Ontologies for geospatial information: Progress and challenges ahead

Christophe Claramunt

Naval Academy Research Institute, Lanvéoc-Poulmic, France

Abstract: Over the past 50 years or so the representation of spatial information within computerized systems has been widely addressed and developed in order to provide suitable data manipulation, analysis, and visualisation mechanisms. The range of applications is unlimited and nowadays impacts almost all sciences and practices. However, current conceptualisations and numerical representations of geospatial information still require the development of richer abstract models that match the complexity of spatial and temporal information. Geospatial ontologies are promising modelling alternatives that might favour the implementation and sharing of geographical information. The objective of this vision paper is to provide a short introduction to the principles behind semantic ontologies and how they can be applied to complex geospatial information, by evaluating their potential and limitations.

Keywords: geospatial ontologies, conceptual and formal models, numerical models, ontology, geospatial information, semantics

1 Introduction

While geographical information systems (GIS) have successfully developed over the past 50 years, it is nowadays recognized that this has been the case without a so far complete formal theoretical support that might encompass the full complexity of many space and time phenomena. In particular, the old cartographical paradigm had a strong influence on the development of computational GIS frameworks, as illustrated by the 'layer' concept often implemented in GIS software solutions. This led to the development of software-oriented GIS solutions oriented towards either raster or object-based representations. The success

of the relational database approach in the 1970s led to the development of geo-relational models where layers were closely associated with relational tables, offering many practical, but nonetheless limited solutions to the integration of additional entity attribute properties. Therefore, most GIS applications were, and still are, dependent on raster or object-based models, despite the limitations of such approaches in capturing the complexity of scientific studies and many real-world phenomena [7].

The big issue that arises is how can we design a conceptual bridge between current GIS technologies and models on the one hand, and the necessary theoretical GIS foundations on the other hand, and how to do so?

This question should lead to a preliminary investigation on how humans conceptualize space and time? What are the roles of language and cognition when doing so? This also stresses the close link between what reality is and how interpretations should materialize it as much as possible in computerized frameworks. In order to understand how people perceive the world, cognitive conceptualizations of geographic features and appropriate abstraction paradigms should be developed to support computerized representations [5].

With many in the GIScience community searching for novel theoretical pathways to re-engineer GIS data models, the concept of ontology re-appeared and offered canonical descriptions of knowledge domains as defined as "a neutral and computationally tractable description or theory of a given domain which can be accepted and reused by all information gatherers in that domain" [14]. An ontology is usually defined as "a formal, explicit specification of a shared conceptualization" [8] providing a non-ambiguous and formal representation of a domain.

While early GIS data models were not really successful in establishing a close link between reality and data representations, ontologies should abstract the world as it is, using formal and primitive entities, and with much more attention to the underlying properties of geographical phenomenon. A sound geospatial ontology should first define in formal terms the constituents of a reality within a given domain and should be soundly defined and logically possible, extensible and implementable [8]. A geospatial ontology should encompass all the categories and modelling abstractions necessary for a meaningful representation of a given real-world domain: from fields to objects, from events to processes as well as causal to qualitative spatial and temporal relations. A geographic ontology will then be specifically oriented to the "necessary and sufficient conditions for something to be a particular kind of entity within a given—geographic—domain and not an abstraction of the formal features that characterize all scientific areas" [12]. In fact, a geospatial ontology should provide a taxonomy, a formal vocabulary that can be computerized at the software engineering level.

2   Ontologies

An ontology should also include axioms to explicitly define the abstractions to represent and reasoning mechanisms [14]. As there will be some approximations in this process, the objective is to minimize the distance between reality and a final domain-based representation [19]. Ontologies should not only explicitly represent concepts and relations abstracted from reality formally, allowing numerical notation using symbolic grammars, they should also favour interoperability and knowledge sharing between different applications [9]. An ontology can be formalized by description logics through definition of classes, relations,

functions, and axioms. In description logic data is represented using a hierarchy of classes, relations and instances. Under the umbrella of the semantic web, ontologies can be implemented according to standard formalisms such as the Web Ontology Language (OWL). OWL offers a formal logic-based semantics and complemented by the Resource Description Framework (RDF) and query standards such as SPARQL whose objective is to provide schema and query mechanisms and reasoning rules to manipulate the represented data. An important property of RDF triples made of subject-predicate-object is that they are easily understandable by machines (though perhaps not so well by humans...). Several formats nowadays support RDF implementations including RDF/XML, N-Triples, JSON-LD, Turtle, and Notation [6].

One of the main advantages of such approaches is that the software level can use and reuse the represented semantic data and rules without rewriting code, thus reducing maintenance and evolution costs. RDF can be thought of as a grammar in which facts about the world are expressed in RDF as triplets of <subject, predicate, object>. RDF, usually in the form of XML, can be embedded in HTML so that browsers, search engines, and other programs can manipulate the represented data and infer additional knowledge. Alternative models to RDF and XML have been suggested, for example the JSON JavaScript Object Notation which is both more compact and easier for humans to read and interpret.

3 Towards geospatial ontologies

The benefits of a sound ontology for geographic information include not only a conceptual, logical and computational bridge between reality and machines, but also a basis for exchange of information and cross-disciplinary collaboration between different domains of science.

When discussing geospatial ontologies, we can differentiate between a unified framework whose objective is to identify high levels geographical concepts [4] and domain-based ontologies. Domain ontologies have been developed for many domains including cadastral applications [17], urban studies [2], or as mediators for knowledge sharing [16, 18] and to the companion domain of remote sensing [1]. Geospatial ontologies share many structural similarities, regardless of the language in which they are expressed. Most ontologies describe individuals, categories, attributes, relations, rules, actions, and events.

The search for a rich geospatial ontology generalizable across many fields and applications is still a major challenge. Geospatial objects are complex abstractions, they have parts and can be constituents of others [15], they have bona fide or fiat boundaries, they are either well or vaguely defined and encompass a large range of spatial relations and are associated to categories and additional semantic. While being potentially defined at different levels of abstraction and granularity they evolve though events and processes and generate multiple relational networks in space and time [3, 20]. Several fundamental challenges have not been completely resolved with respect to the development of geospatial ontologies:

- to provide complete and appropriate representations of real-world phenomena that integrate the four spatio-temporal dimensions and the whole complexity of real-world phenomena;
- to create a formal and computational data model that could provide a sound representation of all the concepts identified at the ontological level;

- whether a general ontological-based might support a formal umbrella that includes all four spatio-temporal dimensions within a unified framework?

A geospatial ontology for the web nowadays offers a series of functionalities towards the geospatial semantic web where a comprehensive set of geographical properties and abstractions can be both understandable by different communities and implemented. Kuhn, Raubal, and Gärdenfors [11] underlined that, in order to better match human cognition, geospatial ontologies should be grounded by establishing meaningful and suitable geographical and semantic primitives and integrate time as well as different levels of abstraction and users' points of view. A good balance should be also made between generic geospatial ontologies and domain-based ontologies [10], since the two views are complementary.

An important recent trend is that the development of geospatial ontologies has been closely addressed in the dual context of standard recommendations from the ISO and OGC, in order to represent geospatial concepts and properties for use on the Web. This is a major trend that of course is also a consequence of the dominance of the Web in the development of novel software engineering solutions, but one might wonder if alternative software engineering options are not possible? Thus, geospatial ontologies on the Web are largely based on several formats to implement RDF triples such as XML, RDFa, and JSON-LD. GeoSPARQL language is a standard RDF SQL-based query that manipulates geospatial RDF data. It provides a GML-based representation of geometrical literals, topological relations, a SPARQL query interface and a rule interchange formal for further inferences. So far several vendor-based software implementations of RDF (e.g., Oracle Spatial) and SPARQL are currently been implemented as well, and are associated with geometrical extensions such as (e.g., KML, GeoJSON). But SPARQL has similar limitations as SQL, queries are not always intuitive to read and understand. GeoNames and LinkedGeoData are examples of datasets that cover a vast part of the world, for instance by allowing integration of large data repository such as OpenStreetMap.

Despite the fact that geospatial ontologies provide sound and formal representation mechanisms, a series of limitations can be still identified and should be considered as major research challenges to address:

- The difficult formalization of expert knowledge is a key issue. Moreover, transferring expert knowledge to classes, relations and rules is not always straightforward, especially as declarative languages are not user friendly. The way triples might represent the full complexity of relational concepts is not always satisfactory, especially for some semantically complex relations and the way triples might be interpreted is another difficult issue. Transferring specialized knowledge from texts or domain experts to abstract and effective concept representations is far from being an easy task and can often lead to misinterpretations and ambiguities.
- Are OWL and RDF sufficient enough to represent and manipulate the whole complexity of geographical and temporal abstractions? Although objects are relatively well represented, image data is not completely represented by RDF triples and GeoSPARQL and temporal abstractions have still to be integrated. Similarly, 3D models and Building Information Models should be fully integrated. Last, but not least the emergence of big geospatial data is likely to bring computational issues as RDF and GeoSPARQL were not designed to deal with massive geospatial datasets.

- Are the functionalities of current model and query languages such as GeoSPARQL rich and understandable enough to provide a high level data manipulation level? Is GeoSPARQL computationally effective as server loads are likely to be costly operations? So far GeoSPARQL functions and queries are far from being intuitive and really far away from what a typical user might expect.
- Geospatial ontologies should be extensible and reusable and possibly cross-domains and communities. Interoperability implies leveraging existing standards and being adaptable to existing data-centric infrastructures.
- The large range of ontology language editors, although some are well established (e.g., Protégé), does not always facilitate interoperability and results in uncoordinated software engineering efforts.
- One of the advantages of formal and numerical representations of ontologies and geospatial ontologies lies in the visibility of the notations. However, this leads to large repositories of data representation and an intermediate level, where users might manipulate such abstractions at a higher level of representation, is lacking.
- As scientific applications are not of high priority within the GIS industry despite the availability of many ontology standards, and as re-engineering existing application will be extremely costly, embedding geospatial ontologies within GIS will be far from straightforward.

Geospatial ontologies offer fundamental resources to remodel geospatial information. However, a good balance should be sought between the need of offering sound and interoperable geospatial infrastructures and that of not re-inventing the wheel as every effort should be made to leverage existing GIS data infrastructures wherever possible.

Finally, geospatial ontologies are very likely to be closely integrated within much broader and large contexts. For instance, the Sustainable Development Goals Interface Ontology [13] is an important example and demonstration of how geospatial ontologies might act as a foundation for sustainable development between more specialized domain ontologies that will offer cross-reference entities of sustainable development knowledge. When developing such an agenda for a sustainable environment it turns out that intimately connecting such efforts to the representation of environmental entities, processes, interconnections with many ecosystems and urban systems is a key challenge that can be addressed by geospatial ontologies in order to creating open representations and standards. This effort is on that should involve many GISscience related communities, researchers and practitioners.

References

[1] ARVOR, D., BELGIU, M., FALOMIR, Z., MOUGENOT, I., AND DURIEUX, L. Ontologies to interpret remote sensing images: why do we need them? *GIScience & remote sensing 56*, 6 (2019), 911–939.

[2] CHEN, Y., SABRI, S., RAJABIFARD, A., AND AGUNBIADE, M. E. An ontology-based spatial data harmonisation for urban analytics. *Computers, Environment and Urban Systems 72* (2018), 177–190.

[3] CLARAMUNT, C., AND THÉRIAULT, M. Managing time in GIS an event-oriented approach. In *Recent advances in temporal databases*. Springer, 1995, pp. 23–42.

[4] COUCLELIS, H. Ontologies of geographic information. *International Journal of Geographical Information Science 24*, 12 (2010), 1785–1809.

[5] FONSECA, F., CÂMARA, G., AND MIGUEL MONTEIRO, A. A framework for measuring the interoperability of geo-ontologies. *Spatial Cognition and Computation 6*, 4 (2006), 309–331.

[6] GANDON, F., KRUMMENACHER, R., HAN, S.-K., AND TOMA, I. The resource description framework and its schema. In *Handbook of Semantic Web Technologies*. Springer, 2011.

[7] GOODCHILD, M. F. Reflections of a past editor. *Geographical Analysis 31*, 4 (1999), 328–331.

[8] GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies 43*, 5–6 (1995), 907–928.

[9] GUARINO, N. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, vol. 46. IOS press, 1998.

[10] JANOWICZ, K. Observation-driven geo-ontology engineering. *Transactions in GIS 16*, 3 (2012), 351–374.

[11] KUHN, W., RAUBAL, M., AND GÄRDENFORS, P. Cognitive semantics and spatio-temporal ontologies. *Spatial Cognition & Computation 7*, 1 (2007), 3–12.

[12] PEUQUET, D., SMITH, B., AND BROGAARD, B. The ontology of fields: Report of a specialist meeting held under the auspices of the VARENIUS project. https://philarchive.org/archive/PEUTOO, 1998.

[13] SDIGO. The sustainable development goals interface ontology (SDGIO) code repository. https://github.com/SDG-InterfaceOntology/sdgio, 2020.

[14] SMITH, B. Ontology: philosophical and computational. In *The Blackwell Guide to the Philosophy of Computing and Information*. 2003.

[15] SMITH, B., AND MARK, D. M. Ontology and geographic kinds. In *8th International Symposium on Spatial Data Handling*. IGU, 1998, pp. 308–320.

[16] STOCK, K., KARASOVA, V., ROBERTSON, A., ROGER, G., SMALL, M., BISHR, M., ORTMANN, J., STOJANOVIC, T., REITSMA, F., KORCZYNSKI, L., BRODARIC, B., AND GARDNER, Z. Finding science with science: evaluating a domain and scientific ontology user interface for the discovery of scientific resources. *Transactions in GIS 17*, 4 (2013), 612–639.

[17] STOCK, K., LEIBOVICI, D., DELAZARI, L., AND SANTOS, R. Discovering order in chaos: Using a heuristic ontology to derive spatio-temporal sequences for cadastral data. *Spatial Cognition & Computation 15*, 2 (2015), 115–141.

[18] SUN, K., ZHU, Y., PAN, P., HOU, Z., WANG, D., LI, W., AND SONG, J. Geospatial data ontology: the semantic foundation of geospatial data integration and sharing. *Big Earth Data 3*, 3 (2019), 269–296.

[19] WAND, Y., AND WEBER, R. Reflection: Ontology in information systems. *Journal of database management 15*, 2 (2004), III.

[20] YUAN, M. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science 28*, 2 (2001), 83–96.