Eliciting the Ancient Geography from a Digital Library of Latin Texts

Maurizio Lana[1] and Timothy Tambassi[2(✉)]

[1] Dipartimento di Studi Umanistici, Università del Piemonte Orientale,
Piazza Roma 36, 13100 Vercelli, Italy
maurizio.lana@uniupo.it
[2] ICUB, University of Bucharest, 1 Dimitrie Brandza St.,
060102 Bucharest, Romania
timothy.tambassi@gmail.com

Abstract. *Geolat – Geography for Latin Literature* is a research project, aimed at making accessible a digital library containing the works of Latin literature (from its origins in 240 BCE to the end of the Roman Empire in 476 CE) through a query interface of geographic/cartographic type representing the geographic knowledge expressed in the Latin texts themselves. A core activity of the project has been the development of the ontology *GO!*, which describes the geographical knowledge contained in the texts of the library. The ontologically annotated texts will allow for a variety of scientifically relevant uses, apart from the geo-based browsing: for example the production of digital and printed critical editions. The project is under development at Dipartimento di Studi Umanistici of Università del Piemonte Orientale, and financially supported by Fondazione Compagnia di San Paolo.

Keywords: Geography · Ontology · OWL · Web · Classical latin texts
Digital library

1 Geolat, GO! and the Plurality of Areas of Research

Geolat – Geography for Latin Literature is a research project, aimed at making accessible a digital library containing the works of Latin literature (from its origins in 240 BCE to the end of the Roman Empire in 476 CE) through a query interface of geographic/cartographic type, representing the geographic knowledge expressed in the Latin texts themselves[1]. The key points and the most relevant aspects of Geolat project are various (LOD, crowdsourcing, Open Access, CC licenses, semantic annotation of geographical references, URIs for annotation of places, and others) but here we want to focus on its ontology.

In particular, the introduction of a specific geo-ontology represents a fundamental innovation compared with similar projects focused on the ancient world. Some questions arise, the main ones being:

– what does ontology mean in this context?

---

[1] M. Lana is specifically responsible for Sects. 1 and 5, and T. Tambassi for Sects. 2–4.

- what kinds of problems does it deal with?
- what are the main objectives and features of *GO!* - the geo-ontology of Geolat?

First of all, we can distinguish three different disciplinary areas which make up this specific geo-ontological domain:

- computer science,
- contemporary geography,
- ancient geography.

In the domain of computer science, ontology is a structure aimed to describe the categorical hierarchy of a specific domain, analysing its basic constituents (entities like objects, events, processes, etc.), the properties characterizing them and the relationships which correlate them - using a language (usually OWL) that is understood both by the machines and the humans. The resulting structured representation of knowledge allows to resolve conceptual or terminological inconsistencies and provides a lexical or taxonomic framework for the representation of knowledge [7, 14, 16].

From a geographical point of view, the aim of a geo-ontology is to analyse the mesoscopic world of geographical partitions in order to:

- establish whether and what kinds of geographical entities exist, their borders, their spatial representation (in maps, software, etc.), their mereological and topological relations, and their location;
- determinate how they can be defined and classified in an ontological system which gather them together;
- argue whether and how the geographic descriptions of reality emerging from common sense can be combined with descriptions derived from different scientific disciplines [2, 3, 5, 8, 9, 15, 17–19].

Finally, if one wants to investigate the geographical knowledge expressed in the Latin texts[2] (which is a component of the Roman culture) some specific problems closely interconnected, and sharing the vagueness of data and information available, have to be taken into account. They can be distinguished in topological, source and methodological problems. Topological problems have to do e.g. with: measurement and measurability of distances (and their different units of measurement), location of places and absolute vs relative distances/coordinates. Problems concerning documentation and sources, have to do e.g. with: lack of reliability and homogeneity of some data, disagreement among authors, difficulty or impossibility of autoptical confirms and isolation of properly geographical contents from the rest of the texts. The third kind of problems is strictly connected with the second ones and refers to methods and to the (multiplicity of) approaches to ancient geographical investigation which involved e.g. heterogeneity of aims, points of view, interpretations and perspectives (sometimes overlapped) through which the information was transmitted, processed and implemented, the importance of imagination (and mental maps), the necessity of folk

---

[2] Until now the Latin texts are offered with no translation because even if translations do exist for them, recent translations are protected by Intellectual Property Rights; and if they are free from IPR they are generally speaking too remote from today sensibility to be usable.

theorizing, in order to understand other's mind and ancient culture [1, 6, 11]. All of these problems are mentioned because they mark the distance between the two (today and ancient) geographies involved in *GO!*.

2 Steps for the Creation of GO!

Given the plurality of domain interests, the creation of GO! [10] imposed a division of work in different steps: a search and analysis of the existing ontologies which can or could be of interest for the geography in order to start to understand to which extent they could be reused for the scope of describing the geographical knowledge contained in the Latin classical texts; this was done through a critical review of contemporary geo-ontologies[3], aimed to identify common classes[4] and properties[5], and also missing classes and properties needed to describe ancient geography - the scope was to to establish if an ontology had to be built *ex novo* or if more simply classes and properties can be selected and imported from other existing ontologies, emphasizing in this way the specific contribution of Geolat ontology to the contemporary debate.

After that the analysis of Latin literature texts started, in order to identify geographical entities, classes, properties and relations; practically speaking it meant that around 15.000 pages of translated Latin texts were read (no problem with the the translation because not concepts but things were searched for), and everything *related to geography in broad sense* was highlighted: proper nouns (e.g. Rome); common nouns (mons, mare); names of populations; space/place indications: above, below, beyond, etcetera; all the verbs having any geography related nuance: build, move, settle, etcetera; properties and relations related to these entities or describing them.

Then all the relevant passages where re-read in Latin in order to check for possible translation problems and the highlighted entities were listed by type and author subsequently agreements and disagreements among the Latin authors were analysed and highlighted, focusing on their basic distinctions; the study of the differences between ancient and contemporary geography, in terms of domains, presuppositions, representations and vagueness; the scope was that of understanding was could be expressed

---

[3] A thorough description of this phase is Tambassi T.: Rethinking Geo-Ontologies from a Philosophical Point of View. Journal of Research and Didactics in Geography (J-Reading) 2(5), 51–62 (2016), https://doi.org/10.4458/7800-04. Geo-Ontologies can be broadly distinguished among geomatics, topological and geometrical ontologies: see e.g. OGC GeoSPARQL, Spatial Schema – ISO 19107, Spatial referencing by coordinates - ISO 19111; physical and natural ontologies: see e.g. NDH Ontology (USGS) and Hydro Ontology (Spanish GeoData); human ontologies: see e.g. FAO Geopolitical Ontology. Well known ontologies like DOLCE, CIDOC-CRM, FRBR were not used because they don't offer a sufficiently detailed characterization of the geographical knowledge.

[4] "Classes are used to group individuals that have something in common in order to refer to them. […] In modeling, classes are often used to denote the set of objects comprised by a concept of human thinking, like the concept person or the concept woman." *OWL 2 Web Ontology Language Primer (Second Edition)*, https://www.w3.org/2007/OWL/wiki/Primer.

[5] "In OWL 2, we denote … relations as *properties*. Properties are further subdivided: *Object properties* relate objects to objects (like a person to their spouse), while *datatype properties* assign data values to objects (like an age to a person)" (*id.*).

using existing ontologies, and what not, and at the end of this phase it was discussed what could be reused form existing ontologies and what had to be created *ex novo*; this produced a reunification of these information in a geo-ontology for Latin literature, based on common sense classes, properties and relations, and folk conceptualizations. It allows to improve the usability of this ontology, making it more compatible with similar ontologies and conceptualizations.

An ulterior check against the Latin texts of the first phase was made to be sure that the conceptualization which was arising was effectively usable to describe the knowledge contained in those texts; and an additional similar check was made with other authors (minor historians – Eutropius, Velleius Paterculus, etc.; plus, some works from Cicero, Seneca, Vergil, Plautus, Catullus, Terentius) to be sure that relevant concepts didn't appear which were forgotten.

3   GO!: A Geo-Ontology for Latin Literature

The result of this work[6] is *GO!*, a geo-ontology which provides a description of the geographical knowledge emerging from Latin literature and an inventory of classes and relation mainly focused towards semantically annotating Latin texts, identifying the places mentioned in these texts, and connecting them with their contemporary equivalents. The fundamental scopes of this ontology are essentially four: informativeness, completeness, reusability and accessibility (both for the scientific community and for general public). The main challenge of the project has been to put together all the different disciplinary areas - which include, at least, Computer Science, Contemporary Geography, Ancient (History and) Geography, Latin Literature, Ontology of Geography - that constitute the domain of this geo-ontology. Accordingly, the idea behind this ontology was that the study of common sense (geographical) conceptualizations - that is the body of knowledge that (ancient) people have about the surrounding geographic world - could constitute a fundamental infrastructure for the ontological representation and for the communication among different areas of research.

Also thinking of the reuse the ontology is built as a collection of four interconnected modules (expressed in OWL2) freely accessible, readable, usable at the following IRIs (now that the PURL services does no more accept new contents):

https://w3id.org/geolit/ontologies/GO-TOP
https://w3id.org/geolit/ontologies/GO-PHY
https://w3id.org/geolit/ontologies/GO-HUM
https://w3id.org/geolit/ontologies/GO-FAR

The modules are open to the use of all the interested people under a CC BY-NC-SA license; modification locally managed are discouraged because they create unnecessary

---

forks of the ontology; requests to the managing team are instead welcome. Graphical representations of the modules can be found at this address: http://vowl.visualdataweb. org/webvowl/index.html#iri=https://w3id.org/geolit/ontologies/GO-TOP (replace GO-TOP with GO-PHY, GO-HUM, GO-FAR for the graphical view of the other ontologies).

4   The Modules and the Modelling Choices

GO-TOP contains 21 classes, 38 object properties, 15 datatype properties, 4 individuals. It is the *top level ontology* which connects all the other modules and contains the most general elements that describe all the geographic entities included in GO!. In particular, all the most general classes and (object and data) properties belong to the GO-TOP and are used by the other three modules.

GO-PHY contains 127 classes, 3 individuals. It imports the GO-TOP module, and includes a taxonomy which represents geographical entities with physical-natural aspects. All the classes of GO-PHY are sub-classed of *astronomical entity*, *physical entity*, *geographic entities*, *natural entities*, *event* and *terrestrial entity* classes of GO-TOP.

GO-HUM contains 204 classes, 8 object properties. It imports the GO-TOP module, and is organized in a taxonomy which constitutes an inventory of geographical entities created by humans. The high level classes imported from GO-TOP are *astronomical entity*, *anthropic entity*, *geographic entity*, *event*, *go entity*, *length*, *non-physical entity*, *physical entity* and *terrestrial entity*, from which GO-HUM defines its specific subclasses. The main specific object properties are: *fought between*, *composed by*, *has stop over*, *has length*, *has path*, *has cultural heritage of* and *won*.

GO-FAR contains 87 classes, 2 object properties. It imports the GO-TOP module, and describes all (and only) the geographic features (including places, people and events) produced by human during ancient times, with particular reference to ancient Rome as the main scope of this ontology is the annotation of Latin texts. Moreover, it includes, among others, some specific entities and classes which describe the Ancient World imported from *ancient entity*, *socio-institutional entity*, *group of people*, *populated place*, and *artifact* classes of GO-HUM, *geographic entity* from GO-TOP. Finally, it has *has real place* among the Object Properties.

An example of the connection of these modules might be represented by the class of *GO! Entities*, which includes, among others, *Natural Entities*, *Anthropic Entities*, *Astronomical Entities*, *Physical Entities*, *Terrestrial Entities* and *Unreal Entities*, as can be seen in Fig. 1.

Each of these subclasses of *GO! Entities* contains further subclasses and is characterized by specific properties and relations (*name*, *location*, *length*, *size*, *spatial relation* and so forth). In this sense, the *GO!* modelling choices allow to express a range of information about geographical places (i.e. their evolution through time as attested
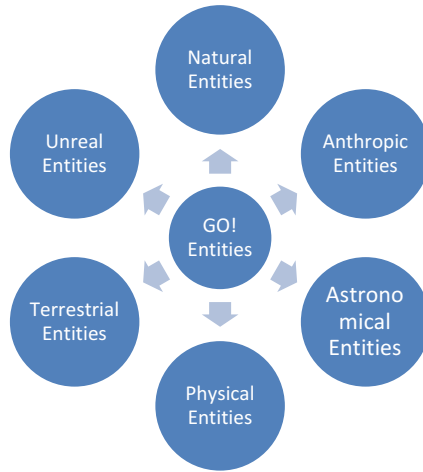
Fig. 1. GO! entities

by the texts which mention them, GPS coordinates, physical and geopolitical descriptions, switch of name, and so forth); they also allow to describe historical events connected with specific places. Connections with places data available in Pleiades are used; moreover GO! can manage imaginary places; the Open Annotation ontology is adopted to cite the passages.

5 Envisioned Uses

In the digital humanities field a geographical turn happened, whose meaning is that the geography is seen no more simply as a specific discipline rather its meaning is that of the environment where most of the activities and knowledge lie. A glue, a substrate connecting most of what exists and happens in the human world [4]. The Geolat project is a product of this idea.

In fact, in Geolat the annotation of texts based on the ontology is conceived for at least two different scopes:

- browsing and searching texts on the basis of the internal geographical content and knowledge, made visible and usable thanks to the ontology;
- production of digital (online) and printed critical editions of new type: geographical critical editions.

The first scope can be exemplified saying that the reader draws an area on a map and the system shows on the map the places mentioned in the texts of the digital library underlying the map. That is the interconnection between the library of digitally annotated texts and the map allows for a truly exploratory (the reader has not to know anything very precise about the area s/he are investigating), yet controlled (the reader can fine tune the texts collection which is analysed: period, authors, genres, …) approach to a given collection of texts. But the same cartographical query can produce lists of authors, works, passages, mentioning places belonging to the area traced by the reader on the map. Or inversely the reader starts from one or more works and can display on the map the places mentioned in those works. The apparently simple scope of searching can become very powerful if one considers that in the annotation of places events and involved people with their roles can be described. This way one can search e.g. for the geographical places where *proconsules* are mentioned. This is not possible with more simple NER approaches where the place name is a string and not - as it happens instead when using an underlying rich ontology - a complex entity of meaning full of connections and implications.

In the second scope the concept of "digital edition of a text" is developed/expanded. The "edition of a text" has today - at least for classical works, but many things are similar for the editions of medieval or modern texts - two main forms: with or without the description of the variants and their evaluation by the curator of the edition in the *apparatus criticus*. In the first case the edition is usually qualified as a *scholarly edition*, in the second one it is usually called a *variorum edition*. In both cases the editor of the texts offers what he judges being the best reconstruction of the true text issued by the hand of the author. And this type of work on the text was consolidated through thousands of years (the first critical editions of texts were produced more than 2000 years ago). The scope of a scholarly edition (be it of the first or of the second type just described) is that of offering those who want to study a given work a rich informational environment. Until these years the main type of information offered is that of an accurate description of the state of the manuscript tradition. Nowadays, tools are available allowing to expand the type of critically assessed information which is offered to the researchers who study a given text, what is both the cause and the effect of new study perspectives. For example, we are *more aware* of the fact that the geography conveys political and ideological meanings: giving names to specific points on the surface of Earth means affirming a property on them; the same is for defining boundaries; and mentioning specific place names in groups is a way of declaring interpretations of facts, or of affirming political or religious positions. This type of research interest needs the support of an edition of the text giving space to this study approach: a geographer will probably appreciate a scholarly edition enriched by a geographical apparatus that is an edition with a good comment of the text plus a full spectrum of documentation about the geographical knowledge contained in the text that is link the source describing in all the possible ways the places which are mentioned (Fig. 2):

1 Se quoque, cum transiret mare, non Ciliciam aut Lydiam—quippe tanti belli exiguam hanc
 esse mercedem—sed Persepolim, caput regni eius, Bactra deinde et Alexandria ultimique
 Orientis oram imperio destinasse. Quocumque ille fugere potuisset, ipsum sequi posse: desineret
 terrere fluminibus, quem sciret maria transisse.
5 Reges quidem haec invicem scripserant. Sed Rhodii urbem suam portusque dedebant Alexandro.
 Ille Ciliciam Socrati tradiderat, Philota regioni circa Tyrum iusso praesidere. Syriam, quae Coele
 appellatur, Andromacho Parmenio tradiderat bello, quod supererat, interfuturus.

----
1 Ciliciam: [http://www.geonames.org/8378491/cilicia.html]
  Lydiam: [http://www.trismegistos.org/place/1269]
2 Persepolim: [https://pleiades.stoa.org/places/922695 I https://vici.org/vici/20445/ I http://gazetteer.dainst.org/place/2043479]
  Bactra: [https://pleiades.stoa.org/places/961886] also known as Zaraspadum and Zariaspa
  Alexandria: Alexandria MSA [Alexandria of Egypt http://pleiades.stoa.org/places/72707 I Alexandria Eschate
  http://pleiades.stoa.org/places/59672], Alalia MSB [https://pleiades.stoa.org/places/472048]
5 Rhodii: inhabitants of Rhodos [https://pleiades.stoa.org/places/590030]
6 Ciliciam: see 1.1
  Tyrum: Tyrum MSB[http://pleiades.stoa.org/places/609564], Tyraion MSC [https://pleiades.stoa.org/places/609564 I
  http://dare.ht.lu.se/places/21528]
  Syriam ... Coele: [https://pleiades.stoa.org/places/991407 I http://www.geonames.org/8378530/syria-coele.html]

Fig. 2. Mock-up of a scholarly geographical edition

while a classical scholar will probably appreciate a scholarly variorum edition enriched
by a geographical apparatus, that is an edition where it is possible to cross-reference the
variants of the manuscript tradition with the geographical knowledge. But other types
of information could be equally interesting: historical information about the events
mentioned in the text, or prosopographical information about the persons involved in
the events, or both types merged. No single printed edition, nor single scholar could
bear the weight of this complexity:

(variorum: yes/no) AND/OR (geographical: yes/no) AND/OR (historical: yes/no)
AND/OR (prosopographical: yes/no)

which can probably only be managed if the edition is conceived as a collaborative work
of different scholars and gives origin to an edition whose content typology is
dynamically generated on the basis of the researcher's interest (strictly philological
AND/OR geographical, etc.). In other words we suggest here that the complexity of the
digital edition be conceived not at the very philological level only (as variorum edition,
with complex discussions about what it must contain and how the content must be
presented also in rapport with the printed version of this type of edition) but also as a
research tool which creates a research environment which configures itself according to
the scope and the interests of the researcher/reader (Fig. 3).

 All the above is possible because the specific value of a digital edition is not only in
the digital form of representation of textual information: dynamic rather than static,
resulting in better visual or practical usability; but it mainly lays in the ability to work
with computational methods on the text and on the information it conveys [12, 13].

 In conclusion, the digital edition of a text should aim to provide adequate data and
functionality to further forms of processing. Hence the idea that the "digital scholarly
edition" until now often identified with the "digital critical edition" also known as
"digital variorum edition", can also take other forms focused on other types of
'scholarly research': from the geographical knowledge contained in the text, to the
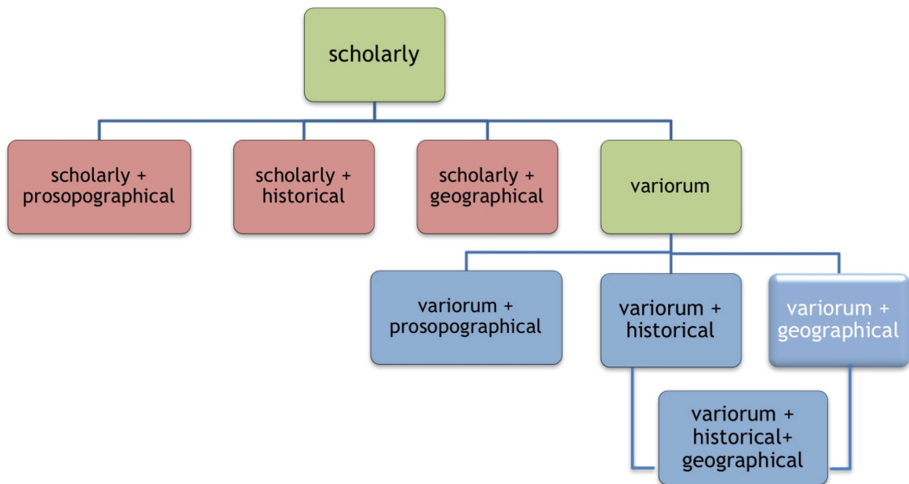
Fig. 3. Various types of edition of a text: actual (green) and foreseeable (other colors) (Color figure online)

historical knowledge (time and events) often inextricably linked with the prosopography, and much more.

So, if the digital critical edition (digital variorum edition) is a type of digital scholarly edition containing an apparatus that analyses and describes the state of the text in the witnesses, then we can conceive e.g.

- the digital scholarly geographical edition of a work – whose apparatus contains an analytical description of the geographical knowledge contained in the place names;
- the digital <u>critical</u> geographical edition (digital variorum geographical edition) whose geographical apparatus is layered over a base critical edition:

To do so the knowledge contained in the text must be expressed in a highly formal manner - the same way that the critical apparatus is a highly formal device – and an ontology is a good and *very complex* means to do that. Here below an abstract sample of a passage of text where the place name Lydia is annotated with reference to the GO! ontology (Fig. 4):

1 Se quoque, cum transiret mare, non Ciliciam aut `<geogName ref=https://w3id.org/ geolit/ontologies/GO-PHY 'Lydia' xml:id='Lydia'>`Lydiam`</geogName>` — quippe tanti belli exiguam hanc esse mercedem—sed Persepolim, caput regni eius, Bactra deinde et Alexandria ultimique Orientis oram imperio destinasse. Quocumque ille fugere potuisset, ipsum sequi posse: desineret terrere fluminibus, quem sciret maria transisse.

Fig. 4. Passage where the word Lydiam is annotated

More references to our studies in this field can be found also in forthcoming publications, namely the proceedings of the conferences DHANT (MSH-Alpes, Grenoble 2015) and Dixit (Köln 2016).

References

1. Bianchetti, S.: Geografia storica del mondo antico. Monduzzi, Bologna (2008)
2. Casati, R., Smith, B., Varzi, A.C.: Ontological tools for geographic representation. In: Guarino, N. (ed.) Formal Ontology in Information Systems, pp. 77–85. IOS Press, Amsterdam (1998)
3. Casati, R., Varzi, A.C.: Parts and Places. MIT Press, Cambridge (1999)
4. Elliott, T., Gillies, S.: Digital geography and classics. In: DHQ: Digital Humanities Quarterly, Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure, vol. 3(1) (2009). http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html
5. Frank, A.: Spatial ontology. In: Stock, O. (ed.) Spatial and Temporal Reasoning, pp. 135–153. Kluwer Academic, Dordrecht (1997)
6. Geus, K., Thiering, M. (eds.): Feature of Common Sense Geography. Implicit Knowledge Structures in Ancient Geographical Text. Lit Verlag, Wien-Berlin (2014)
7. Guarino N.: Formal ontology and information systems. In: Proceedings of FOIS 1998. Trento, Italy, 06–08 June 1998, pp. 3–15. IOS Press, Amsterdam (1998)
8. Kavouras, M., Kokla, M., Tomai, E.: Comparing categories among geographic ontologies. Comput. Geosci. **31**(2), 145–154 (2005)
9. Kuhn, W.: Ontologies in support of activities in geographical space. Int. J. Geogr. Inf. Sci. **15**(7), 613–631 (2001)
10. Lana, M., Borgna, A., Ciotti, F., Tambassi, T.: Ontologies and the cultural heritage. The case of GO! In: Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage Co-Located with 13th Extended Semantic Web Conference (ESWC 2016) Heraklion, Greece, 30th May 2016, CEUR Workshop Proceedings, vol. 1595, pp. 7–18 (2016). urn:nbn:de:0074-1595-0, ISSN 1613-0073
11. Momigliano, A.: Alien Wisdom. The Limits of Hellenization. Cambridge University Press, Cambridge (1975)
12. Monella, P.: Why are there no comprehensively digital scholarly editions of classical texts? In: IV Meeting of Digital Philology, Verona, 15 September 2012. http://www1.unipa.it/paolo.monella/lincei/files/why/why_paper.pdf
13. Pierazzo, E.: Digital Scholarly Editing. Theories, Models and Methods. Ashgate, Farnham Burlington (2015)
14. Smith, B.: Ontology. In: Floridi, L. (ed.) The Blackwell Guide to the Philosophy of Computing and Information, pp. 155–166. Blackwell, Malden (2004)
15. Smith, B., Mark, D.M.: Ontology and geographic kinds. In: Poiker, T.K., Chrisman, N. (eds.) Proceedings of the Eighth International Symposium on Spatial Data Handling, pp. 308–320. International Geographical Union, Burnaby (1998)
16. Smith, B., Mark, D.M.: Geographical categories: an ontological investigation. Int. J. Geogr. Inf. Sci. **15**(7), 591–612 (2001)
17. Tambassi, T.: Rethinking geo-ontologies from a philosophical point of view. J. Res. Didactics Geogr. (J-Reading) **2**(5), 51–62 (2016). https://doi.org/10.4458/7800-04
18. Tambassi, T., Magro, D.: Ontologie informatiche della geografia. Una sistematizzazione del dibattito contemporaneo. Rivista di estetica **58**, 191–205 (2015)
19. Turner, A.J.: Introduction to NeoGeography. O'Reilly Media Inc., Sebastopol (2006)