

METHOD

Laying the Foundation for Community-Driven, Open Cultural Gazetteers

Randa El Khatib

University of Victoria, CA
khatib@uvic.ca

Geospatial humanities projects rely on information found in gazetteers to supply the infrastructure for projects. However, a majority of spatial gazetteers provide place names and geographical coordinates but lack contextualizing information that give meaning to a place, making them insufficient resources for humanities inquiry. In this article, I explore contemporary approaches to data collection and models for cultural gazetteers set forth by early modern chorographical traditions to lay the foundation for building community-driven, open cultural gazetteers. Concurrently, the role of the public in providing Volunteered Geographical Information (VGI) by harnessing user-friendly tools is explored.

Keywords: cultural gazetteers; topophilia; volunteered geographical information; geospatial humanities; prototyping

Introduction

Geospatial humanities, a subfield of digital humanities that applies Geographical Information Systems (GIS) technologies to the humanistic study of spatiality in texts, relies heavily on information found in geographical databases for supplying the backbone for projects (Kallaher and Gamble 2017). Currently, a majority of digital resources that supply geodata for humanities projects are spatial gazetteers which, in addition to place names and their geographical coordinates, conventionally provide information about the type of entity, population count, and elevation of a location. In some isolated cases, gazetteers contain additional information such as alternative place name spellings; although this type of information is useful for humanities research, especially for those working with historical material, existing methodologies for automatically extracting that information are not always straightforward. Concurrently, contemporary spatial repositories are overflowing with information from volunteered geodata, a phenomenon expedited by a combination of user-friendly technology for recording and uploading spatial information, and an increased participation of the public in data contribution. Information collected through these devices is often centered on cartographic exactitude and a detached representation of space, while lacking data that animates and contextualizes place name occurrences. In order to build cultural gazetteers that would inform humanities research, it is worthwhile to assimilate models from early modern chorographies using contemporary approaches to data collection.

Volunteered Geographical Information

Volunteered geographical information (VGI) is among the most widespread user-generated content of the last decade. Apps such as *Waze* (Waze 2006), a community-driven navigation app that aids drivers by supplying real-time information about traffic, roadblocks, and other road-related information uploaded by drivers, are plentiful and are used by millions daily. In this way, VGI makes up for unavailability of geographical data and counters restrictions imposed by commercial providers. An example is OpenStreetMap (OSM), a project driven by VGI aimed at creating a free, editable map of the world. The platform supplies the infrastructure for a number of other projects, such as the open-source JavaScript library for interactive maps called *Leaflet* (Agafonkin 2017). Aggregated by millions of users and expanding daily, OSM contains geographical points of interest from users around the world. Two of OSM's primary hosts are University College London and Imperial College London, making it a noteworthy example of a technology-driven open knowledge initiative

founded on collaboration between the university and members of the public. At the same time, VGI is increasingly being employed in the study of linguistic landscapes, a multidisciplinary area that draws on linguistics, sociology, geography, and anthropology (Ben-Rafael, Shohamy & Barni 2010), to study how '[t]he language of public road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings combine to form the linguistic landscape of a given territory, region, or urban agglomeration' (Landry and Bourhis 1997, 25). These projects are increasingly relying on user-friendly apps to record and upload geographical information in real time.

Digital Gazetteers

Digital gazetteers are the knowledge bases that supply infrastructural data for geohumanities research (Mostern and Johnson 2008, 1091). Gazetteers are curated by global, transnational, and national web services that provide geographical information; some gazetteers also include VGI. For instance, *GeoNames*, one of the largest open source digital gazetteers, is compiled through a combination of web services and user correction and expansion. Some libraries in *GeoNames* are created entirely by members of the community and are exportable in a number of computer languages for reuse. *GeoNames* contains over nine million unique entries and five and a half million alternate place names, making it a valuable geographical database for historical research. According to Gao et al. (2017), '[e]xisting GIS and spatial databases are mature in representing space, but limited in representing place' (173). Elaborating on the conceptual difference between space and place is Yi-Fu Tuan (1979), who argues that place not only signifies a unit of location within the larger frame of space, but also 'a reality to be clarified and understood from the perspectives of the people who have given it meaning' (387). Consequently, a place begins to emerge by extending the focus beyond the place name mention to the different meanings given to it through narratives. In 'On Historical Gazetteers,' Ruth Mostern, Humphrey Southall, and Lex Berman (2016) argue that humanities research often requires information about places, the locations of which have not been determined with certainty; how places and names have evolved over time; and descriptions of places that contextualize them. This type of information is not typically supported by conventional spatial gazetteers, which necessarily excludes an entire dimension of subjective or ambiguous spatial information that is so central to humanities research. As a response, the authors call for building cultural gazetteers that will move beyond the geographical description of a place to additionally provide information about its historical significance and identity. Bringing together contemporary methodologies, such as extracting information from gazetteers, relying on user-friendly apps, and building on volunteered geodata, can lay the foundation for open, community-based, cultural gazetteers that focus on a more in-depth representation of places.

Literary Topophilia and the Role of the Public

Apart from the technological support for transmitting geodata through mobile navigation devices, a crucial motivation for VGI and other forms of volunteered data appears to stem from a new form of social responsibility that has been expanding in the last decades of the digital age. This phenomenon can be witnessed in the overwhelming amount of contributions to global knowledge on the Internet. But how can we channel this practice into motivating users to contribute information about places to enrich cultural gazetteers? In a sense, members of the public already are—and have always been—inquisitive about places of their literary topophilia. The term 'topophilia' is adopted from Yi-Fu Tuan's *Topophilia: A Study of Environmental Perception, Attitudes, and Values* (1972), and is also discussed at length by Gaston Bachelard in *Poetics of Space*. Topophilia is broadly described by Tuan (1972) as the 'affective bond between people and place or setting' (4); this general definition is adopted here to describe the love of places or settings in texts and fiction, broadly categorized as 'literary.' For Bachelard, topophilia occupies a more poetic space that can be applied to places we love and feel a connection to, such as the places where we grew up. Bachelard argues that spatial associations are so strong, in fact, that memory functions in a spatial rather than temporal dimension (1994); according to him, remembering a moment in the past first happens by remembering the setting, and then the time and context. Literary topophilia does not conform to the corporeal dimension outlined by Tuan and Bachelard and is more of an abstract association, although this form of topophilia is not necessarily perceived as less intimate or real. Entire volumes have been written on places of literary topophilia, such as *The Dictionary of Imaginary Places* by Alberto Manguel and Gianni Guadalupi (1999), a 755-page book that names and describes thousands of imaginary places. Places of literary topophilia are ironically separate from academia and are often supplied by popular or public culture. Some examples of places that have culturally induced literary topophilia are J.K. Rowling's Hogwarts, J.R.R. Tolkien's Middle Earth, Arthur Conan Doyle's Baker Street, Charles Dickens' London, James Joyce's Dublin, and Jane Austen's

Pemberley, among many others. These fictional places occupy such a significant part of our imagination that some of them have been commercially recreated in the real world, such as the Shire in New Zealand, or The Wizarding World of Harry Potter chain in several countries. Places that figure in texts, whether they are entirely fictional or not, can and have become places of literary topophilia.

Contemporary Cultural Gazetteers

Cultural spatial gazetteers first appeared in print form; with the advent of the digital age, researchers have been experimenting with ways to collect, store, present, and reuse data in their digital counterparts. Some of these projects have been successfully carried out in the humanities, such as *Pelagios Commons*, *Pleiades* (Bagnall et al. 2016), and 'The *MoEML* Gazetteer of Early Modern London' (Jenstad and McLean-Fiander n.d.). *Pelagios Commons* is an infrastructure and community for linking open geodata. It provides the platform and tools for researchers to semi-automatically annotate place name occurrences in texts using a tool called *Recogito*, and to link them internally and externally to other works. This platform supports the multilayering of information in a user-friendly interface; for example, it records the type of place, its source, languages in which it appears, a timeline for when it was mentioned, and other useful place-specific information. *Pelagios Commons* draws its geodata from gazetteers such as *Pleiades* (Bagnall et al. 2016) and *GeoNames*. *Pleiades* (Bagnall et al. 2016) is an open, community-built gazetteer that includes information about ancient places, is navigable by individual users, and supplies geodata for computational extraction. 'The *MoEML* Gazetteer of Early Modern London' is the first digital gazetteer for place names in early modern London, circa 1550–1650, with categories comprising Variant Toponym, Authority Name, @xml:id, Agas Map Reference, Other Variant Names and Spellings, and Location Type. The aforementioned resources are notable examples of community-driven platforms and gazetteers that provide valuable information and enable research questions; currently, all three are necessarily limited to specific time periods based on their disciplinary foci.

An alternative example of a community-driven open cultural gazetteer is *Wikipedia* (Wikipedia, The Free Encyclopedia 2001), one of the most comprehensive single resources for spatial data—that includes real and fictional places and covers an indefinite span of time. *Wikipedia* is linked to *GeoHack*, a gazetteer that provides geographical data in multiple formats and includes corresponding maps. The primary advantage of *Wikipedia* as a community-driven cultural gazetteer is that it supplies extensive notes about places, counting place name spellings, place significance, historical context, change over time, and other contextual information that users contribute. Such discursive information can help identify a given geographical entity, which is especially useful when working with historical texts—not least because many place names correspond to locations that are not determined with certainty and because it is not always clear what location a place name is referring to. In these cases, instead of suspending ambiguous geographical entries altogether, as most spatial repositories would, *Wikipedia* retains this data and editors often provide a map pointing to the general area where the place might have been located. Retaining this information adds an entire dimension of data that enables robust research over cherry-picking available geographical entries. It is *Wikipedia's* less rigid structure that supports the capacity for continuously enriching entries that gives it the opportunity to augment. However, this open structure is precisely its drawback when attempting to automatically extract data from *Wikipedia* for large projects: namely that categories vary from page to page rather than adhering to a standardized format. This configuration makes *Wikipedia* a valuable resource for conducting research on historical places, and less useful for automatic extraction that relies on standardized categories. Attempts are being made by the Wikimedia Foundation to address this issue. *Wikidata* (Wikidata 2012) is a collaboratively edited knowledge base published under a Creative Commons license that is intended to supply open data for Wikipedia and other Wikimedia Foundation Projects, as well as for anyone who wants to use their structured information. However, Wikidata is still at too early a stage (with respect to *Wikipedia*) to rely on for geodata.

Revisiting Early Modern Chorographical Traditions

Enriched gazetteers align more closely with the model set out by earlier humanities traditions in works that focus on recording and describing places. Mostern, Southall, and Berman (2016) provide an overview of the history of these traditions and trace their collapse to the nineteenth century. This geographically-oriented genre was pioneered by early modern writers roughly around the sixteenth century, at a time that was marked with a heightened attention towards spatiality, which is palpably evident in the abundance of published atlases, chorographies, and travelogues. One example of this genre is William Camden's *Britannia*, first published in 1586 (Camden 1633) and followed by numerous editions. *Britannia* is the first chorographical

survey describing the islands of Great Britain and Ireland in which Camden traces the historical roots of the Silures, Brigantes, Scots, Saxons, and Normans in an attempt to tie Britain to its antiquity and to portray how they were unified through the same language and land. Consider *Britannia's* entry on Dobuni; Camden writes:

The part, that lyeth more West beyond Severne, (which the Silures in old time possessed) along the revier Vaga or Wye, that parteth England and Wales, was wholly bespred with thicketall woods: we call it at this day, Deane-forrest: The Latin writers some name it of the Danes Danice Sylva the Danes wood, others with Girald, the Wood of Danubia. But I would thinke...it was called Deane, for Arden. Which term both Gauls and Britains in ancient times may seeme to have used for a wood. (1833, 358)

In this entry, Camden provides information about places, alternative place names and place name spellings, the possible origins of their names, and how they were called by different communities. Additionally, he explains that numerous crimes used to occur in those woods and that '[f]or the reigne of Henry the sixt... there were lawes made by authority of the Parliament, for to restraine them. But since that Mines of Iron were heere found out, those thicke woods began to wax thin little by little' (1833, 358). This passage contains historically significant information about natural resources and deforestation in that area, which gives readers an idea about what the landscape looked like before the unearthing of natural resources began and under whose reign it occurred. Camden's entry captures a more localized and specific representation of a place through its narrative style. *Britannia* and many other chorographies and itineraries of the time focus on topographical features, while simultaneously recording anthropological information. Instead of implying that all the information listed in Camden's Dobuni entry needs to be included in cultural gazetteers, the example is meant to show that cultural gazetteers were fairly common and that we can learn a great deal about modelling current digital cultural gazetteers by exploring that tradition.

Conclusion

Projects that rely on VGI for building gazetteers exist. We need to decide which models work best for geospatial humanities research, and what type of information should be included in cultural gazetteers. For example, should we model gazetteers based on *Wikipedia*? If so, do we keep contributing to *Wikipedia*, since it has accumulated a tremendous amount of information already, and focus our energies on making it more reliable for data extraction and manipulation instead? *Pelagios Commons*, *Pleiades*, and 'The *MoEML* Gazetteer of Early Modern London' provide strong models as well; the next steps would involve deciding on how to augment contextual information about places and aggregate them for a wider disciplinary span. Building on the notion of literary topophilia, how can volunteers harness mapping platforms to actively participate in data generation in an engaging way, such as through gamification or connecting to existing platforms? At this point, we need more experimental prototyping to develop engaging ways to contribute volunteered data to repositories; we also need to consider sustainable ways to house this information on a single open-source platform. Whichever path the development of these cultural gazetteers takes, they will not be limited to real places, but will also accommodate imaginary places of literary topophilia.

Competing Interests

The author has no competing interests to declare.

References

- Agafonkin, Vladimir. 2017. "Leaflet." *OpenStreetMap Contributors*. <https://leafletjs.com>.
- Bachelard, Gaston. 1994. *The Poetics of Space*. Boston: Beacon Press.
- Bagnall, Roger, et al. (eds.) 2016. *Pleiades: A Gazetteer of Past Places*. Accessed July 3, 2018. <http://pleiades.stoa.org>.
- Camden, William. 1633. *Camden's Britannia: Brittania*. London: Edmund Gibson.
- Gao, Song, Linna Li, Wenwen Li, Krzysztof Janowicz, and Yue Zhang. 2017. "Constructing Gazetteers from Volunteered Big Geo-Data based on Hadoop." *Computers, Environment and Urban Systems*, 61(B): 172–186. DOI: <https://doi.org/10.1016/j.compenurbysys.2014.02.004>
- GeoNames. n.d. *GeoNames*. <http://geonames.org/>.
- Jenstad, Janelle, and Kim McLean-Fiander. n.d. "The *MoEML* Gazetteer of Early Modern London." *The Map of Early Modern London*, Janelle Jenstad (ed.). Victoria: University of Victoria Accessed July 3, 2018. http://mapoflondon.uvic.ca/gazetteer_about.htm.

- Kallaher, Amelia, and Alyson Gamble. 2017. "GIS and the Humanities: Presenting a Path to Digital Scholarship with the Story Map App." *College & Undergraduate Libraries*, 24(2–4): 559–573. DOI: <https://doi.org/10.1080/10691316.2017.1327386>
- Landry, Rodrigue, and Richard Y. Bourhis. 1997. "Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study." *Journal of Language and Social Psychology*, 16(1): 23–49. DOI: <https://doi.org/10.1177/0261927X970161002>
- Manguel, Alberto, and Gianni Guadalupi. 1999. *The Dictionary of Imaginary Places*. New York: Harcourt, Inc.
- Mostern, Ruth, Humphrey Southall, and Merrick Lex Berman. (eds.) 2016. "On Historical Gazetteers." In: *Placing Names: Enriching and Integrating Gazetteers*. Bloomington, IN: Indiana University Press.
- Mostern, Ruth, and Ian Johnson. 2008. "From Named Place to Naming Event: Creating Gazetteers for History." *International Journal of Geographical Information Science*, 22(10): 1091–1108. DOI: <https://doi.org/10.1080/13658810701851438>
- Pelagios Commons. n.d. *Commons In A Box*. Accessed July 5, 2018. <http://commons.pelagios.org/>.
- "Recogito." n.d. *Pelagios Commons*. Accessed July 5, 2018. <http://recogito.pelagios.org/>.
- Shohamy, Elana, Eliezer Ben-Rafael, and Monica Barni. (eds.) 2010. *Linguistic Landscape in the City*. Bristol: Multilingual Matters. DOI: <https://doi.org/10.21832/9781847692993>
- Tuan, Yi-Fu. 1972. *Topophilia – A Study of Environmental Perceptions, Attitudes, and Values*. New York: Columbia University Press.
- Tuan, Yi-Fu. 1979. "Space and Place: Humanistic Perspective." In: *Philosophy in Geography*, 387–427. Dordrecht: Springer Netherlands.
- Waze. 2006. *Waze Mobile*. <https://www.waze.com/>.
- Wikidata. 2012. *The Wikimedia Foundation*. <https://www.wikidata.org/>.
- Wikipedia, The Free Encyclopedia. 2001. *The Wikimedia Foundation*. <http://wikipedia.org/>.

How to cite this article: El Khatib, Randa. 2019. Laying the Foundation for Community-Driven, Open Cultural Gazetteers. *KULA: knowledge creation, dissemination, and preservation studies* 3(1): 21. DOI: <https://doi.org/10.5334/kula.53>

Submitted: 07 June 2018

Accepted: 25 September 2018

Published: 27 February 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



KULA: knowledge creation, dissemination, and preservation studies is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access logo, which is a stylized 'a' inside a circle.