

## 11. Putting the Eighteenth Century on the Map: French Geospatial Data for Digital Humanities Research

Katherine McDonough

Databases at the heart of the Mapping Print, Charting Enlightenment (MPCE) project at Western Sydney University capture information from archival documents about where books were printed, sold, purchased, and censored. The MPCE team at Western Sydney working on these databases (Simon Burrows, Jason Ensor, Rachel Hendery, Laure Philip, and myself) faces a challenge in transforming this information about place (most commonly in the form of the name of a city) into a location (normally expressed with a coordinate system).<sup>1</sup> Traditionally, when working with fairly small datasets, scholars have chosen to look up locations of places one by one, adding latitude and longitude as a new column in their spreadsheet. A crucial advantage of this method was that it allowed an expert to pick just the right location for a historical place. One obvious disadvantage was the amount of time required to make these selections. Scholars working in some other periods and places do not face this problem at all: they have access to digital gazetteers that facilitate matching a modern or historical place to its location on the earth. Historians working on 19<sup>th</sup> and 20<sup>th</sup> century US history, for example, can reliably use existing gazetteers like GNIS (Geographic Names Information System) for most places. Historians of modern France can access locations and boundaries for French communes, departments, etc. through the IGN portal.<sup>2</sup> Those of us who work on France before the Revolution and the invention of this political geography, face a special challenge in finding a timely way of accurately locating places because we lack a period-appropriate gazetteer.

As a part of MPCE's work to make archival evidence about the eighteenth-century book trade more accessible in a digital format, I am working toward the creation of a gazetteer of the early modern world. The long-term goal is to develop a digital resource that early modernists can use to leverage the geospatial semantic web in the same way that scholars of other periods already can. Gazetteers are central to making connections among historical evidence that includes information about place. If a gazetteer is a tool for 'obtaining uniquely identifying descriptions of things referred to in discourse', an early modern gazetteer must include references to period discourse.<sup>3</sup> No existing resource does this, however. There is an operational gap between, on the one hand, gazetteers that do include unique identifiers (such as coordinates) but have no relation to early modern sources and, on the other hand, those early modern sources

---

<sup>1</sup> *Mapping Print, Charting Enlightenment* [<http://fbtee.uws.edu.au/mpce/>] is funded by the Australian Research Council and is directed by Simon Burrows (Western Sydney University). Other Chief Investigators include Jason Ensor and Rachel Hendery (Western Sydney University), Dan Edelstein (Stanford University) and Angus Martin (University of Sydney). We are project partners with the MEDATE project directed by Alicia Montoya at Radboud University [<http://mediate18.nl/>]. Matje van de Camp is a computational linguist collaborating on the ARTFL Encyclopédie research [<http://www.taalmonsters.nl/>].

<sup>2</sup> Accessible at [<http://www.ign.fr/>].

<sup>3</sup> Ryan Shaw, 'Gazetteers enriched: a conceptual basis for linking gazetteers with other kinds of information', in *Placing names: enriching and integrating gazetteers*, ed. Ruth Mostern, Humphrey Southall, and Merrick Lex Berman (Bloomington, Indiana University Press, 2016), p.69.

that have no unique identifiers (as we define them in the twenty-first century).<sup>4</sup> A growing collection of digital humanities projects focus on the early modern period (ca. 1400–1800) and do include information about place. Early Modern Letters Online (EMLO), for example, has embraced Linked Open Data opportunities for identifying places named in their metadata with GeoNames record IDs. The challenge here is that GeoNames records primarily refer to contemporary places rather than historical entities. A gazetteer for early modern studies would provide the kind of connective tissue that classical scholars have with the resources of Pelagios Commons, which allows them to identify the location of ancient places (with the Pleiades gazetteer), annotate digitized texts with digital gazetteer record links (with Recogito), and identify the locations of classical texts, objects, etc. (Peripleo).<sup>5</sup> Now that the Pelagios initiative is expanding to include resources for other periods and places, the time is ripe for an early modern gazetteer that makes intelligent use of existing gazetteers and joins them in a network of interoperable place-name directories.

We use Geographic Text Analysis (GTA), a relatively new method for humanities scholars, to work toward the gazetteer. As more and more texts from the eighteenth century are digitized, we can envision new methods for querying, comparing, analyzing, and visualizing their geographic content. By linking Named Entity Recognition (from the suite of Natural Language Processing tools) with GIS processes like geolocation, we can study new sets of sources previously inaccessible to the individual or small group of researchers during the life span of a single research project.<sup>6</sup> In the field of early modern history, for example, we can learn more about the reference sources at the heart of the Enlightenment project (like Diderot and d’Alembert’s *Encyclopédie*) and use these sources to unpack spatial questions about other sources. These spatial questions matter for the eighteenth century, and the early modern period more broadly. Capturing a digital record of early modern spatial knowledge—what places were known to whom, in what context, and through which sources, for example—not only matters in terms of helping researchers identify historical places based on contemporary records. The act of making a gazetteer that digitizes printed historical knowledge gives us time to reconsider the early modern corpus of geographic reference works as a portal into historical representations of space complementing and yet distinct from contemporary map production.

Of the many digital humanities projects focused on the eighteenth century, MPCE is not alone in struggling to deal with questions about historical places.<sup>7</sup> EMLO at Oxford, Mapping the Republic of Letters at Stanford, MEDATE at Radboud, ARTFL at Chicago (all of which

---

<sup>4</sup> For a survey of recent literature on gazetteers, see Ian N. Gregory, ‘Further reading: from historical GIS to spatial humanities: an evolving literature’, in *Toward spatial humanities: historical GIS & spatial history* (Bloomington, University of Indiana Press, 2014), p.191-92.

<sup>5</sup> Available at [<http://commons.pelagios.org/>].

<sup>6</sup> Ian Gregory, Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson, ‘Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial Humanities Research,’ *International Journal of Humanities and Arts Computing* 9, no. 1 (2015): 1-14.

<sup>7</sup> While MPCE is (and FBTEE was) predominantly focused on eighteenth-century France, the range of geographical sources we consult for the gazetteer will cover approximately 1500 to the present. To establish a list of ‘early modern’ places, we will narrow coverage to 1500-1790 (just before France was reorganized into the *départements* that it retains today for administrative jurisdictions). The genre of reference sources with geographical coverage gathered steam in the second half of the seventeenth century and continued until the Revolution. Our coverage extends geographically beyond France (and Europe) because the source often had ‘universal’ (global) coverage.

are described in this volume), and the large network of European projects under the umbrella of the Reassembling the Republic of Letters Project are resources that could potentially all be linked through their common place-name data. Like the scholars affiliated with all these projects, those behind the French Book Trade in Enlightenment Europe (FBTEE), the first database in what will become the MPCE ecosystem, most often dealt with historical geographical issues using tried and true methods of hand matching place names with locations on historical maps and then with latitudinal and longitudinal coordinates. Subject experts completed this data-entry and verification labor. An early, modern gazetteer, however, should help to automate the tasks associated with locating places mentioned in historical texts, to facilitate links between projects dealing with the same places, and to open up the study early modern cultures of geographic knowledge through the sources analyzed.

Project interoperability for early modern history can be thought of in terms of the major types of information in our datasets: people, institutions, objects (letters, books), and places. At the first Digitizing Enlightenment conference at Western Sydney University, directors and researchers affiliated with the most robust DH projects in eighteenth-century French studies came together to discuss how we might offer links between datasets and visualization tools with different institutional homes. From just the first in what is now planned to be a series of Digitizing Enlightenment gatherings, the takeaway for many of us was the importance of developing metadata as Linked Open Data (LOD).<sup>8</sup> With so many geographic information specialists and humanities scholars recognizing the potential of LOD for data matching, we aim to think through the challenges of dealing with geospatial referents in early modern sources. To a large degree, these challenges stem from a lack of standardized methods for locating and creating metadata about locations in early modern datasets.<sup>9</sup>

We are not the first to combine the tools of computational linguistics and geographic science for historical research, but we are the first to use these tools for French-language texts and for a corpus printed largely before 1800.<sup>10</sup> These novelties present their own challenges. In this chapter, I lay out how FBTEE, MPCE, and similar projects have dealt with geographic information in the past, what I hope to do in the near future, what sources I am using and why, and how initial tests with GTA tools inform this work. Underlying this discussion is the issue of labor in spatial humanities projects. GTA and the prospect of LOD for early modernists should be considered in these terms, as should the significant intellectual gains resulting from having access to digitized historical geographical sources.

---

<sup>8</sup> The second Digitizing Enlightenment conference was held in June 2017 in the Netherlands.

<sup>9</sup> For a summary of research directions ca. 2012 in dealing with interoperability issues with geospatial data, see Krzysztof Janowicz, Simon Scheider, Todd Pehle, and Glen Hart, 'Geospatial semantics and linked spatiotemporal data—past, present, and future,' *Semantic web* 3.4 (2012), p.321-32.

<sup>10</sup> Ian Gregory, David Cooper, Andrew Hardie, and Paul Rayson, 'Spatializing and analyzing digital texts: corpora, GIS, and places,' in *Deep maps and spatial narratives*, ed. David J. Bodenhamer, John Corrigan, and Trevor M. Harris (Bloomington, Indiana University Press, 2015), p.152-53. In a separate paper, Gregory and Patricia Murrieta-Flores have named the combination of Natural Language Processing, Corpus Linguistics, and GIS 'Geographic Text Analysis'. I also use this term to describe our work with early modern texts. See Ian Gregory and Patricia Murrieta-Flores, 'Further frontiers in GIS: extending spatial analysis to textual sources in archaeology', *Open archaeology* 1.1 (2015), p.168-69.

To determine best practices for early modern GTA, we should consider the following questions: First, how can we engage critically with early modern sources so that we do not sacrifice their contextual specificity as we mine them for place-name data. Second, how do we adapt tools that have been designed to work with English and modern texts? Third, how do we leverage existing digital resources that have been created for research on other periods? Though we lack a gazetteer for early modern history, many existing resources can help us build a database of places and their locations relevant to this period. These include digital and print gazetteers and other geographic reference sources from the sixteenth century to the present. Working with early modern geographical sources and building digital tools to play nicely with these French texts breaks new ground in gazetteer design.

### *Mirroring One Archive, Connecting Many Archives*

FBTEE was the brainchild of Simon Burrows. The database organizes materials from the archive of the Swiss publishing house the *Société Typographique de Neuchâtel* (STN), active between 1769 and 1794. This is a famous archive for historians of the French Enlightenment and Revolution. Robert Darnton was the first to see the STN archive as a gold mine for questions about the relationship between reading and revolution.<sup>11</sup> For those who agreed with Darnton's work, it was hailed as a representative sampling of the French book trade and the renewal of book history in French studies. Others had concerns that the STN was only a narrow set of publications catering to a particular, skeptical audience, or that Darnton overemphasized the 'Grub Street' aspects of the Enlightenment. Regardless, Darnton's work brought books to the forefront of scholarship on the Revolution and propelled a new generation of scholars to keep digging in the archives. To get a fuller picture of the STN archive in particular, Burrows and colleagues at the University of Leeds painstakingly transformed the collection of manuscript documents into a database organized around tables of people, books, and the 'events' tying them together. People included the publishing agents, booksellers, and clients. Books were divided into 'superbooks' (the conceptual *work*) and editions, tagged using two categorization systems: those used by eighteenth-century booksellers and those useful to twenty-first century researchers. Events included transactions between the STN and its clients as well as internal events (like printings). The original database included 450,000 books.

Place-related data from the STN archive included client and bookseller towns (516 total places). In addition to these locations, Vincent Hiribarren and Burrows created geographic filters and regional boundaries through which users could view the dissemination of books or the location of booksellers, client professions, book category keywords, or book languages. Town-based data can be aggregated at the regional level to compare sales and purchases across larger spaces. Faced with the challenges of early modern jurisdictional overlap and fuzziness, FBTEE gave users choices in terms of how results were visualized: as sovereign territories (e.g., France), lower territories (French provinces like Normandy, or Spanish Intendancies), modern countries and administrative units (e.g., French *départements*), broader geographic

<sup>11</sup> Robert Darnton, *The Literary Underground of the Old Regime* (Cambridge, Harvard University Press, 1985), *The Corpus of Clandestine Literature in France, 1769-1789* (New York, W. W. Norton, 1995), and *The Forbidden Best-sellers of Pre-Revolutionary France* (New York, W. W. Norton, 1996). For a full list of Darnton's STN related publications, see [<http://www.robertdarnton.org/publications>].

'zones' (Helvetic, Scandinavian, Iberian, etc.), or other select place groupings such as higher education towns, imperial free cities, or ecclesiastical lands. Hiribarren drew borders in QGIS by hand, starting from Natural Earth data and consulting eighteenth-century maps of France.<sup>12</sup> Users could not only visualize the data in the web-based interface; they could also download it. FBTEE modeled open access research by making both the transcribed archival data and the GIS data developed for the visualizations freely available.<sup>13</sup> The different border and zone datasets were valuable contributions to HGIS resources for early modern Europe. These filters and regions make it possible for the user to move beyond the 'town' scale to begin interpreting the results from a query. Figure 1 displays the results from a search of where books published in Amsterdam were sold. The provinces and lower-division territories on the map are colored according to the number of books sold to clients living in town within those regions.

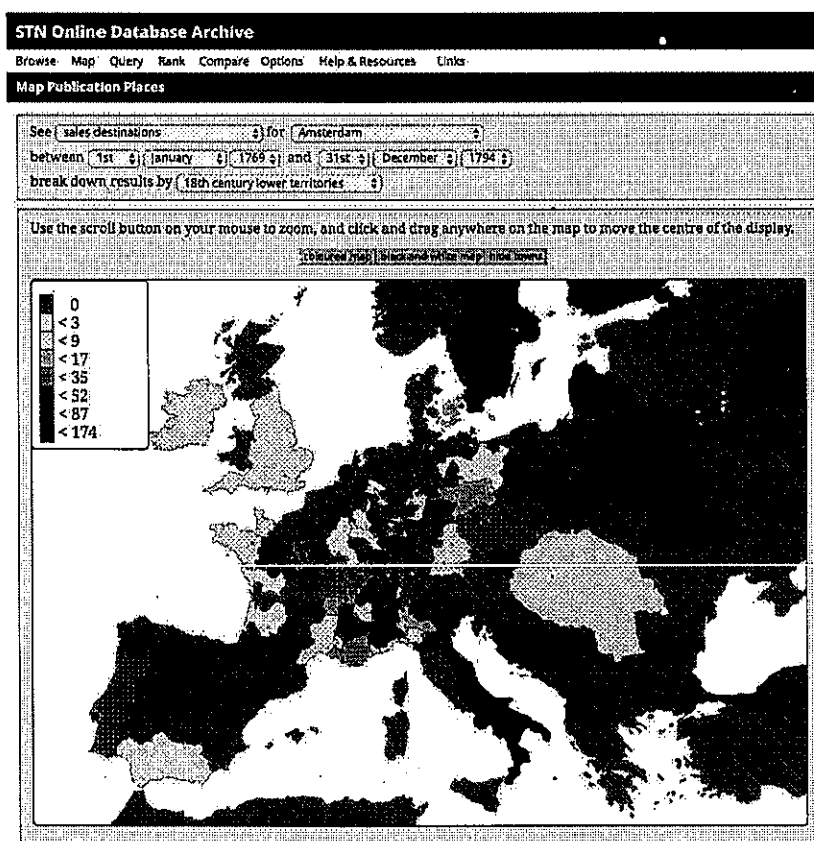


Figure 1. Screenshot of a query in the FBTEE database [<http://fbtee.uws.edu.au/stn/interface/>] for sales destinations of book published in Amsterdam. Borders represent '18<sup>th</sup> century lower territories' and the data is visible by town.

FBTEE was innovative in digitizing a fairly inaccessible archive that continues to be at the center of debate about the nature of the Enlightenment and the origins of the French Revolution. It tackled major problems in book history by trying to handle multiple editions,

<sup>12</sup> Maps by Guillaume Delisle (1675-1726) were consulted for borders. Reported in conversation with Vincent Hiribarren on July 21, 2016.

<sup>13</sup> See [<http://fbtee.uws.edu.au/main/download-maps/>] and [<http://fbtee.uws.edu.au/main/download-database/>]

printings, and pirate copies of single works. It created a way to see where books were going, when they were circulating, and who was ordering them. Though it gave a fuller picture of the STN archive than earlier work did, FBTEE still accounted for only one publishing house. Fast-forward to 2016, and FBTEE is the foundation for the MPCE project. Burrows wanted to expand the database to include information about the full range of production, sales, dissemination, policing, and reception of French books. This is what MPCE undertakes.<sup>14</sup> My colleagues and I are working to add new databases,<sup>15</sup> to improve metadata for existing and new databases, and to re-envision how users can access these datasets and those of other eighteenth-century French digital humanities projects.

These last two points are related. As we develop our metadata alongside partner projects, we are developing the possibility for other researchers to easily access the data via place queries. The broader community of place-name and gazetteer scholars has been developing and improving standards for integrating historical gazetteers with major semantic web resources like GeoNames and DBpedia.<sup>16</sup> MPCE metadata can interact with the semantic web in three ways: through records of books, people, and places. MPCE has already committed to thinking through the complex issues of entity relationships between different editions of works using the FRBR model for bibliographic records. (FRBR establishes that a “work is realized through [an] expression is embodied in [a] manifestation is exemplified by [an] item.”<sup>17</sup> With partners at the MEDiate project at Radboud University who are working with many of the same eighteenth-century French books, we are establishing methods for creating authority records for books—both at the *work* and *edition* levels. The authority record makes references to objects, places, and people translatable across projects (in the same way that subject, title, and author listings act as authority records in library records). Our Stanford partners, Dan Edelstein and Nicole Coleman (Humanities + Design Lab) are working toward a metadata model for early modern people, drawing on VIAF records and other linked biographical data.<sup>18</sup> During the course of my involvement with the MPCE project, I will be focused on linking our shared place records through metadata redesign.<sup>19</sup>

<sup>14</sup> Simon Burrows, Jason Ensor, Per Henningsgaard, and Vincent Hiribarren, ‘Mapping print, connecting cultures’, *Library & information history* 32.4 (2016), p.259-71.

<sup>15</sup> MPCE continues to focus on books in French and primarily on the book trade within French, but MEDiate expands the scope of our bibliometric evidence north into the Netherlands. As we continue to make links with other projects, we aim to examine historical bibliometrics on a European scale.

<sup>16</sup> Merrick Lex Berman, Joham Ahlfeldt, and Marc Wick, ‘Historical gazetteer system integration: CHGIS, Regnum Francorum, and GeoNames’, in *Placing names*, 13-24.

<sup>17</sup> Barbara Tillett, ‘What is FRBR? a conceptual model for the bibliographic universe’, Library of Congress Cataloguing Distribution Service, 3. Available online at [<https://www.loc.gov/cds/downloads/FRBR.pdf>]. FRBR (Functional Requirements for Bibliographic Records) was developed by a working group from the International Federation of Library Associations and Institutions. It defines the relationships between entities, which include works, expressions, manifestations, and items. MPCE is using the work, manifestation, and item categories to differentiate between different printings and establish linked a ‘superbook’ record for the work.

<sup>18</sup> Maria Teodora Comsa, Melanie Conroy, Dan Edelstein, Chloe Summers Edmondson, and Claude Willan, ‘The French enlightenment network’, *The journal of modern history* 88.3 (2016), p.495-534. VIAF: [<https://viaf.org/>].

<sup>19</sup> Gazetteer standards options and a brief history can be found in Raj Singh, ‘International standards for gazetteer data structures’, in *Placing Names*, 79-91.

### *Metadata for Early Modernists*

If FBTEE focused on boundary issues related to visualizing datasets, in MPCE we are returning to the metadata to do more with the places recorded in sources. This change in focus puts the emphasis on improving our dataset rather than building out visualizations. Place names and their locations are important for historical research. As Peter K. Bol has put it,

The interface humans have created between themselves and the physical world . . . is created through the process of naming. Naming—of a mountain or a river, town or a building—maintains an intelligible interface between the geophysical world and human culture; the name makes it a place. Naming pertains to all aspects of human life. But, like everything in human culture, names are not stable. They are changed, abandoned, forgotten, fabricated.<sup>20</sup>

Place names are central to early modern European history, since at that time world geography was being recorded, organized, and described in new ways.

Linked geodata is a key component leading toward project interoperability. As the number and size of DH projects grows, this becomes ever more important so that work does not become siloed in the Internet wasteland. Existing national historical GIS projects have devoted enormous resources to establishing historical boundaries, but for the early modern period (in Europe and elsewhere) this is extremely difficult, if not outright misleading. By focusing our attention on place names rather than boundaries, we can begin to tackle the (somewhat) less complicated issues of defining metadata around a fixed point. Historical boundaries for the early modern period could be redrawn almost every year. Place-name metadata allows for a greater degree of uncertainty about temporal attributes, alternate spellings, and so forth. The Pelagios Gazetteer Interconnection Format is already used by a number of humanistic gazetteers, and because we hope to draw on the work already done to locate classical places, we aim to join this group of interoperable gazetteers.<sup>21</sup> Furthermore, place name authorities developed from gazetteers are, as Humphrey Southall has highlighted, one key to linking into the GLAM (Gallery, Library, Archive, Museum) world.<sup>22</sup> Major repositories for early modern research such as the Bibliothèque nationale de France have long had subject, person, and place authorities managed internally.<sup>23</sup>

Unlike scholars of the nineteenth and twentieth centuries who, with fewer challenges, can use geodata from government geospatial agencies or open source data providers—and unlike even ancient and medieval scholars, who have a plethora of digital geographical resources completed or in development—early modernists lack period-specific digital infrastructure.<sup>24</sup> If we use (exclusively) official French geographical data and open data from

<sup>20</sup> Peter K. Bol, 'On the cyberinfrastructure for GIS-enabled historiography', *Annals of the Association of American Geographers* 103.5 (2013), p.1089.

<sup>21</sup> See [<https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>].

<sup>22</sup> Humphrey R. Southall, 'Applying historical GIS beyond the academy: four use cases for the Great Britain HGIS', in *Toward Spatial Humanities: Historical GIS & Spatial History*, ed. Ian N. Gregory and Alistair Geddes (Bloomington, Indiana University Press, 2014), p.97-104.

<sup>23</sup> On Interarc metadata:

[[http://www.bnf.fr/fr/professionnels/f\\_interarc/s.format\\_interarc\\_biblio.html](http://www.bnf.fr/fr/professionnels/f_interarc/s.format_interarc_biblio.html)].

<sup>24</sup> On medieval GIS and spatial history, see Gregory, 'Further Reading', p.199-200. The realm of ancient DH projects building geodata resources includes Pleiades, Hestia, GAP Vis, and now the

services like GeoNames, we assume that information about place names as documented in these datasets matches the information in our historical sources. Setting aside the work of matching a place named in a source to a place record in these modern databases, we still have to take for granted that (1) place names are spelled the same, (2) differences in spellings have no temporal attributes; (3) place names have not changed entirely; (4) communities retain administrative integrity, never experiencing shifting borders; and (5) a record actually exists for the historical place one wants to locate. Working with digital gazetteers created for administrative or LOD purposes is therefore not ideal for the early modernist. For scholars of other periods, it is this type of realization that has led to the creation of period- and place-specific gazetteers either on their own or as a part of historical GIS projects. The early modern period so far lacks this infrastructure, but, drawing on the resources available both through LOD and through gazetteers for the ancient and medieval world, we can build a new gazetteer for early modern studies.

Because of these challenges, early modern DH has for the most part depended on individually created databases with hand-located places. The most successful tools designed by and for early modernists, like the Palladio visualization tool from Stanford's Humanities + Design Lab, reinforce this self-service methodology. You can bring your own data to the party, so to speak, but you must already have located geographical data. For the scholar working with archival or print materials that include a wealth of geographical information, locating historical places by hand can take hundreds (if not thousands) of hours. This kind of method and dataset design does not lend itself to using place as a link to other research. In effect, this is what FBTEE did for the first generation of its database. Places (towns) were each assigned a latitude and longitude.<sup>25</sup> But these coordinates are isolated within the FBTEE 'universe'. They do not make reference to any LOD. When every project hand locates a place, and they do not share a common method, the chance that 'Paris', 'Athens', and 'Montréal' will have precisely consistent coordinates is virtually nil. Datasets could be matched across shared text ('Paris' written here equals 'Paris' mentioned elsewhere), but the word 'Paris' falls short of being a unique identifier. One example of the problem relying on place names in early modern studies is spelling variance over time and across languages. 'Paris' is spelled the same way in English and French, but 'Athens' and 'Athènes' present a difficulty. An early modern gazetteer record would account for many different spellings attested in different sources, avoiding this problem. A model enriched digital gazetteer depends on a Uniform Resource Identifier (URI) associated with a range of information (spellings, coordinates, source citations, etc.).

If we can transform MPCE metadata to include references to place-name records with URIs in preexisting gazetteers and in the new early modern gazetteer, we open the database to an array of other users and uses. As other projects create geodata that also uses URIs for shared gazetteers, we will see the kind of abundant data that scholars of the ancient world are beginning to see via Peripleo.<sup>26</sup> Peripleo pulls together onto a map any resource that includes metadata meeting Pelagios interconnection standards so that researchers can browse data where 'places

---

Pelagios Commons project, which seeks to move past the classical world and foster a community and set of tools (like the place-name annotator Recogito) that can apply to other historical periods and places.

<sup>25</sup> See the FBTEE Designer Notes:

[[http://fbtee.uws.edu.au/stn/interface/dsgnotes.php#\\_Toc327948853](http://fbtee.uws.edu.au/stn/interface/dsgnotes.php#_Toc327948853)].

<sup>26</sup> Available at [<http://pelagios.org/peripleo/map>].



are the common (linking) element'.<sup>27</sup> A comparable future outcome depends on collaboration between early modern DH projects to agree on a minimum of metadata that would be shared across datasets.<sup>28</sup>

### *Mining Early Modern Geographic Knowledge*

How might we arrive at the point of having a curated list of eighteenth-century place names related to book history? First, I am compiling a list of all relevant historical geographical texts in French.<sup>29</sup> From these, we are identifying texts that have already been digitized and are available either as image or text files (usually via the French National Library or the Internet Archive). We will select nondigitized texts to scan and run through an OCR (Optical Character Recognition) preprocessor developed by Matje van de Camp (who worked as an academic technologist on the MEDiate project), and then we will run them through OCR to create our own text files. Figure 2 illustrates how we might use various GTA tools to move from one of these text files toward the goal of the gazetteer. The diagram is actually representative of a circular process: information about locations we receive from one text will be fed back into our working lexicons of place names and most relevant records when we run another text through the GTA tools. Before explaining these tools in more detail, it is necessary to introduce the geographical texts that are our source material for identifying place names. One of the reasons that GTA is a useful method for early modern studies is the centrality of geographical thinking to the period.

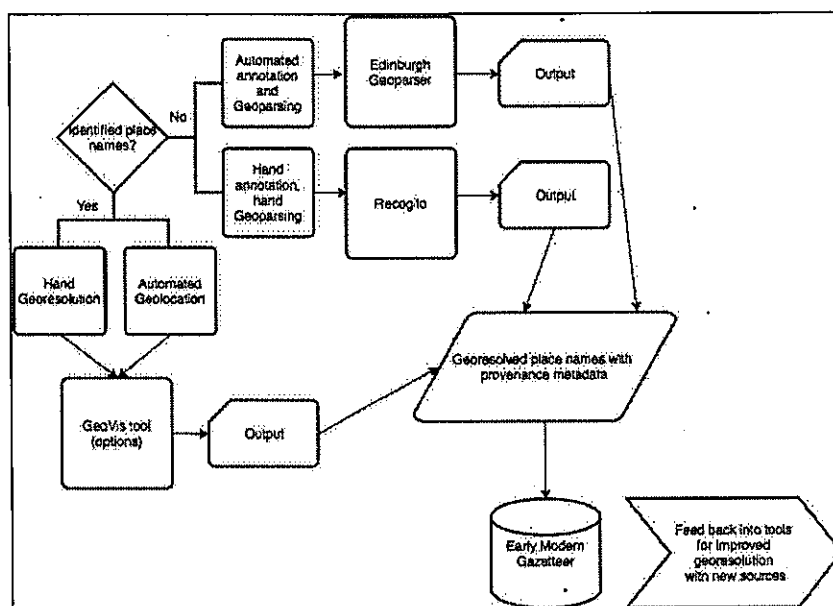


Figure 2. Workflow example for Early Modern Geographical Reference Sources.

<sup>27</sup> On Peripleo: [<http://commons.pelagios.org/about-the-data/>].

<sup>28</sup> Collaboration could usefully be done within the framework of the Pelagios Commons, where the researchers are keen to incorporate modern geodata with their earlier focus on the ancient and medieval periods.

<sup>29</sup> This expands on my dissertation research, where I studied similar texts for the French province of Brittany in the eighteenth century. Katherine McDonough, 'Building the roads: expertise, labor, and politics in provincial France, 1675-1791' (PhD diss., Stanford University, 2013).

The dedicated author of the *Bibliographie historique et topographique de la France* spent 25 years combing through France libraries to compile a list of about 12,000 printed texts (not including maps) related to French geographical history. Together in one reference source, he included works from antiquity up to 1845.<sup>30</sup> In contrast to many of his colleagues, he eschewed the distinction between ‘scientific’ and ‘nonscientific’ and lumped pre-nineteenth-century publications in with the most recent ones. Like this author, I conceptualize these sources as a continuum, and I see the early modern gazetteer that we hope to develop as a product of a range of sources from a range of periods. Each type of source has its role to play in allowing us to connect recent locational data to the descriptions of places in early modern sources.

The GeoNames of the nineteenth century was the directory of jurisdictions printed by the state. In the eighteenth century, writing geographical dictionaries and accounts of voyages preoccupied administrators, philosophes, religious men, and travelers. Prior to that, traditions of the geographical description derived from Pliny (an early example would be Sebastian Münster’s *Cosmographia universalis* [1544] in German) and lists of places following Ptolemy’s *Geographia*.<sup>31</sup> These different cultures of geographic knowledge nonetheless share common ground in the places they describe and locate. We can bridge the gaps by acknowledging the links between early modern works, the scholarly and administrative gazetteers that followed in the nineteenth century, digitized print resources in the late twentieth century, and data from Global Positioning System satellites in the last 30 years. The challenge is not that we have slim pickings for developing an early modern gazetteer from period sources. On the contrary, we must develop a methodology for sampling from these different categories of geographical texts.

The seventeenth- and eighteenth-century geographical dictionaries constitute a crucial set of texts among the broader collection. They form the basis for our core list of place names in the gazetteer. They were part of the explosion of reference works ordering knowledge in early modern Europe, and are an important repository of different ideas about what counted as authoritative geographical knowledge for their reading publics. Their source material was a mangled combination of passages from competitors’ and volumes, classical works, descriptions of the content of maps, as well as firsthand experiences or conversations recounted. Conrad Gesner’s *Onomasticon* (1544) and Charles Estienne’s *Dictionarium historicum ac poeticum* (1553) were important trendsetters in the dictionary genre and were already spin-offs of earlier works.<sup>32</sup> The seventeenth- and eighteenth-century texts will thus have links to works produced before and after them. Our database of such historical texts can eventually be mined for

<sup>30</sup> Girault de St-Fargeau, *Bibliographie historique et topographique de la France* (Paris, Libraire de Firmin Didot Frères, 1845), p.ii.

<sup>31</sup> Sebastian Münster, *Cosmographia. Beschreibung aller Lender durch Sebastianum Münsterum: in welcher begriffen aller Voelker, Herrschaften, Stetten, und namhafftiger Flecken, herkommen: Sitten, Gebreüch, Ordnung, Glauben, Secten und Hantierung durch die ganze Welt und fürnemlich Teütscher Nation* (Getruckt zu Basel, durch Henrichum Petri 1544).

<sup>32</sup> Ann M. Blair, *Too Much to Know: Managing Scholarly Information Before the Modern Age* (New Haven, Yale University Press, 2010), p.123-24. Conrad Gesner, *Onomasticon propriorum nominum, virorum, mulierum, sectarum, populorum, idolorum, syndrum, ventorum, vabium, marium, fluviorum, montium, et reliquorum, ut sunt vici, promontoria, stagna, paludes, etc.* (s.n.: Basiliae, 1544). Charles Estienne, *Dictionarium historicum ac poeticum* (cura ac diligentia Caroli Stephani, 1553).

commonplaces and citations.<sup>33</sup> We can track the content of these dictionaries in relation to each other and the rest of the corpus of geographical texts about French place names. This will allow us to analyze, for example, what kinds of information the compilers' privileged, what rhetoric they used, and what they omitted.

As sources of geographic evidence, the significance of these geographical dictionaries is tied to the historiography of the Enlightenment. These texts—regional, national, global in scope—embody the cosmopolitan, networked sensibilities of the period. And yet they have been somewhat forgotten among the great creations of the early modern 'age of information overload'.<sup>34</sup> In tracing which place names are listed in which works (and with what associated content), we can learn a great deal about the ways that compilers handled geographic information, for what purposes geographic information was printed, and how these practices and uses changed over time and space. Charles Withers, author of many important works on the geography of the Enlightenment, has written that 'the Enlightenment took place somewhere. Thinking about the Enlightenment as a locatable and not just datable thing—as dynamic, mobile, as cosmopolitan, not just static and national—is to enlarge our own consciousness as to what the Enlightenment is and was and how it should be thought of'.<sup>35</sup> 'People in the eighteenth century', Withers emphasizes, 'understood their world to be changing as a fact of geography, and as the result of processes of geographical inquiry'.<sup>36</sup> If the gazetteer is a means to providing accessible, field-specific locational data for early modernists, creating the gazetteer is a process that lends itself to the examination of 'geographical inquiry'.<sup>37</sup> The project unites questions about historical bibliometrics with the intellectual history of early modern Europe on a grand scale.

With these two simultaneous aspirations in mind, we are working with a test text to identify pitfalls related to geolocating places named in early modern texts. In recent ventures to develop a workflow for identifying place name locations in the growing collection of databases linked to the MPCE and partner projects, the driving questions are: What is the most effective way for us to use digitized early modern French texts to create a gazetteer of early modern locations? What is it possible to automate, and what must be done by hand? What roles do expert users and volunteers play in this process?

---

<sup>33</sup> With more information about the geographic content of the *Encyclopédie* and cognate texts, we will be able to address the kinds of text appropriation questions discussed in Glenn Roe, Dan Edelstein, and Robert Morrissey, 'To Quote or Not to Quote: Citation Strategies in the *Encyclopédie*', *Journal of the history of ideas* 74.2 (2013), p.213-36.

<sup>34</sup> Blair covers dictionaries broadly, but does not explore geographical dictionaries as a separate genre.

<sup>35</sup> Charles Withers, *Placing the Enlightenment: Thinking Geographically about the Age of Reason* (Chicago, University of Chicago Press, 2008), p.21.

<sup>36</sup> Withers, *Placing the Enlightenment*, p.5.

<sup>37</sup> Here I am influenced by Withers's other work, for example, his review article with Robert J. Mayhew, 'Geography: space, place and intellectual history in the eighteenth century', *Journal for eighteenth-century studies* vol.34, no. 4 (2011), 445-52, and Neil Safier's intellectual histories of Enlightenment geography in the South American context. See his recent article, 'The tenacious travels of the torrid zone and the global dimensions of geographical knowledge in the eighteenth century', *Journal of early modern history* 18 (2014), p.141-72.

### *Experimenting with the Encyclopédie*

My tests so far have explored three main methodologies for geolocating places mentioned in eighteenth-century French texts. These tests have been run on the text of the *Encyclopédie*, the great Enlightenment project edited by Diderot and d’Alembert. We are lucky to have an exceptionally clean digital text of the *Encyclopédie* thanks to the decades of work by the ARTFL project.<sup>38</sup> To assist with this initial research, ARTFL shared data on every geographical entry in the *Encyclopédie*.<sup>39</sup> The editors had divided the 14,547 entries as either *Géographie ancienne* or *Géographie moderne*. First, we separated headwords (entry titles) from the full text. Because every headword is a place name, we could move right into the work of trying to locate these places. With the full text, we needed to first identify place names within the text before locating them.

Ensor, van de Camp, and I wanted first to document how inadequate GeoNames is for the purposes of early modern research. Even though GeoNames includes place names written in other languages, it is a substantial challenge to match a headword to the correct GeoNames record. After some initial efforts to perform that matching by transforming these into MySQL tables and running a PHP script to select specific data from each table, we were able to use a Perl script to perform a string match on the GeoNames data and our headwords text file. To improve the performance of this script, van de Camp devised normalization rules for our headwords, making them more ‘readable’ for the GeoNames match. She created a simple but useful application for us to visualize the results (<http://geoviz.taalmonsters.nl/>, or GeoViz). On the left-hand side of the screen, GeoViz allows us to see the original headword text, the number of derived, or normalized headwords, and the number of matches those normalized headwords have in GeoNames. After a headword is clicked, on the right-hand side appears a map of all the matching GeoNames records’ locations, the full text from the *Encyclopédie* for that headword, and a list of the derived place names matched to a GeoNames record.

We plan to further develop this application to take into account contextual details in the full text of the source. For example, in the full text of the *Encyclopédie* (see, in fig. 2, the text below the map), there are other geographical features that situate the place we are attempting to locate (the continent it is on, a river nearby, the closest major city). Often the *Encyclopédie* text includes latitudinal and longitudinal coordinates. While the longitude is usually wildly incorrect, the latitude is more reliable (see fig. 4). We should be able to set a limit on the latitude of results to match this, thus improving the speed and quality of our matches. GeoViz is a useful tool for exploring where GeoNames fails, and it is a prototype for future collaborative and crowdsourcing work with our book history datasets. With this tool, we can see how important having contextual details is to determining a historical location. Given only a list of the GeoNames matches, it would be difficult to pick the correct coordinates.

<sup>38</sup> American and French Research on the Treasury of the French Language. *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers, etc.*, ed. Denis Diderot and Jean le Rond d’Alembert; University of Chicago: ARTFL Encyclopédie Project (Autumn 2017 edition), ed. Robert Morrissey and Glenn Roe, [<http://encyclopedia.uchicago.edu/>]. See also chapter 1 of the present volume.

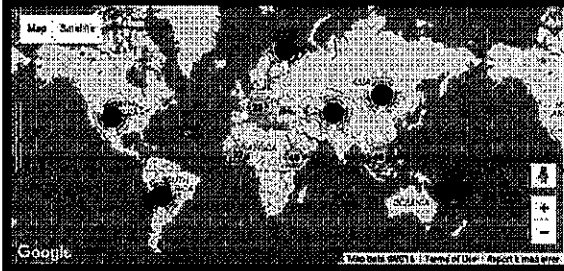
<sup>39</sup> Access provided by Glenn Roe (Australian National University, and Senior Project Manager for ARTFL and Associate Editor for the ARTFL edition of the *Encyclopédie*).

Geo Viz

### Head words

← Previous 1 2 ... 13 14 15 16 17 18 19 20 21 ... 589 590 Next →

Original headword	# Derived place names	# Linked Geonames locations
APT	1	3
AQUILA	1	10
AQUINO	1	11
AQUITAINE	1	2
ARA ou HARA	2	168
ASA ou ARA	2	168
ARABA	1	15
ARAB	1	36
ARABIE	1	3
ARACA	1	36



Map Satellite

Google

ASA ou ARA \* ASA ou ARA, (Géog. arabique), ville de la région d'Égypte.

Derived place names  
 Ara IDENTIFIED AS Ara  
 Ara IDENTIFIED AS Ara

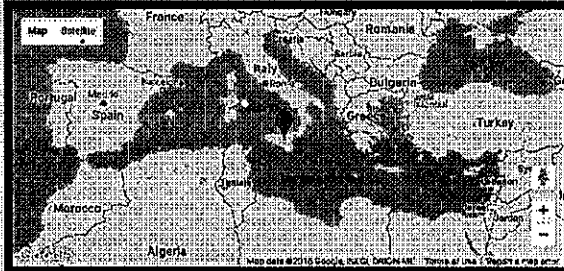
Figure 3. Screenshot of GeoViz. String match returns 168 GeoNames record locations for 'Asa'.

Geo Viz

### Head words

← Previous 1 2 3 4 5 6 7 8 9 ... 589 590 Next →

Original headword	# Derived place names	# Linked Geonames locations
ADEN	1	21
ADERNO	1	1
ADIGE	1	2
ADJA ou AGGA	1	3
ADOM ou ADON	2	15
ADOUR	1	3
ADRA	1	16
ADRIA ou HADRIA	2	15
ADRIEN	1	1
AERSCHOT	1	2



Map Satellite

Map date: 8/21/16 Google, LLC © 2016 All rights reserved. Terms of Use | Report a map error

ADERNO \* ADERNŌ, (Géog. mod.) ville de Sicile dans la vallée de Demone.

Derived place names  
 Aderno IDENTIFIED AS Aderno

Figure 4. Screenshot of GeoViz. String match returns 1 correct match for 'Aderno'. Note the latitude and longitude in the Encyclopédie full text below the map.

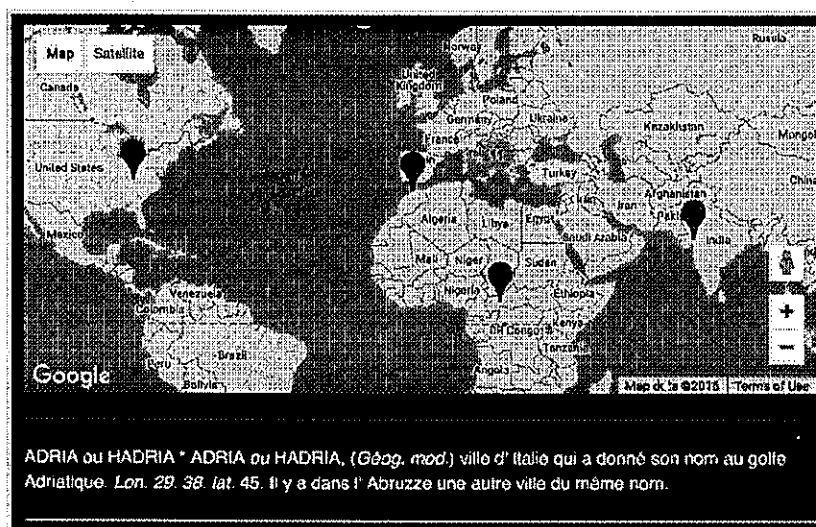


Figure 5. Detail of GeoViz. String match returns several results. We plan to use the latitude information in the source text to bound the results.

Assuming GeoViz is useful in selecting the correct gazetteer file to match to a place name reference in historical texts, we still need to develop ways of recognizing place names within those texts. Switching from working with the *Encyclopédie* headwords to locating place names within the full text meant finding tools and processes that could deal with structured language, and do so in historical French. There are three main ways that we could approach these challenges. First, we could hand-annotate the full text to select every place name, and then feed these into GeoViz as we did with the headwords. Second, we could use the Recogito tool developed by the Pelagios project to hand locate places with their preselected gazetteer locations.<sup>40</sup> Third, we could use the Edinburgh Geoparser to automate both place-name recognition and geolocation.

Working with Recogito is a good option for text files that are already clean.<sup>41</sup> In its late 2016 release, it is possible to tag place names (some automatically identified, some hand identified) with GeoNames records in addition to ancient and medieval gazetteers. Recogito is a powerful tool for automatically adding LOD metadata to place-name mentions in historical documents. It is possible to add additional information, such as a source citation or a temporal attribute, to each tagged place name. We could use Recogito both to sift through documents by hand and associate early modern places with existing gazetteer records. This kind of data emphasizes links across gazetteers and to source material. Extremely time consuming, it is useful when it is important to compare results to automated findings with the geoparser or to create new lexicons for the geoparser to improve the results of the named-entity recognition scripts.

<sup>40</sup> Rainer Simon, Elton Barker, Leif Isaksen, and Pau de Soto Cañamares, 'Linking early geospatial documents, one place at a time: annotation of geographic documents with Recogito', *e-Perimtron* 10.2 (2015), p.49-59.

<sup>41</sup> Available at [<http://recogito.pelagios.org/>].

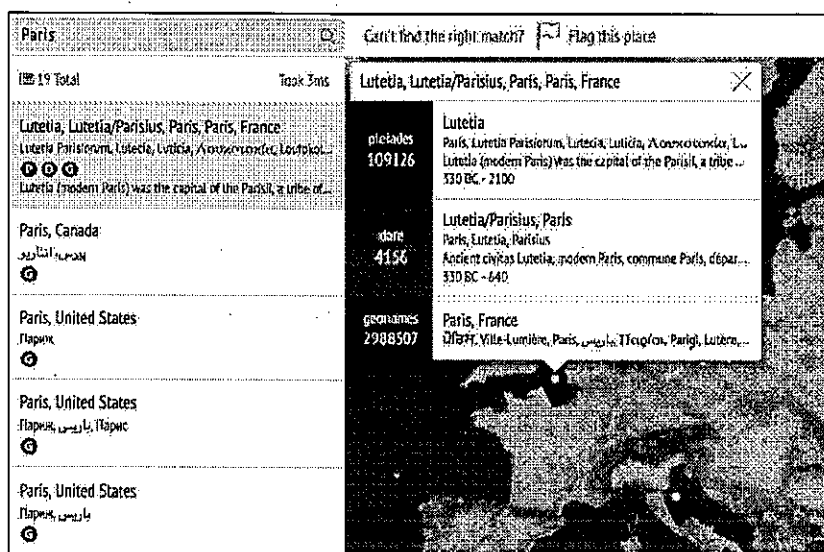


Figure 6. Screenshot of search for Paris when annotating Jaucourt, 'Paris', *Encyclopédie*.

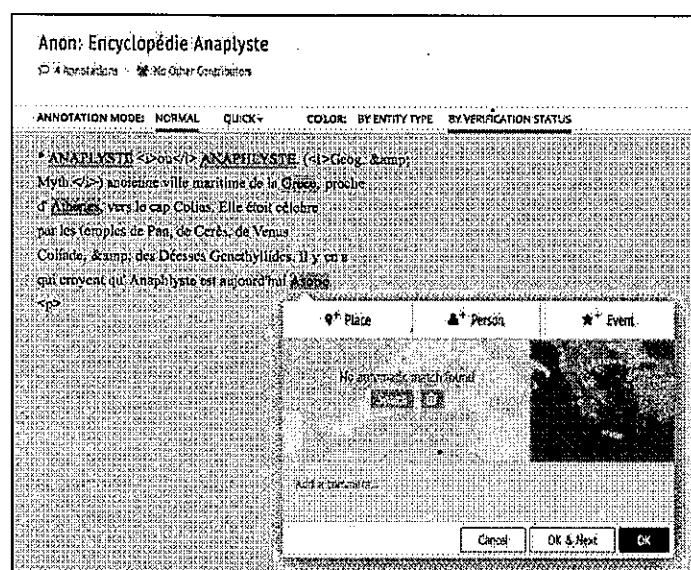


Figure 7. Screenshot of 'Anaplyste' *Encyclopédie* article uploaded to Recogito. One of the challenges that arises with Recogito for early modern data is dealing with French spellings. Another is the inability to suggest locations for places that do not pull up a gazetteer record, but this is more a limitation for us given our wish to alter the gazetteer rather than most users who will rarely find a place they cannot match. Diderot, 'Anaplyste', *Encyclopédie*.

The most experimental work we have done to date has been with the Edinburgh Geoparser.<sup>42</sup> To date, the geoparser has only been used with English documents.<sup>43</sup> By combining the tools of natural language processing and geo-resolution, it can automate the identification and location of place names in texts. This capability means that we could use the

<sup>42</sup> Available at [<http://www.ltg.ed.ac.uk/software/geoparser/>].

<sup>43</sup> Claire Grover has made adaptations to the geoparser part-of-speech tagger scripts to work with French texts. No other adaptations have yet been made to switch languages.

geoparser to scan clean text files to accurately match place names to locations that have records in existing gazetteers (similar to Recogito).

Once operational, the geoparser runs two processes, ideally on short chunks of texts. (Rather than feeding it one giant file of all the geography entries from the *Encyclopédie*, for example, we broke down the file into one new file for each entry, or 14,547 files.) It deploys the Named Entity Recognition (NER) capabilities of the Stanford NLP tools to tag certain kinds of proper nouns in the texts. The Part-of-Speech (POS) tagger that is part of this process is the only script that has been adapted to accommodate French language so far.<sup>44</sup> Other parts of the process, perhaps most importantly the Named Entity Recognition (NER) scripts, have not been adapted to French. The Stanford NLP tools do not include a French NER package, but other open source NLP software does.<sup>45</sup> We hope to try these out in the near future to improve NER results in the geoparser for French texts. In the meantime, van de Camp and I are each hand annotating places in the same subset of 100 randomly selected *Encyclopédie* entries (in Recogito). We will be able to compare results from hand annotations to the geoparser outputs and decide whether (1) to retrain the NER scripts on eighteenth-century French (versus the recent NER French training data) and/or (2) to improve the lexicon of place names we provide the NEW. The first strategy helps the NER find place names in the context of full text (and, ideally, period texts), while the second merely tells the NER that a particular word is in fact a place. Both of these may be necessary and our initial comparison with the 100 entries will provide better information about NER performance.

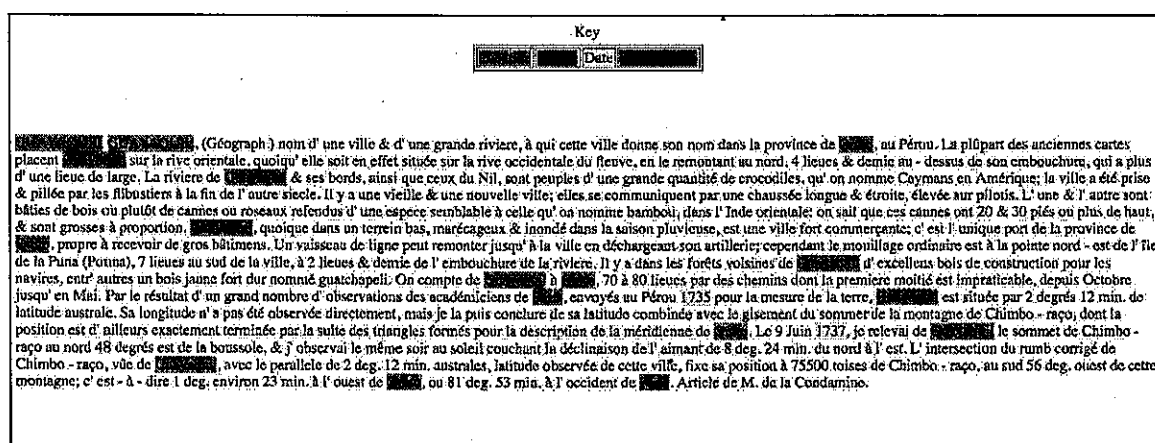


Figure 8. Screenshot of output file after NER process for La Condamine, 'Guayaquil', *Encyclopédie*, [http://encyclopedia.uchicago.edu/].

Once place names are identified, the next set of scripts calls local or web-based gazetteers and then tries to match each place name to one or several gazetteer record(s). The

<sup>44</sup> The Stanford Natural Language Processing Group provides open source NLP software including the Part-of-Speech Tagger trained on French at [http://nlp.stanford.edu/software/tagger.shtml].

<sup>45</sup> See [http://www.opener-project.eu/] and [https://opennlp.apache.org/]. On Apache OpenNLP, see Andoni Azpeitia, Montse Cuadros, Seán Gaines, and German Rigau, 'NERC-fr: Supervised Named Entity Recognition for French', *Text, speech and dialogue: 17<sup>th</sup> international conference*, ed. Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala (Springer, 2014), p.158-65.



geoparser is already capable of connecting to different gazetteer web services, including GeoNames, Pleiades+, the Digital Atlas of the Roman Empire (DARE), and DEEP (Digital Exposure of English Place-Names). Currently, as shown in Figure 9, the geoparser only returns one match from GeoNames. Given our preference for matching to French (and Latin and Greek) spellings, we aim to customize some gazetteer calls to express a preference for French or Latin results. In GeoNames, for example, this is possible because all alternate names in the database are tagged by language. We plan to customize local GeoNames and Pleiades+ gazetteers with language and similarity preferences. The latter will aid the toponym recognition process to take place names mentioned within the same entry into account in selecting the most relevant coordinate matches.

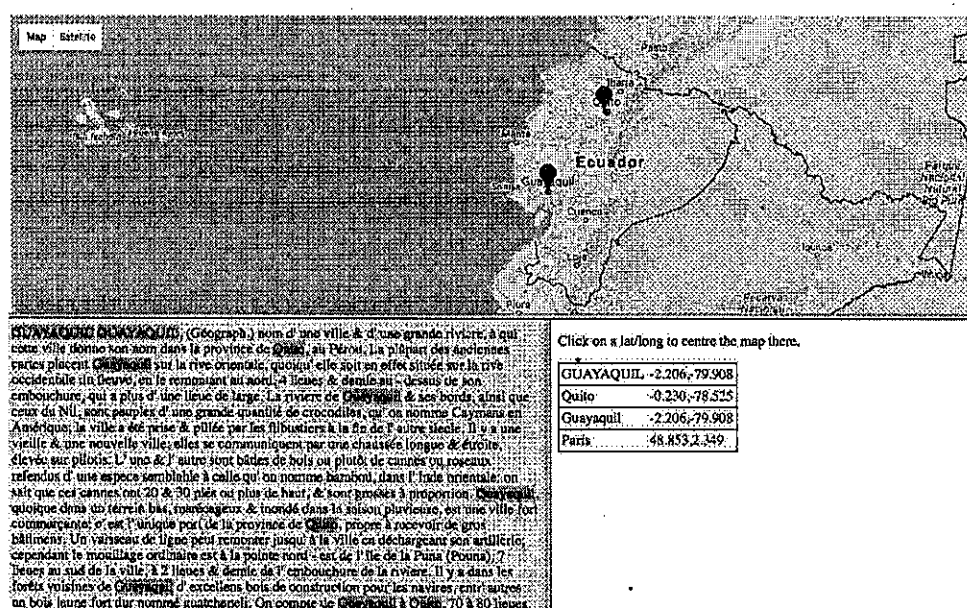


Figure 9. Screenshot of output file after toponym resolution process for Guayaquil article.

Like the recently completed DEEP project, we hope to use the geoparser to create our early modern gazetteer. We can then feed correct matches between place names stated in historical sources and GeoNames records back into a local version of Recogito or the geoparser so that it is available to other researchers.<sup>46</sup> Other humanities projects have created customized versions of the geoparser for corpora of historical texts. The Lake District Travel literature project, the Trading Consequences Project, and the Google Ancient Places project have all

<sup>46</sup> Beatrice Alex, Kate Byrne, Claire Grover, and Richard Tobin, 'Adapting the Edinburgh Geoparser for historical georeferencing', *International journal of humanities and arts computing* vol.9, no.1 (2015), p.27. See also Claire Grover and Richard Tobin, 'A gazetteer and georeferencing for historical English documents', *Proceedings of the 8<sup>th</sup> workshop on language technology for cultural heritage, social sciences, and humanities (LaTeCH) @EACL 2014*, p.119-27. DEEP can be searched at [<http://placenames.org.uk/>] in addition to being a gazetteer option with the Edinburgh Geoparser. DEEP itself grew out of the collective scholarship and attention to historical English place names. See, for an overview, Jayne Carroll, 'Perceiving place through time: English place-name studies, 1924-2013', in *Perceptions of place: twenty-first-century interpretations of English place-name studies*, ed. Jayne Carroll and David N. Parsons (English Place-Name Society, 2013), p.xiii-xxxvii.

worked closely with the developers to prepare their texts, refine the NLP tools, and write rules for the NER-tagging and georesolution processes.<sup>47</sup> Our texts are most similar to the Lake District project, and so we are seeing in initial tests the same kinds of errors Gregory outlined recently: (1) place names not being used as place names but tagged as such (e.g., the ‘King of France’ is a person, not a place); (2) place names are not being matched to a record in the gazetteers; and (3) place names are being matched to the incorrect record with wrong coordinates.<sup>48</sup>

Even though this is the ‘automatic’ option for annotating and georeferencing historical texts, it still requires dozens, if not hundreds, of hours of handwork by multiple people. Is it worthwhile?<sup>49</sup> Initial results for the Lake District, GAP, and Trading Consequences project are positive. In the Trading Consequences project, for example, 85 percent of the toponyms selected by the geoparser were within 5 kilometers of a gold-standard text prepared by hand annotation.<sup>50</sup> As Jim Clifford and his colleagues note, it was the give and take between the historians and the geoparser developers that allowed improvements to be made over time to the Trading Consequences results. The experience of working with historical documents of different kinds has prepared the geoparser team to adapt more readily when new corpora become candidates for geographical text analysis.<sup>51</sup> It would be folly to assume that the geoparser will automate one hundred percent of data entry. Instead, the geoparser permits us to take a broader view on a large amount of text so that we can better choose which parts of the text to explore more carefully. Paired with the work to feed lexicons of early modern place names (such as the *Encyclopédie* headwords) and potentially the retraining of the NER tool on hand-annotated texts, we should produce a useful, and unprecedentedly large collection of metadata about place names occurring in early modern texts. Output from the geoparser as well as any hand matching we perform will generate the early modern gazetteer.

Because we will develop the gazetteer simultaneously with MPCE databases, we are developing workflows for locating early modern places as a core part of MPCE tasks. For example, in the work to enter evidence from booksellers’ catalogs about the sale of novels, we

---

<sup>47</sup> See Jim Clifford, Beatrice Alex, Colin M. Coates, Ewan Klein, and Andrew Watson, ‘Geoparsing history: locating commodities in ten million pages of nineteenth-century sources’, *Historical Methods: a Journal of Quantitative and Interdisciplinary History* 49.3 (2016), p.115-31, Uta Hinrichs, Beatrice Alex, Jim Clifford, Andrew Watson, Aaron Quigley, Ewan Klein, Colin M. Coates, ‘Trading consequences: a case study of combining text mining and visualization to facilitate document exploration’, *Digital Scholarship in the Humanities* 30.1 (2015), p.i50-i75, C. J. Rupp, Paul Rayson, Alistair Baron, Christopher Donaldson, Ian Gregory, Andrew Hardie, and Patricia Murrieta-Flores, ‘Customising geoparsing and georeferencing for historical texts’, *IEEE international conference on big data* (2013), p.59-62, Ian Gregory et al., ‘Spatializing and analyzing digital texts’, Ian Gregory and Christopher Donaldson, ‘Geographical text analysis: digital cartographies of Lake District literature’, in *Literary Mapping in the Digital Age*, ed. David Cooper, Christopher Donaldson, and Patricia Murrieta-Flores (Routledge, 2016), p.67-87, and Claire Grover, Richard Tobin, Kate Byrne, Matthew Woollard, James Reid, Stuart Dunn, and Julian Ball, ‘Use of the Edinburgh Geoparser for Georeferencing Digitized Historical Collections’, *Philosophical Transactions of the Royal Society A*, 368 (2010), p.3875-89.

<sup>48</sup> Gregory et al., ‘Spatializing and analyzing digital texts’ p.158. Beyond these kinds of errors, one also has to account for OCR errors prevalent in scans of historical texts. Trading Consequences documented this as one of their challenges. Clifford et al., ‘Geoparsing history’, p.122-23.

<sup>49</sup> Gregory et al., ‘Spatializing and analyzing digital texts’, p.159.

<sup>50</sup> Clifford et al., ‘Geoparsing history’, p.123.

<sup>51</sup> Clifford et al., ‘Geoparsing history’, p.123.

would like to include a process for verifying publication locations and bookseller locations. Prior to opening up our crowdsourcing tool, I will curate a list of bookseller places and publisher places which have been prematched to GeoNames records. Rather than asking research assistants or crowdsourcing participants to pin down a location on a map, the metadata editor (with an interface similar to GeoVis) will present a series of options for locations from which the participant could select the most relevant choice. These pre-located options would already include interconnection-friendly metadata. By incorporating this selection task into the process of entering data about booksellers' catalogs, we avoid having to hand search for locations and individually enter coordinates for new each place and each catalog and aids us in georesolving places that other scholars are likely to encounter. The book history locations derived from *Encyclopédie* results in the geoparser will be our first 'mini gazetteer'.

Working with the *Encyclopédie* data has raised some issues that are specific to our sources. Early modern texts thought nothing of mixing real places with mythical, fantastical, biblical, and no-longer-existent places. How will we deal with these? The proliferation of ancient and biblical locations in pre-nineteenth-century reference sources makes this a nonnegligible question. For the *Encyclopédie* at least, we can rely somewhat on the internal categorization of places as ancient or modern. In other texts, we should be able to rely on contextual clues to make this distinction. Ideally, our metadata will account for these different types of places and support links to the relevant classical gazetteers. For purely imaginary places, we may develop other means of locating and mapping based on computational linguistics methods for determining the statistical linguistic frequencies of these locations named in sources.<sup>52</sup>

Though we are only at the beginning of the MPCE project, our aims to redesign our metadata, foster collaboration with other early modern projects, and reassess how we can use newly digitized early modern geographical texts are, to be sure, ambitious. Geographic Text Analysis on early modern French materials should result in both an invaluable, openly accessible gazetteer and a range of new interpretations of early modern geographic cultures. GTA requires patience. Plenty of time is needed to get to the point where the program runs. Preparing clean OCR scans, customizing the scripts, preparing gazetteers, lexicons, and training data, and analyzing the output files—all told, working with the 'automated' processes is anything but fast. Just as we learn about our sources when we sit for days in the archives, we also benefit from the time spent debugging scripts. MPCE will continue to depend heavily on expert time, but as we create sample sets of located places (thanks to the *Encyclopédie* tests), we hope to experiment with allowing volunteers to select for the best GeoNames (or another gazetteer) match. Granted, our volunteers will have to be fluent in French. They will likely be other French studies scholars or advanced students. This may not quite qualify as crowdsourcing, but perhaps our slight widening of the net for recording metadata is nonetheless useful in spreading the word about the possibilities of LOD for early modern research.

Improved GTA techniques for semi-automating the identification and (where possible) georesolution of attestations of places in early modern texts is the necessary first step towards

---

<sup>52</sup> M. M. Louwerse and N. Benesh, 'Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth', *Cognitive science* 36 (2012), p.1556-69.

the construction of this gazetteer.<sup>53</sup> In addition to considering the advantages of building gazetteers as LOD that will give digital humanities projects a shared metadata foundation, there is a broader consequence of this infrastructure work. Access both to these GTA tools (a pipeline of Natural Language Processing components designed specifically for early modern French language) and the gazetteer will make spatial information in a range of sources more discoverable and accessible. Open source GTA tools finely tuned for specific periods and document types create opportunities for others to identify and locate place names in non-geographic reference type texts (newspapers, diaries; government documents, business records). Not just points on a map or entries in a gazetteer, GTA results are data that can be used to explore the rhetoric of place and the relationship between place and other types of entities (people, objects, events, institutions). Collating these results—whether as part of archival or library metadata, or during scholarly projects—would allow for spatial views of source material impossible before the advent of Named Entity Recognition.

Finally, expanded use of GTA tools generates conversations about the historical and current ethics of place naming because, for example, of the choices one must make to define the links between different gazetteer records. GTA can help early modernists to frame digital approaches to spatial history with an awareness of the multiplicity of local and foreign (Eurocentric) naming conventions. Geospatial data built from early modern French sources will foster spatially-driven research questions and addresses the politics of naming. Making the tools and outputs of this research available are the goals of our future research.

---

<sup>53</sup> K. McDonough and M. van de Camp, 'Mapping the Encyclopédie: Working Towards an Early Modern Digital Gazetteer,' *Proceedings of the 1<sup>st</sup> ACM SIGSPATIAL Workshop on Geospatial Humanities* (2017), 16-22.