**Invisible Interpretations: Reflections on the Digital Humanities and Intellectual History**

**Mark J. Hill**

*Department of Government, London School of Economics, London, UK*

Houghton Street, London, WC2A 2AE, 0207 852 3700, m.j.hill@lse.ac.uk

# Invisible Interpretations: Reflections on the Digital Humanities and Intellectual History

Much has been made of the digital humanities, yet it remains an underexplored field in relation to intellectual history. This paper aims to add to the little literature which does exist by offering a survey of the ideas and issues facing would-be practitioners. This includes: an overview of what the digital humanities are; reflections on what they offer intellectual history and how they may be problematic in regard to, first, accessing texts, and second, analysing source material; a conclusion with three reflections on future best practices – to be sceptical of digital sources, to be reflective of methodologies and how they may need to be modified when engaging with the digital humanities, and to embrace more directly the methodological, statistical, and technical aspects behind digital humanities. The aim is not to provide all the answers – at this stage that is impossible – but to be part of an emerging and ongoing discussion.

## 1. Introduction

While the ideas, tools, and methods which make up the digital humanities have increased in prominence, the uptake amongst traditional scholars has been limited. For evidence of this we need only look to two recent texts: *The Digital Humanities Manifesto 2.0*, which argues 'print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated' and instead 'digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences,' and Guldi and Armitage's *The History Manifesto*, a call to arms to embrace, and even create, these digital tools.[1] However, were these views to be largely shared there would be no need to write a manifesto.

A key reason for this lag is the lack of knowledge necessary to make use of these new tools.[2] While the digital humanities may offer exciting new possibilities, they have

---

[1] *Digital Humanities Manifesto*, 2-3.

[2] The word lag, rather than lack of interest, is purposeful. In a non-scientific survey distributed online, with responses largely made up of postdocs and early career academics, a majority of self-described intellectual historians note that they have used digital primary source material (86%). In regard to the use analytical tools

also developed their own disciplinary methods and practices. Thus, for those interested in engaging, serious learning and reflection are necessary before we can come to understand *how* we engage. The goal of this paper, then, is to offer some initial thoughts from the perspective of intellectual history on these issues.[3] This is done in four sections: first, a very brief introduction to digital humanities is presented. This is followed by a survey highlighting the methodological implications of big data and computational research for intellectual historians in two parts; first on accessing sources, and second on analysing data. Finally, some thoughts on best practices for intellectual historians are offered.

## 2. What are the Digital Humanities?

The digital humanities can be summed up as the application of algorithms to facilitate in search, retrieval, and analysis of data traditionally worked on in the humanities. What this may look like in practice includes authorial attribution, stylometry, metaphor tagging, GIS-based spatial analysis, distant-reading, quantitative text-analysis, digitization, social network analysis, discourse analysis, and more. However, as foreign as these ideas may seem from intellectual history, the history of the two disciplines is tied.

Jesuit priest Roberto Busa is credited with founding the field. Interested in Aquinas' concept of 'presence,' Busa began his research like many scholars – by working through tables and subject indexes which referenced the idea of interest. In doing this, however, he drew two conclusions:

> I realized first that a philological and lexicographical inquiry into the verbal system of an author has to precede and prepare for a doctrinal interpretation

---

the opposite was true (80%). However, a clear majority responded that they are interested in these tools (66%). The reason for this disparity may be highlighted in one respondent's comments: 'I don't think I have ever been told about these, or about their relevance to my research. I don't know what they do. I would be interested.' A Google Scholar search for the concurrence of 'intellectual history' and 'digital humanities' may highlight why there is so little awareness. At this time there are perhaps six works worth noting: David Armitage's "What's the Big Idea?" and the already mentioned *History Manifesto,* co-authored with Jo Guldi, are interesting, albeit introductory, reflections. Marie Leca-Tsiomisa's "The Use and Abuse of the Digital Humanities" is a valuable contribution to what can go wrong. In terms of intellectual history research which uses the digital humanities, Peter De Bolla's *The Architecture of Concepts* stands out. Dan Edelstein's "Intellectual History and Digital Humanities" is a helpful review essay of both *The Architecture of Concepts* and *The History Manifesto*. Jennifer A. London's "Re-imagining the Cambridge School in the Age of Digital Humanities" offers some further reflections (although the emphasis on 'Digital' and 'Cambridge School' is, at times, exaggerated). For historians generally, Robertson's "Differences between Digital Humanities and Digital History" is a useful resource offering an overview of digital history projects.

[3] Many of the issues noted apply to disciplines beyond intellectual history, thus works from outside the discipline are – and should be – made use of. However, it is also the case that many of the methodological issues which concern intellectual historians are less pressing in other fields.

of his works. Each writer expresses his conceptual system in and through his verbal system, with the consequence that the reader who masters this verbal system, using his own conceptual system, has to get an insight into the writer's conceptual system. The reader should not simply attach to the words he reads the significance they have in his mind, but should try to find out what significance they had in the writer's mind.

Second, I realized that all functional or grammatical words (which in my mind are not 'empty' at all but philosophically rich) manifest the deepest logic of being which generates the basic structures of human discourse. It is this basic logic that allows the transfer from what the words mean today to what they meant to the writer.[4]

The first concern is shared by intellectual historians today. The second, a more nuanced theory, led Busa to imagine creating 'a concordance of all the words of Thomas Aquinas, including conjunctions, prepositions and pronouns.'[5] Daunted by the, in excess of, 10 million words in the *Index Thomisticus*, however, Busa turned to the computer. In 1949 he began working with IBM to create a punch-card index, the beginning of a process which resulted in the first printed volume (of 56) being released in 1974, a CD-ROM in 1989, a CD-ROM with hypertext in 1992, and a website in 2005.

As access to computers increased, and scholarly engagement with digitized text became more frequent, the Oxford Text Archive (OTA) was born. Founded in 1976, the initial *raison d'etre* was to collect and maintain electronic texts to prevent the duplication of efforts. Today, the OTA is recognized as the beginnings of the digital library. This was followed by moves to standardize digital data by the Text Encoding Initiative (TEI), followed by new ways of engaging with digital text, such as HyperText Markup Language (HTML) and Extensible Markup Language (XML). Optical Character Recognition (OCR) has allowed for the digitization of texts and archives with more ease, and digital photography has allowed for further advances amongst art historians and others who engage with facsimiles of documents. All of this has contributed to a situation in which texts are now 'born digital,' and 'digital scholarly editions' have become an important form of academic work.[6]

---

[4] Busa, "Annals of Humanities Computing," 83.
[5] Busa, "Annals of Humanities Computing," 83.
[6] For more on the digital scholarly edition and its relevance to intellectual history see Evrigenis' "Digital Tools and the History of Political Thought." For more general reflections see Price's "Edition, Project, Database,

However, digitization is only one part of the digital humanities. Busa's work was complimented by an increase in interest in the application of statistical methods to text. Mosteller and Wallace's *Inference and Disputed Authorship: The Federalist* (1966), for example, used Bayesian statistics to identify the most probable author of the disputed *Federalist* articles. As data has become easier to engage with and manipulate, more sophisticated mathematical investigations have emerged.[7]

There have also been contributions to fields that are less Western-centric. While lag remains an issue between projects which focus on Europe and the rest of the world, there are a number of sources and tools which have emerged which make reading and analysing historical objects from a global perspective more feasible.[8]

All of this – from digitization to analysis – has brought us to a point where the digital humanities have become an established part of the university, with its own degrees and departments, conferences and journals, and regional and international associations. However, its emergence as a discipline may have had a negative impact. Segregating research into specific digital humanities outputs has made it less likely for traditional scholars to engage. In addition, more specialized works – those relying on advanced statistical modelling or computer coding – have become inaccessible to traditional humanities scholars. The result of this has been, as Hockey argues, a 'less critical appraisal of methodologies than might be desirable' and the siloing of research into less mainstream publications.[9] It is the intention of this paper to examine this problem in relation to *doing* intellectual history.

Since ideas are not created by autonomous and atomistic individuals they cannot be understood as if they were uttered by such. Therefore, intellectual historians argue that textual meaning must be reconstructed through a contextual understanding of the specific intellectual world from which a given text comes. Only in this way can one grasp the intricacies of meaning, the available conceptual frameworks behind words, shifts in meaning, and potentially the illocutionary acts contained. To understand historical ideas, then, we must understand how those ideas would have been created and understood by historical people. The way one does this may differ

Archive, Thematic Research Collection" and "Electronic Scholarly Editions," and the MLA's *Considering the Scholarly Edition in the Digital Age*.

[7] See also: Morton's *Authorship of the Pauline Epistles*. For more recent work in authorial attribution see Kestemont et al's "Authorship Verification with the Minmax Metric."

[8] See: Lie et al's "Textual Analysis for Studying Chinese Historical Documents and Literary Novels" and Long and So's work at the Chicago Text Lab.

[9] Hockey, "The History of Humanities Computing," 10.

between researchers, but Mark Bevir's general description of historical work is helpful here:

> Historians have before them various relics left over from the past. There are government documents, newspapers, the natural landscape, paintings, tools and utensils, books, etc… Historians study relics from the past. They use them to reconstruct historical objects, or, perhaps we should say, to construct historical objects. Historians of ideas [or intellectual historians] study relics from the past in order to recover historical meanings.[10]

Taking on this distinction as a guide we can then identify two roles for digital humanities: aiding access to relics and facilitating the reconstruction of historical meaning.

## 3. Access to historical relics

Providing new ways to access historical texts is an area in which the digital humanities has made an obvious – and in many ways unreflected upon – impact on historical research. For example: *Early English Books Online* (*EEBO*), *Eighteenth-Century Collection Online* (*ECCO*), and the *British Library 19th Century* collection collectively hold 350,000 texts; *Perseus Digital Library* provides access to over 160 million words; *Making of the Modern World* has 61,000 texts; the *Analyse et Traitement Informatique de la Langue Française* (*ARTFL*) project has 3,500 texts; *Electronic Enlightenment* has 67,875 documents; *Early Modern Letters Online* contains another 98,450 letters. While much of this data is segregated, there are also movements to merge databases and share metadata (*Historical Texts*), or create aggregators (*Connected Histories*). There are also an increasing number of databases of non-Western source material, such as the *China Biographical Database Project*, the *Database for the Study of Modern Chinese Thought and Literature*, the *Aozora Bunko Digital Library*, and the *Digital Islamic Humanities Project*. Finally, there are ever increasing – and exciting – digitization projects (or scholarly editions) dedicated to specific authors, topics, or works, such as the *Newton Project*, *Charles Booth's London*, and the *Grub Street Project*. The fluid and, importantly, economical world of digital publication is likely to ensure this trend continues. Evrigenis, who is working on a digital edition of Bodin's *Six livres*, writes:

---

[10] Bevir, *The Logic of the History of Ideas*, 31. Although a description of the history of ideas, it is as applicable to intellectual history.

[T]he plasticity of web design allows for flexibility with regard to the structure of an edition that is simply impossible on paper… [O]ne can place two or more versions of a text side-by-side, in order to compare them, and still maintain the ability to increase or reduce the number of texts being compared. At the same time, users can have the ability to call up photo-reproductions of the pages, so as to be able to compare the text with its source and one original with another. The same is true of all other relevant resources. Editions can now link to maps and gazetteers, biographies, histories, primary and secondary sources, databases, multimedia, and dictionaries. With links to these resources, a user will not only be alerted to the fact that a passage contains a reference to another work; she will also be able to link to the appropriate passage in that work. Dictionaries and grammar tools allow even users who cannot read the original language the ability to approach the text. Similarly, the side-by-side display of different editions of the text allows readers to see not only how the text may have changed from one version to another, but also how it was translated in another language. Alignment between different versions means that a user will need reading knowledge of just one language in order to access the text, making it possible to introduce primary material in a variety of languages to students at early stages of their studies.[11]

He adds that editions of this kind offer another advantage: they remain works in progress which can be easily amended and corrected.

All of this has resulted in the weakening of traditional academic gatekeepers, both through 'open access' digital projects, and by making physical presence in libraries and archives unnecessary. While simply making use of these resources is not *doing* digital humanities *per se*, their existence is an outcome of the digital humanities, and therefore, as sources, they should not be glossed over in a discussion on this topic. Especially as, for a discipline which places emphasis on deep and wide reading, the increase in digitized texts may be seen as a boon. However, there are a number of concerns that those who use these sources (which, today, would be the vast majority of intellectual historians) may not be aware of.

### 3.1. Data overload and 'cognitive bottlenecks'
Again, thanks to digitization projects, intellectual historians today have access to ever increasing numbers of sources. Beyond ease, there is, as Edelstein notes, the hope

---

[11] Evrigenis, "Digital Tools and the History of Political Thought," 196.

that sources 'forgotten for generations can now be rediscovered thanks to the magic of search engines.'[12] The reality, however, is that the proliferation of sources has resulted in a situation where we are now presented with more information than it is possible to work with. For example, a search on *Historical Texts* for 'Plato' returns 39,522 hits; 'free will' provides 27,593; and 'natural law' and related concepts result in 40,426 hits. The physical limits a humanist faced before were humane limits; one *could* read everything they had access to. We are now in a position in which we must ask, as Crane has, 'what do you do with a million books?'

This is not bad in essence, but it means that we must develop strategies to choose what sources – or paths of sources – to follow, and these decisions could become methodologically problematic as researchers make choices as to what is, and perhaps more importantly, what is *not* read. We must be on guard so as to not put limitations on research, or worse, allow algorithms to invisibly impact our interpretations.

### 3.2. Exploratory and decontextualized engagements with data

A key problem is that the ways we access data has not changed despite radical shifts in the type and amounts of data. Mussell notes how there has been little change in the structure of digital resources, and there continues to be a preference for engagements which privilege searching over browsing – digital libraries are reflections of physical libraries in which users are expected to know what they are looking for.[13]

This has resulted in researchers developing strategies for exploring digital sources, such as Keywords-in-context (KWIC) searches, hoping to efficiently decipher some sense of context. However, this desire – and need – for efficiency may result in us focusing on subjects we are already familiar with, rather than discovering forgotten texts or new areas of research.

These simplistic engagements with complicated data may be causing larger methodological problems, however. Bullard notes that, while for eighteenth century scholars, *ECCO* 'has become an indispensable resource, always open on the desktop, the oracle to which every research question is first referred,' its 'searchability' has 'in fact distanced scholars from these documents and has accelerated a tendency towards information-grabbing.'[14] That is, rather than hermeneutic engagements with

---

[12] Edelstein, "Intellectual History and Digital Humanities," 1.
[13] Mussell, "Doing and Making," 83.
[14] Bullard, "Digital Humanities and Electronic Resources," 754-755.

texts, or careful reconstructions of context, one collects diverse pieces of data. To use Bevir's 'relic' metaphor, one creates mosaics rather than restoring a vase. This is particularly problematic for intellectual historians who generally favour broader readings, which these databases are immensely useful for, but also detailed – and time consuming – engagements.

### 3.3. Incorrect data

Regardless of the methods used to explore data, we run into larger issues if the data itself is inaccurate. Two issues in this regard can be noted using *ECCO* – the world's largest database of its kind – as a case study: comprehensiveness and accuracy.

*ECCO* holds over 32 million pages of text, 182,898 titles, which it claims covers 'every significant English-language and foreign-language title printed in the United Kingdom between the years 1701 and 1800.'[15] However, as Bullard notes, this is only slightly more than half of the titles listed in the *ESTC* (*English Short Title Catalogue*). As he notes, 'Gale's decision to leave the database at this size reflects the tacit opinion of most customers that its coverage is adequate.'[16] However, adequate does not mean comprehensive, and the decision as to what is 'significant' is one that some may wish to challenge. As Spedding notes, the microfilm series on which *ECCO* is based is not 'random' but instead based on a criteria that is rarely mentioned, and includes ease of access.[17] This may not be a problem if one wants access to a text which they know exists, but for those interested in forgotten texts or the statistical analysis of entire corpora, this may preclude *ECCO*.

The second issue is accuracy. To start, OCR remains imperfect, and historical variations in spelling and typeset are an issue. Bullard, who has made a survey of *ECCO*'s accuracy, concluded that it is potentially highly problematic: results ranged from 12.5 words recognized for every word missed to a 1.3.[18] There are some search engines which try to mitigate these issues by performing fuzzy, variant, and metaphone searches, but for many it is not initially clear what these are or how they are represented in results.[19] Spedding, after hitting a dead-end in his own research into the history of condoms, offers some reflections on this issue. Having tried 'any of the many alternate spellings or circumlocutions,' he turned to 'searching for texts

---

[15] *Eighteenth Century Collections Online (ECCO)*. http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx.
[16] Bullard, "Digital Humanities and Electronic Resources," 755.
[17] Spedding, "'The New Machine,'" 440-441.
[18] Bullard, "Digital Humanities and Electronic Resources," 755-756.
[19] More advanced search methods are emerging, such as Word2Vec, which make use of 'neural networks' and 'deep learning' to provide results. See Goldberg and Levy's "word2vec Explained."

that mention venereal diseases by name […], or the symptoms of venereal diseases […], or for references to the writers who have discussed venereal diseases and condoms.' While helpful advice, it is still problematic: first, it requires deep contextual awareness which precludes its use in many cases, and second, as Spedding concluded, 'the results are for the most part still disappointing.'[20]

Thus, despite there being many who take the position (at least implicitly) 'that anything ECCO does not know about the eighteenth century is not really knowledge,' the most comprehensive database of its kind does not accurately reproduce or represent the texts it contains.[21]

### 3.4. Neutrality of data

It should be noted that 'more' data – or ideas represented as data – is in itself not a positive thing. In many ways intellectual history is premised on this point: a successful idea does not mean it is a superior idea. That is to say, wide-spread belief is not evidence of logical superiority. Instead, intellectual historians aim to understand ideas on their own – historical – terms; what they meant and how they could mean it. While for more statistically situated disciplines more data can provide a better representation of 'reality' (even if not statistically necessary), for the intellectual historian there is likely to be diminishing returns as we read more texts. As we add more data we may simply be adding noise.

What is more, text is not neutral despite it being treated as such once digitized. From an ontological perspective this makes sense. As Latour wrote, digitization leads to 'the greater merging of figures, numbers and letters, merging greatly facilitated by their homogenous treatment as binary units in and by computers.'[22] Treating all sources as equal (or ontologically equivalent), however, may lead to interpretations that are less than robust in a discipline which is concerned with intentions, power relationships, contributors and interlocutors, translations, and many other potential variables which are difficult to digitize.[23]

This leads to the issue of gatekeepers again: Berry notes: 'Technology enables access to the databanks of human knowledge from anywhere, disregarding and bypassing the traditional gatekeepers of knowledge in the state, the universities and the market.' However, these new sources may 'provide destabilising amounts of

---

[20] Spedding, "'The New Machine,'" 444.
[21] Bullard, "Digital Humanities and Electronic Resources," 754-755.
[22] Latour, "Visualisation and Cognition," 16.
[23] Forms of digital discourse analysis may be of use here. See Herrera and Braumoeller's "Symposium: Discourse and Content Analysis."

knowledge and information that lack the regulating force of philosophy.'[24] That is to say, while gatekeepers of knowledge (paywalls, for example) are not necessarily a good thing, training is. When history becomes something which one engages with without guidance (be it historical, statistical, or philosophical) the results will suffer.

## 3.5. (Re)Presentation of data

A final issue worth noting in this section is the false equivalency made between original sources and digitized versions. When digitized a number of aspects of a source are lost, most commonly noted bring physical attributes, such as size, number of pages, marginalia, and illustrations. However, there are less obvious examples, such as the quality of penmanship in a private letter, which may tell us something about the author or recipient, and the choice of font used in a pamphlet, which may say something about its intended audience.[25] Even periodicals lose something when divorced from their material form.

*The Spectator* was Joseph Addison and Richard Steele's early eighteenth century contribution to polite culture, responsible in part, Habermas claims, for the emergence of the public sphere.[26] Issues were written and published for patrons of London's coffeehouses with the expectation of them being shared, discussed, and commented upon.[27] Its physical format – a single folio half-sheet – is a testament to this. They were cheap to publish, quickly read, and easily distributed. Its perishability contains contextual meaning: 'Except when used for a broadside ballad, the lowest genre of all, the single sheet was virtually useless from a literary point of view.'[28] This implicit understanding of the status of their own work was made explicit in issue 529 of *The Spectator*, in which the author mocks himself for being author of such a trivial form of literary endeavour. However, the physical quality of the work belied its content, and its seemingly low provenance may have been necessary for its audience to engage with, and remark upon, the ideas presented. Yet, when one reads *The Spectator* today it is part of a multi-volume collection, printed on bible-thin paper, and dotted with scholarly notations. Even an online version, *The Spectator Project*, is simply a facsimile of an out of print edition. When engaged with it this way it becomes a series of tomes representing lessons on politeness; its immediacy as a contributor to an important public discussion is lost, and this is certainly one reason

---

[24] Berry, "The Computational Turn," 8.

[25] Thornton, *Handwriting in America: A Cultural History;* Kinsley, "Meaning and Format."

[26] Habermas, *Structural Transformation*, 43.

[27] *The Spectator* was not unique in this way. Some publications integrated this 'public' aspect physically – *The Flying-Post* and *Dawk's News-Letter* had blank spaces in which comments and additional news could be written before being passed on.

[28] Kinsley, "Meaning and Format," 487.

for the changing reaction readers have had to the text.[29] This is an issue not strictly related to the digital humanities, but is one which both digitization projects, and those engaging with digitized text, must be aware of.

One potential solution is the digital thematic collection which aims to bring together relevant contextual information, both primary and secondary, to encourage a more thorough understanding (see Evrigenis' description above). However, these – or any – collections can also inadvertently impact research outcomes. To offer one example: Jones notes that the nation-state as a central research topic in the nineteenth-century may have emerged because 'historical methods meant methods of archival research, and the archives were the archives established by modern states. So there was a natural affinity between the modern state and the emergent historical profession.'[30] Thus, as we create thematic collections or digital archives, we need to be aware that what is chosen to be included may directly influence the research that is done. As Marshal McLuhan wrote, 'the medium is the message.'[31] In our case, this works in two ways: we may miss an aspect of the message, or the message may be contorted into something which does not reflect reality.

## 4. Reconstruction of historical meaning

While the goal until now has been to demonstrate that the digital humanities are not neutral in their relationship to text, there is more to the digital humanities than digitization projects: in addition to providing new ways of accessing historic relics, it promises to aid in the reconstruction of historical meaning.

For many this is the exciting part. It is the magic box which one can input data and extract exciting new research. While this perspective has been routinely – and correctly – tempered, there are opportunities for researchers willing to invest time. As Bullard has noted: 'the most successful resources of recent years have been analytical in function, using database software to present texts that are too information-rich to be readable in any straightforward way, such as diaries, registers, library catalogues, or serials published by parliaments, say, or by courts of law.'[32] Thus, in addition to providing us with new ways of interacting with our sources, these tools may help overcome many of the issues identified in the previous section. Tools,

---

[29] See C.S. Lewis' dismissive claim that in *The Spectator* 'Everything the moderns detest, all that they call smugness, complacency, and bourgeois ideology, is brought together' ("Addison," 167). It can only be seen as being "brought together" if it is read together.

[30] Jones, "Towards a Genealogy of Intellectual History," 3-4.

[31] Marshall McLuhan, *Understanding Media: The Extensions of Man*, 1964.

[32] Bullard, "Digital Humanities and Electronic Resources," 749.

such as Paper Machines, MALLET, Gephi, Quanteda, and MARKUS, and resources, such as the Six Degrees of Francis Bacon, may allow us to sift through data in more efficient – and perhaps surprising – ways.[33] However, there are a number of issues intellectual historians need to be aware of before embracing these tools, many of which may have been designed for use in other fields.

## 4.1. Misuse of methods and tools

The primary problem facing those interested in using the tools of the digital humanists is misuse due to a lack of understanding. To offer an example we can look to MALLET:

> MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text… It includes sophisticated tools for document classification: efficient routines for converting text to 'features', a wide variety of algorithms (including Naïve Bayes, Maximum Entropy, and Decision Trees), and code for evaluating classifier performance using several commonly used metrics.[34]

Meeks and Weingart highlight the problem tools built upon specialized knowledge introduce: 'Were a critic of digital humanities to dream up the worst stereotype of the field, he or she would likely create something very much like this, and then name a popular implementation of it after a hammer.'[35] This is a real issue; the traditional intellectual historian is going to struggle to make sense of what these tools do, let alone how they work or what the outputs may mean. De Bolla's *Architecture of Concepts*, a text which explicitly engages with these issues, offers some insights into these issues.

The aim of his work, he writes, is to 'develop a way of generating data that throw[s] light on the strength of attraction one concept may have had to another.' This, what he calls, 'orbital drag' can be used to develop insights into the historical structure of a concept. For example: 'over a long period of time in the English language context, "democracy" pulls within its orbit "liberty." In operating the first concept, one cannot but be operating the second.'[36] To trace these 'ideational concepts' De Bolla turns to *ECCO* to find co-occurrences with the term 'rights' between 1700 and 1800. By

---

[33] See: Crymble, "Review of Paper Machines"; Graham and Milligan, "Review of MALLET"; Bastian, Heymann, and Jacomy, "Gephi"; Benoit, "quanteda"; Warren et al, "Six Degrees of Francis Bacon."

[34] *MAchine Learning for LanguagE Toolkit*. http://mallet.cs.umass.edu/.

[35] Meeks and Weingart, "The Digital Humanities Contribution to Topic Modeling," 2.

[36] De Bolla, *Architecture of Concepts*, 44.

noting how these words shift and change over time he hopes to come to understand when and how certain concepts came into play – specifically the shift from 'rights of man' to 'human rights.' While a valuable project, there are issues, as De Bolla is first to admit.[37]

First, due to limitations with *ECCO*'s search engine, co-occurrences had to be noted by hand. Therefore, De Bolla chose the words which he felt were of particular interest, or he noticed occurring most frequently. This method may preclude outcomes which would be unexpected. Second, there is no data normalization – the fact that he has access to fewer books at the beginning of the eighteenth century than at the end, or that many books are catalogued multiple times in *ECCO*'s database, impacts the results. De Bolla accepts this, but notes that with 'statistical extrapolation' there is little distortion.[38] This may be the case, but it nonetheless weakens the evidence for his conclusions. Finally, as Edelstein has noted, there seems to be confusion between '*frequency* in word collocation' and 'number of *documents* containing the proximate terms.'[39] That is to say, the distance of co-occurring words is not taken into account, resulting in documents being returned as relevant if they contain the search words *anywhere* in the text rather than within a certain distance from each other. Again, this is largely due to limitations with *ECCO*, but it nonetheless leads to conclusions which are problematic.[40]

Importantly, not all reviewers have picked up on these issues. As noted, Edelstein – an intellectual historian and digital humanist in his own right – has written a review which engages with the methodology as much as with historical claims. Jonathan Israel, however, simply concludes: 'By means of a systematic digital study […] he produces fully convincing and highly innovative conclusions.'[41] This is not a criticism of Israel or an indictment of his review, but an important point is raised by the incongruity: unless we are trained in the tools and methods being used, aware of the databases sourced, and familiar with the nuances of statistical analysis, not only may we make mistakes, but we may be unable to recognize issues with other's analysis.

Weingart highlighted a similar issue when describing the software package Gephi. In his introduction to social network analysis he notes that Gephi (at the time of publication) would accept data, load bimodal networks with that data, run a

---

[37] De Bolla, *Architecture of Concepts*, 7-10.
[38] De Bolla, *Architecture of Concepts*, 9.
[39] Edelstein, "Intellectual History and Digital Humanities," 242.
[40] De Bolla now has a research group in Cambridge, The Concept Lab, which has solved some of these, and other, issues.
[41] Israel, "Review of Peter De Bolla," 709.

centrality test, and produce outputs without any visible issues. However, the outputs would be nonsense as the centrality algorithm used had been written for single mode networks.[42] Because Gephi is open source, Weingart was able to detect this. The reality, however, is that most intellectual historians do not have the training to recognize the initial error, let alone the programming skills to discover the cause.

One possible solution is to place emphasis on methods of validation. As Gimmer and Stewart note 'the blind use of any method without a validation step' must be avoided, especially when it is 'difficult to know if the methods […] are optimal for the particular problem at hand.'[43] Thus falsifiability must be built into intellectual history research which makes use of these tools and ideas.

It should perhaps also be noted that the inclination to embrace tools one may not be familiar with is indicative of a larger issue facing scholars as a whole: the pressure to push out outputs, or expedite research. While not a pressure unique to intellectual history, it may be particularly harmful to a discipline which aims towards investigations which are both deep and broad.[44] However, it is also worth noting that this should not lead one to dismiss those doing research in this area. While the tools being used by digital humanists have underlying scientific and methodological theories which intellectual historians may not be familiar with, and therefore, errors are likely to occur – and in some cases errors that may not look like errors – this does not need be the case.

### 4.2. Reliance on tools

There is another issue: one may understand a tool and its underlying methods perfectly but lack the understanding of historical context to make sense of the data. Allen et al's 'Plundering Philosophers: Identifying Sources of the *Encyclopédie*' is an example of this.

The article claims that by using Vector Space Models they are able 'to identify articles in the *Encyclopédie* that may have been borrowed, in whole or in part' from the *Dictionnaire de Trévoux* (1743), 'an intellectual rival of the *Encyclopédie* compiled by French Jesuits in the first half of the 18th century.' Specifically, they write that their research shows '5.32 percent of the examined articles (80 - 85% of the entire work) in the *Encyclopédie* were borrowed.' Furthermore, by using Pairwise Alignment of Intertextual Relations, they claim to detect passages borrowed from Louis Moréri's

---

[42] Weingart, "Demystifying Networks," 12.
[43] Gimmer and Stewart, "Text as Data," 5.
[44] For a view which sees digital humanities as a potential solution to these problems see Armitage's "What's the Big Idea?" and Guldi and Armitage's *History Manifesto*.

*Grand dictionnaire historique*. In this case they claim to have 'identified some 580 possible shared n-gram sequences of varying lengths. Upon close inspection, we determined that about 72% of these instances (418 n-gram sequences) were indeed common to both works.'[45]

While the results highlight research possibilities, there are a number of issues highlighted by Leca-Tsiomis. First, the authors ignored older sources in their analysis which could have been originators for both later sources – in fact, in one case a clear citation to this older source is ignored. Leca-Tsiomis notes that 'if they had consulted the existing specialist studies, they would have seen that we have long known that borrowing is the very essence of dictionaries… [It is] a law of lexicographical work and what is interesting to study is not the similarities, which are obvious, but the differences.'[46] Second, they compare the 1759 edition of the *Dictionnaire historique* with the *Encyclopédie* despite the latter's first volume being published in June 1751, and having a total of seven volumes published by November 1757. Leca-Tsiomis concludes that their historical analysis overstates, confuses, and ignores a number of issues which weaken the entire exercise. She writes that: 'if numerical research is to be taken seriously the historical grounding for this research must be solid. It is therefore to be regretted that the authors of the article did not pay sufficient attention to this aspect.'[47] We may wish to add to this: knowing how to use a tool does not insure that the results mean what we think they mean when accurately placed in their historical context.

### 4.3. Visualizations are not objective

Thus far we have touched on the problems which emerge when tools are used to analyse text. However, one should also reflect on how the results can be presented – especially when it comes to visualizations.[48]

First, visualizations can be meaningless to the uninitiated. A cross-tab outputted from a piece of statistical software to many is nonsense. A box-plot may be perfectly clear to those who engage with them regularly, but perhaps less so to a historian of medieval religious thought. When chi-squared tests or P-values are included the problem is compounded. This is not simply an issue of information conveyance – it may also be a hindrance to getting work published in traditional sources where

---

[45] Allen et al, "Plundering Philosophers."
[46] Leca-Tsiomis, "Use and Abuse," 471-472.
[47] Leca-Tsiomis, "Use and Abuse," 470.
[48] This is a topic of increasing interest. For more see: Drucker's *Graphesis*; Gatto's "Making Research Useful"; and Stone's "Information Visualization: Challenge for the Humanities."

audiences are less quantitatively literate. This leads to a slightly different problem: data visualization can be a form of rhetoric itself – a type of obfuscation through science.

Data is not fact – it reports very specific ideas and categories chosen by individuals for reasons often only known to the researcher. This is a point Drucker makes in *Graphesis: Visual Forms of Knowledge Production*. While we may know this explicitly, we often fail to recognize it implicitly, thus explicit reflection on this would is useful as our ways of engaging with and presenting data continues to change. What is more, there are potentially political implications to this. D'Ignazio, who has done work into the rhetorical power of data visualization, data literacy, and 'the politics and ethics of representation,' has written on this topic for the MIT Center for Civic Media: 'Even when we rationally know that data visualizations do not represent "the whole world", we forget that fact and accept charts as facts because they are generalized, scientific and seem to present an expert, neutral point of view.' We must remember that 'all knowledge is socially situated and that the perspectives of oppressed groups including women, minorities and others are systematically excluded from "general" knowledge.'[49] This is, of course, indicative of a larger problem, and one which faces practitioners in many disciplines, but it is something intellectual historians may need to reflect on more as they begin to embrace these new forms of scholarship.

### 4.4. The Quantitative Shift

The issues discussed in this paper are part of a larger change – a quantitative shift is taking place as qualitative texts are taken, transformed, and treated as data. The result is often referred to as distant reading, and the goal is to discover statistical patterns or relationships which are undetectable, have gone unnoticed, or are too labour intensive by traditional means.

Moretti's *Distant Reading* is the text one may initially come into contact with when researching on these ideas. A reflective exercise, the book offers its readers the development of Moretti's own ideas on turning literature into quantitative data. While intellectual historians may find these ideas and methods, to put it mildly, problematic, they should not be immediately dismissed. The "Slaughterhouse of Literature," to point to one examples, demonstrates the labour intensive work which digital approaches may be able to overcome, while also highlighting an investigation which intellectual historians may be particularly impressed by: the emergence,

---

[49] D'Ignazio, "What Would Feminist Data Visualization Look Like?"

imitation, and success of a particular concept (in this case, the 'clue' in mystery novels).

While Moretti's approaches have been exploratory – and in many ways, much of what we learn of these methods emerges because of these explorations – much more formal engagements exist in the social sciences. Grimmer and Stewart offer an overview of these methods in 'Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Text.' However, again, the methods may initially appear problematic for the intellectual historians. For example, they note that the most 'shocking' step is to discard word order and sentence structure in the documents they investigate.[50] That is, these methods rely on a key assumption: words can be treated as quantitative data which can transfer meaning without reference to sentence construction – they are, essentially, a 'bag of words.' For an example in practice we can turn to Lever, Benoit, and Garry's 'Extracting Policy Positions from Political Texts Using Words as Data,' a study which 'treats texts not as discourses to be understood and interpreted but rather, as data in the form of words.'[51] This is done by creating a 'document-feature matrix,' a dataset which can be analysed using statistical methods to create 'valid estimates' of the underlying characteristics of interest. That is, a table is created in which every unique word (or word sequence) is summed in its own column, while each text or document is represented as rows. This data can then be interrogated like any other quantitative datasets. Interestingly, this method allowed them to correctly replicate published estimates of policy positions as judged by humans, therefore making their method capable of making predictions in other cases 'without the substantial costs of time and labor that these require.'[52]

Blaydes, Grimmer, and McQueen are currently working on a project more directly relevant to intellectual historians titled 'Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds.' Although unpublished, it has recently been described by London:

> After compiling a large collection of 21 texts from Islamic contexts and 26 from Christian European contexts, the authors convert these texts into vectors… [They] then apply the statistical model they have developed, which

[50] Grimmer and Stewart, "Text as Data," 272.

[51] Laver, Benoit, and Garry, "Extracting Policy Positions," 311. For further examples see Slapin and Proksch's "A Scaling Model"; Quinn et al's "How to Analyze Political Attention"; and Blumenau and Lauderdale's forthcoming "Never Let a Good Crisis go to Waste."

[52] Laver, Benoit, and Garry, "Extracting Policy Positions," 312.

isolates what they call a 'hierarchy of topics' identifying general themes that span these texts and also finer, 'more granular' themes nested within those topics. Specifically, the model identifies four broad themes across mirrors-for-princes texts: the role of the king and his ideal practices of governance; his private life and personal virtues; the role of religion and its place in political, social, and private life; and geographical concerns that pertain to politics. The model also discovers key words that distinguish these broad themes.[53]

Again, many intellectual historians will object immediately to these novel methods. These are texts written by different authors, in different languages (translated into English), at different times between the eight to seventeenth century, in different places, and amidst radically different political situations. That is to say, historical context seems to be abandoned, and authorial intention dismissed – all that is left is words.

The practitioners of these methods may argue that nothing is lost – the words in the bag are no different from those which the author initially put to paper. Instead, they are simply engaged with those words in a different way. Yet, we must ask whether the outputs from a topic modelling algorithm would be recognizable to the original author. What is more, how do such methods interpret Swift's *Modest Proposal* or Machiavelli's *Prince*?

To put it simply, the problem with this quantitative shift is the way in which texts are turned into data which is then treated as empirical fact – a perspective which Skinner long ago argued we had moved away from: 'Scarcely anyone nowadays believes in the possibility of building up structures of factual knowledge on foundations purporting to be wholly independent of our judgements.'[54] However, it must be admitted that practitioners of these types of analysis refuse to take a fully positivistic turn.[55] Grimmer and Stewart write: 'We emphasize that the complexity of language implies that automated content analysis methods will never replace careful and close reading of texts. Rather, the methods that we profile here are best thought of as *amplifying* and *augmenting* careful reading and thoughtful analysis.'[56] However, even this position can be problematized, as London does: 'the objective of starting with

---

[53] London, "Re-imagining the Cambridge School," 364.

[54] Skinner, *Visions of Politics*, 1.

[55] To claim that those doing this type of research are ignorant of these issues would be a disservice. Jurafsky and Martin (*Speech and Language Processing*) highlight many shared concerns.

[56] Grimmer and Stewart, "Text as Data," 2.

text-as-data findings and then moving to traditional textual analysis should be revised, to question what is being "amplified and augmented" in the first place.'[57]

## 5. Looking towards best practises

The use of digitized source material has become a part of most intellectual historians working process, and there is little to suggest that the tools discussed in this paper will have any less of an impact. However, integration of these tools and ideas must not take place without reflection. Although this survey has largely been critical, the aim has not been to discount the digital humanities. Instead, these are problems which must be faced head-on by intellectual historians using the outputs of, or wanting to venture into, the digital humanities. It is a starting point rather than a conclusion. Thus, to end this survey, three reflections on best practises for moving forward are offered.

First, we need to be aware of the limits of digital sources, and aim to influence their development with more robust usability in mind. As De Bolla has shown, databases as they currently exist are largely unsatisfactory for the types of engagement many may hope for. We cannot rely on metadata, accuracy, comprehensiveness, interoperability, or even the ability to extract raw text. However, that does not mean digital text is of no use. For now, however, those who want to engage with text as data may need to create their own datasets.

Second, there needs to be methodological reflectivity as we, as intellectual historians, move forward. This is a position which has become increasingly important to those already working in the digital humanities.[58] However, it is just as important for intellectual historians, who are just beginning to delve into the digital, to keep in mind.

The implications of this are perhaps most clear when we reflect on sources themselves: digital sources are not the same as print sources, and we need to take this fact into consideration. As Gordon noted:

> Intellectual history is by definition not merely a description of perceptible objects but an inquiry into meaning. From this basic premise intellectual

---

[57] London, "Re-imagining the Cambridge School," 362. London argues for 'collaborative work between Cambridge School theorists and political scientists mining textual data would allow them to analyze implications of text-as-data findings' (Ibid., 365).

[58] Drucker is an important example of this type of reflective work in digital humanities. Nowviskie offers a review of the 'hack' and 'yack' debate in the digital humanities which may be informative for intellectual historians.

historians have recognized their own hermeneutic implication in the meanings they study, and they have developed sophisticated theories so as to explain the way that any access to past meaning must of necessity involve the historian in an act of interpretation rather than objectivistic description.[59]

As the way we access the past changes we need remind ourselves of this. New issues may emerge as we engage with digital texts – issues which others working in the digital humanities have yet to come into contact with.

At the same time, we must reflect on the ways computers aid in engaging with texts, or extract information from them. That is to say, we should not assume that the methods developed in other disciplines can, or should, be applied to intellectual history in a way which is compatible with our own methodological concerns. In fact, these tools, techniques, and methods may be entirely incompatible. However, computers may be capable of engagements in ways which we have not yet imagined. De Bolla's *Architecture of Concepts*, and Betti and van den Berg's theory of computationally modelling ideas, and applying this model to historical texts, are examples of attempts to do this. However, there is more than just text in historical studies.

There are new tools and methods emerging which intellectual historians may be able to make us in their quest to reconstruct contextual information. Spatial Humanities, and tools making use of Geographic Information System (GIS), for example, may be used to glean new insights into historical relationships. As a recent volume by Stock and Kümin highlights, the study of the 'emplacement, distribution, and connection of entities, actions, and ideas' can be used 'to understand the societies, cultures, and mentalities of the past.'[60] One positive of example of research in this field is Stanford's Mapping the Republic of Letters project. However, there are certainly many more potential uses, and a number of interesting projects are worth reflecting on, such as the 3D recreation of seventeenth century Pudding Lane in London, and Mylène Pardoen's simulation of the audio experience of eighteenth century Paris.[61] While, as Knowles has noted, the intellectual substance of some of these methods

---

[59] Gordon, "Contextualism and Criticism," 43.

[60] Stock and Kümin, *The Uses of Space in Early Modern History*, 2.

[61] The Spring 2014 issue of the *Journal of Digital Humanities* is dedicated to "Exploring and Designing Virtual Worlds" through digital models, which includes Dempsey et al's "Pudding Lane: Recreating Seventeenth-Century London." The CNRS has an article on Pardoen's work, as well as the audio, available at https://news.cnrs.fr/articles/sound-18th-century-paris.

remains to be fully demonstrated, they are, at the very least, valuable in generating new questions.[62]

Third, and following from the previous point, we need to rethink the training and knowledge appropriate for intellectual historians. As scholarship becomes increasingly entangled with technology, and as the quantitative shift continues, we need to make visible what these digital tools and statistical models are doing. This means that we have a lot of learning to do.

While what needs to be learned can be mitigated to an extent by taking part in collaborative work with those who have these skills, and there are certainly benefits to interdisciplinary research, we should not become a discipline that relies on others. Thus, while there is much to be said for London's idea of 'revolutionary, collaborative conversations in our field that could unite empirical and hermeneutical scholars to identify and compare central concepts and themes across varied traditions,' Guldi and Armitage's position – that historians have their own unique insights which could be enhanced by history departments which are willing to 'train designers of tools and analysts of big data' – should not be dismissed.[63] Again, this is not a call to 'throw the baby out with the bathwater.' There is much which is to be learned from statisticians, computer scientists, methodologists, and others engaged in the digital humanities. The point is, however, that *we* should take the time to learn these things.

As Edelstein has noted, however, this is not an easy task:

> [T]his is an ambitious project, particularly if we are also to be designing and building tools (and presumably studying a little history on the side). While very laudable and in some respects necessary, Guldi and Armitage's proposals do raise the question of opportunity costs… More practically, *when* is the budding historian to master all these skills? In college? Graduate school? As an assistant professor? After tenure?[64]

For those interested in these ideas, the simple answer is now. However, with regard to the discipline, it is worth noting that most intellectual history PhD programmes already require a working knowledge of at least two languages. It may not be far in

---

[62] Knowles, "Why We Must Make Maps," 24. Von Lünen follows from Knowles' point in "So long, and thanks for the GIS: Digital Spatial History," concluding that history is specifically lagging in its use of these tools, but potential exists nonetheless.

[63] London, "Re-imagining the Cambridge School," 366; Guldi and Armitage, The *History Manifesto*, 107.

[64] Edelstein, "Intellectual History and Digital Humanities," 246.

the future that, as Kirschenbaum suggests, this could be supplemented with a computer language.[65] What is more, departments may wish to begin offering methodological training in statistical analysis, as is often done in other disciplines. Again, it is not enough to know how to *do* topic modelling (for example) – in fact, making it easier to do so may not be a positive thing at all. Instead, as scholars we need to understand how data is being manipulated and transformed. As Mullen – and others – has argued: 'One of the fundamental principles of [Digital Humanities] is that you should let your research questions, not the tools, drive your work. This is very difficult to do with tools which embed assumptions. It is easier to do with a programming language, which is far more malleable.'[66] For these reasons, those interested in these types of analysis should perhaps look to R or Python, rather than MALLET or QDA Miner.[67]

While the future is bright for intellectual historians interested in the digital humanities, we need to take a more active role in its development. There are certainly tools which would be of value to intellectual historians, but do not yet exist because we have not created them (or been involved in their creation). For example, a method for extracting contextual information from a corpus would certainly be appreciated by many intellectual historians, but in other fields such outputs could be considered noise – and thus not a particularly useful avenue of exploration. Again, for such a tool to be developed we need to be involved in their development. As the founder of the digital humanities noted: 'the use of computers in the humanities has as its principal aim the enhancement of the quality, depth and extension of research and not merely the lessening of human effort and time.'[68] Thus, the way forward for intellectual historians interested in these ideas may be the opposite of what we have been doing: we should be more sceptical of easy-to-access digital primary sources – though we should not abandon them – and we should be more willing to learn about, engage with, and perhaps create, new tools and methods.

## Acknowledgements

---

[65] Kirschenbaum, "Hello Worlds."

[66] Mullen, *Digital History Methods*. See also: Mullen, "The Programming Historian."

[67] For those interested in quantitative text analysis, Ken Benoit's Quanteda package for R is a useful starting point.

[68] Busa, "Annals of Humanities Computing," 89.

# Bibliography

Allen, Timothy, Charles Cooney, Stéphane Douard, Russell Horton, Robert Morrissey, Mark Olsen, Glenn Roe, and Robert Voyer. "Plundering Philosophers: Identifying Sources of the *Encyclopédie*." *History and Computing* 13, no. 1 (2010). http://hdl.handle.net/2027/spo.3310410.0013.107.

Armitage, David. "What's the Big Idea? Intellectual History and the *Longue Durée*." *History of European Ideas* 38, no. 4 (2012): 493-507. doi:10.1080/01916599.2012.714635.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks." Paper presented at the International AAAI Conference on Weblogs and Social Media, San Jose, California, May 17-19, 2009.

Benoit, Ken. "quanteda: Quantitative Analysis of Textual Data." Github. https://github.com/kbenoit/quanteda.

Berry, David M.. "The Computational Turn: Thinking About the Digital Humanities." *Culture Machine* 12 (2011): 1-22. http://www.culturemachine.net/index.php/cm/article/view/440/470.

Betti, Arianna and Hein van den Berg. "Modelling the History of Ideas." *British Journal for the History of Philosophy* 22, no. 4: 812-835 (2014). doi:10.1080/09608788.2014.949217.

Bevir, Mark. *The Logic of the History of Ideas*. Cambridge: Cambridge University Press, 2002.

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

Bullard, Paddy. "Digital Humanities and Electronic Resources in the Long Eighteenth Century." *Literature Compass* 10, no 10 (2013): 748–760. doi:10.1111/lic3.12085.

Busa, Roberto. "The Annals of Humanities Computing: The *Index Thomisticus*." *Computers and the Humanities* 14, no. 2 (1980): 83-90. http://www.jstor.org/stable/30207304.

Champion, Erik Malcolm. "Digital humanities is text heavy, visualization light, and simulation poor." *Digital Scholarship in the Humanities* (2016). doi:10.1093/llc/fqw053

*Considering the Scholarly Edition in the Digital Age: A White Paper of the Modern Language Association's Committee on Scholarly Editions*. Modern Language Association, 2015. https://scholarlyeditions.commons.mla.org/2015/09/02/cse-white-paper/.

Crane, Gregory. "What Do You Do with a Million Books?" *D-Lib Magazine* 10, no. 3 (2006). doi:10.1045/march2006-crane.

Crymble, Adam. "Review of Paper Machines, produced by Chris Johnson-Roberson and Jo Guldi." *Journal of Digital Humanities* 2, no. 1 (2012). http://journalofdigitalhumanities.org/2-1/review-papermachines-by-adam-crymble/.

Dempsey, Joe, Daniel Hargreaves, Daniel Peacock, Chelsea Lindsey, Dominic Bell, Luc Fontenoy, and Heather Williams. "Pudding Lane: Recreating Seventeenth-Century London." *Journal of Digital Humanities* 3, no. 1 (2014). http://journalofdigitalhumanities.org/3-1/pudding-lane-recreating-seventeenth-century-london/.

*The Digital Humanities Manifesto 2.0*. http://manifesto.humanities.ucla.edu/.

Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press, 2014.

Edelstein, Dan. "Intellectual History and Digital Humanities." *Modern Intellectual History* 13, no. 1 (2016): 237-246. doi:10.1017/S1479244314000833.

Evrigenis, Ioannis D.. "Digital Tools and the History of Political Thought: The Case of Jean Bodin." *Redescriptions: Political Thought, Conceptual History and Feminist Theory* 18, no. 2 (2015): 181-201. doi:10.7227/R.18.2.4.

Findlen, Paula, Dan Edelstein, and Nicole Coleman. "Mapping the Republic of Letters." Stanford University. http://republicofletters.stanford.edu/.

Gatto, Malu A. C.. *Making Research Useful: Current Challenges and Good Practices in Data Visualisation*. University of Oxford Reuters Institute for the Study of Journalism, 2015. http://reutersinstitute.politics.ox.ac.uk/publication/making-research-useful.

Goldberg, Yoav, and Omer Levy. "word2vec Explained: Deriving Mikolov et al.'s Negative-sampling Word-embedding Method." *arXiv* 2014. http://arxiv.org/abs/1402.3722.

Gordon, Peter. "Contextualism and Criticism in the History of Ideas." In *Rethinking Modern European Intellectual History: Reappraisals and New Perspectives for the Twenty-First Century*, edited by Darrin M. McMahon and Samuel Moyn. New York: Oxford University Press, 2013. doi:10.1093/acprof:oso/9780199769230.003.0002.

Graham, Shawn, and Ian Milligan. "Review of MALLET, produced by Andrew Kachites McCallum." *Journal of Digital Humanities* 2, no. 1 (2012). http://journalofdigitalhumanities.org/2-1/review-mallet-by-ian-milligan-and-shawn-graham/.

Grimmer, Justin and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21, no 3 (2013): 267-297. doi:10.1093/pan/mps028.

Guldi, Jo, and David Armitage. *The History Manifesto*. Cambridge: Cambridge University Press, 2014. doi:10.1017/9781139923880.

Habermas, Jürgen. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger. Cambridge, MA: MIT Press, 1992.

Herrera, Yoshiko M. and Bear F. Braumoeller. "Symposium: Discourse and Content Analysis." *Qualitative Methods* 1, no. 2 (2004): 15-19. https://www.maxwell.syr.edu/uploadedFiles/moynihan/cqrm/Newsletter2.2.pdf.

Hockey, Susan. "The History of Humanities Computing." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004. http://www.digitalhumanities.org/companion/.

D'Ignazio, Catherine. *What Would Feminist Data Visualization Look Like?* (blog). *MIT Center for Civic Media*. https://civic.mit.edu/feminist-data-visualization.

Israel, Jonathan. "Review of Peter de Bolla, *The Architecture of Concepts*." *Critical Inquiry* 41, no. 3 (2015): 709-710. doi:10.1086/680201.

Jones, H. S.. (2010). "Towards a Genealogy of Intellectual History." Lecture, University of Manchester, Manchester, March 24, 2010. https://www.escholar.manchester.ac.uk/jrul/item/?pid=uk-ac-man-scw:78554.

Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Englewood Cliffs, New Jersey: Prentice-Hall, 2009.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans. "Authorship Verification with the Minmax Metric." Paper presented at Digital Humanities Benelux, Luxembourg, June 9-10, 2016.

Kirschenbaum, Matthew. "Hellow Worlds." *The Chronicle of Higher Education* 55, no. 20 (2009). http://chronicle.com/article/Hello-Worlds/5476.

Kinsley, William. "Meaning and Format: Mr. Spectator and His Folio Half-Sheets." *ELH* 34, no. 4 (1967): 482-494. doi:10.2307/2872181

Knowles, Anne K.. "Why We Must Make Maps: Historical Geography as a Visual Craft." *Historical Geography* 42 (2014): 3-26.

Latour, Bruno. "Visualisation and Cognition: Drawing Things Together." In Vol. 6 of *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, edited by H. Kuklick. Greenwich, Connecticut: JAI Press, 1986. http://www.bruno-latour.fr/node/293.

Laver, Michael, Kenneth Benoit, and John Garry. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97, no. 2 (2003): 311-331. doi:10.1017.S0003055403000698.

Leca-Tsiomis, Marie. "The Use and Abuse of the Digital Humanities in the History of Ideas: How to Study the *Encyclopédie*." *History of European Ideas*, 39, no. 4 (2013): 467-476. doi:10.1080/01916599.2013.774115.

Lewis, C. S.. "Addison." In *Selected Literary Essays.* Cambridge: Cambridge University Press, 1969.

Liu, Chao-Lin, Guan-Tao Jin, Hongsu Wang, Qing-Feng Liu, Wen-Huei Cheng, Wei-Yun Chiu, Richard Tzong-Han Tsai, and Yu-Chun Wang. "Textual Analysis for Studying Chinese Historical Documents and Literary Novels." *arXiv* 2014. doi:10.1145/2818869.2818912.

London, Jennifer A.. "Re-imagining the Cambridge School in the Age of Digital Humanities." *Annual Review of Political Science* 19 (2016): 351-373. doi:10.1146/annurev-polisci-061513-115924.

Long, Hoyt and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." *Critical Inquiry* 42, no. 2 (2016): 235-267. doi:10.1086/684353.

von Lünen, Alexander. "So long, and thanks for the GIS: Digital Spatial History." In *Spatial Cultures: Towards a New Social Morphology of Cities Past and Present*, edited by S. Griffiths, S. and A. von Lünen. London: Routledge, 2016.

McLuhan, Marshall. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill, 1964.

Meeks, Elijah, and Scott B. Weingart. "The Digital Humanities Contribution to Topic Modeling." *Journal of Digital Humanities* 2, no. 1 (2012): 2-6. http://journalofdigitalhumanities.org/2-1/.

Moretti, Franco. *Distant Reading*. London: Verso Books, 2013.

Morton, A. Q.. *The Authorship of the Pauline Epistles: A Scientific Solution*. Saskatoon: University of Saskatchewan, 1965.

Mosteller, Frederick, and David Lee Wallace. *Inference and disputed authorship: The Federalist*. Reading, Massachusetts: Addison-Wesley, 1964.

Mullen, Lincoln. *Digital History Methods in R*. http://lincolnmullen.com/projects/dh-r/introduction.html.

Mullen, Lincoln. "The Programming Historian." *Journal of American History* 103, no. 1: 299-301. doi:10.1093/jahist/jaw172.

Mussell, Jim. "Doing and Making: History as Digital Practice." In *History in the Digital Age*, edited by Toni Weller. Oxford: Routledge, 2013.

Nowviskie, Bethany. "On the Origin of 'Hack' and 'Yack'." *Debates in the Digital Humanities 2016*: 66-70. http://dhdebates.gc.cuny.edu/debates/text/58.

Price, Kenneth M.. "Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?" *Digital Humanities Quarterly* 3, no. 3 (2009). http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html.

Price, Kenneth M.. "Electronic Scholarly Editions." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004. http://www.digitalhumanities.org/companion/.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54, no. 1 (2010): 209-228. doi:10.1111/j.1540-5907.2009.00427.x.

Robertson, Stephen. "The Differences between Digital Humanities and Digital History." In *Debates in Digital Humanities 2016*, edited by Matthew K. Gold and Lauren F. Klein. Minnesota, University of Minnesota Press, 2016. http://dhdebates.gc.cuny.edu/debates/text/76.

Slapin, Jonathan B., and Sven-Oliver Proksch. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52, no. 3 (2008): 705-722. doi:10.1111/j.1540-5907.2008.00338.x.

Skinner, Quentin. *Visions of Politics: Regarding Method*. Cambridge: Cambridge University Press, 2002.

Spedding, Patrick. "'The New Machine': Discovering the Limits of ECCO." *Eighteenth-Century Studies* 44, no. 4 (2011): 437–53. doi:10.1353/ecs.2011.0030.

Stone, Maureen. *Information Visualization: Challenge for the Humanities*. Council on Library and Information Resources, 2009. https://www.clir.org/pubs/resources/promoting-digital-scholarship-ii-clir-neh/stone11_11.pdf.

Terras, M., J. Nyhan, and E. Vanhoutte. *Defining Digital Humanities: A Reader*. Farnham: Ashgate, 2013.

Thornton, Tamara Plakins. *Handwriting in America: A Cultural History*. New Haven: Yale University Press, 1996.

Warren, Christopher N., Daniel Shore, Jessica Otis, Lawrence Wang, Mike Finegold, and Cosma Shalizi. "Six Degrees of Francis Bacon: A Statistical Method for Reconstructing Large Historical Social Networks." *Digital Humanities Quarterly* 10, no. 3 (2016). http://www.digitalhumanities.org/dhq/vol/10/3/000244/000244.html.

Weingart, Scott B.. "Demystifying Networks, Parts I & II." *Journal of Digital Humanities* 1, no. 1 (2011): 9-21. http://journalofdigitalhumanities.org/1-1/.