Space and Time in the Republic of Letters: A working paper

Ian Gregory, Patricia Murrieta-Flores and Catherine Porter

The *Republic of Letters* is made up of the surviving pan-European correspondence that developed as a result of the development of the postal network between 1500-1800. The correspondence allowed men and women to create a knowledge-based civil society that helped lead to the era's intellectual breakthroughs and contributed to the development of many modern European values and institutions. However, where this correspondence survives, it is fragmented through different collections across Europe and beyond. These collections are increasingly available in digital form, but these are again fragmented geographically and institutionally. The standards and comprehensiveness of what has been digitised and to what standards also varies widely. The aim of the *Reassembling the Republic of Letters* action is to identify how this material can be brought together using modern digital technologies to enable new scholarly methods and research questions.

This working paper draws on the discussions that took place at a meeting held at Lancaster University on 4th December 2015 under the auspices of Working Group 1 (WG1) "Space and Time" of COST action IS1310 "Reassembling the Republic of Letters, 1500-1800." It aims to explore exemplar projects that exploit space and time in the humanities, to explore the challenges of exploiting spatial and temporal information that is currently in the *Republic of Letters* corpus, and to identify the opportunities and potential strategies for exploiting spatio-temporal information within the corpora that make up the *Republic of Letters*.

1. Georeferencing sources in the humanities

Although technologies such as geographical information systems (GIS) are well established in the social sciences, their use in the humanities has, to date, been much more limited. This ultimately is due to the nature of sources and data. The social sciences work in a (relatively) data-rich world where large amounts of data, typically in tabular form, can be linked to precisely defined spatial units such as counties or districts. The humanities, by contrast, tend to use unstructured texts or their associated metadata in which locational information is given by place-names that refer to places that are frequently settlements or regions whose locations are less precisely defined than modern administrative units and may, in some cases, have been lost or changed considerably over time. Recently, there has been an upsurge in interest in how these types of sources can be brought into GIS and other geographical technologies, and in the opportunities that this opens up to ask new research questions of the past. There are a number of websites that provide lists of humanities GIS, spatial humanities or geo-humanities projects (the terms refer to broadly similar things) and a number of edited volumes that explore the field and the projects that make it up.[1]

[1] For lists of projects in these fields see, for example: http://anterotesis.com/wordpress/mapping-resources/dh-gis-projects; http://www.geohumanities.org/gis; http://spatial.scholarslab.org/project; http://www.hgis.org.uk/resources.htm. For reviews of the field and exemplar projects see: Bodenhamer D.J., Corrigan J. and Harris T.M. (2015, eds) *Deep Maps and Spatial Narratives.* Indiana University Press: Bloomington; Bodenhamer D.J., Corrigan J. and Harris T.M. (2010, eds) *The Spatial Humanities: GIS and the*

Ultimately GIS and other mapping technologies require coordinate-based locations to provide them with the spatial information that allow features to be mapped. Associating datasets with relevant coordinates is usually referred to as *georeferencing*. A major challenge for the humanities is how to georeference place-names. This is usually achieved using a gazetteer which, in its simplest form, is a database table that lists place-names alongside the coordinates for that place usually as latitudes and longitudes. As there is a single coordinate pair for each location it can be mapped using a point which is effective for features such as settlements, mountains and individual buildings but less effectively for linear features such as roads or rivers, or large areas such as countries, regions, seas and lakes.[2] In theory the geometries of these line or polygon-based representations can be found in gazetteers but points are most commonly used.

If a source contains a list of place-names then georeferencing it using a gazetteer is relatively straight-forward. A relational join, a standard piece of database functionality, is used to automatically match the place-names in the source with those in the gazetteer. This results in a version of the source in which coordinates have been added. A database table that includes coordinates can be converted into a point layer within GIS software with only a few mouse clicks.[3] There are, however a number of issues with this process. First, a relational join will only work with exact matches in spellings with even minor variations, such as the difference between "St" and "St." and "Sca-Fell" and "Sca Fell", failing to match. Fuzzy matches can be used to assist in resolving such problems but can lead to false positives. Colloquial forms and abbreviations, such as "Soton" for "Southampton" will also fail to match. Variations in spellings are common with historical sources due to both genuine variations in spelling and naming conventions and due to digitisation errors. Ambiguity, where more than one place has the same name, for example "Newcastle" or "Sutton", is an additional problem which is not easily resolved unless by user intervention of automated means using additional information.

Second, the gazetteer is critical to the success of this process. Most obviously, places that are not included in the gazetteer cannot be allocated a coordinates. The original purpose of the gazetteer, which is normally closely associated to the scale covered, is crucial. A gazetteer designed to cover a country or continent will include different features to one representing a single city. Connected to this there is the issue of what can be considered a place, and therefore, whether it was included in a gazetteer. Thus are, for example, mountains, regions, rivers and streets included in the gazetteer? The accuracy of the coordinates within the gazetteer is also a potential issue and will again be related to scale. Broader scale gazetteers are likely to be less accurate in their coordinates than

---

future of humanities scholarship. Indiana University Press: Bloomington; Dear M., Ketchum J., Luria S. and Richardson D. (2011, eds) *GeoHumanities: Art, history, text at the edge of place* Routledge: London; Gregory I.N. and Geddes A. (2014, eds) *Toward Spatial Humanities: Historical GIS and Spatial History*. Indiana University Press: Bloomington.

[2] For more on gazetteers see: Southall, H., Mostern R. and Berman M.L. (2011) "On historical gazetteers" *International Journal of Humanities and Arts Computing*, 5, pp. 127-145; Mostern, R. and Johnson, I. (2008) "From named place to naming event: creating gazetteers for history" *International Journal of Geographical Information Science,*22, pp. 1091- 1108; Hill L.L. and Janee G. (2004) "The Alexandria Digital Library Project: Metadata development and use" in Hillmann D.I. and Westbrooks E.L. (eds.) *Metadata in Practice*. American Library Association: Chicago, pp. 117-138; Larson R.R. (2003) "Placing cultural events and documents in space and time" in Duckham M., Goodchild M.F. and Worboys M.F. (eds.) *Foundations of Geographic Information Science* Taylor & Francis: London, pp. 223-239

[3] Gregory I.N., Donaldson C., Hardie A. and Rayson P. (in press) "Modelling space and time in historical texts" in Flanders, J. and Jannidis F. (eds.) *Data Modelling in the Digital Humanities*. Ashgate

more detailed gazetteers but all coordinates will have some levels of positional inaccuracy associated with them. A number of gazetteers are available with Geonames being perhaps the most widely used, freely available one with global coverage.[4] However, none of these are comprehensive for Early Modern Europe so, any gazetteer or gazetteers used for georeferencing Early Modern sources, will need extending and enhancing.

A third set of problems are related to the fact that the process relies on the source containing lists of identified place-names. Data in this form will be found in, for example, correspondence metadata that includes fields to say where a letter went to and from. If, however, the source consists of unstructured text, such as the letter itself, the place-names will need to be identified within the text. Here a process called *geoparsing* can be invaluable.[5] This is a two-stage process in which natural language processing techniques are first used to identify words in the text that are likely to be place-names and, in the next stage, these are then matched to a gazetteer. This results in the text being georeferenced automatically, a process that can be implemented using, for example, the Edinburgh Geoparser.[6]

An additional layer of complexity is the diversity of languages in which the Republic of Letter's material was written. These include a wide variety of languages including French, Dutch, Latin and German as well as English. Much of the work on both gazetteers and geoparsing to date has emphasised English. Two strategies could be developed to cope with this. One would be to develop entirely separate gazetteers for different languages, the other would be to create a single multi-lingual gazetteer that would identify the same place in multiple languages.

The use of gazetteers, with or without geoparsing, makes it relatively easy to produce a database table or a GIS point layer, this may not be what is required based on the technical expertise of the people working with the data. Many users may want to work with KML (Keyhole Markup Language) data, an XML (eXtensible Markup Language) implementation that can be opened in Google Earth. While Google Earth is arguably the easiest software to use to map and explore data, getting data into KML format either requires the user to export the data from a piece of GIS software, or to write a program to convert the table into KML. Both of these require a certain level of technical expertise. KML can also be used to insert the data into a webpage that uses a Google Maps-based interface. Building such an implementation requires Javascript programming and the ability of the user to then explore the data will be dependent on the resulting interface. Thus, even once the data have been georeferenced, there are subsequent challenges in how they will be visualised and analysed.

---

[4] The Geonames gazetteer is at http://www.geonames.org; other gazetteers include: the Getty Thesaurus of Geographical Names (http://www.getty.edu/research/tools/vocabularies/tgn/index.html) and Natural Earth (http://www.naturalearthdata.com). The Quattroshapes gazetteer has been built to contain polygon data (http://quattroshapes.com). More generally, spatial data are available in Wikidata (http://www.wikidata.org) and DBpedia (http://wiki.dbpedia.org).

[5] Grover C., Tobin R., Byrne K., Woollard M., Reid J., Dunn S. and Ball J. (2010) "Use of the Edinburgh geoparser for georeferencing digitized historical collections" *Philosophical Transactions of the Royal Society A* 368 pp. 3875-89

[6] For the Edinburgh Geoparser see: https://www.ltg.ed.ac.uk/software/geoparser. Other work in this area includes: DeLozier G., Baldridge J. and London L. (2015) "Gazetteer-Independent Toponym Resolution Using Geographic Word Profiles" In Proceedings of Assoc. for the Advancement of Artifical Intelligence 2015. Austin, TX; and Speriosu M. and Baldridge J. (2013) "Text-driven toponym resolution using indirect supervision" In Proceedings of the Association for Computational Linguistics 2013. Sofia, pp. 1466-1476.

Historical Source	Gazetteer
Spelling variations	Purpose
Place Ambiguity	Scale
Abbreviations	Choice of Geographic Coordinate System
Features being referred as places	Language
Variety of languages	Geographic Transformations when joining datasets

Table 1 Summary of issues to take into account in the georeferencing of historical corpora

The gazetteer is, therefore, of central importance in getting a source that contains placenames into a position where it can be explored spatially. While gazetteers make this process achievable, they are by no means unproblematic for reasons summarised in table 1. The creation or enhancement of an effective gazetteer or gazetteers will be essential to the effectiveness by which space and time can be explored in the *Republic of Letters*.

2. Georeferenced data in the humanities: Examples of the state of the art

A number of projects that make use of georeferenced humanities sources were presented at the meeting. The first of these was the *Trading Consequences* project.[7] This project aims to create a better understanding of nineteenth century global commodity trading associated, particularly, with the British Empire. This drew on digitised version of the House of Commons Parliamentary Papers and other sources which were text mined and geoparsed to find coincidences of place-names with a large lexicon of commodity names. One particular issue in developing this was the problem of digitisation errors associated with the use of Optical Character Recognition (OCR) software to convert image scans of the sources into machine-readable text. While these were problematic, strategies could be developed to overcome them. Once the data had been captured, they were disseminated through a visualisation system that allows patterns of trade of individual commodities to be explored through a map-based interface and also using a time-line. In this way a series of very general unstructured texts were converted into a resource for exploring global trade in space and time.

A second, contrasting example, is the *Palimpsest* project at the University of Edinburgh.[8] This project takes a major corpus of literary writing focusing on the city of Edinburgh. As with the *Trading Consequences* project these were geoparsed to identify place-names and locate them in space. This allowed the creation of an interactive map-based website and the development of an app that allows users to explore the source on location in the city using a smartphone or tablet. As it is a local-scale project, focussing on a single city, the *Palimpsest* project requires a very different gazetteer to *Trading Consequences*. Many of the features of interest are streets that cannot be effectively represented with points. To resolve this they created a bespoke gazetteer based heavily on Open Street Map,[9] crowd-sourced spatial data initially developed to provide data to GPS and satnav users that is free of copyright from national mapping agencies or commercial companies. This data was repurposed to allow historical streets and buildings to be mapped. As with other gazetteers, there are issues associated with renaming over time, spelling variations that need to be

[7] http://tradingconsequences.blogs.edina.ac.uk
[8] http://www.ed.ac.uk/literatures-languages-cultures/english-literature/research-activities/palimpsest, see also http://litlong.org
[9] https://www.openstreetmap.org

resolved and additional features that needed to be added. Geoparsing is important in this as it identifies potential place-names in the sources for which no gazetteer entry is available and which thus need to be found.

Geoparsing is also being used to georeference a number of large collections of digital texts including the Text Creation Partnership (TCP) versions of Early English Books Online (EEBO) and Eighteenth Century Collections Online (ECCO), as well as the British Library's Nineteenth Century books collection which consists of around 65,000 books. This again poses significant challenges to both the gazetteers and to the disambiguation stage. Nevertheless, geoparsing does open up the possibility of mapping and spatially analysing very large corpora.

An alternative approach to georeferencing data was presented by the *Pelagios* project.[10] This was originally conceived of a tool to explore geographies of the Ancient World by connecting online resources including texts, images, archaeological sites and museum records based on their references to place. This is implemented as linked open geodata where each place has its own Universal Resource Identifier (URI). These were originally developed from the Pleiades gazetteer which, in turn, was developed from a paper atlas of the Ancient World. [11] More recently, *Pelagios* has been extended to include medieval Christian, Islamic and Asian geographies. The basic idea of the system is that it allows the user to type in the name of a place, or to click on a map, and all of the records within the community identified as being associated with that place will be returned. As these are based on URIs and open data a large community of relevant projects is being developed that are all able to benefit from and contribute to the system's content.

A different approach again was taken by the Mapping the Lakes project.[12] This project had taken two relatively short accounts of early tours round the English Lake District and had georeferenced them by hand. This allowed new research questions to be asked of the texts concerning what places the authors named, where they actually went, how the two tours compared geographically, how the authors responded to the landscapes they were in, and so on. It also showed how additional georeferenced data, for example on height or population density, could be integrated to provide additional information about the places.[13] Subsequent work has developed this using additional texts geoparsed using the Edinburgh Geoparser.[14] What marks this approach out from the ones above is that this is more concerned with the results of an analysis of georeferenced texts, whereas the projects above were concerned primarily with georeferencing sources and delivering them to a user, typically through a web interface.

While different approaches to the use of gazetteers has been used to georeference all of these sources, the temporal aspect has usually been managed using more conventional approaches. In

[10] http://pelagios-project.blogspot.co.uk/p/about-pelagios.html
[11] T. Elliot and S. Gillies (2009) "Digital Geography and Classics" *Digital Humanities Quarterly*, 3 (1). See: http://digitalhumanities.org/dhq/vol/3/1/000031/000031.html
[12] http://www.lancaster.ac.uk/mappingthelakes
[13] Cooper D. and Gregory I.N. (2011) "Mapping the English Lake District: A literary GIS" *Transactions of the Institute of British Geographers*, 36, pp. 89-108; Gregory I.N. and Cooper D. (2009) "Thomas Gray, Samuel Taylor Coleridge and Geographical Information Systems: A Literary GIS of Two Lake District Tours" *International Journal of Humanities and Arts Computing*, 3, pp. 61-84
[14] Donaldson C., Gregory I. and Murrieta-Flores P. (2015) "Mapping 'Wordsworthshire': A GIS study of literary tourism in Victorian England" *Journal of Victorian Culture*, 20, pp. 287-307

particular, metadata from when a book or volume was published provides the temporal information to most of these projects and allows the exploration of temporal as well as spatial change. It may also be possible to use techniques similar to geoparsing to extract temporal references. More sophisticated work, similar to geoparsing, can also be done to extract temporal references from unstructured texts.[15]

3. Spatio-temporal information in the *Republic of Letters*

The period covered by the *Republic of Letters* was affected by major upheavals, such as the Reformation and the Thirty Years War, and longer-term, more gradual processes such as migrations of people and diffusions of ideas. The material that makes up the *Republic of Letters* has three entities that are located in space and time: the letters themselves, the people that wrote the letters, and the topics that they were writing about which include both personal information and a wide range of academic material.. The peregrinations of many of the more prolific writers are already fairly well known, however, truly understanding the letters and what they tell us requires an understanding of the systems by which knowledge was exchanged. In particular, this refers to: the postal network and how it developed, the evolution of the book trade,[16] the academic networks (for which matriculation records provide invaluable information), and the wider communities and diasporas of the time which, in turn, lead to linguistic diversity and complexity.

Alexandre Tessier has been working on developing a better understanding of the Early Modern postal network. This draws inspiration from the Orbis system that attempts to calculate travel routes, times and distances across the Roman Empire.[17] Creating a similar system for the Early Modern postal network is made possible from a number of sources: postal treaties between counties which give the rules of international exchanges, and particularly the charges and mechanisms for crossing borders. Postal charges are sometimes written out in documents with postal advertisements potentially providing information on times and costs. Diplomatic correspondence can supplement this as it provides information about how long a letter would spend in transit as it frequently give the dates of both sending and arrival. Similar information is available in the Hartlib Papers. While these sources provide a large amount of information on the time and cost of sending a letter, modelling postal routes is more difficult. British maps clearly show postal roads but for other parts of the continent these are more difficult to discover. The architecture required to create a database to model the Early Modern postal system requires a knowledge of both where post offices were located and the stages that postal roads would follow. Potentially, however, a system could be developed that would allow the time, distance, cost and an estimation of the route that a letter would take to travel between any two post offices.

While this potentially allows the physical methods of communication to be modelled, Charles van den Heuvel presented on a project that is looking at networks and dynamics of examples of correspondence that help to explore some of the academic networks. He does this drawing on a range of approaches from geographical information systems and network analysis. Some of this

[15] An example of this is the Heideltime system (https://github.com/HeidelTime/heideltime). See also: Joho H., Jatowt A. and Blanco R. (2015) "Temporal Information Searching Behaviour and Tactics" *Information Processing and Management*, 51, pp. 834-850

[16] https://atlas.lib.uiowa.edu

[17] http://orbis.stanford.edu

work has been based on correspondence metadata, however, he stresses the importance of extracting information from the texts themselves as this can provide information on all of the books, people, places and subjects associated with a person in a way that metadata searches cannot. Topic modelling has proved very useful in doing this.[18]

Ruth Whelan presented on the ways in which spaces of shared learning within wider communities could be explore through correspondence. She stresses the importance of understanding the different networks of men and women. Friendship books can be used in addition to correspondence to assist with this. She is also interested in the role of religion and ethnicity in defining and disrupting networks and the importance of movement, particularly those caused by diasporas, in understanding these.

4. Conclusions

Gazetteers are clearly of critical importance to locating the *Republic of Letters* in space and time. As identified above, there are a significant number of gazetteers already in existence and there is already too much reproduction across them, however, no one gazetteer is suitable for the *Republic of Letters* as a whole. In particular, most of the gazetteers currently in existence have been developed for the modern period and are Anglo-centric in their focus. This means that rather a gazetteer construction, what is required is a gazetteer enhancement process in which relevant information from existing gazetteers is ingested and enhanced. A second issue is that no one gazetteer will be suitable for all tasks within the *Republic of Letters*. One gazetteer for the whole of Europe (and perhaps beyond) might be suitable, however this should not include detailed local-scale information for individual cities such as streets and buildings. Instead, separate gazetteers can be built for cities.

A good starting to developing a broad-scale gazetteer of Europe (or wider) is place-names in existing *Republic of Letters* metadata and potentially place-names extracted from *Republic of Letters* texts using geoparsing or topic modelling techniques or identified by hand. The precise architecture of the resulting gazetteer would need careful thought, it would need to be able to manage (at a minimum): multilingual issues, name changes over time, spelling variations (actual and common digitisation issues), a typology of places, textual descriptions of the place, a source for the name information, a location and a source for the location.

More local gazetteers for important cities such as Paris, London, Berlin, Florence and Athens may be simpler as the multilingual issues will be reduced and should probably start with an existing resource which might be historic, for example potentially the Map of Early Modern London,[19] or modern sources such as Open Street Map. These are relatively simple to develop and should probably be built in response to specific project needs although an exemplar may help.

A second issue is how to populate the gazetteer. Where information is ingested from existing sources frequently different gazetteers will give slightly different locations for the same place thus a precedence needs to be established. Crowdsourcing could be used to investigate place-names that can not easily be located from existing sources, however, this then raises quality control issues.

[18] Blei D.M. and Lafferty J.D. (2009) "Topic Models" in Srivastava A. and Sahami M. (eds.) *Text Mining: Classification, Clustering, and Applications* CRC: Boca Raton, pp. 71-94

[19] https://mapoflondon.uvic.ca

Automated techniques may also be usable to match existing records with close variants but these again have quality control issues.

A third issue is the question of national and regional boundaries and also contemporary background maps. These are useful for context and, in the case of boundary files, can assist in more complex queries such as "what letters were sent from Hanover?" These can probably be sourced from historical atlases however issues concerned with how boundaries change over time, especially in modern Germany, could become a very large project very quickly.

A fourth issue is that of time which is important but perhaps poses fewer challenges than space at this stage as it is usually encoded in most metadata and poses fewer challenges than space. Finally, while creating the resources required to add spatio-temporal information to *Republic of Letters* resources is important and challenging, the overall aim of conducting applied research using this information, and how to go about doing this, must not be forgotten.

5. Recommendations

1. A sub-group should be set up to comprehensively investigate current gazetteers and their applicability to the *Republic of Letters* should be set up with the aim of producing a short report. Murrieta-Flores volunteered to lead this.
2. A pilot needs to be conducted on georeferencing one or more existing corpora. Suggestions on suitable corpora, particularly from Whelan, were gathered at the meeting.
3. Further investigation of whether the *Pelagios* architecture could be repurposed for use within the Republic of Letters should be conducted. Isaksen has volunteered to assist in this.
4. Although temporal information is important, priority should be given to spatial information in the short term as it is more complex while relatively simple, metadata-based approaches can be used to handle time.
5. Further thought needs to take place about how resources that have been located in space and time are going to be analysed. This needs to take place in collaboration with other working groups, particularly Visualisation & Communication (WG 6) and People & Networks (WG 2). As well as the technical aspects of this, the applied research questions that this work enables should also be considered.

Acknowledgements: This paper was produced under COST Action IS1310 "Reassembling the Republic of Letters, 1500-1800." It reflects the findings of a meeting held at Lancaster University, UK on the 4th December 2015. This meeting was attended by: Howard Hotson (Chair of Republic of Letters Cost Action); Ian Gregory (Chair of WG1); Patricia Murrieta-Flores (Co-chair of WG1); Paul Atkinson; Alexander Butterworth; Claire Grover; Leif Isaksen; Bruno Martins; Catherine Porter; Sinai Rusinek; Alexandre Tessier; Charles van den Heuvel; Ruth Whelan. We would like to thank them all for their contributions.