

# Ontologies and the Cultural Heritage. The case of GO!

Maurizio Lana<sup>1</sup>, Alice Borgna<sup>1</sup>, Fabio Ciotti<sup>2</sup>, Timothy Tambassi<sup>1</sup>

<sup>1</sup>Università del Piemonte Orientale, Dipartimento di Studi Umanistici,  
Piazza Roma 36, 13100 Vercelli, Italy  
{maurizio.lana, alice.borgna, timothy.tambassi}@uniupo.it

<sup>2</sup>Università di Roma Tor Vergata, Dipartimento di Studi letterari, filosofici e di Storia dell'arte,  
Via Columbia n. 1 - 00133 Roma  
fabio.ciotti@uniroma2.it

**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 to the end of the Roman Empire) where the geographic knowledge expressed in the Latin texts themselves can be reused. To do so the core of Geolat is the GO! ontology developed *ad hoc* to describe the geographical knowledge contained in the texts of the library. The semantic annotation of geographical knowledge allows to highlight that the perception of the space that ancient populations had in their minds is a type of cultural heritage; and thanks to the annotation this type of knowledge can produce new types of cultural objects (interactive maps, editions of texts, etc.). 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, interactive maps, digital critical editions.

## 1 The case for ontology in Digital Heritage sciences

The huge digitalization projects in the Humanities and Heritage sciences that took place in the last years have brought to the creation of many extensive, high-quality archives of texts belonging to different linguistic traditions and cultures. In order to leverage these digital archives to enhance texts analysis and interpretation practices we believe that they should be enriched by:

- computational methods and tools for distributed and cooperative annotation of digital resources;
- models and tools apt to represent and process "semantic" levels of digital information, which allow knowledge transfer and sharing within the digital environment, namely formal ontologies and linked data services.

Although these technologies and the underlying methodologies are now widely used, there are some relevant theoretical points that should be emphasized in this context:

- formal ontologies have the dual capacity of fixing prior knowledge of what is (in the domain), and, simultaneously, of enabling the discovery of new knowledge;
- multiple ontological modeling can be connected with the same (passage of) text, such as knowledge and cultural contents layers that overlap with the textual layer, thus uncovering its complexity;
- such stratified texts can be re-used in different fruition contexts, as by professional scholars' or culturally curious users' who are attracted by the potential text mash-ups.

When a community of scholars annotate a set of texts and populate the ontology, parts of the texts are converted into rich data, whether because they are organized in a conceptual framework describing the relationships with others data and documents in the digital networked ecosystem, or because they are related with the point in the text which they belong to. As a result, it is possible to use those data for advanced processing and activities such as:

- application of computational inferences and reasoning methods to deduce hidden information contained in the ontology and in the documents domain it describes;
- visual representations of the texts content;
- text and maps integration/interaction;
- re-use in different context and connection with the Linked Data Cloud.

These enriched data sets, in addition, open a space in which also 'non-experts' may enter and well-known cultural objects such as maps, thematic maps, digital devices for interaction with the physical world, are cognitively enriched, allowing the user – whether "expert" or "non-expert" – to access to the world they come from, by reading the text which the textual datum has been extrapolated from. Thus, the text (partially) transformed into data discloses its information wealth just as the data on reuse reveal the text's interest. In this approach, the traditional experts' literary, aesthetic, historical-critical reading / interpretation is no longer exclusive or dominant.

The Geolat (Geography for Latin literature) project and the GO! Ontology are rooted in this theoretical and methodological framework.

## **2 Geolat and the GO! ontology**

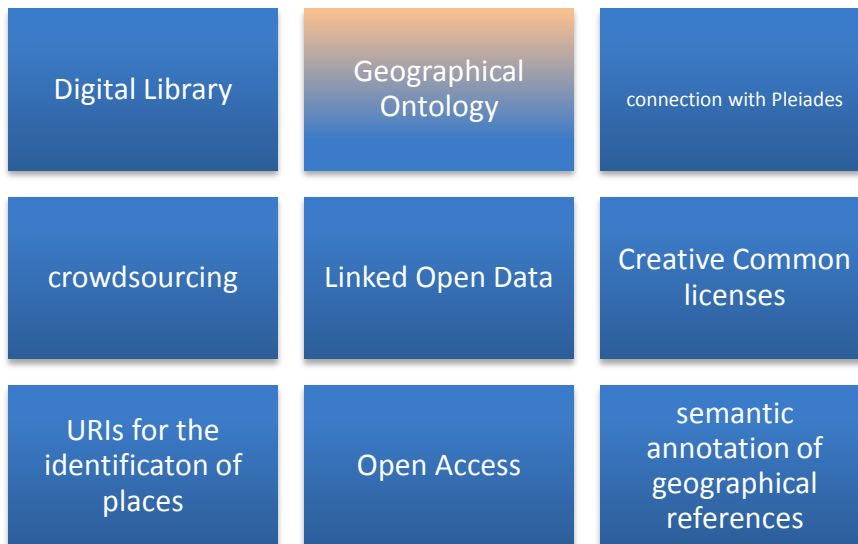
The Geolat project scope is the semantic annotation of the geographical knowledge related to placenames present in the classical Latin texts. This type of formal annotation will allow for a number of new research approaches to the classical Latin texts, also thanks to new ways of building the search inside the texts, with the most innovative being this one: it will be possible to draw an area on a map and obtain by the system various lists showing the placenames pertaining to that area and contained in the underlying digital library; the titles of the works containing those placenames; the authors of the works.

Geolat is one of the outcomes of what has been called “the spatial turn in the Humanities”, determined by the invention of GIS: that is the discovery that geography

can be expressed also through positional data (latitude, longitude) which can be computed, elaborated and used [4]. But it is also, in some ways, an expression of what has been called “neogeography” or collaborative geography. One of its most famous products is OpenStreet Map. The survey of places in the case of Geolat is not done on the terrain but on the texts. The surveyed places are actually placenames contained in classical Latin texts.

One of the problems to face is “how can be the geographical knowledge formally expressed and reused?” and one robust answer is the recourse to an ontology: it allows to formally describe a knowledge domain in a way understandable and usable both for humans and for machines.

There are many aspects of geolat which would deserve attention and description:



**Fig. 1.** Main components of Geolat project

but here we want to focus onto the ontology also because the creation and adoption of a specific geo-ontology represents a fundamental innovation compared with similar projects focused on the ancient world. So we will briefly outline:

- which is the meaning of an ontology in this context
- what kinds of problems can be faced in the creation of an ontology
- which 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 and ancient geography. From a computer science point of view, ontology is a structure aimed to analyse and 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)

which is understood both by the machines and the humans. The resulting structured representation of knowledge allows to resolve conceptual or terminological inconsistencies, providing a dictionary of terms formulated in a canonical syntax and with commonly accepted definitions. Moreover, it provides a lexical or taxonomic framework for the representation of knowledge, shared by different communities of information systems that can range across several domains [7] [15] [13]. From a geographical point of view, the aim of a geo-ontology is to analyze 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] [14] [16,17,18].

As it has been said:

«Mesoscopic geography deals mostly with qualitative phenomena, with phenomena which can be expressed in the qualitative terms of natural language; the corresponding scientific disciplines, in contrast, deal with the same domain but consider features which are quantitative and measurable. GIS thus requires methods that will allow the transformation of quantitative geospatial data into the sorts of qualitative representations of geospatial phenomena that are tractable to non-expert users» [14].

Specifically investigating Roman culture and geography 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 are and, therefore, are related to:

- measurement and measurability of distances (and their different units of measurement);
- location of places;
- organization, shapes, sizes and boundaries of the inhabited world (and its parts);
- representations of the world itself;
- absolute vs relative distances/coordinates.

Problems concerning documentation and sources, are in particular:

- lack of reliability and homogeneity of some data (even when they are direct evidences);
- disagreement among different authors;
- difficulties (in some cases, impossibility) of autoptical confirms;
- isolation of properly geographical contents from the rest of the texts.

The third kind of problems is strictly connected with the second one and refers to methods and to the (multiplicity of) approaches to ancient geographical investigation.

They concern:

- the heterogeneity of aims, points of view, interpretations and perspectives (sometimes overlapped) through which the information was transmitted, processed and implemented;
- the use of assumptions and models representing cosmos and aimed to demonstrate or support some specific thesis;
- the attempts to make the data more consistent, putting them in different (or opposite) geographical conceptions;
- the selection processes of sources;
- the definition of a “geographical unit”;
- places whose existence is only theorized;
- the importance of imagination (and mental maps);
- the use of geometrical and mathematical models that speak a universal (and shared) language and aimed at analysis and organization of space that surrounds us;
- the necessity of folk theorizing, in order to interpret other's mind and ancient culture [1] [6] [10].

The creation of GO!, given the plurality of domain interests imposed a division of work in different steps:

1. the analysis of Latin literature texts, in order to identify geographical entities, classes, properties and relations;
2. the study of the differences between ancient and contemporary geography, in terms of domains, presuppositions, representations and vagueness;
3. a critical review of contemporary geo-informatics ontologies, aimed to identify common classes and properties, and then to specify missing classes and properties in order to describe ancient geography. So it is possible to establish if an ontology must 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;
4. 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.

The result of this work is *GO!*, a geo-ontology aimed at providing:

- a description of the geographical knowledge emerging from Latin literature (the first development starts from the analysis of Caesar, Sallust, Tacitus, Livy, Ammianus Marcellinus texts);
- 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.

with some general aims: openness, accessibility (both for the scientific community and for general public), informativeness, completeness, interoperability and reuse.

Also thinking of the reuse, the ontology is built as a collection of four interconnected modules (expressed in OWL2) freely accessible 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 described in the following paragraphs. Their live graphical representations 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). The figures available in the next pages don't substitute the personal analysis and observation which we encourage to do using the above mentioned URLs, but can offer a useful hint about the complexity of the ontologies which were created.

### 3 Ontological modelling

The *GO!* modelling choices allow to express:

- a range of information about geographical places (i.e. their evolution through time, GPS coordinates, physical and geopolitical descriptions, switch of name, and so forth);
- descriptions of historical events connected with specific places;
- connections with places data in Pleiades;
- management of imaginary places;
- connection with Barrington Atlas;
- connections with the Open Annotation ontology to cite the passages;
- textual sources where ancient places had been marked-up (with philological reference).

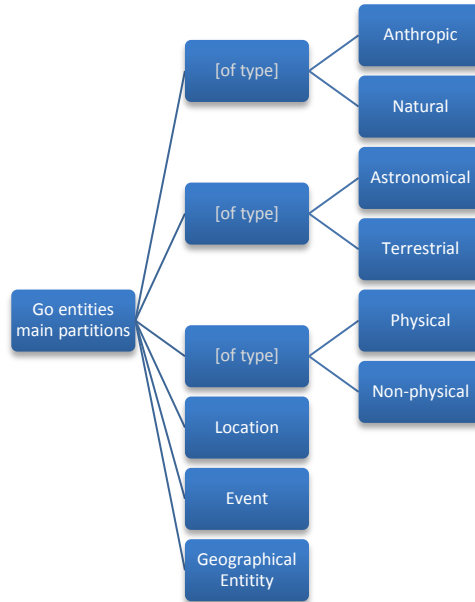
#### 3.1 GO-TOP

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 (all the other three ontologies) 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! indirectly imports some other ontologies:

- *geosparql* (<http://www.opengis.net/ont/geosparql>)
- *sf* (<http://www.opengis.net/ont/sf>)
- *gml* (<http://www.opengis.net/ont/gml>)
- *skos* (<http://www.w3.org/2004/02/skos/core>)

The main conceptual structure of the GO! geographical entities can be summarized as follows, showing classes from GO-TOP:



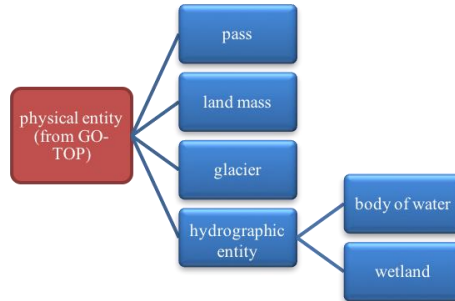
**Fig. 2.** Main structure of GO-TOP

Among the most relevant Properties describing the content of the classes of GO-TOP we can mention the following ones:

selected Object Properties	<p><b>Spatial dimension:</b> above, below, borders, has location, has SRID, in place, in SRID, is under, left of, nearby, right of, visible from, beginning place, ending place, part of</p> <p><b>Time dimension:</b> after, before, occur in</p> <p><b>Nominal size:</b> derive from, has name, name of</p> <p><b>Size of the actors:</b> becomes, owns, involves, object of, subject of, to, by, passes through, controls, belongs to.</p> <p><b>Size measures:</b> measured by.</p>
selected Data Properties	<p><b>Spatial dimensions:</b> latitude, longitude.</p> <p><b>Time dimension:</b> beginning period, temporal period, ending period, in date, order, valid since, valid until.</p> <p><b>Area of nouns:</b> etymology.</p> <p><b>Area of measures:</b> has value.</p>

### 3.2 GO-PHY

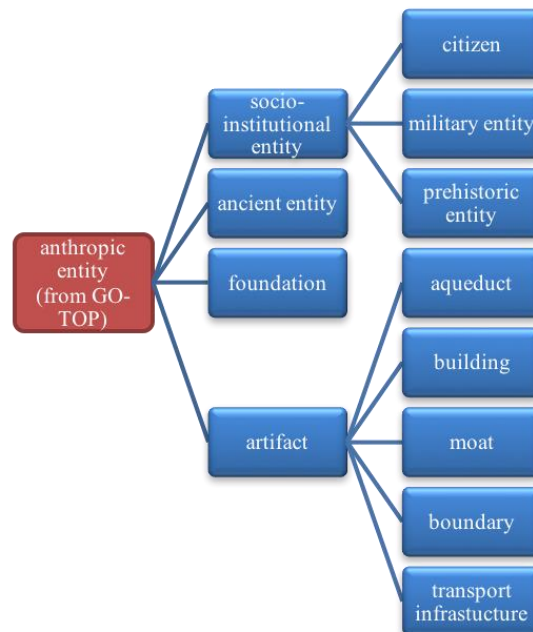
GO-PHY contains 127 classes, 3 individuals. It imports the *GO-TOP* module, and includes a taxonomy which represents geographical entities with physical-natural aspects. All the classes of *GO-PHY* are sub-classed of *astronomical entity*, *physical entity*, *geographic entities*, *natural entities*,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



**Fig. 3.** Sample classes of GO-PHY from physical entity of GO-TOP

### 3.3 GO-HUM

*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*.



**Fig. 4.** Main structure of classes of GO-HUM which expands the anthropic entity of GO-TOP



### 3.4 GO-FAR

*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 humans 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.

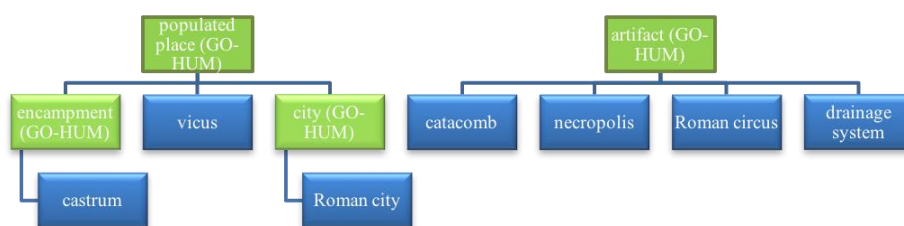


Fig. 5. Sample classes of GO-FAR from GO-HUM which expand populated place and artifact

## 4 How to work with the ontology

The work of annotation of the texts with the ontology is starting. It involves but an automated approach and another human intelligence-intensive. Through NER processes starting from the parsing of Latin morphology and lexicon the placenames will be searched for, recognized and automatically annotated taking from Pleiades the names in the contemporary languages if available, and the GPS coordinates. This automated phase will need a good amount of work to be properly set up and would never be really apt at recognizing rare placenames such as Uxellodunum which will require a human intervention to be annotated. But the human work of annotation will be specifically needed to enrich the annotation using the most of the opportunities offered by the ontology which allows for the indication of historical events connected with a specific place, and of the name(s) of the person(s) involved in the event (think of: Rubicon, act of trespassing, Caesar). This is the reason why the project sees the collaborative building of the annotation as an indispensable approach: the type of annotation which the scholars (the humans) are able to deliver thanks to their knowledge of the context. The annotation will be partly inline and partly offline: in the text there will be simply a link towards the external ontology, which will be populated by the annotations done on the texts: the class river will be populated by all the names and data referring to the river names effectively annotated in the texts. Building the ontology was, for many reasons, a work of years which only now is satisfactorily (even if not perfectly yet) concluded; so only in the next months the very work of the annotation will start.

## 5 Envisioned uses

As it was recalled at the opening of this paper, in the digital humanities field a geographical turn happened, whose meaning is that the geography is seen no more simply as a specific discipline; its meaning is rather 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. Geolat is the product of this idea [4].

In fact in Geolat the annotation of texts based on the ontology allows for at least two different uses:

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

The first use 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 investigated area), yet controlled (the reader can fine tune the texts collection which is analyzed: 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.

In the second use the concept of "digital edition of a text" is developed / expanded. 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 [10, 11]. Therefore 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 said "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 historical knowledge (time and events) often inextricably linked with the prosopography, and much more [12].

So if the *digital critical edition* (digital variorum edition) is a type of *digital scholarly edition* containing an apparatus that analyzes 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 placenames;
- the *digital critical geographical edition* (*digital variorum geographical edition*) whose geographical apparatus is layered over a base critical edition:

In order to achieve this scope the knowledge contained in the text must be expressed in a highly formal manner - the same way that the critical apparatus is a high-

ly formal device – by means of an ontology.

In synthesis the availability of a digital library of texts semantically annotated (we spoke of geography, but the same could be done for events and persons) opens completely new opportunities of managing texts (we are working with classical texts but the same approach can be applied to texts of any period) activating the connective potential they intrinsically own but which is left today the study and research activity of the scholars – while the exposition of the ontological annotation in form of Linked Open Data is intentionally aimed to allow for the interconnection of elements of knowledge all over the web.

## 6 Conclusions

The semantic annotation of the geographical knowledge thorough an *ad-hoc* ontology has also a meta-meaning: highlighting that the cultural heritage of ancient texts is represented not only by the textual and literary data and information that come from philology or archaeology, *but also by the perception of the space that ancient populations had in their minds*, which is reflected in their literary works. In particular, each historical period looked at the geography from a specific point of view, depending on the repositioning of the known world, the discovery of marginal lands, or the changes of the perceptible center (for example from Rome to the various cities, like Sirmium, Trier, Milan and Constantinople, where emperors lived).

In such an heterogeneous context, ontology can play the role of a unifying theory of knowledge representation and organization, and its creation is probably one of the best and most pertinent ways to explain and value this peculiar perception of space. An ontology requires a great investment in time, work of many people with many different competencies, and money but it is the condition to work in the field of “meaning” in broad sense in a highly formal – hence reusable and shareable – manner.

Moreover the act of annotating the text in the way here drafted means that the knowledge “we” have about the text and its content is poured into the text itself in a way which makes shareable, reusable, and available in context to which it pertains the knowledge which was implicit, not directly accessible. One side one could say “nothing new, the glossae were the same thing” but really here we have a type of glossae whose content is shared and can be (re)used both by the scholar and by the computers: and this opens the table to completely new plays.

## 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*. IOS Press, Amsterdam (1998) pp. 77-85.
3. Casati R., Varzi A. C., *Parts and Places*, MIT Press, Cambridge (MA) (1999).
4. Elliott. T, Gillies S., *Digital Geography and Classics*. In DHQ: Digital Humanities Quar-

- terly, Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure. Winter 2009, Volume 3, Number 1. <http://www.digitalhumanities.org/dhq/vol1/3/1/000031/000031.html>
5. Frank A., Spatial ontology. In Stock O. (Ed.), *Spatial and Temporal Reasoning*. Kluwer Academic, Dordrecht (1997) pp. 135-153.
  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, Proceedings of FOIS '98*. Trento, Italy. 1998, June 06-08. IOS Press, Amsterdam (1998) pp. 3-15.
  8. Kavouras M., Kokla M., Tomai E., Comparing categories among geographic ontologies. In *Computers and Geosciences*, vol. 31, n. 2, 2005, pp. 145-154.
  9. Kuhn W., Ontologies in support of activities in geographical space. In *International Journal of Geographical Information Science*, vol. 15, n. 7, 2001, pp. 613-631.
  10. Momigliano A., *Alien Wisdom. The Limits of Hellenization.*, Cambridge University Press, Cambridge (1975).
  11. Monella P., Why are there no comprehensively digital scholarly editions of classical texts?, IV Meeting of digital philology, Verona September 15, 2012, [http://www1.unipa.it/paolo.monella/lincei/files/why/why\\_paper.pdf](http://www1.unipa.it/paolo.monella/lincei/files/why/why_paper.pdf)
  12. Pierazzo E., *Digital Scholarly Editing. Theories, Models and Methods*, Ashgate, Farnham Burlington (2015)
  13. Smith B., Ontology. In Floridi L. (Ed.), *The Blackwell Guide to the Philosophy of Computing and Information*. Blackwell, Malden (MA) (2004) pp. 155-166.
  14. 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*. International Geographical Union, Burnaby, British Columbia (1998) pp. 308-320.
  15. Smith B., Mark D. M., Geographical categories: an ontological investigation. In *International Journal of Geographical Information Science*, vol. 15, n. 7, 2001, pp. 591-612.
  16. Tambassi T., *Rethinking Geo-Ontologies from a Philosophical Point of View*. In J-Reading, 2016, forthcoming.
  17. Tambassi T., Magro D., *Ontologie informatiche della geografia. Una sistematizzazione del dibattito contemporaneo*. In *Rivista di estetica*, vol. 58, 2015, pp. 191-205.
  18. Turner A. J., *Introduction to NeoGeography*. O'Reilly Media Inc., Sebastopol (CA) (2006).