 1930. A revised chronology of internationalism shows a peak in the 1970s, followed by a period of frustration.

New findings that draw from long timescales can border on the breathtaking. Perhaps the most surprising finding in Ihalainen and Sahala’s investigation of long-term discourses of internationalism was the longevity of ideas. Ihalainen found that down to the ‘leave’ side of Brexit, many of the ways of describing the promise and threat of internationalism remained unchanged since discussions of the League of Nations in the interwar period, on both the left and the right.

Much of the work that was once carried out by the ‘close reading’ of the cultural turn in the 1980s and 1990s can now be achieved with greater precision and efficacy by algorithms designed to discern, and to measure, similarities and differences of expression and sentiment, allowing the tight comparison of decades, institutions, political parties and individuals. The work in this volume aptly demonstrates that the work of the cultural and linguistic turns—concerned with the shift of lexicons and the insight this provides about historical identities and communities—are now, on their cutting edge, digital in method.

There are real challenges facing the prosecution of the *longue durée* as well. In Chapter 3, Jari Eloranta, Pasi Nevalainen and Jari Ojala point to the overwhelming scale of the archives that still await digitisation in their domain, that of the modern business history of Finland. They cite the archives of Finland’s 20th-century government administration since the 1970s, now held in the National Archives: ‘roughly 200 shelf-kilometres’ of documents essential to

understanding the history of the welfare state and neoliberalism. These archival materials would have to be scanned first before they were analysed by digital means, and they are not scheduled for digitisation. A more practical approach, for the moment, is the kind of sampling recommended by Claire Lemerrier and Claire Zalc in their *Quantitative methods in the humanities*.<sup>9</sup> The facts of this dynamic raise challenges to archival institutions, grant-writing bodies in the humanities and the institutions of democracy themselves. A survey *could* be conducted by sampling documents out of that 200 kilometres of shelves, but what would be left out? Surely citizens of each region in Finland deserve the tools to monitor the history of how changes in government organisation affected their own landscape. Serving citizens requires an infrastructure and the building of tools. If such an infrastructure needs to be built for other purposes, and the archives need to be digitised, then historians can rise to the challenge of asking questions about the *longue durée* of that 200 kilometres of shelves and its analysis. Indeed, it is possible, as Ted Underwood has lately argued, that trends otherwise invisible on the short *durée* would emerge from such an analysis.<sup>10</sup>

Some of the guesses hazarded by *The history manifesto* went against the actual course of scholarship, in particular the continuing importance of research into gender, race and class. Prognostications that identity politics would be displaced by a larger, shared concern over economic inequality proved short-sighted in the face of the rise of right-wing movements around the world and the renewed relevance of identity-based activism. To counter those movements, scholars, journalists, lawyers, artists and ordinary citizens have returned to the *longue durée* of injustice in a powerful way; for instance, through the demands for reparations for slavery in Britain and America, or in the public controversy over monuments to confederate generals across the American South. Indeed, until racism, sexism and nationalism are abolished, historians are bound to ask questions about where they came from.

Thus, what we see in the new practice of digital history is not so much the displacement of critical theory by digital history, as the integration of the questions posed by critical theory on longer time spans, addressed with methods that allow the historian to fully integrate the methods of the cultural, social and linguistic turns. The natural outgrowth of these dynamics is a kind of digital history that fixes on identity and empire as its subject, exemplified by the many projects gathered and reviewed in Roopika Risam's *Postcolonial digital humanities* (2019). The present volume gathers studies in the history of Karelian borderlands and Finnish feminism. It is unsurprising that any cultural or political event could be traced at scale and in depth by digital means.

### Implications and Future Directions

What are the implications of such a study for other scholars, if not to mark out a definitive set of algorithms or revisions for others? I would argue that



the output of such a volume as this is significant in that it sketches out the computational best practices of a moment. It also speaks frankly to some of the challenges ahead. It is to the work of coming challenges that I will write for the remainder of this conclusion.

*The need for theorising the bridge between distant and close readings*

The possibility of informing close reading with the power of synthetic tools is one of the major promises of digital history. Of his century-long examination of collocation in parliamentary discourse, Ihalainen explains, 'distant reading reveals peculiar political points that might have gone unnoticed in mere close reading or full-text keyword searches of the same documents'.

While a few of the chapters in this book begin with distant reading and end by examining particular documents, there are few examples today of historical or literary practice that moves from the distant overview down to the level of authors or other categories and into particular passages in the book, critically examining the results of a search based on the close reading of the page. Like much of the work in the contemporary digital humanities, results of distant reading are frequently given in a summary chart or single finding. Two notable exceptions offer a meditation on the 'bridge' between the distant and the close, and provide historians with a way forward. In a rich meditation on the history of close reading among historians of women, in Chapter 9, Heidi Kurvinen draws a contrast between the tools she studied in her training and the topic modelling she applied to a study of Finnish suffragettes; in the process, she offers her reflections on using the topic model as an index of different episodes, and compares some of the findings of close and distant reckoning. Similarly, in Chapter 7, Johan Jarlbrink descends from a project that surveys the effects of media on cities to critically examine the sample of cities recognised by the computer. Bridging close and distant reading becomes, for Jarlbrink, an opportunity to recognise problems in the measurements supplied by algorithmic tools on 'dirty' data. As we can see from the two examples above, the practice of distant and close reading is evolving, and new hybrids are being forged that unlock insights in the archives and highlight shortcomings in the technology.

Future approaches to the bridge between close and distant reading may do well to follow the pattern set by Andrew Piper in his recent survey of modern literature, *Enumerations*, which proceeds from distant readings of the themes and trends across poetry and the novel, down to particular authors, poems and passages, as guided by the tools of distant reading.<sup>11</sup> After all, the same tools that draw our attention to words can be used to compare individual speakers as well as parties, and indeed to draw attention to the particular paragraphs and sentences that the computer discerns to be the most exemplary cases of a particular concept and collocation pattern. That is, where a scholar learns that an important collocate of 'internationalism' is 'nation', it would be useful to learn



next which individual speaker in Parliament pairs the two words together most frequently, and the individual speech in which those concepts are juxtaposed the most. Additional measures such as these may lend confidence that particular sentences given as examples of collocation are not merely cherry-picked for their familiarity, but actually offered foundational instances of the making or re-use of the concept.

More to the point, as historians engage in moving from the small example to the big question, and from the big overview back to individual speech acts, the process of movement itself is open to methodological argument, questions of interpretation and over-interpretation. It is important that historians begin to describe their choice of exemplary passages. Thick description of the process of extraction affords an opportunity to articulate the work of human interpretation and machine contextualisation. As we examine this frontier, we will begin to understand better the application of distant reading to work on different scales, including the scale of the corpus, the author, the work, the paragraph and the word.

### *Theorising the difference between AI and statistical measures*

In the work here, the terms ‘mutual information’ and ‘neural nets’ appear on the same page. Their basis could not be more different, however. The former is statistical and can be described, mathematically, at every step; the latter has been developed mainly from computer-science departments aiming to mirror human processes, and essentially represents a black box of pattern recognition. Some scholars in the computer sciences herald a day where autonomous intelligence will obviate human supervision in most domains, including education. Colleagues in other parts of the university are more skeptical, arguing that AI, in most cases, relies on the labour-intensive hand-tooling of research questions to algorithms. More precise and transparent answers, they suggest, are to be had from old-fashioned statistics.

Humanists are far from the centre of these debates, but our testing of algorithms and our successes and failures have implications beyond our own discipline. As Eloranta and his colleagues point out, business and economic historians have a long history of critical engagement with statistical measures such as regression and event analysis that could easily contribute to a rigorous comparison between statistical measures and unsupervised machine learning. Elsewhere, in Chapter 16, Välimäki and his colleagues rely on neural networks (the epitome of unsupervised, black-box AI), while another advances a preference for mutual information: the epitome of advanced statistics, where equally useful clusters are formed on the basis of a relatively transparent clustering formula. Because we know our textual corpora and their historical context so well, as the scholars’ research on authorship illustrates, historians are often in a better position to ‘train’ and ‘test’ the scripts of AI and to comment on how

well they work. These findings could usefully be published in the *Papers of the National Academy of Sciences* or *Science*, to the edification of other disciplines and the credit of our own.

To the degree to which historical methods border on questions of the success of AI or a preference for transparent statistics, historians have an opportunity to theorise about the stakes of one choice over another. As the discipline of digital history becomes more advanced, we should expect more work of this kind.

*Transparent documentation of the choice of algorithm, text and result  
in the practice of critical search*

Elsewhere, I have argued that scholarly engagement, both traditional and digital, in general tends towards the critical examination of the choices that inform a research project.<sup>12</sup> Whether the choices made in an archival visit are informed by the reading of critical theory, or whether political interest in the *longue durée* drives a scholar towards particular algorithms, critical thinking about the motives and limits of particular kinds of research is always being called upon to inform the constraints of the research process. I generalised the digital research process into a pattern I called ‘critical search’, and I characterised major opportunities for critical thinking inherent in any research process. Instances of critical search in an article might include discussing the choice of keywords and algorithms and how initial choices reveal and conceal aspects of a corpus later disclosed by adjustments to the initial search. The point, in any case, is that these are forms of critical reflection that scholars are prone to in general, and through these critical reflections, the entire community of readers comes to consensus about the uses of particular algorithms, the multiple dimensions of digital archives and the interpretative questions that govern digital research.

The chapters in this volume offer freely evolved examples of critical search, in the sense that they reflect upon the process of creating knowledge with digital tools. In their search for how 19th-century Austrian newspapers described the emergent principle of humanism, Hakkarainen and Iftikhar describe multiple iterations of different kinds of topic models, configurations, with or without stop words, and how they ultimately decided on a combination of topic models with a corpus little prepared for analysis. The process of curation, decision-making and interpretation is shown to be at the heart of scholarly digital work.

In Chapter 12, Ihalaainen and Sahala explain their persuasive use of information theory’s concept of ‘mutual information’ to examine lexical change over time. They use pointwise mutual information (PMI) to identify the most regular combinations of words used to describe the international, the global and the transnational in British parliamentary debates in the 20th century. Their method begins with computational work, but traces back digital findings to the text, excerpting compelling passages illustrating the rising tide of interwar optimism about international relations.

We should expect more critical engagement with the search process in the future. As a community, we are learning how to better highlight the distance between *interpretive* work and *computational* work in each research process. In making the choices behind an algorithmic deployment transparent, digital scholars acknowledge that an algorithm isn't a toaster oven, into which a neophyte puts texts in order to achieve an automatic result. Rather, the process of curation, critical inquiry, secondary reading and interpretation remain at the heart of scholarly inquiry.

*Engagement with new standards of scholarship from the institutions of historical and cultural knowledge-making*

Several of the chapters in this volume implicitly call for a deeper level of participation by historians and their national and international societies in following standards of data preservation, sharing and transparency. In Chapter 10, Maiju Kannisto and Pekka Kauppinen offer a detailed description of copyright issues that had to be overcome for data analysis and sharing in the domain of contemporary media analysis. In Chapter 5, Jessica Parland-von Essen argues that the institutions of cultural analysis should take care to preserve, describe and make accessible multiple layers of data, including its mark-up, analyses, tools, descriptive metadata, consent, rights and attributions of labour. In the process, she describes a potential mountain of cultural labour to be executed by the libraries, archives and IT centres of the world.

Parland-von Essen's chapter suggests a precise charge to the meetings of national historical associations and other learned societies. All of our meetings should have not merely panels for presenting new work in the digital humanities, but also panels for discussing the standards of data presentation, annotation and interoperability.

Cultural institutions (for instance, the Swedish and Finnish literary societies) have a particular role to play in setting out standards for data that is transparent and accessible. Were they to engage the questions raised in this book, in meetings, pamphlets, conferences and hiring, they would have the opportunity to shape how the caretakers of data document the many kinds of labour that have shaped the collection, as well as how practising scholars indicate that they have used data with origins elsewhere. National and international historical meetings offer an important opportunity for inviting cultural institutions and providers of data, from our museums to the private Elseviers of the world, to cooperate with scholars in following these mandates.

On the one hand, directives from national historical associations and conferences are profoundly needed. Even high-profile infrastructure initiatives such as Europeana do not currently provide workflows suited to historians' needs. On the other hand, groups of historians concerned with these issues have already assembled over generations. Petri Paju's Chapter 2 discovers a long-standing tradition for digital-historical institutions and collaborations: for example, the

Helsinki Corpus of English Texts with its origins in the 1980s, the Association for History and Computing, and the Electronic Center for History Research (later Agricola), launched in 1996 by scholars, librarians and archivists; and later the Helsinki Centre for Digital Humanities, or HELDIG, was established at the University of Helsinki in 2016. If these institutions have historically had little influence over mainstream practice, then a new generation of national historical associations and other learned societies could usefully mine the affiliates list of the digital centres for faculties already invested in the important issues of data preservation, documentation and accessibility.

### *International collaborations and tool sharing*

The beauty of computational research is its interoperability: once a technique has been discovered for author attribution in medieval Latin, it should work nearly as well on any early Latin texts whatsoever. The same is true for parliamentary discourse: the collocates machine described by Ihalaenen and Sahala should apply to any modern body of text, and the particular tactics for engaging party-political differences of lexicon should be immediately applicable to digitised records of the French *débats*, the debates of the European Union, the Canadian and Australian Hansards and the debates of the City Council of New York City, to name a few active projects. Despite these opportunities, however, there are relatively few examples of international collaborations that take advantage of the astonishing interoperability of algorithms by drawing together scholars working on similar genres of texts.

One initiative that successfully crosses these boundaries is the Oceanic Exchanges project,<sup>13</sup> to which six nations (including Finland, represented by Hannu Salmi at Turun Yliopisto) contribute their historical newspapers and their technical expertise. In the case of newspapers, text-recognition technology applied in one nation can rapidly be adopted to newspapers elsewhere, so the international collaboration represents a massive virtuous cycle of exchanges. All of these efforts contribute to making the international infrastructure of future research, such that one day soon, we should expect that high-school students of history in Finland, America and Mexico will be able to keyword search the newspapers of their national traditions and compare debates about democracy and markets across nations over 200 years.

It is harder to explain why there is not such an international collaboration for the novel, for debates of democratic bodies, for the records of courts of law, for religious texts, for plays, for musical notation, for stylometric authorship attribution and for other genre-specific questions where scholars have similar questions related to form: comparing authors and chapters of the novel, for instance; comparing speakers, parties and constituencies in democratic debate; comparing kinds of charges, prosecution and defence; or judges, juries and defendants in the courts of law. In each of these genres, an interna-

tional pooling of technical expertise could result in the rapid creation of new knowledge about best methods for turning raw visual images into readable text, for turning text into data annotated with appropriate metadata categories and for deriving meaning over time. The only bar against such collaborations is one of organisation, effort and collegiality: creating useful partnerships depends on planning for international visits, sharing plans of work and the willingness to engage in a process of mutual discernment about where grants and research projects overlap.

The benefits of such collaborations are, of course, tremendous: they can result in the pooling of common goals and strategies, the most efficient use of grant money, the joint discovery of new methodologies and even the joint funding of new infrastructure that makes clean, accessible data available to all.

## Notes

- <sup>1</sup> Stilgoe 1982.
- <sup>2</sup> Arendt 1973 [1958].
- <sup>3</sup> Risam 2019.
- <sup>4</sup> Michel et al. 2011.
- <sup>5</sup> Stone 1979.
- <sup>6</sup> Klingenstein, Hitchcock & DeDeo 2014; Barron et al. 2018; Kraicer & Piper 2019.
- <sup>7</sup> Blair 2010; Murdock, Allen & DeDeo 2017.
- <sup>8</sup> Weeks 1982; Gay 1986; Schwartz 2001; Brown 2013.
- <sup>9</sup> Lemercier & Zalc 2019.
- <sup>10</sup> Underwood 2019.
- <sup>11</sup> Piper 2019.
- <sup>12</sup> Guldi 2018.
- <sup>13</sup> The website of the project is <https://oceanicexchanges.org/>.

## References

- Arendt, H.** (1973 [1958]). *The human condition*. Chicago, IL: University of Chicago Press.
- Barron, A. T. J., Huang, J., Spang, R. L., & DeDeo, S.** (2018). Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, 115(18), 4607–4612. DOI: <https://doi.org/10.1073/pnas.1717729115>
- Blair, A.** (2010). *Too much to know: managing scholarly information before the modern age*. New Haven, CT: Yale University Press.
- Brown, C. G.** (2013). *Postmodernism for historians*. London: Routledge. DOI: <https://doi.org/10.4324/9781315836102>

- Gay, P.** (1986). *Freud for historians*. Oxford: Oxford University Press.
- Guldi, J.** (2018). Critical search: a procedure for guided reading in large-scale textual corpora. *Journal of Cultural Analytics*. DOI: <https://doi.org/10.22148/16.030>
- Klingenstein, S., Hitchcock, T., & DeDeo, S.** (2014). The civilizing process in London's Old Bailey. *Proceedings of the National Academy of Sciences*, 111(26), 9419–9424. DOI: <https://doi.org/10.1073/pnas.1405984111>
- Kraicer, E., & Piper, A.** (2019). Social characters: the hierarchy of gender in contemporary English-language fiction. *Journal of Cultural Analytics*. DOI: <https://doi.org/10.22148/16.032>
- Lemercier, C., & Zalc, C.** (2019). *Quantitative methods in the humanities: an introduction*. Charlottesville, VA: University of Virginia Press.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A. & Lieberman Aiden, E.** (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. DOI: <https://doi.org/10.1126/science.1199644>
- Murdock, J., Allen, C., & DeDeo, S.** (2017). Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition*, 159, 117–126. DOI: <https://doi.org/10.1016/j.cognition.2016.11.012>
- Piper, A.** (2019). *Enumerations: data and literary study*. Chicago, IL: University of Chicago Press.
- Risam, R.** (2019). *New digital worlds: postcolonial digital humanities in theory, praxis, and pedagogy*. Evanston, IL: Northwestern University Press. Retrieved from <http://www.jstor.org/stable/10.2307/j.ctv7tq4hg>
- Schwartz, V. R.** (2001). Walter Benjamin for historians. *American Historical Review*, 106(5), 1721–1743.
- Stilgoe, J. R.** (1982). *Common landscape of America, 1580 to 1845*. New Haven, CT: Yale University Press.
- Stone, L.** (1979). The revival of narrative: reflections on a new old history. *Past & Present*, 85[11], 3–24.
- Underwood, T.** (2019). *Distant horizons: digital evidence and literary change*. Chicago, IL: University of Chicago Press.
- Weeks, J.** (1982). Foucault for historians. *History Workshop*, 14[3], 106–119.