Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research

Gregory, Ian; Donaldson, Christopher; Murrieta-flores, Patricia; Rayson, Paul

[Link to publication on Research at Birmingham portal](#)

GEOPARSING, GIS, AND TEXTUAL ANALYSIS: CURRENT DEVELOPMENTS IN SPATIAL HUMANITIES RESEARCH

Ian Gregory, Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson

Introduction

The spatial humanities constitute a rapidly developing field that has the potential to revolutionize humanities scholarship. As yet, however, research in the spatial humanities has only begun to deliver the applied contributions to knowledge that will prove its significance. Demonstrating the potential of innovations in technical fields is, almost always, a lengthy process, as it takes time to create the required datasets and to design and implement appropriate techniques for engaging with the information those datasets contain. Beyond this, there is the need to define appropriate research questions and to set parameters for interpreting findings, both of which can involve prolonged discussion and debate. The spatial humanities are still in early phases of this process. Accordingly, the purpose of this special issue is to showcase a set of exemplary studies and research projects that not only demonstrate the field's potential to contribute to knowledge across a range of humanities disciplines, but also to suggest pathways for future research. Our ambition, at this level, is both to demonstrate how the application of exploratory techniques in the spatial humanities offers new insights about the geographies embedded in a diverse range of texts (including letters, works of literature, and official reports) and, at the same time, to encourage other scholars to integrate these techniques in their research.

To date, a standard definition of the spatial humanities has yet to be determined (a fact which, no doubt, has much to do with the pioneering nature of the field); but Richardson et al's 2011 description of GeoHumanities as a 'rapidly growing zone of creative interaction between geography and the humanities' is an excellent place to start.[1] Although there is nothing explicitly digital about this definition, such 'creative interaction' certainly owes much to the widespread application of geographical technologies within the digital humanities.[2] In this way, one can begin to trace the relation between GeoHumanities and what Gregory and Geddes describe when referring to the spatial humanities as 'a field' that employs 'geographical technologies to develop new knowledge about the geographies of human cultures past and present.'[3] For Gregory and Geddes, the spatial humanities has its origins in historical geographical information systems (HGIS), from which it has developed both on account of technological advances in geographic information science (GISc) and on account

of the increased acceptance of digital technologies as tools for knowledge creation across a range of humanities disciplines. In their seminal collection, *The Spatial Humanities*, Bodenhamer et al advance a similar conception of the field, presenting the spatial humanities both through critical engagements with specific spatial technologies and through more general evaluations of the benefits these technologies can bring.[4] Implicit in each of these accounts is the notion that the spatial humanities has come about through the creative adaptation and application of geographical technologies in humanities research. Taking this basic definition of the spatial humanities as a starting point, this issue draws attention to important developments that are shaping the field.

1. Digital collections and geographical technologies

Information technology is currently advancing in ways that make both non-quantitative source materials and geographical technologies increasingly available and accessible. Whereas geographical technologies were once seen as primarily, if not exclusively, suited to processing quantitative data, they are now commonly used to work with information from a wide range of formats, including unstructured texts. Significantly, this development has occurred in tandem with a proliferation of the digital textual material available to researchers, including digital collections of historical books and reports such as the Old Bailey Online, Early English Books Online, and the British Library's Nineteenth-Century Newspaper collection,[5] which comprise several million – and, increasingly, several billion – words. The digitisation of these corpora (as large collections of digital texts are called) was principally driven by the desire to make them available online with keyword-search facilities. With recent advances in text analysis software, however, it is becoming apparent that researchers can gain new insights about these corpora through more complex forms of analyses. In addition, researchers have also begun to recognize the manifest importance of 'born digital' material, including corpora compiled from websites, email archives, and social media, which can be analysed in many of the same ways.

Alongside these developments, the last several years have also witnessed unprecedented advances in the design and distribution of geographical technologies. Once these were the preserve of geographical information systems (GIS) platforms, such as ArcGIS, which, although powerful, were both expensive and highly specialised, and therefore only accessible to a small number of specialists.[6] Thanks to the increasing number of open-source GIS software packages, however, technologies that were once available only to a select few have

begun to reach a mass audience.[7] Moreover, with the ubiquity of satnavs and virtual globes (such as Google Earth), interest in and familiarity with digital mapping has increased exponentially. Indeed, as Bodenhamer et al. proffer, 'we are more aware than ever of the power of the map to facilitate commerce, enable knowledge discovery, [and] make geographic information visual and socially relevant.'[8]

When we draw these threads together – the proliferation of digital corpora, the increasing availability and accessibility of geographical technologies, and the growing awareness of the potential contributions to knowledge that these technologies can make – the potential of the spatial humanities to advance humanities scholarship is readily apparent. This is especially so when we consider the recent advances made in the development of automated geoparsing.

2. Georeferencing and automated geoparsing

Creating georeferenced databases (where each item of data is assigned geographic coordinates, allowing it to be mapped and spatially analysed) has long been identified as one of the main challenges in implementing GIS and other geospatial technologies in humanities research.[9] Much of the foundational work in quantitative HGIS was completed by projects that built major databases of administrative boundaries. These projects entailed time-consuming research, and frequently millions of dollars in funding, which led to the digitisation of historical administrative boundaries that were linked to databases of statistical tables.[10] The challenge facing humanities researchers today is rather different and, arguably, much more straightforward. In general, the sources used by humanities researchers are unstructured texts in which geographical information is present in the form of specific named-entities such as toponyms or place-names. In order to georeference an unstructured text, one needs to identify the place-names it contains and then to assign each place-name the coordinates that represent its location. The first of these steps can be achieved by implementing Natural Language Processing (NLP) techniques that are capable of automatically recognizing the place-names within a text. Completing the second step entails pairing these place-names with coordinate data from a gazetteer, such as GeoNames or GNIS.[11] This two-step process is known as geoparsing.[12] Geoparsing provides a solution to georeferencing texts; however, the next (and arguably the more important) task is to decide what to do with the fully georeferenced text. The software and the analytic techniques developed for working with georeferenced databases have been developed to handle quantitative sources, usually from scientific or social science paradigms. But how does one

analyse a georeferenced text in ways that are sensitive to the complex nature of humanities sources? Moreover, what contributions to knowledge can we expect from these approaches? This special issue addresses these questions by presenting series of studies and research projects that demonstrate not only the opportunities afforded by geoparsing and working with georeferenced texts, but also the challenges this presents and the implications of these challenges for spatial humanities research.

3. Trends in spatial humanities research

The early days of HGIS were characterised by projects that used quantitative data to create comprehensive models for studying spatial patterns in fields such as historical demography and environmental and economic history. Over the past decade, the potential of these approaches to make contributions to knowledge has become increasingly apparent,[13] and scholars in disciplines across the humanities and social sciences have begun to incorporate GIS, and cognate geospatial technologies, in their research. In what follows, we offer a brief survey of these developments, beginning with archaeology and history and then moving on to consider literary studies and, finally two closely related areas, corpus and computational linguistics.

3.1. Archaeology and history

Outside of geography, archaeology was the first discipline to integrate geospatial technologies in its methods and to apply those technologies in its research. That this was the case is largely because the study of the material past is inherently spatial in nature. In order to make sense of artefacts and sites, archaeologists need to understand their spatial contexts. They need, in other words, to determine how specific objects, features, and structures relate not only to the places where they are found, but also to the wider landscapes those places comprise. It is only by understanding the spaces inhabited by past cultures that archaeologists are able to reconstruct the customs, beliefs, and institutions that defined those cultures. At this rudimentary level, all forms of archaeology can be characterised by a preoccupation with spatial thinking. Given this, it comes as little surprise that the methodologies and theories that have shaped the discipline over the past fifty years have continued to emphasize the spatial dimension of archaeological practice. The emergence of 'spatial archaeology' during the 1970s can be seen as paradigmatic in this respect; for, although not uncontroversial, its influence is apparent not only in the widespread use of spatial analysis for studying excavated artefacts to landscape compositions, but also in recent trends in archaeological computing.[14]

Archaeologists first began to utilise GIS during the 1980s.[15] Since that time, other spatial technologies such as Remote Sensing and GPS have also been incorporated in almost every branch of the discipline.[16] By contrast, the spatial study of textual sources in archaeology is a much more recent phenomenon. The reasons for this are varied; but, for the present purposes, it suffices to say that the use of textual sources in archaeology is different than in most humanities disciplines. This is one of the reasons why digital archaeology has often seemed to have such a different 'scene' during the emergence of digital humanities.[17] Indeed, pioneering projects have only recently begun to harness the potential of spatial technologies to investigate text corpora relevant to archaeological research. Some of these projects have focused on the application of techniques from corpus and computational linguistics to mine grey literature reports, extracting potential spatial and contextual information for archaeological interpretation.[18] Others projects have focused on the exploration of textual sources through using enhanced visualisations. For example, HESTIA[19] and Google Ancient Places[20] have developed Web-based visual resources to facilitate the exploration of places of archaeological interest mentioned in ancient literature.[21]

These projects stand at the forefront of textual research in digital archaeology, and they are leading the way in providing both scholars and the general public with the means to envisage texts geographically. Although this work provides a foundation for assessing the geographies underlying such texts, the methodologies that can go beyond data exploration and enter the realm of corpus analysis have yet to materialise in archaeology. We think, however, that this will not take long. With the combination of the experience in spatial methods and thinking from archaeology, and the diverse approaches developed in history, corpus and computational linguistics, we expect to see advanced forms of textual spatial analysis in archaeology in the near future.

In the case of history, the application of geospatial methods is a more recent phenomenon. Here, as was the case in archaeology, the integration of GIS came about largely in response to the need to find methodologies to answer specific research questions in the wider context of the discipline. HGIS, the result of these endeavours, has over the past fifteen years become a diverse and dynamic subfield. Although, as noted above, HGIS projects initially concentrated on the quantitative exploration of economic and political data,[22] more recent studies have focused on the application of GIS in the qualitative analysis of historical documents.[23] Exploratory HGIS scholarship has also recently undertaken experimental research combining spatial and corpus analysis,[24] and, in some cases, even integrating GIS

with serous gaming engines to facilitate the 3-D virtual modelling of historical places and landscapes.[25] In all of its manifestations, HGIS continues to distinguish itself as an agenda-setting area of spatial humanities scholarship that, in testing innovative interdisciplinary approaches, is still blazing new pathways for research.

3.2. Literary studies

A number of ground-breaking research projects have also recently identified the transformative potential of GIS, and related technologies, for the discipline of literary studies. Led by research teams at centres of excellence in Britain, Europe, the United States, and Australia, these projects have proven that GIS and its cognates have the power to revolutionise how we interpret the material, imaginative, and discursive geographies not only of individual novels, poems, and plays, but also of large corpora of literary works. In doing so, these projects have helped reinvigorate both literary geography (the study of the spatiality of literary works) and, more generally, what might be called the geography of literature (the study of the place-bound nature of the acts of writing, publishing, and reading). Even more remarkably, they have also suggested the potential of wholly new modes and practices for literary scholarship.

A key development here has been the emergence of digital literary atlas projects, such as ETH Zurich's Literary Atlas of Europe, the University of Queensland's Cultural Atlas of Australia, Trinity College Dublin's Digital Literary Atlas of Ireland, and the New University of Lisbon's LITSCAPE.PT (which is featured in this collection). Taking their cue from the pioneering work of Franco Moretti, Matthew Jockers, and the Stanford Literary Lab, the creators of these atlases have embraced the idea that literary critics can use maps as 'analytical tools: that dissect the text in unusual ways, and bring to light relations that would otherwise remain hidden.'[26]

Underlying this methodological premise is a conviction in the value of the map as a form of abstraction that, in reducing the object of study – the literary text or corpus – to a few particulars, both defamiliarizes it and, in the process, helps to generate new research questions and to guide critical inquiry. Equally operative here is a new, contested paradigm for literary hermeneutics, variously called 'distant reading'[27] or 'macroanalysis'[28], which has sought to supplement more traditional critical approaches through the aggregate analysis of large literary corpora. These new practices, as Jockers explains, are poised to take advantage

of 'the massive digital-text collections' available on the World Wide Web and, in the process, to launch literary studies into a new age:

> Today, in the age of digital libraries and large-scale book-digitization projects, the nature of the evidence available to us has changed, radically. Which is not to say that we should no longer read books looking for, or noting, random 'things,' but rather to emphasize that massive digital corpora offer us unprecedented access to the literary record and invite, even demand, a new type of evidence gathering and meaning making.[29]

In other words, instead of simply identifying, isolating, and analysing the features of a handful of 'representative' texts, literary scholars today should strive to create new knowledge about those features and texts by contextualizing them in relation to the large text corpora that are now available to them.

The creation of digital literary atlases, such as the Literary Atlas of Europe and LITSCAPE.PT, have demonstrated how GIS and spatial analysis can assist in facilitating this sort of contextualization. But, even though these projects are successful in their own terms, they have also stimulated important debates about the merits of such macro-mapping activities for literary scholarship. A key issue here is the widely perceived incongruity between the methodologies of GISc – with their reliance on precise, quantifiable data – and the kinds of equivocal or, what Bushell calls, 'slippery' information with which literary scholars typically engage.[30] Put bluntly, tools and techniques designed to measure absolute Euclidean space often prove inadequate for modelling the complex, contingent, and, at times, contradictory geographies of literary works of art.

Consequently, many literary scholars have dismissed the application of GIS, and other digital tools, in the macro-mapping of literary corpora as problematically instrumentalist and reductive, noting that this approach tends to flatten out and suppress the differences that distinguish literary works from one another. Other scholars have been more conciliatory, expressing interest in the results of such projects whilst advising that the value of mapping as a critical practice depends largely on the nature of texts being studied. Notably, Hewitt councils that the analysis of literary works 'is more revealing when sensitivity is shown to the approaches of individual texts and authors,' and that, accordingly, that 'mapping … cannot be the first step in a mass hermeneutic process,' but should come 'after an exploration of … evidence of a work's engagement with spatial concerns.'[31]

Another common criticism of large-scale, literary atlas projects is that they problematically

conflate the real world and the world of the literary text. For, although acknowledging that 'the geography of fiction follows its own distinctive rules, since literature can create its own space, without physical restrictions,' in the end these projects are often simply predicated on the positivist assumption 'that a large part of fiction indeed refers to the physical/real world.'[32] As Bushell reminds us, 'the points where [a literary map] does not correspond directly to the world of the book may be more interesting than the points where it does.'[33]

These are major impediments for the integration of spatial humanities approaches in literary studies. For although it is clear that GIS and distant reading are clearly compatible and can be usefully combined to build spatial models of specific narrative structures, their efficacy in aiding textual study is, as yet, limited. The problem here, as one research team has noted, is both epistemological and technological: both a consequence of differing research cultures and of the limitations of current research tools.[34] In order to overcome these deficits, researchers within the field of the digital humanities – including literary scholars, computer scientists, and GISc specialists – need to work together not only to develop new practices and frameworks for interdisciplinary collaboration and creative exchange, but also to produce substantial works of scholarship that demonstrate how those practices and frameworks contribute to close study and analysis of specific literary texts.

Encouragingly, this is precisely the direction in which much work in the field of literary cartography is moving. Notably, research centres such as the Stanford Literary Lab have continued to pioneer innovative interdisciplinary approaches to the study of literary spaces, most recently by modelling the use of digital crowdsourcing to construct an 'emotional map' of London based on a corpus of eighteenth- and nineteenth-century novels.[35] At Lancaster University, moreover, a team of literary scholars, GIScientists, quantitative historians, corpus and computational linguists is using data mining in conjunction with GIS to investigate literary representations of and responses to the English Lake District.[36] Furthermore, an interdisciplinary team of scholars at the University of Edinburgh is currently using immersive-mapping and mobile-computing technologies to enable users to explore the Edinburgh cityscape through geo-located extracts of literary works from the early modern period to the twentieth century.[37] Alongside these projects, other integrative forms of geospatial technology are informing the development of deep mapping, a newly emerging concept that an increasing body of literature discusses in detail.[38]

3.3 Corpus and computational linguistics

Enabling the geospatial analysis of qualitative texts is crucial to the enterprise of the spatial humanities. At a minimum, such analysis entails geoparsing texts to extract the place-name references they contain and then plotting them on a map to facilitate distant readings. However, with the addition of specialised techniques from the closely related areas of corpus and computational linguistics, spatial humanities researchers can also begin to combine distant reading and close textual analysis.

Computational linguistics or NLP techniques drawn from computer science allow the automatic summarisation of meaning or extraction of a variety of patterns from text. One exemplary NLP technique is named-entity extraction, which can, to a certain level of accuracy, find all mentions of personal names, organisations, place names, dates and times in a text. As demonstrated in this special issue, combining this technique with toponym resolution[39] to locate the extracted place-names on a map forms one of these bi-directional links from analysis to text. Another large class of NLP techniques enable the automatic annotation of words or phrases within a text at various levels of linguistic detail. Part-of-speech (POS) annotation enables highly accurate (97-98%) identification of major word classes in text, such as adjectives, nouns, verbs and adverbs. Once these categories are marked in the corpus, they can be searched alongside the word forms, such as proper nouns, which are particularly useful for identifying place-names and personal names. Adjectives are a good source when searching for evaluative descriptions of landscape features for example. A second level of tagging, called semantic annotation,[40] adds meaning or conceptual labels to words and phrases in a text. This enables searching by concepts, such as health and disease, finance and money, or education within the corpus.

The second family of related techniques stems from corpus linguistics, which is a method or collection of methods for text analysis stemming from the discipline of linguistics. With the increase in power and storage capacity of computers in the 1970s and 1980s, a set of methods emerged which enabled large quantities of machine-readable text to be analysed and explored semi-automatically for language description purposes. Similar developments in the digital humanities can be traced back to Father Roberta Busa's working with IBM to produce his computer-generated *Index Thomisticus* of the writings of Thomas Aquinas. In parallel, dictionary publication was revolutionised in the 1980s with the creation of machine-readable corpora such as COBUILD alongside new searching and analysis software. At least five corpus linguistics methods are worthy of mention here. In combination, they provide a semi-automatic approach to data-driven exploratory analysis that can uncover patterns within the

data that are otherwise difficult or impossible to extract by more manual analyses. First, frequency lists show all the different word types in a text, and how often they occur, allowing the researcher to focus their efforts on the most represented features in a corpus. Frequency lists can also be extended to show how well a word is dispersed within a corpus and this is key to understanding the salience of a word in a corpus. Second, concordances show every occurrence of a word in a text with a small amount of context, usually four to five words on either side. This enables the researcher to look for patterns and meanings by sorting the surrounding text. Third, the keywords method compares two or more frequency lists to identify words which are statistically more represented in one text relative to another sub-corpus or a much larger reference corpus. This can show what a text is about and highlight interesting terms for further analysis. Fourth, n-grams (sometimes called lexical bundles or clusters) shows repeated consecutive sequences of words of a given length (n) which extends the single word frequency lists. Finally, the collocation technique assists in finding which words regularly co-occur in close proximity in texts. In the context of the spatial humanities, the collocation technique has proven useful for discovering which topics are discussed in relation to different places that are mentioned in a corpus. In addition, the combination of keywords methods, semantic analysis and collocation means that it is possible to uncover and visualise the topics associated with particular place-names by connecting GIS databases to their semantic collocates, so called 'visual GISting'.[41]

4. The essays in this issue

The contributors to this special issue draw attention to how extracting the geographies embedded within texts can contribute to research in the humanities. The first essay, by Alex et al, is principally concerned with geoparsing, and with how the automated extraction of geographic information can aid the analysis of large text corpora. The main focus here is the Edinburgh Geoparser, a state of the art Web-based tool, which has been adapted to facilitate the georeferencing of historical texts. In order to illustrate the power and flexibility of the Edinburgh Geoparser, Alex et al present three brief, but contrasting, case studies: one concerned with nineteenth-century trade, another concerned with the ancient world, and a third concerned with a historical gazetteer of English place-names.

The four essays that follow Alex et al move forward from geoparsing to consider how working with georeferenced corpora can inform humanities research. In the first, Schwartz engages with a field that is near the traditional heartland of HGIS: environmental history.

Drawing on georeferenced texts from the British Parliamentary Papers, his chapter examines nineteenth-century reports on fish stocks in British waters. In order to do this, Schwartz combines computer-assisted qualitative data analysis (CAQDAS) methods with GIS to assess changing perceptions about the decline of fish stocks by comparing texts from Royal Commissions in 1863 and 1893. Although Schwartz is keen to stress that CAQDAS and GIS are reductionist approaches, his essay convincingly demonstrates how they can aid more traditional forms of historical analysis by generating research questions and guiding critical inquiry.

Alves and Queiroz's essay, which focuses on the application of geospatial tools in literary studies, also suggests how GIS-based distant reading can complement more traditional close reading practices. Here the focus is on LITSCAPE.PT, a digital literary atlas of historical and modern Portuguese literature. Rather than using geoparsing software, as Alves and Queiroz explain, LITSCAPE.PT relies on a creative combination of crowdsourcing and relational databases to facilitate the mapping and analysis of excerpts from a wide variety of literary works. Using this mixed-methods approach, the project has successfully catalogued more than 6,000 excerpts, which are being used to study literary representations of mainland Portugal as well as social and environmental history. As examples, Alves and Queiroz present two case studies: a comparative examination of the evolving physical and literary geographies of Lisbon and a transhistorical assessment of the declining presence of wolves in Portuguese literature.

The fourth essay, by Purves and Derungs, also addresses the representation of landscapes in writing; however, they are more concerned with the way that texts represent landscape than what those representations reveal about the texts themselves. As Purves and Derungs argue, engaging with written representations of space can help human geographers move beyond the GIS-facilitated measurement of Euclidean space towards a more nuanced conception of place. As proof-of-concept, Purves and Derungs focus on two apparently contrasting corpora related to mountain landscapes in Switzerland and Britain: the Text+Berg archive of the *Yearbooks* of the Swiss Alpine Club and georeferenced photographs from Flickr. Although the *Yearbooks* (a historical corpus documenting 150 years of Alpine mountaineering) seem like a typical resource to draw on in such an analysis, Flickr (a web 2.0 site that allows users to upload photographs) is a much less obvious choice. Yet, as Purves and Derungs explain, the metadata added to Flickr photos make them an equally excellent resource for understanding how places are experienced and perceived.

Van den Heuvel's essay, which concludes this special issue, also emphasizes the need to move beyond a purely Euclidean conception of space, which, he contends, is inadequate for modelling the sorts of networks and systems of knowledge exchange in which scholars in the humanities are often interested. As his example, van den Heuvel concentrates on the Republic of Letters, suggesting that spatial humanities approaches can aid us in reconstructing the geographical distribution of documents and drawings that defined this epistolary intellectual community. Taking inspiration from the notion of 'deep maps', van den Heuvel, concludes by positing the idea of 'deep networks' as a means for visualising how knowledge was communicated and disseminated in early modern Europe.

Taken together, these essays offer a representative sample of the sorts of projects currently being pursued within the spatial humanities. They call attention to new developments and present different perspectives on the challenges and the potentials of research in the field. In doing so, they affirm that the creative adaption and application of geographical technologies has the potential to revolutionize scholarship across a number of different humanities disciplines. Our hope is that, in modelling the use of innovative methods they will encourage other scholars to integrate these approaches in their research.

Acknowledgements

End Notes

[1] D. Richardson, S. Luria, J. Ketchum and M. Dear, 'Introducing the geohumanities', in M. Dear, J. Ketchum, S. Luria and D. Richardson, eds., *GeoHumanities: Art, history, text at the edge of place* (Abingdon, 2011), 3-4. Cited here at 3.

[2] See, for instance, K. Offen 'Historical geography II: Digital imaginations', *Progress in Human Geography*, 37, no. 4 (2013), 564-77.

[3] I.N. Gregory and A. Geddes, 'Introduction: From Historical GIS to Spatial Humanities: Deeping scholarship and broadening technology', in I.N. Gregory and A. Geddes, eds., *Towards Spatial Humanities: Historical GIS and Spatial History* (Bloomington, IN, 2014), ix-xix. Cited here at xv.

[4] See: D.J. Bodenhamer, J. Corrigan, and T.M. Harris, 'Introduction', in D.J. Bodenhamer, J. Corrigan and T.M. Harris, eds., *The Spatial Humanities: GIS and the future of humanities scholarship* (Bloomington, 2010), vii-xv; and D.J. Bodenhamer, 'The potential of spatial humanities', in D.J. Bodenhamer, J. Corrigan, and T.M. Harris, eds., *The Spatial Humanities: GIS and the future of humanities scholarship* (Bloomington, 2010),14-30.

[5] See: *Early English Books Online*, http://eebo.chadwyck.com/home, last accessed 5 Aug 2014; *British Newspapers, 1600-1950*, http://gale.cengage.co.uk/product-highlights/history/19th-century-british-library-newspapers.aspx, last accessed 5 Aug 2014; *The Proceedings of the Old Bailey – London's Central Criminal Court, 1674 to 1913*, http://www.oldbaileyonline.org, last accessed 5 Aug 2014.

[6] See *ArcGIS*, <http://www.esri.com/software/arcgis> [accessed 5 Aug 2014].

[7] See: *Google Earth*, <http://earth.google.com> [accessed 5 Aug 2014]. Examples of free and open source GIS software include: *QGIS: A free and open source geographic information system*, http://www.qgis.org, last accessed 5 Aug 2014; and *MapWindow*, http://www.mapwindow.org, last accessed 5 Aug 2014.

[8] Bodenhamer et al. (2010), vii.

[9] I.N. Gregory and P.S. Ell, *Historical GIS: Technologies, methodologies, scholarship* (Cambridge, 2007), Chap 3.

[10] A.K. Knowles, ed. 'Reports on National Historical GIS projects' *Historical Geography*, 33 (2005), 293-314 provides a review.

[11] *GeoNames*, <http://www.geonames.org> [accessed 5 Aug 2014]; *GNIS*, <http://nhd.usgs.gov/gnis.html> [accessed 5 Aug 2014].

[12] C. Grover, R. Tobin, K. Byrne, M. Woollard, J. Reid, S. Dunn and J. Ball, 'Use of the Edinburgh geoparser for georeferencing digitized historical collections', *Philosophical Transactions of the Royal Society A*, 368 (2010), 3875-3889.

[13] I.N. Gregory, 'Further reading: From historical GIS to spatial humanities: An evolving literature' in I.N. Gregory and A. Geddes, eds., *Towards Spatial Humanities: Historical GIS and Spatial History* (Bloomington, IN, 2014), 186-202; A.K. Knowles, ed., *Placing History: How maps, spatial data, and GIS are changing historical scholarship* (Redlands, CA, 2008).

[14] J. Huggett, 'What Lies Beneath: Lifting the Lid on Archaeological Computing' in A. Chrysanthi, P. Murrieta Flores, and C. Papadopoulos, eds., *Thinking Beyond the Tool: Archaeological Computing and the Interpretative Process* (Oxford, 2012) 204–214.

[15] M. Aldenderfer, H. Maschner, and M. Goodchild, *Anthropology, space, and geographic information systems* (New York, 1996); A.S. Fotheringham, C. Brunsdon, M. and Charlton, *Quantitative geography: perspectives on spatial data analysis* (Thousand Oaks, CA, 2000); and D. Wheatley and M. Gillings, *Spatial Technology and Archaeology. The Archaeological Applications of GIS* (London, 2002).

[16] R.N. Parker and E.K. Asencio, *GIS and spatial analysis for the social sciences: coding, mapping, and modelling* (New York, 2008); and M.F. Goodchild and D.G. Janelle, 'Toward critical spatial thinking in the social sciences and humanities', *GeoJournal* 75, no. 1 (2010), 3-13.

[17] J. Huggett, 'Core or periphery? Digital Humanities from an archaeological perspective', *Historical Social Research,* 37, no. 3 (2012) 86-105.

[18] J. Richards, S. Jeffrey, S. Waller, F. Ciravegna, S. Chapman, and Z. Zhang, 'The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing', in E. C. Kansa, S. Whitcher Kansa, & E. Watrall, eds., *Archaeology 2.0. New Approaches to Communication and Collaboration* (Los Angeles, 2011), 31–56.

[19] E. Barker, S. Bouzarovski, C. Pelling, and L. Isaksen, 'Mapping an ancient historian in a digital age: the Herodotus Encoded Space-Text-Image Archive (HESTIA)', *Leeds International Classical Studies*, 9, no. 1 (2010), 1-24.

[20] *Google Ancient Places,* <http://googleancientplaces.wordpress.com/> [accessed 5 Aug 2014].

[21] E. Barker, K. Byrne, L. Isaksen, E. Kansa, and N. Rabinowitz, Google Ancient Places, <http://googleancientplaces.wordpress.com/2012/02/25/ the-story-continues/> [accessed 5 Aug 2014].

[22] J. A. Galloway, 'Reconstructing London's distributive trade in the later middle ages: the role of computer-assisted mapping and analysis', in M. Woollard, ed., *New Windows on London's Past: Information Technology and the Transformation of Metropolitan History* (Glasgow, 2000), 1-24; I.N. Gregory, 'Longitudinal analysis of age- and gender- specific migration patterns in England and Wales: a GIS-based approach', *Social Science History* 24 (2000), 471-503; *Vision of Britain through Time*, <http://www.visionofbritain.org.uk/> [accessed 5 Aug 2014].

[23] B. Donahue, 'Mapping Husbandry in Concord: GIS as a Tool for Environmental History' in A. K. Knowles and A. Hillier, eds., *Placing history: how maps, spatial data, and GIS are changing historical scholarship* (Redlands, CA, 2008), 151-77; A. K. Knowles, W. Roush, C. Abshere, L. Farrell, A. Feinberg, and T. Humber, 'What could Lee see at Gettysburg', in A. K. Knowles and A. Hillier, eds., (2008), 235-66; R.M. Schwartz, I. Gregory, and J. Marti-Henneberg, 'History and GIS: railways, population change, and agricultural development in late nineteenth century Wales', in M. Dear, et al. (2011), 251-266.

[24] I. Gregory and A. Hardie 'Visual GISting: bringing together corpus linguistics and Geographical Information Systems', *Literary and Linguistic Computing* 26, no. 3 (2011), 297-314; P. Murrieta-Flores, A. Baron, I. Gregory, A. Hardie, and P. Rayson, 'Automatically analysing large texts in a GIS environment: the Registrar General's reports and cholera in the nineteenth century', *Transactions in GIS* (2014).

[25] See, for example, *UCLA RomeLab*, <http://romelab.etc.ucla.edu/> [accessed 5 Aug 2014]; *Virtual St Paul's Cross Project*, <http://vpcp.chass.ncsu.edu/> [accessed 5 Aug 2014]; and T.M. Harris, L.J. Rouse, and S. Bergeron, 'Humanities GIS: Adding place, Spatial storytelling and Immersive visualization into the Humanities', in M. Dear, et al. (2011), 226-240.

[26] F. Moretti, *Atlas of the European Novel, 1800-1900* (London, 1998), 3.

[27] F. Moretti, 'Conjectures on World Literature', *New Left Review*, 1 (2000), 54–68.

[28] M. L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Illinois, 2013).

[29] Jockers (2013), 7-8.

[30] S. Bushell, 'The slipperiness of literary maps: Critical cartography and literary cartography', *Cartographica*, 47, no. 3 (2012), 149-60.

[31] R. Hewitt, 'Mapping and Romanticism', *Wordsworth Circle*, 42, no. 2 (2011), 157-65.

[32] B. Piatti and L. Hurni. 'Editorial: The cartographies of fictional worlds', *The Cartographic Journal*, 48, no. 4 (2011), 218-23: 218-19.

[33] Bushell (2012), 154.

[34] D.J. Bodenhamer, T.M. Harris, and J. Corrigan, 'Spatial Narratives and Deep Maps: A Special Report', *International Journal of Humanities and Arts Computing*, 7, nos. 1-2 (2013), 170-75.

[35] See *Stanford Literary Lab*, <http://litlab.stanford.edu/?page_id=13> [accessed 5 Aug 2014].

[36] See D. Cooper and I. Gregory, 'Mapping the Lakes District: A Literary GIS', *Transactions of the Institute of British Geographers*, 36, no. 1 (2011), 89-108; see also *Spatial Humanities: Texts, GIS, Places*, <http://www.lancaster.ac.uk/fass/projects/spatialhum.wordpress/> [accessed 5 Aug 2014].

[37] See *Palimpsest*, <http://palimpsest.blogs.edina.ac.uk/> [accessed 5 Aug 2014].

[38] See D.J. Bodenhamer, T.M. Harris, and J. Corrigan, *Deep Maps and Spatial Narratives* (Bloomington: Indiana University Press, forthcoming 2014).

[39] See J. L. Leidner. 'Toponym resolution in text: annotation, evaluation and applications of spatial grounding'. *SIGIR Forum* 41, 2 (2007), 124-126.

[40] Rayson, P., Archer, D., Piao, S. L., McEnery, T. 'The UCREL semantic analysis system'. In *proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004),* 25th May 2004, Lisbon, Portugal, (2004) pp. 7-12.

[41] Gregory and Hardie (2011).