

Models for Space Unite! ^{*}

The Need and Opportunities for Domain–transcendent Modelling of Spatial Data

Frans Knibbe^[0000–0003–3789–2260]

fjknibbe@gmail.com

Abstract. This position paper makes a case for a unified way of representing spatial information. For spatial data it should not matter from which knowledge domain they originate. Whether data come from GIS (Geographic Information Systems), CAD (Computer Aided Design), CG (Computer Graphics), the Web, or any other present or future domain in which data have some kind of spatial characteristic, they share common ground. Space is a fundamental aspect of the real world, and making it so in the models that govern the digital world will yield an invaluable reward: system interoperability through data harmonisation. Globalisation, the increasing need for interdisciplinary cooperation and an increasing demand for transparency are some of the developments that fuel ongoing research and development in the fields of Semantic Web, Linked Data and graph–based data. This kind of progress — opening up and defragmentation of information — would be well served by a process of integration of disparate models for spatial information. The current work on updating GeoSPARQL could be a fitting starting point.

Keywords: Space · Geometry · Model Unification · Interoperability · GeoSPARQL · Ontology

1 The Problem, and How It Came To Be

Space is everywhere. It is an elementary aspect of our perception and it is present in almost all domains of humanity’s knowledge and endeavours. People share a common idea of space in the real world. Yet in the digital world, where ideas and observations take the form of data, space has far less of a common understanding. Numerous ways of modelling spatial/geometric data have come into existence and a large amount of data structures for encoding and exchanging spatial data have been devised. Until this day the diversity in working with spatial data is overwhelming. Even within the niche of GIS there is a staggering amount of data structures and software to deal with. Although this is understandable when looking at historic development, it is generally undesirable.

In the later part of the twentieth century, digitalisation and globalisation have been drivers for domain communities to develop and establish standards

^{*} "Copyright ©2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)."

for data storage and exchange. These were primarily focused on how to model and structure data within domains. This has led to industry standards for spatial data in domains like geography, industrial design and manufacturing and computer graphics. That development came with a good amount of competition between businesses trying to give their own inventions the upper hand, leading to many similar or greatly functionally overlapping solutions for the same task. Understandably, this caused a certain amount of frustration at the users' side of the matter. Fortunately the surge of ways to work with spatial data was somehow channeled and made more manageable by the emergence of industry standardisation organisations like the OGC and the W3C, and involvement of ISO, followed by efforts at data harmonisation by international collaborations like the UN and the EU. However, working with spatial data still is a very domain-dependent affair. With an ever growing need to share data between commercial and non-commercial organisations, national and local governments, and regular people, this domain dependency becomes increasingly burdensome. Although working with data in GIS, CAD, or CG does become doable for domain experts — after a few years of practice — a growing number of people, Web developers for instance, would like their data to be readily available and interpretable.

Even within domain communities the need to cross domain boundaries increases. For example, GIS people want to use computer graphics and Web graphics technology for visualisation of geographic features and are hampered by the lack of true 3D capabilities in GIS. And with an increasing amount of man-made structures on the earth's surface, combined with an increasing level of detail in the gathering of geographic data, there is a mounting need to efficiently be able to use systematic geometries like circles, splines or cuboids in GIS, next to the unruly shapes that are preferred by nature. Another example: because geographic data can tell so many interesting stories, their visual and analytical power is much appreciated among non-experts. But then things often go wrong because of incorrect use of earth based coordinate reference systems (CRS), which are hard to simplify because our planet's surface changes all the time. Yet another example is the tracking and tracing of objects like people, vehicles, or sets of keys. When objects travel from the outside to the inside of buildings, or the other way around, they too travel between different industry domains: GIS and BIM (Building Information Modelling). In short: present day technological challenges increasingly have cross-domain characteristics.

Not only domain fixation is to blame for discord in models and structures for spatial data. There is also a tendency to resolve problems with methods that do alleviate niche difficulties, but fail to simplify the realm of spatial information as a whole. Although such initiatives could make a number of use cases easier to accomplish in the short term, in the long term it tends to sow more chaos. As a general principle, knowledge representation and information modelling, at least within the international and cross-organisational scope of the World Wide Web, should not be application-driven. Yes, use cases could and should be sourced from applications, and they are a valuable input for information models. But they

can not be expected to fully encompass all present and future requirements, nor can they ensure a faithful and as-simple-as-possible representation of real world phenomena. What is needed for optimal information modelling is a thorough understanding of the basic elements at play.

Non-interoperability of data and semantics is a huge socio-economic problem. Even in areas where lives are at stake, like health care, public safety, traffic management and climate change, combining data from all relevant sources in a timely, meaningful and error-free fashion is notoriously difficult. The basic interoperability problem is recognized by governmental organisations, and on several levels programs are being set up to combat the phenomenon. For example, the UN has the [Data Interoperability Collaborative](#)¹, the EU has the [European Interoperability Framework](#)² and [INSPIRE](#)³, the UK has the [National Data Strategy](#)⁴ and the USA has several domain-specific initiatives such as [USCDI](#)⁵. But a lack of interoperability is not limited to government and public services. Perhaps not always recognised as such, and rarely publicly admitted, the problem recurrently arises within commercial organisations — within departments, between departments, and in dealings with other organisations.

Not always does a lack of data and semantic interoperability cause clear system failures. Its more common influence is an ever present friction in basic procedures that are necessary to achieve more high-level goals, silently eating away time and money, as well as people's work pleasure.

Of course a lack of interoperability is not limited to spatial data. Nevertheless, space does stand out as an area in which much progress can be made.

2 The Solution, and How It Could Take Form

At the start of the present century, a simple but potentially very powerful solution to information fragmentation was put forward: the twin concepts of the Semantic Web and Linked Data, offering both semantic harmonisation of data and metadata, and a means to interlink data in different datasets. They form the obvious paradigm to achieve interoperability of spatial data. Even so, that has not really happened yet.

Interestingly enough, there is an example of a similar challenge that *has* proven to be feasible. The phenomenon of time has much in common with space. It is something ubiquitous in our daily lives, and therefore it is present in a large share of our digital systems. And there is a profound need for exchange of temporal data, without any misunderstandings. Contrary to space, time already has its domain-independent Web ontology: [OWL-Time](#)⁶. Supporting temporal relations and arbitrary temporal reference systems, this time ontology can be

¹ <https://www.data4sdgs.org/initiatives/data-interoperability-collaborative>

² <https://ec.europa.eu/isa2/eif>

³ <https://inspire.ec.europa.eu>

⁴ <https://www.gov.uk/government/publications/uk-national-data-strategy>

⁵ <https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>

⁶ <https://www.w3.org/TR/owl-time>

used to facilitate exchanging and linking temporal data from very diverse fields of knowledge and enterprise.

Why a similar domain-independent Web ontology for spatial data has not come into existence yet has several reasons. One has to do with complexity. While spatial data can encompass one to three dimensions, with additional measures in the case of linear referencing, time is a phenomenon that is basically one-dimensional. And while different temporal reference systems exist, they tend to be less complex than geographical reference systems. So for spatial data, there is much more ground to cover to arrive at a fully functional cross-domain Web ontology.

Another reason why a universal space ontology has not been developed yet is the aforementioned entrenchment of space in disparate domains. Spatial information, like temporal information, can be encountered in many domains. But there are several, such as GIS, CAD and BIM, Computer Graphics and Web Graphics, in which spatial aspects show a dominance that is not comparable to the weight time has in any domain. Models and data structures cater for domain-specific requirements, and through many years of evolution do serve well enough when application requirements stay within domain bounds.

The fact that space-heavy domains have their own characteristics dictating domain information models and data structures is worthy of attention when contemplating universal Web semantics for spatial data. Existing idiosyncrasies are well-rooted in domain requirements and constraints. Computer graphics are served with capabilities for fast processing of relatively large amounts of 3D data. Hence a preference for uniformly structured data, like meshes of triangles or clusters of cubes. CAD and BIM (in which CAD methodology is used for spatial data), are concerned with flexibility in design while maintaining efficient data structures. Hence a focus on primitive shapes that can be used to make more complex shapes, by sweeps along curves and by Boolean operations between shapes. In LIDAR, data structures are governed by data capture hardware, making point clouds the typical data structure. GIS historically has focused on 2D maps of mainly natural features on the Earth's surface, resulting in relatively simple ways of using vector data and raster data. In Web Graphics other modelling constraints exist: depiction of spatial data always takes place on a Web page, and ideally fits in with existing Web standards like HTML, XML, JSON and CSS.

Yet, for all the differences between different ways of modelling and encoding spatial data, all models and methods share the same roots. They describe Euclidean space, and use the same mathematical principles in definitions of model elements and operations. Mathematics is the foundation of all spatial models, which makes it the common language that all models speak, and an assured way of allowing any kind of translation between disparate models and data structures. If there is no more direct mutual understanding, or mode of data transformation, one can always turn to mathematics as a lingua franca. This means that a cross-domain ontology for space will have to have a semantic foundation drilling down to universal mathematics.

So, if a universal ontology for spatial data should be both deep (with semantics going to mathematical roots) and wide (with capabilities for diverse domain requirements), will that mean it is bound to become an unwieldy monstrosity? That does not need to be the case, thanks to the principle of modularity. It is a defining feature of the Semantic Web, where ontologies can always be used on a need-to-know basis. Only ontologies that are needed to encode a certain set of data need to be invoked, and large parts of those ontologies can be ignored if they are not required for the job. Cherry picking is the go-to mode of adding meaning to data on the Semantic Web, making this technology very appropriate for providing practical semantics for spatial data.

Modularity can also be put to good use *within* a complex Web ontology, with well thought out separation of concerns aiding both development and use. In case of a space ontology, a few topics immediately stand out as candidates for modularisation. Take CRS — for some domains, defining a CRS is nothing more than defining an origin for Cartesian coordinates. In GIS they can be complicated and tricky, and often cause unintentional data errors. To a certain extent that is because CRS definitions, like metadata, are compartmentalised and weakly linked in GIS. Housing CRS definitions in a Web ontology, or a module of a spatial ontology, would make it easier to jump from coordinates to the CRS definition that describes how they should be interpreted. And sharing the same semantic roots would make it easier to combine CRS, a task increasingly required with data being used across domains and across scales: the CRS of a door handle can sit within the CRS of a door, which sits within the CRS of a building, which sits within the CRS of a land plot, which sits within the CRS of a national cartographic grid, which sits within a geographic CRS, which sits within a CRS of the solar system, which sits within a CRS of the galaxy.

Different modules could also exist for popular ways of modelling spatial things. Meshes, point clouds, volume-based and boundary-based models, or cell-based models each have their own specialised capabilities and functionality that usually are sufficient for use. But as modules of one overarching ontology, included will be the benefit of sharing common principles like geometric primitives, geometric operators and functions, CRS and topology (the latter also probably being worthy of its own semantic module). Whatever further modularisation can be envisaged, it is important to know that helpful semantic data from other modules is always accessible through graph traversal.

Once a basic Web ontology for space is in place, with increased capabilities to share and link spatial (meta)data across domains, it could reduce the amount and complexity of software that is used for storage and processing of spatial data, and increase its interoperability. That would be welcome, because data transformation accounts for a lot of friction in working with spatial information. It eats up resources and often incurs data getting lost in translation. The currently diverse models for spatial knowledge representation underpin a much larger amount of data structures (e.g. file and database formats). Many data structures were designed with specific use in mind, and have optimisations for those specific cases, so there will still be a need for them. But it should be pos-

sible to much easier transform data from one data structure to another if those structures are defined in terms of the same core model. Also, next to making data structures more interoperable, a core ontology for spatial data could make native formats for graph data on the Web more powerful. For instance, with full freedom of expression of Semantic Web data, JSON-LD could play a much bigger role in encoding any type of spatial data. With more versatile structures for spatial data in place, additional capabilities for software that is intended for direct user interaction (e.g. analysis and visualisation) and supporting software libraries (e.g. CGAL, OpenGL, WebGL, the Point Cloud Library, GDAL/OGR and JTS/GEOS) should be attainable.

3 A Possible Starting Point

No matter the gains society could reap from a universal spatial ontology, its development and establishment will be a daunting task. Not only is there much knowledge to be encoded in ontology terms, also its development will require close collaboration between domain experts who are not used to working together.

One source of relief is — again — modularisation. Developing a Web ontology for spatial data could and should be done step by step, with each step resulting in ontological modules that are useful on their own. And to a certain extent, different ontological modules could be developed in parallel.

Another circumstance mitigating the amount of work to be done is that progress has already been made. GeoSPARQL, OGC's Semantic Web standard, could be a suitable kernel for development of a wider scoped ontology for spatial data. Its custodian, the OGC, is well aware of increasing demand for interdisciplinary exchange of spatial data and the benefits of combining geography and Linked Data. Its collaboration with the W3C in the [Spatial Data on the Web Working Group](#)⁷ can attest to that. Furthermore, the OGC has compiled a large body of tried and tested specifications for spatial data that have not found their way into GeoSPARQL yet, but are quite ready for that. In fact, the sparseness of the current version of GeoSPARQL (1.0) is a benefit to its potential to grow into a broader ontology for more than only geographic information. And what's more, new versions of GeoSPARQL are currently in development — in an open manner, which allows spatial enthusiasts of any background to pitch in.

The name GeoSPARQL, and its title, 'A Geographic Query Language for RDF Data', suggest the specification is intended for geographic data only. But looking past that, the ontology itself is quite fit for non-geographic data. Its root class (`SpatialObject`) covers all things that are spatial in some way. So in essence all the different ways of modelling space can be included. Its two currently defined subclasses, `Feature` and `Geometry`, neither are limited to geography. But the way the `Geometry` class is effectuated in GeoSPARQL 1.0 does play in favour of geography. It is not really possible to describe a geometric object (a collection of coordinates in a certain coordinate system) purely in RDF at the

⁷ https://www.w3.org/2015/spatial/wiki/Main_Page

moment. Description takes place by means of geometry literals, using one of two serialisations that are rooted in GIS: GML and WKT. Nevertheless, there does not seem to be a semantic impediment for extending GeoSPARQL in ways that allow the use of geometric models from other domains. Moreover, should one wish to use other means than numerical coordinates to describe a spatial thing, it possible to specialise `SpatialObject` in other directions.

This is not to say that GeoSPARQL is the only starting point to consider for an all-encompassing spatial Web ontology. X3D (a standard for three-dimensional spatial data on the Web), for example, has a paired [X3D Web ontology](#)⁸ currently in development. And IFC (Industry Foundation Classes), the main standard in BIM, has an ontology for the Web too: [ifcOWL](#)⁹. Given that IFC allows more kinds of spatial data than comparable OGC standards (OGC's Simple Features can be regarded as a subset of IFC), it could be viewed as further on the way towards a universal ontology for spatial data. But ifcOWL currently is a result of a direct and automated translation of the IFC schemas, resulting in an ontology that is not very understandable to those without expert knowledge of IFC. GeoSPARQL, on the other hand, was handcrafted from the ground up, making it friendlier to implementers and users, and more capable of utilising the unique abilities of Semantic Web technology. Furthermore, GeoSPARQL's rather limited semantics for geometry could be viewed as a weakness, but it does provide ample opportunities for further growth in domain-independent functionality.

Irrespective of the starting point for a universal Web ontology for space, developers and maintainers of existing spatial ontologies ideally will have to cooperate, inspire each other and learn from each other. This is also true for the possible evolution of GeoSPARQL from a bare bones standard for geographic data to a multidisciplinary ontology for spatial data with different origins. The process will need multidisciplinary involvement of people with different origins. All viewpoints and contributions, however narrow of focus or overly broad they may be thought to be, are valuable. Development of next versions of GeoSPARQL takes place at <https://github.com/opengeospatial/ogc-geosparql>. Participation is open to everyone.

⁸ <https://www.web3d.org/x3d/content/semantics/semantics.html>

⁹ <https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl>