

# Open Geospatial Consortium

Submission Date: 2019-11-21

Approval Date: <yyyy-mm-dd>

Publication Date: <yyyy-mm-dd>

External identifier of this OGC® document: <http://www.opengis.net/doc/{doc-type}/{standard}/{m.n}>

Internal reference number of this OGC® document: 19-078r1

Category: OGC® White Paper

Editor: Joseph Abhayaratna, Linda van den Brink, Nicholas Car, Rob Atkinson, Timo Homburg, Frans Knibbe, Kris McGlinn, Anna Wagner, Mathias Bonduel, Mads Holten Rasmussen, Florian Thierry

## OGC Benefits of Representing Spatial Data Using Semantic and Graph Technologies

### Copyright notice

Copyright © 2019 Open Geospatial Consortium

To obtain additional rights of use, visit <http://www.opengeospatial.org/legal/>

### Warning

This document is not an OGC Standard. This document is an OGC White Paper and is therefore not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard. Further, an OGC White Paper should not be referenced as required or mandatory technology in procurements.

Document type: OGC® White Paper

Document subtype:

Document stage: Draft

Document language: English

## License Agreement

Permission is hereby granted by the Open Geospatial Consortium, ("Licensor"), free of charge and subject to the terms set forth below, to any person obtaining a copy of this Intellectual Property and any associated documentation, to deal in the Intellectual Property without restriction (except as set forth below), including without limitation the rights to implement, use, copy, modify, merge, publish, distribute, and/or sublicense copies of the Intellectual Property, and to permit persons to whom the Intellectual Property is furnished to do so, provided that all copyright notices on the intellectual property are retained intact and that each person to whom the Intellectual Property is furnished agrees to the terms of this Agreement.

If you modify the Intellectual Property, all copies of the modified Intellectual Property must include, in addition to the above copyright notice, a notice that the Intellectual Property includes modifications that have not been approved or adopted by LICENSOR.

THIS LICENSE IS A COPYRIGHT LICENSE ONLY, AND DOES NOT CONVEY ANY RIGHTS UNDER ANY PATENTS THAT MAY BE IN FORCE ANYWHERE IN THE WORLD.

THE INTELLECTUAL PROPERTY IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE DO NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE INTELLECTUAL PROPERTY WILL MEET YOUR REQUIREMENTS OR THAT THE OPERATION OF THE INTELLECTUAL PROPERTY WILL BE UNINTERRUPTED OR ERROR FREE. ANY USE OF THE INTELLECTUAL PROPERTY SHALL BE MADE ENTIRELY AT THE USER'S OWN RISK. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR ANY CONTRIBUTOR OF INTELLECTUAL PROPERTY RIGHTS TO THE INTELLECTUAL PROPERTY BE LIABLE FOR ANY CLAIM, OR ANY DIRECT, SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM ANY ALLEGED INFRINGEMENT OR ANY LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR UNDER ANY OTHER LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH THE IMPLEMENTATION, USE, COMMERCIALIZATION OR PERFORMANCE OF THIS INTELLECTUAL PROPERTY.

This license is effective until terminated. You may terminate it at any time by destroying the Intellectual Property together with all copies in any form. The license will also terminate if you fail to comply with any term or condition of this Agreement. Except as provided in the following sentence, no such termination of this license shall require the termination of any third party end-user sublicense to the Intellectual Property which is in force as of the date of notice of such termination. In addition, should the Intellectual Property, or the operation of the Intellectual Property, infringe, or in LICENSOR's sole opinion be likely to infringe, any patent, copyright, trademark or other right of a third party, you agree that LICENSOR, in its sole discretion, may terminate this license without any compensation or liability to you, your licensees or any other party. You agree upon termination of any kind to destroy or cause to be destroyed the Intellectual Property together with all copies in any form, whether held by you or by any third party.

Except as contained in this notice, the name of LICENSOR or of any other holder of a copyright in all or part of the Intellectual Property shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Intellectual Property without prior written authorization of LICENSOR or such copyright holder. LICENSOR is and shall at all times be the sole entity that may authorize you or any third party to use certification marks, trademarks or other special designations to indicate compliance with any LICENSOR standards or specifications. This Agreement is governed by the laws of the Commonwealth of Massachusetts. The application to this Agreement of the United Nations Convention on Contracts for the International Sale of Goods is hereby expressly excluded. In the event any provision of this Agreement shall be deemed unenforceable, void or invalid, such provision shall be modified so as to make it valid and enforceable, and as so modified the entire Agreement shall remain in full force and effect. No decision, action or inaction by LICENSOR shall be construed to be a waiver of any rights or remedies available to it.

# Table of Contents

1. Introduction	6
2. Semantic Technologies	7
3. Graph Technologies	8
4. Beneficiaries and benefits	9
4.1. Beneficiaries	9
4.1.1. Beneficiary 1: Data consumers	9
4.1.2. Beneficiary 2: Data publishers	9
4.2. Benefits	9
4.2.1. Benefit B1: Disambiguation	9
4.2.2. Benefit B2: Dereferenceability	9
4.2.3. Benefit B3: Reasoning	9
4.2.4. Benefit B4: Extendability	10
4.3. Use Cases	10
4.3.1. Use Case 1: Data integration	10
4.3.2. UC1D Mapping Authorities: <i>LINDAS Initiative</i>	11
4.3.3. Use Case 2: Data Product metadata	12
4.3.4. Use Case 3: Recording Provenance	13
4.3.5. Use Case 4: Data analysis	14
4.3.6. Use Case 5: Diagnoses	15
4.3.7. Use Case 6: Simplified Access to Heterogeneous Data	15
4.3.8. Use Case 7: Integrating Aspatial and Spatial Data	16
4.3.9. Use Case 8: Data Mining	18
4.3.10. Use Case 9: Improving Search	19
5. Current capabilities	20
5.1. GeoSPARQL	20
5.1.1. Requirements addressed	20
6. Learnings from use	24
6.1. Proposed extensions to GeoSPARQL 1.0	24
6.1.1. Extension 1: Extend expressive power of ontology for encountered use cases	24
6.1.2. Extension 2: Make underlying definitions available on the semantic web	24
6.1.3. Extension 3: Include all functions already described in Simple Features for SQL	25
6.1.4. Extension 4: Add semantics for CRS	25
6.1.5. Extension 5: Explicitly expect significant digits in numerical data	26
6.1.6. Extension 6: Widen the scope to all spatial data	26
6.1.7. Extension 7: Availability in JSON-LD format	27
6.1.8. Extension 8: Extending GeoSPARQL by defining more vector literal types	28
6.1.9. Extension 9: Extending the GeoSPARQL ontology with support for raster data	28
6.1.10. Extension 10: Extending the GeoSPARQL ontology with raster data query capabilities	29

6.1.11. Extension 11: Extending the GeoSPARQL ontology with full-featured 3D support . . . . .	29
6.1.12. Extension 12: Extending the GeoSPARQL ontology with functions to handle 3D geometries . . . . .	30
6.1.13. Extension 13: Extending the GeoSPARQL ontology with support for M and T coordinates . . . . .	30
6.1.14. Extension 14: Propose how authoring metadata should be included . . . . .	31
6.1.15. Extension 15: Add concepts for accuracies and tolerances . . . . .	31
6.1.16. Extension 16: Enable semantic descriptions of the applied geometry representation contexts . . . . .	32
6.1.17. Extension 17: Allow multiple modelling levels for connection patterns between objects and geometry descriptions . . . . .	32
6.1.18. Extension 18: Need for a universal linking method between objects and geometry descriptions in any existing schema . . . . .	33
6.1.19. Extension 19: Terminology needed to express relations between geometry descriptions . . . . .	35
6.1.20. Extension 20: Support for parametric geometry descriptions via links between a non-geometric property and geometric properties . . . . .	37
6.2. Defects reported in GeoSPARQL 1.0 . . . . .	38
6.2.1. Bug 1: Corrections of example data and queries . . . . .	38
6.2.2. Bug 2: GeoSPARQL Schema v1.0.1: hasDefaultGeometry . . . . .	38
6.2.3. Change 1: Decouple CRS and WKT . . . . .	39
7. Other Related Activities . . . . .	40
7.1. W3C Linked Building Data Community Group . . . . .	40
7.1.1. Scope of Activity . . . . .	40
7.1.2. Web Site(s) . . . . .	40
7.1.3. Source Repositories . . . . .	40
7.1.4. Liaisons . . . . .	40
7.2. ISO/IEC JTC 1/SC 32/WG 3 . . . . .	40
7.2.1. Scope of Activity . . . . .	40
7.2.2. Web Site(s) . . . . .	40
7.2.3. Source Repositories . . . . .	41
7.2.4. Liaisons . . . . .	41
Annex A: Revision History . . . . .	42
Annex B: Bibliography . . . . .	43

## **i. Abstract**

<Insert Abstract Text here>

## **ii. Keywords**

The following are keywords to be used by search engines and document catalogues.

ogcdoc, OGC document, Linked Data, Semantic Web, Graph Data

## **iii. Preface**

### **NOTE**

Insert Preface Text here. Give OGC specific commentary: describe the technical content, reason for document, history of the document and precursors, and plans for future work. > Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

This purpose of this document is to outline the benefits of representing geospatial data using semantic and graph technologies. It aims to provide motivation for OGC members to consider the publication of geospatial data using these technologies.

## **iv. Submitting organizations**

The following organizations submitted this Document to the Open Geospatial Consortium (OGC):

Organization name(s)

## **v. Submitters**

All questions regarding this submission should be directed to the editor or the submitters:

Joseph Abhayaratna (PSMA Australia) Linda van den Brink (Geonovum)

# Chapter 1. Introduction

The spatial aspect of data is a core data dimension. Spatial information provides the capability to answer the question "where", providing additional context to whatever it relates to. The authors of this White Paper believe that semantic and graph technologies, together with strategies for Linked Data, will increase the value that can be extracted from (geo)spatial data.

Although many current and past work within the OGC involves semantic and graph technologies, there is one main applicable OGC standard: GeoSPARQL. GeoSPARQL offers both semantics for representing geospatial data and it specifies how geospatial data that are expressed using the Resource Description Framework (RDF) can be queried using SPARQL, the query language for RDF.

This paper does four things. Firstly, it describes the benefits of representing geospatial data using semantics, graph, and web technologies. Secondly, it gives an overview of the current capabilities of the GeoSPARQL standard, showing that many benefits of semantic and graph technologies are already within reach. Thirdly, it outlines some shortcomings of the existing GeoSPARQL implementation specification that, if addressed, would unlock its potential to a greater extent, and could significantly increase its user base. Finally, it identifies other related activities that are current at the time of editing this paper. In doing so, it establishes liaison's between the different activities in an attempt to achieve alignment.

The purpose of this paper is to provoke further thought about a best course for further development of the GeoSPARQL standard, and to invite active involvement in that development. Particularly, the involvement of people and organisations that until now have not been able to put GeoSPARQL to good use, either because of perceived limitations or because of unfamiliarity with the standard, will be highly valued. Also, since one development under consideration is to make provisions for use of GeoSPARQL with non-geographic spatial data, those that see opportunities for using spatial data in a broad sense together with the aforementioned technologies are cordially invited to share their views.

# Chapter 2. Semantic Technologies

Semantic technologies, in the scope of this document, relate to the use of formal ontologies to describe concepts: the relationships between "things" and categories of "things". There are a number of goals for using these technologies, which often relate to providing the capability to reason over the data and simplified dealing with heterogeneous data sources. Example goals include enabling machine readability, and federated searching. A major area in which semantic technologies are used is the Semantic Web, the part of the World Wide Web dealing with semantics.

A concept closely related to the Semantic Web is Linked Data. Linked Data is the name given to the idea of using common Web communication protocol HTTP(S) to link raw data from different sources, much in the same way as HTML documents can cross-reference each other using hyperlinks. Data sets as well as individual data elements are identified by HTTP(S) IRIs, which act both as unique and persistent identifiers and allow direct retrieval of raw data. This leads to a web of data that can be tapped into by both human users and software. The concept is applicable to open data as well as data that are not supposed to be accessed by everyone.

# Chapter 3. Graph Technologies

Graph technologies, in the scope of this document, relate to storage and querying of data items stored as nodes and edges. Each node represents an entity or an instance (e.g., a person, place, or thing). Each edge represents a relationship between nodes. Graph databases are optimised for querying graph data structures making them useful for goals like computing the shortest path between two nodes or finding nodes related by particular patterns of relationship. This contrasts with the strict schema of the model used by relational databases that imposes limitations on how relationships can be queried. Graph systems also define their schema "as they go" since all nodes and edges are defined and new ones can always be added with their definitions. This makes data model expansion easy but still manageable which contrasts with both more fixed relational models and un-governed noSQL systems.



# Chapter 4. Beneficiaries and benefits

This section describes the beneficiaries and benefits of representing data, including geospatial data, using semantic and graph technologies. Furthermore, a collection of use cases demonstrate how semantic and graph technologies are used together with spatial data to tackle real world problems.

## 4.1. Beneficiaries

### 4.1.1. Beneficiary 1: Data consumers

Data consumers use data, including geospatial data. Their key needs are:

1. to access and understand data from multiple sources;
2. to integrate data - geospatial and non-geospatial; and
3. to get a data quality evaluation of the given geospatial data.

### 4.1.2. Beneficiary 2: Data publishers

Data publishers have data, including geospatial data. Their key needs are:

1. to maximise use of their data; 2. to manage their data as simply as possible; 3. to indicate possible usage contexts of the geospatial data; and 4. to quality-assure their geospatial data.

## 4.2. Benefits

The benefits of semantic and graph technologies are outlined below.

### 4.2.1. Benefit B1: Disambiguation

Using formal ontologies to describe entities and their relationships removes definitional ambiguity. This allows for accurate item reuse or new item creation (when things can't be reused). When definitions are coupled with uniform resource identifiers (URIs), global easy reuse and differentiation (are X & Y really the same thing?) are enabled as the global DNS namespace is used for everything.

### 4.2.2. Benefit B2: Dereferenceability

OWL and RDF-compliant graphs are composed entirely of nodes and edges identified by URIs or literal values, enabling any entity or relationship to have its own metadata, and for that metadata to be downloaded as required by a data consumer, whether human or machine. This 'dereferenceability' makes accessing and understanding models and content easy: it's accessible by normal Internet tools.

### 4.2.3. Benefit B3: Reasoning

Semantic Web graph systems use formal *description logic* to define and relate elements. This allows for rule-based or other computational reasoning to occur. New information may be calculated from

simple base data by following ontology *axioms* which leads to new knowledge, including multi-dataset traversals and validation of data.

#### 4.2.4. Benefit B4: Extendability

Graphs "carry their schema with them" since the data model is defined with each node and edge's definition. This allows data models to be expanded easily without becoming unmanageable. Whatever model extensibility is made, general graph traversal will be able to follow it and the definitions of the new information able to be extracted in predictable ways.

### 4.3. Use Cases

The benefits above have been successfully realised in a number of different industries, where semantic and graph technologies were implemented for a number of different use cases. Many of these cases have been taken almost verbatim from reference [3].

#### NOTE

Links should be created between benefits and use cases so as to highlight which benefits are realised through each use case.

#### 4.3.1. Use Case 1: Data integration

##### UC1A: Government and Public Administration: *SIRIS Academic*

To promote more transparent and inclusive governance in the Tuscany region SIRIS Academic, a small Spanish company specialized in providing data management solutions, has developed Tuscany's Observatory of Research and Innovation portal. They integrate Open Data in the Higher Education & Research field, including official Italian student and researcher data coming from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), and European data on FP7 and H2020 research projects. They follow the VKG approach and use the platform University Analytics (UNiCS) developed by SIRIS Academic. The platform uses Ontop to integrate open data repositories and to make them available via a dedicated SPARQL endpoint which is a query service for graph data. Then the platform shows the data as an interactive dashboard hosting data visualisations, with underlying data supplied by the UNiCS SPARQL endpoint.

##### *Benefits*

- B1: Disambiguation

##### UC1B Government and Public Administration: *Constitute Project*

Over the last 200 years, countries have replaced their constitutions on average every 19 years, and some have amended them almost yearly. A basic problem in the drafting of these documents is the search and analysis of model text deployed in other jurisdictions. In the Constitute Project [8], Ultrawrap was used to integrate the world's constitutions into a single unified semantic endpoint for contextual searching. The project was launched at the General Assembly of the United Nations in 2013 and continues to integrate over 196 current databases of all of the world's constitutions on the Web. Countries throughout the world can take advantage of this free service to modify and develop their constitutions.

## *Benefits*

- B1: Disambiguation

### **UC1C Health Care: HL7**

Semantic interoperability is essential when carrying out post-genomic clinical trials where several institutions collaborate, since researchers and developers need to have an integrated view and access to heterogeneous data sources. The work of Clinical data access shows how to query clinical data in HL7 RIM-based relational models using the Morph system. It presents a solution that uses R2RML mappings that relate an integrating ontology to the underlying relational database implementations. morph-RDB is used to expose a SPARQL endpoint to access the data in Semantic Web form too.

## *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

### **4.3.2. UC1D Mapping Authorities: LINDAS Initiative**

It is in the best interest of national agencies for cartography to provide services for other national authorities covering a wide range of topics. Usually, these topics are displayed using thematic maps (e.g. <https://www.bkg.bund.de/DE/Produkte-und-Services/Shop-und-Downloads/Landkarten/Karten-Downloads/Themenkarten/themenkarten.html>) which are created with respect to different demands of the general public, other national agencies or by the government. Thematic maps always highlight certain characteristics of a dataset (e.g. school accessibility) for which at least those characteristics should be firstly available and secondly in a usable state. Very often, attributes are required which are not collected by any governmental agency, so that crowdsourced data is correlated with already existing governmentally-administered data. This of course poses a big integrational problem which many agencies for cartography would like to solve by setting up linked data repositories which can be interlinked to further crowdsourced elements. To that end Switzerland launched its LINDAS initiative (<https://www.egovernment.ch/de/umsetzung/e-government-schweiz-2008-2015/lindas/>). In Germany, the Federal Agency for Cartography and Geodesie is aiming to create national ontology standards and to set up a linked data infrastructure in cooperation with the University Of Applied Sciences Mainz. (<http://i3mainz.de/de/projekte/intelligente-datenerfassung-oeffentliche-Verwaltung>) The European Union supports such initiatives by defining appropriate INSPIRE vocabularies (<https://github.com/inspire-eu-rdf/inspire-rdf-vocabularies>).

## *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

### **UC1E Government and Public Administration: Italian Public Debt Directorate**

The Italian Public Debt Directorate is responsible for various matters, such as issuance and management of the public debt, and analysis of the problems inherent to its management. The

Directorate is organized into offices that deal with specific aspects, and each sub-unit has an understanding of a particular portion of the public debt domain. However, a shared and formalized description of the relevant concepts and relations in the whole domain was missing, since data were managed by different systems in different offices, and their structure had been heavily modified and updated to serve specific application needs. There was a clear need to coordinate and integrate the data of the various sub-units. The work of the Italian Public Debt Directorate presented a project for addressing this issue. They developed the Public Debt Ontology to formalize the whole domain of the Italian public debt. The VKG system Mastro Studio has been used to provide a comprehensive software environment. Users can take advantage of the wiki-like documentation of the ontology to access both its graphical representation and its OWL2 specification.

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

### **4.3.3. Use Case 2: Data Product metadata**

#### **UC2 Construction:** *Semantic Construction Project Engineering (SCOPE) Project*

With the heterogeneous environment of the construction sector, providing suitable product descriptions for any use case and software application is hard to achieve. While open source exchange formats, i.e. IFC and STEP, can be used to describe products in a uniform manner to realise a communication across domains, the amount of required geometric detail is not addressed. For example, lights to indicate emergency exits are needed in different geometric detail. The electrical engineer only needs to know the position of the lighting fixture, whereas the architect requires the bounding box to consider for the design and safety engineers want to know the material, colour and shape of the lighting fixture to ensure that is clearly visible. On the other hand, the manufacturer needs to model the product in its highest geometrical detail for their own production chain.

If the manufacturer provides the highest geometrical detail, the product description will become too large to be handled if multiple instances are placed within the model. Hence, the geometrical detail needs to be broken down, ideally individually in respective of singular use cases, resulting in multiple geometry descriptions for the same object. By applying Linked Data, the attachment of multiple geometry descriptions to a singular object can be realised easily, maintaining means to differentiate between the descriptions and identify singular ones to connect them to their respective use cases. Yet, if the original geometry description changes, the derived geometry descriptions must be identified and updated, as well. This topic is, amongst others, considered in the Semantic Construction Project Engineering (SCOPE) research project funded by the German government and conducted by Ed. Züblin AG, Technische Universität Darmstadt and Fraunhofer Institute for Solar Energy (<https://www.projekt-scope.de/>).

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

- B4: Extendability

#### 4.3.4. Use Case 3: Recording Provenance

##### **UC3A Environmental Science:** *Australian Bioregional Assessments Programme*

To assemble the lineage of data processed by multiple systems and perhaps also by humans, manually, a consistent yet flexible lineage/provenance model is needed. Consistency of patterning is needed to ensure interoperability for information from multiple sources and yet flexibility is needed to accommodate different granularities of processing steps recorded. The PROV Data Model [6] is a graph-based generic, but easily extensible/specializable model for provenance representation. PROV information can be sampled (queried) to aggregate detailed low-level provenance, or drilled into for deeper details where they exist. The standard RDF format used by ontology variants of PROV allow for its storage in standard Semantic Web systems and accessibility via standard SPARQL queries. The strong definitions within PROV prevent unknown log formats being encountered in the future. The Australian Bioregional Assessments Programme [7] used PROV to record both dataset-level provenance (what the ancestors of data sets are) and also fine-grained processing steps for individual data elements within data sets meaning this very varied provenance can, nonetheless, be stored in one system and accessed sensibly.

##### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability
- B4: Extendability

##### **UC3B Libraries and Museums:** *German National Library and British Museum*

To preserve the national heritage of countries, libraries and museums have the task to collect information about artifacts, relate artifacts to other similar artifacts in different museums and to create a historic context for people to understand the artifacts provenance. Those tasks are more and more frequently achieved using linked data technologies and ontologies modeling the necessary data using appropriate vocabularies. One example is the German National Library which since many years develops the "Gemeinsame Normdatei" GND ontology (<https://d-nb.info/standards/elementset/gnd>) including a geospatial component designed to locate the artifacts origins and the origins of their creators. The British museum created a SPARQL endpoint based on Blazegraph which contains similar information about the artifacts displayed in the British museum.

##### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

##### **UC3C Architecture, Engineering and Construction:** *Niras*

During the design stages of a construction project, the building's design changes quite rapidly, and often there are derived consequences of these changes. The cooling demand of a zone is dependent on the solar heat gain through windows and if the windows change, so does the cooling demand.

This affects the capacity requirement of the fan coil in the room and potentially the size of the pipes supplying this fan coil, the pump circulating the cooling water and the size of the chiller.

The danish consulting engineering company Niras uses Linked Data to model these interdependencies. The architect's BIM model is translated from its internal data model through the vendor supplied Revit API into an AEC knowledge graph [13] described with the Building Topology Ontology (BOT) [14]. A direct communication between the BIM authoring tool and an OPM-REST API (<https://github.com/MadsHolten/OPM-REST>) ensures that property changes are captured and described using the Ontology for Property Management (OPM). Small task specific web applications access and extend the knowledge graph through SPARQL queries and uses OPM to relate a derived property to the properties that will affect it. In the current setup, 2D-geometry is extracted as WKT literals and 3D-geometry as OBJ literals. Geometry changes are registered by string comparison. In the UI, the state of the model geometry at the beginning and end of a given time interval is visualised.

#### *Benefits*

- B2: Dereferenceability
- B4: Extendability

### **4.3.5. Use Case 4: Data analysis**

#### **UC4 Oil and Gas Industry: Equinor**

One of the common tasks for geologists at Equinor (Norway) is to find new exploitable accumulations of oil or gas in given areas by analyzing data about those areas in a timely manner. However, gathering the required data is not a trivial task since it is stored in multiple complex and large data sources, including EPDS, Recall, CoreDB, GeoChemDB, OpenWorks, Compass and NPD FactPages. Construction of complex queries is sometimes beyond Equinor geologists, so they have to communicate their needs to IT specialists who then turn them into queries. This drastically affects the efficiency of finding the right data to back decision making. The work of Equinor describes how the data access and integration challenges in Equinor have been addressed by adopting the VKG-based system Optique, which relies on the following tools:

1. the bootstrapper BootOX to create ontologies and mappings from relational databases in a semi-automatic fashion;
2. the VKG system Ontop to perform query reformulation;
3. the federator Exareme to evaluate the reformulated queries over the federated DBs; and
4. the query formulation module OptiqueVQS to support query construction for engineers with a limited IT background.

#### *Benefits*

- B1: Disambiguation
- B3: Reasoning

### 4.3.6. Use Case 5: Diagnoses

#### UC5A Industrial Machinery: *Siemens*

Siemens Energy runs several service centers that remotely monitor and perform diagnostics for several thousand appliances, such as gas and steam turbines, generators and compressors installed in power plants. For performing reactive and predictive diagnostics at Siemens, data access and integration of both static data (e.g., configuration and structure of turbines) and dynamic data (e.g., sensor data) are particularly important but very challenging. The work of Siemens addressed these data access requirements by using the Optique platform as a VKG solution, similar to the Equinor use case.

##### *Benefits*

- B1: Disambiguation
- B3: Reasoning

#### UC5B Health Care: *Diagnosis of Diabetes*

Improving health care for people with chronic conditions requires clinical information systems that support integrated care and information exchange. The adoption of an approach based on semantic information simplifies the use of multiple and diversified Electronic Health Records (EHRs). Within the work described in E-health data access, a Diabetes Mellitus Ontology (DMO) has been developed, and has been used to diagnose patients with diabetes, and automatically identify them by analyzing EHRs. Specifically, by using Ontop, the EHR data from a general practice (with almost 1,000 active patients) could be queried via SPARQL. The accuracy of the algorithm for automatic identification of patients with diabetes was validated by performing a manual audit of the EHRs, and considered good enough for the purpose. Not surprisingly, the accuracy of the automatic method was influenced by data quality, such as incorrect data due to mistaken units of measurement, unavailable data due to lack of or wrong documentation, and data management errors.

##### *Benefits*

- B1: Disambiguation
- B3: Reasoning

### 4.3.7. Use Case 6: Simplified Access to Heterogeneous Data

#### UC6A Digital Humanities: *EPNet Project*

Historians, especially in Digital Humanities (DH), are starting to use new data sets to aggregate information about history. These are collections of data, information and knowledge that are devoted to the preservation of the legacy of tangible and intangible culture inherited from previous generations. In the project Production and distribution of food during the Roman Empire: Economics and Political Dynamics (EPNet), the work of EPNet project presents a framework that eases the access of scholars to much food information during the Roman Empire, distributed across different data sources. The proposed approach relies on the VKG paradigm to integrate the following data sets:

1. the EPNNet relational repository;
2. the Heidelberg Epigraphic database; and
3. Pleiades, an open-access digital gazetteer for ancient history.

An ontology provides the historians with a clear point of access and a unified and unambiguous conceptual view over these data sets.

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

#### **UC6B Archaeology:** *Archaeology and the Semantic Web*

Digital Archaeologists working in DH deal with a lot of heterogeneous data, which is not standardised at all. Semantic technologies and the use of Linked Open Data promises to revolutionise the digital workflow [<https://eprints.soton.ac.uk/206421/>]. As the most digital semantic DH project they are referenced by the International Committee for Documentation (CIDOC) Conceptual Reference Model (CRM) [<http://www.cidoc-crm.org/>] and its extensions, especially CRMgeo[<https://link.springer.com/article/10.1007/s00799-016-0192-4>]. Famous data collections which model object types in their domain and publish them as LOD are nomisma (coins) [<http://nomisma.org/>], kerameikos (ancient ceramics) [<http://kerameikos.org/>], Open Context [<https://opencontext.org/>], the iDAI world [<https://idai.world/>] of the German Archaeological Institute, finds.org [<https://finds.org.uk/>], and Regnum Francorum Online [<http://francia.ahlfeldt.se/index.php>]. Furthermore, Linked Data networks of the Computer Applications and Quantitative Methods in Archaeology (CAA) conference – Little Minions, Data Dragons – and of the Linked Past Community (related to the LOD Pelagios Commons network[<http://commons.pelagios.org/>] – Linked Pipes – try to build up a LOD network of tools, workflows and data of the CH domain[<http://squirrelnator.squirrel.link/>]. Moreover, smaller projects are publishing tools, e.g. for modelling vagueness in graphs like the Academic Meta Tool [<http://academic-meta-tool.xyz/>] to enable the scientific community to handle fuzzy (geographical) relations [[http://unold.net/research/p\\_dls\\_20170320.pdf](http://unold.net/research/p_dls_20170320.pdf)].

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

### **4.3.8. Use Case 7: Integrating Aspatial and Spatial Data**

#### **UC7A Maritime security:** *Real-time Maritime Situational Awareness System*

The maritime security domain presents a need for efficient combining and processing of dynamic (real-time) and static vessel data that come from heterogeneous sources. The project Real-time Services for the Maritime Security (EMSec) needed to integrate static, real-time and geospatial data, including:

1. static vessel metadata;



2. open data like GeoNames and OpenStreetMap;
3. large radar and satellite images; and
4. real-time vessel data (approximately 1,000 vessel positions are acquired per second).

To address this objective, the system Real-time Maritime Situation Awareness System (RMSAS), which relies on the VKG technology, has been developed. RMSAS uses Ontop (with the Ontop-spatial extension) to expose the data mentioned above as SPARQL endpoints. The Web-based tool Sextantis then used to visualize the results on temporally-enabled maps combining geospatial and temporal results from different (Geo)SPARQL endpoints.

### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability
- B3: Reasoning
- B4: Extendability

### **UC7B Heritage:** *Flemish Cities in Transition*

During built heritage projects (e.g. restoration, maintenance, historical research) a large amount of stakeholders collaborate. Each stakeholder assembles and generates a wide variety of data, including 2D and 3D geometries ranging from survey geometry (e.g. a point cloud or complex mesh), over 2D plans and maps (historical situation, previous restorations, derived from survey data, etc.) to volumetric 3D models. These geometries are used to get an overview of the historical and existing situation of the building, for communicating the location of damages or valuable historical elements in the building or to express the intention of the restoration design. Because of the wide variety of geometric data, a large amount of common geometry schemas (text-based and binary, open and proprietary) are currently used in practice.

Instead of developing RDF-based geometry schemas for each existing geometry schema (OBJ, E57, X3D, STEP, WKT, etc.), alternative methods such as the application of RDF literals are considered. These literals can embed geometry descriptions (similar to GeoSPARQL 1.0 but for any geometry schema) or reference external geometry files in their original geometry schema. In that case, the usage of existing geometry schemas and their tools can be continued. Built heritage stakeholders need to be able to link such geometry descriptions to building elements, damages and building spaces they describe. Each described object can have multiple geometry descriptions (different geometry schema, describing an object at multiple moments in time, different amount of detailing/resolution, etc.), potentially coming from different stakeholders. Geometry metadata (accuracy, author, resolution, derived geometry descriptions, file size, etc.) is necessary to reuse the geometry in a collaborative setting as it gives an indication of the geometry provenance. Other metadata (used geometry schema, coordinate system, etc.) might help users in the automatic processing of the data by their geometry applications.

Three domain independent ontology modules have been developed in previous collaborative research and are applied in a built heritage PhD research project named “Flemish Cities in Transition” [9]. These ontologies include the Ontology for Managing Geometry (OMG - <https://w3id.org/omg#>) [10], the File Ontology for Geometry formats (FOG - <https://w3id.org/fog#>)

[11] and the Geometry Metadata Ontology (GOM - <https://w3id.org/gom#>).

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability

#### **UC7C Buildings: *Prime2***

Geometric data plays a central role in the geospatial domain, architectural design and construction industry. For upcoming, new approaches on how to store building data, such as the Semantic Web, however, no universal common agreement exists on the combination of geometric and non-geometric data. Thus, it can be unclear to users on how to represent their geometries, leading to a decelerated application and advancement of making building data available over the web. This gap can only be bridged if a common approach on the representation of geometries on the web is achieved.

In Ireland, the Ordnance Survey Ireland (OSi) has a substantial dataset (over 50 million objects), called Prime2, which includes not only GIS data (polygon footprint, geodetic coordinate), but also additional building-specific data (form and function). The ADAPT research centre working with the Ordnance Survey Ireland has begun publication of their geospatial data using GeoSPARQL [12], with a subset of their buildings data (building name, geolocation, and form and function) in the county of Galway now being available as RDF (<http://data.geohive.ie/downloadAndQuery.html>).

This provides authoritative URIs for Irish buildings which can be used to interlink building data from other domains, such as products, sensors, energy, etc. The potential also exists to support the conversion of their 2D building footprints into a simple 3D geometric model, given some additional properties (height). An existing schema such as the Industry Foundation Classes, and ifcOWL serialization can be supported, but tend to be overly verbose (use of lists for each vertex in a point for example) and geometric and non-geometric data are overly entwined. The possibility to define 3D geometries using less complex geometry schemas would be a huge advantage within the building information modelling domain. This is an important step towards the iterative integration of ever more complex BIM models which can support a range of different use cases into the wider web of data.

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability
- B4: Extendability

### **4.3.9. Use Case 8: Data Mining**

#### **UC8 Cybersecurity: *EBITmax***

Process mining techniques are able to extract knowledge from event log data, which is often available in today's information systems. Process mining tools normally assume that the data to be analyzed are already organized in some specific textual (XML based) format, notably IEEE standard

for eXtensible Event Stream (XES) for achieving interoperability in event logs and event streams. However, in practice, many companies have custom IT infrastructure that maintains the data relevant for process logs, e.g., in relational databases, and hence in forms not compliant with the XES standard. To cope with this kind of problem, the approach proposed exploits a VKG based framework and associated methodology for the extraction of XES event logs from relational data sources. This approach is implemented in OnProm, which provides a complete tool-chain that:

1. allows for describing event logs by means of suitable annotations of a conceptual model of the available data;
2. exploits the Ontop system for the actual log extraction; and
3. is fully integrated with the well-known ProM process mining framework.

It has been tested in EBITmax, an Italian company that provides consultancy services in program management and business process management for small and large enterprises, and that has incorporated process mining to complement its standard consultancy services. The experimentation has shown the added value and flexibility of an approach based on semantics for the semi-automatic generation of process logs from legacy data.

#### *Benefits*

- B1: Disambiguation
- B2: Dereferenceability
- B4: Extendability

### **4.3.10. Use Case 9: Improving Search**

#### **UC9 Smart Cities: DALI**

Smart City applications rely on large amounts of data retrieved from sensors, social networks or government authorities. Open data and data from existing enterprise systems are two valuable data sources. However, open data are often published in a tabular form with little or incomplete schema information, while enterprise applications typically rely on complex relational schemas. There is a clear need to make city-specific information easy to consume and combine at low cost, but this proves to be a difficult task. The work of IBM Ireland presents the system DALI, which exploits Linked Data to provide federated entity search and spatial exploration across hundreds of information sources containing open and enterprise data pertaining to cities. Ontop is used as the VKG solution, and mappings are created using a rule and pattern-based entity extraction mechanism to detect different kinds of entities. The DALI system has been evaluated in two scenarios:

1. Data Engineers bring together public and enterprise data sets about public safety; and
2. Knowledge Engineers and domain-experts build a view of health and social care providers for vulnerable populations.

#### *Benefits*

- B1: Disambiguation

# Chapter 5. Current capabilities

## 5.1. GeoSPARQL

GeoSPARQL is the most common geospatial extension of SPARQL. It was accepted as an OGC standard in 2012.

According to the standard document, "The OGC GeoSPARQL standard supports representing and querying geospatial data on the Semantic Web. GeoSPARQL defines a vocabulary for representing geospatial data in RDF, and it defines an extension to the SPARQL query language for processing geospatial data".

### 5.1.1. Requirements addressed

GeoSPARQL addresses the following requirements.

#### **Requirement 1: Integrate with SPARQL**

Implementations shall support the SPARQL Query Language for RDF [W3C SPARQL], the SPARQL Protocol for RDF [W3C SPARQL Protocol] and the SPARQL Query Results XML Format [W3C SPARQL Result Format].

#### **Requirement 2: Represent Spatial Objects in SPARQL**

Implementations shall allow the RDFS class `geo:SpatialObject` to be used in SPARQL graph patterns.

#### **Requirement 3: Represent Features in SPARQL**

Implementations shall allow the RDFS class `geo:Feature` to be used in SPARQL graph patterns.

#### **Requirement 4: Represent Simple Features Spatial Relationships in SPARQL**

Implementations shall allow the properties `geo:sfEquals`, `geo:sfDisjoint`, `geo:sfIntersects`, `geo:sfTouches`, `geo:sfCrosses`, `geo:sfWithin`, `geo:sfContains`, `geo:sfOverlaps` to be used in SPARQL graph patterns

#### **Requirement 5: Represent Egenhofer Spatial Relationships in SPARQL**

Implementations shall allow the properties `geo:ehEquals`, `geo:ehDisjoint`, `geo:ehMeet`, `geo:ehOverlap`, `geo:ehCovers`, `geo:ehCoveredBy`, `geo:ehInside`, `geo:ehContains` to be used in SPARQL graph patterns

#### **Requirement 6: Represent RCC8 Spatial Relationships in SPARQL**

Implementations shall allow the properties `geo:rcc8eq`, `geo:rcc8dc`, `geo:rcc8ec`, `geo:rcc8po`, `geo:rcc8tpi`, `geo:rcc8tpp`, `geo:rcc8ntpp`, `geo:rcc8ntppi` to be used in SPARQL graph patterns

### **Requirement 7: Represent Geometry in SPARQL**

Implementations shall allow the RDFS class `geo:Geometry` to be used in SPARQL graph patterns

### **Requirement 8: Integration of Geometry with aspatial data**

Implementations shall allow the properties `geo:hasGeometry` and `geo:hasDefaultGeometry` to be used in SPARQL graph patterns

### **Requirement 9:**

Implementations shall allow the properties `geo:dimension`, `geo:coordinateDimension`, `geo:spatialDimension`, `geo:isEmpty`, `geo:isSimple`, `geo:hasSerialization` to be used in SPARQL graph patterns

### **Requirement 10: Conformance Requirements for Geometry Encoding**

All RDFS Literals of type `geo:wktLiteral` shall consist of an optional URI identifying the coordinate reference system followed by Simple Features Well Known Text (WKT) describing a geometric value. Valid `geo:wktLiterals` are formed by concatenating a valid, absolute URI as defined in [RFC 2396], one or more spaces (Unicode U+0020 character) as a separator, and a WKT string as defined in Simple Features [ISO 19125-1]

### **Requirement 11: Adopt WGS84 as Default CRS (change to more explicit code)**

The URI <http://www.opengis.net/def/crs/OGC/1.3/CRS84> shall be assumed as the spatial reference system for `geo:wktLiterals` that do not specify an explicit spatial reference system URI.

### **Requirement 12: Inherit axis order from Spatial Reference System**

Coordinate tuples within `geo:wktLiterals` shall be interpreted using the axis order defined in the spatial reference system used

### **Requirement 13: Interpret empty RDFS as empty Geometry**

An empty RDFS Literal of type `geo:wktLiteral` shall be interpreted as an empty geometry

### **Requirement 14: Represent Geometries as WKT**

Implementations shall allow the RDF property `geo:asWKT` to be used in SPARQL graph patterns

### **Requirement 15: Conform with GML**

All `geo:gmlLiterals` shall consist of a valid element from the GML schema that implements a subtype of `GM_Object` as defined in [OGC 07-036]

### **Requirement 16: Interpret empty GML Literals as empty Geometry**

An empty `geo:gmlLiteral` shall be interpreted as an empty geometry

### **Requirement 17: Document GML profiles**

Implementations shall document supported GML profiles

### **Requirement 18: Represent Geometries as GML**

Implementations shall allow the RDF property `geo:asGML` to be used in SPARQL graph patterns

### **Requirement 19:**

Implementations shall support `geof:distance`, `geof:buffer`, `geof:convexHull`, `geof:intersection`, `geof:union`, `geof:difference`, `geof:symDifference`, `geof:envelope` and `geof:boundary` as SPARQL extension functions, consistent with the definitions of the corresponding functions (distance, buffer, convexHull, intersection, difference, symDifference, envelope and boundary respectively) in Simple Features [ISO 19125-1]

### **Requirement 20: Support Simple Features `getSRID` in SPARQL Queries**

Implementations shall support `geof:getSRID` as a SPARQL extension function.

### **Requirement 21:**

Implementations shall support `geof:relate` as a SPARQL extension function, consistent with the relate operator defined in Simple Features [ISO 19125-1]

### **Requirement 22: Support Simple Features Spatial Relationship Functions in SPARQL Queries**

Implementations shall support `geof:sfEquals`, `geof:sfDisjoint`, `geof:sfIntersects`, `geof:sfTouches`, `geof:sfCrosses`, `geof:sfWithin`, `geof:sfContains`, `geof:sfOverlaps` as SPARQL extension functions, consistent with their corresponding DE-9IM intersection patterns, as defined by Simple Features [ISO 19125-1]

### **Requirement 23: Support Egenhofer Spatial Relationship Functions in SPARQL Queries**

Implementations shall support `geof:ehEquals`, `geof:ehDisjoint`, `geof:ehMeet`, `geof:ehOverlap`, `geof:ehCovers`, `geof:ehCoveredBy`, `geof:ehInside`, `geof:ehContains` as SPARQL extension functions, consistent with their corresponding DE-9IM intersection patterns, as defined by Simple Features [ISO 19125-1]

### **Requirement 24: Support RCC8 Spatial Relationship Functions in SPARQL Queries**

Implementations shall support `geof:rcc8eq`, `geof:rcc8dc`, `geof:rcc8ec`, `geof:rcc8po`, `geof:rcc8tppi`, `geof:rcc8tpp`, `geof:rcc8ntpp`, `geof:rcc8ntppi` as SPARQL extension functions, consistent with their corresponding DE-9IM intersection patterns, as defined by Simple Features [ISO 19125-1]

### **Requirement 25: Support RDFS Entailment**

Basic graph pattern matching shall use the semantics defined by the RDFS Entailment Regime [W3C SPARQL Entailment]

**Requirement 26:**

Implementations shall support graph patterns involving terms from an RDFS/OWL class hierarchy of geometry types consistent with the one in the specified version of Simple Features [ISO 19125-1]

**Requirement 27:**

Implementations shall support graph patterns involving terms from an RDFS/OWL class hierarchy of geometry types consistent with the GML schema that implements GM\_Object using the specified version of GML [OGC 07-036]

**Requirement 28: Support RIF Entailment for Simple Features Spatial Relationships**

Basic graph pattern matching shall use the semantics defined by the RIF Core Entailment Regime [W3C SPARQL Entailment] for the RIF rules [W3C RIF Core] `geor:sfEquals`, `geor:sfDisjoint`, `geor:sfIntersects`, `geor:sfTouches`, `geor:sfCrosses`, `geor:sfWithin`, `geor:sfContains`, `geor:sfOverlaps`

**Requirement 29: Support RIF Entailment for Egenhofer Spatial Relationships**

Basic graph pattern matching shall use the semantics defined by the RIF Core Entailment Regime [W3C SPARQL Entailment] for the RIF rules [W3C RIF Core] `geor:ehEquals`, `geor:ehDisjoint`, `geor:ehMeet`, `geor:ehOverlap`, `geor:ehCovers`, `geor:ehCoveredBy`, `geor:ehInside`, `geor:ehContains`

**Requirement 30: Support RIF Entailment for RCC8 Spatial Relationships**

Basic graph pattern matching shall use the semantics defined by the RIF Core Entailment Regime [W3C SPARQL Entailment] for the RIF rules [W3C RIF Core] `geor:rcc8eq`, `geor:rcc8dc`, `geor:rcc8ec`, `geor:rcc8po`, `geor:rcc8tppi`, `geor:rcc8tpp`, `geor:rcc8ntpp`, `geor:rcc8ntppi`

# Chapter 6. Learnings from use

This section provides an overview of feedback received on the current version of the GeoSPARQL standard (version 1.0). This feedback helps to identify some of the barriers to use, and to outline requirements that have not been addressed that may encourage greater uptake.

## 6.1. Proposed extensions to GeoSPARQL 1.0

### 6.1.1. Extension 1: Extend expressive power of ontology for encountered use cases

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=601](http://ogc.standardstracker.org/show_request.cgi?id=601)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/15>

#### Category

Semantic improvement

#### Description

We request extensions to the GeoSPARQL OWL ontology that includes details it doesn't cater for such as indicating spatial resolution, areas, proportional relations (e.g. area overlaps) and roles for multiple Geometry objects relating to a single Feature.

We have packaged the extensions we are already using in an ontology that is mostly built on GeoSPARQL but it also uses elements from QUDT and the some supporting datatype vocabularies.

The "GeoSPARQL Extensions Ontology" is online, and can be found here: <https://github.com/CSIRO-enviro-informatics/geosparql-ext-ont/>.

### 6.1.2. Extension 2: Make underlying definitions available on the semantic web

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=600](http://ogc.standardstracker.org/show_request.cgi?id=600)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/14>

#### Category

Accessibility



## Description

GeoSPARQL is currently lacking in semantics that are available on the semantic web. For example, the concept of Geometry is not explained in GeoSPARQL itself, but has in its text definition a pointer to ISO 19107: "This class is equivalent to the UML class GM\_Object defined in ISO 19107". Should one look up that document on the web, a paywall from ISO is hit. Such fundamental definitions should be directly accessible to consumers, both human and machine, on the semantic web, either as part of GeoSPARQL or as another web vocabulary.

### 6.1.3. Extension 3: Include all functions already described in Simple Features for SQL

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=599](http://ogc.standardstracker.org/show_request.cgi?id=599)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/13>

#### Category

Geospatial processing

## Description

GeoSPARQL describes a number of SPARQL functions that can be implemented by developers of SPARQL endpoints. It would be beneficial for working with spatial data in SPARQL if more ways of interacting with spatial data were available. The routines and functions described in the Simple Features for SQL specification (06-104r4) are well known and have been implemented in different libraries and data storage systems. It could be a relatively simple effort to transpose the existing definitions to RDFS/OWL, so developers of SPARQL systems can implement those often very useful functions, either by using existing libraries, or by new development.

### 6.1.4. Extension 4: Add semantics for CRS

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=598](http://ogc.standardstracker.org/show_request.cgi?id=598)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/12>

#### Category

Semantic improvement

## Description

Descriptions of Coordinate Reference Systems (CRS) are important in correct interpretation of spatial data, and in working with that kind of data. However, CRS descriptions are not readily available as Linked Data. Therefore, it would be good to have GeoSPARQL use a standard ontology for describing CRS, either as an extension or module of GeoSPARQL itself, or an endorsed external ontology.

### 6.1.5. Extension 5: Explicitly expect significant digits in numerical data

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=597](http://ogc.standardstracker.org/show_request.cgi?id=597)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/11>

#### Category

Semantic improvement

## Description

Numerical data published using GeoSPARQL often have an unrealistic amount of digits. At the same time, there is no standardized way to express spatial resolution yet. Both these problems can be addressed when numerical data, published using GeoSPARQL, are explicitly expected to have only significant digits.

### 6.1.6. Extension 6: Widen the scope to all spatial data

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=594](http://ogc.standardstracker.org/show_request.cgi?id=594)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/9>

#### Category

Increased scope

## Description

GeoSPARQL's scope is geographic data, as the name says. Less explicit, GeoSPARQL is only about vector data. However, there is a need for a web ontology that can be used to work with all kinds of spatial data. GeoSPARQL seems to be the best candidate for realization of a domain independent ontology for spatial data.

A universal, or domain independent ontology for spatial data is needed because space is a

phenomenon that exists everywhere and is present in many kinds of human endeavour. Traditionally, universal phenomena like time and space have been modelled in different domains, according to domain specific requirements. Linked Data and the semantic web now offer a way to share data with many different perspectives, in a domain independent way. A domain independent ontology for time already exists: 0 . The time has now come for space to have a similar ontology. Practically, this will greatly increase interoperability of spatial data. Not only on the web: offline systems (e.g storage systems and libraries) could also benefit from having a single root model to depend on.

GeoSPARQL is a good candidate for evolving into a general ontology for spatial data because: 1. The Semantic Web allows direct open and modular access to all definitions. 2. OGC has a large canon for spatial data modelling ready for re-use. Existing OGC models have sound mathematical foundations that are applicable outside the geography domain. 3. OGC has been broadening its scope. Broadening the scope of GeoSPARQL should fit in nicely with that development. Examples of domains that are using different ways of working with spatial data, but increasingly do need to interoperate with geographic data are building information modelling (BIM) and 3D visualization. 4. OGC is an esteemed authority for standard specifications (although further collaboration with W3C would be beneficial). Widening the scope of GeoSPARQL would certainly mean the ontology becoming much bigger. Further modularization should prevent the ontology becoming unwieldy and users becoming overwhelmed with information which is not required for their purposes. Modularization can also be used to make distinctions between vector and coverage data, where required, but to share fundamentals too.

This subject has been discussed in the Spatial Data on the Web Working Group and is a project proposal in the Spatial Data on the Web Interest Group, found here: <https://github.com/w3c/sdw/issues/1095>

### 6.1.7. Extension 7: Availability in JSON-LD format

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=591](http://ogc.standardstracker.org/show_request.cgi?id=591)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/8>

#### Category

Accessibility

#### Description

The GeoSPARQL ontology is available online in XML and TTL formats. JSON-LD could be added as an additional publication format, supported by content negotiation of course. This will allow easier consumption of the ontology by web pages. This, in turn, allows easier consultation of the ontology by humans. For example, parts of the ontology could be visualized as diagrams, or definitions of terms could be rendered as tooltips on web pages.

## 6.1.8. Extension 8: Extending GeoSPARQL by defining more vector literal types

### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=585](http://ogc.standardstracker.org/show_request.cgi?id=585)

### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/6>

### Category

Geospatial encoding

### Description

GeoSPARQL currently offers WKT and GML literal types which can be used to encode vector geometries. However, geospatial data formats are very heterogeneous and a variety of other data formats deserve to be encodable in GeoSPARQL in our opinion.

In the following we list the most likely candidates: - GeoJSON: Very common throughout the Web - GeoHash: Common has representations of vector data - GPX: GPX Format used in GPS trackers - KML Format by Google - (H)(E)WKB/TWKB: Binary serializations of WKT often used as an internal storage format in databases

The following formats could be considered, but are in our opinion optional: - LatLonText: Common format to display points in e.g. Wikidata or OSM - GeoURI: Defacto standard for mobile phone geo urls - Geobuf Format - OSM Format: OSM XML - Polyshape/EncodedPolyline: Format developed by Google to encode polylines/polyshapes - SVG: Web standard for graphics in general - X3D: Standard to visualize 3D geometries

Implementations of most of the described literals can be seen in an extension for rdf4j and an extension for Jena, and in a proposed ontology here: <https://github.com/i3mainz/geosparql2.0>

Possibly, other literal implementations are useful and could be discussed.

This PDF provides links to all the dataformat specifications and proposes how literal representations could look like: <https://github.com/opengeospatial/geosemantics-dwg/blob/master/CR585attachment.pdf>.

## 6.1.9. Extension 9: Extending the GeoSPARQL ontology with support for raster data

### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/41>

#### **Category**

Geospatial encoding

#### **Description**

GeoSPARQL is currently incapable of encoding and dealing with raster data. However, raster data is essential for many geospatial applications and supported by many of relational geospatial databases such as POSTGIS. Raster data even provides semantics, as interpretations of raster data can be given by interpreting the color codes of raster bands. In order to integrate raster data into GeoSPARQL the following requirements need to be fulfilled: - Extending the GeoSPARQL ontology to include support for GridCoverages - Extending the GeoSPARQL ontology with raster literal types such as CovJSON,GMLCOV,GeoTIFF - Extending the GeoSPARQL ontology with vocabularies to describe raster data content

### **6.1.10. Extension 10: Extending the GeoSPARQL ontology with raster data query capabilities**

#### **Standards Tracker Change Request URI**

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/42>

#### **Category**

Query language

#### **Description**

If the GeoSPARQL ontology is able to support raster data, new query capabilities are needed in order to use raster data in daily applications. In particular the following query capabilities which are the norm in relational GIS databases should be adopted: - Raster algebra operations - Raster relation functions (ST\_Within, ST\_Covers...) - Vectorization and Rasterization capabilities - Raster modification capabilities (e.g. ST\_AddBand)

### **6.1.11. Extension 11: Extending the GeoSPARQL ontology with full-featured 3D support**

#### **Standards Tracker Change Request URI**

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/43>

#### **Category**

Ontology

#### **Description**

GeoSPARQL currently only explicitly supports 2D geometries. However, work has already been done in defining ontologies for 3D geometries (e.g. <https://github.com/w3c-geom-cg/geom> or <https://www.web3d.org/working-groups/x3d-semantic-web/charter>). These ontologies should be checked and integrated into, or merged with, the GeoSPARQL ontology.

### **6.1.12. Extension 12: Extending the GeoSPARQL ontology with functions to handle 3D geometries**

#### **Standards Tracker Change Request URI**

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/44>

#### **Category**

Ontology

#### **Description**

The GeoSPARQL query language currently only supports functions capable of dealing with 2D geometries. However, with the emergence of standards such as CityGML which could be supported as GML literals, specific 3D-aware functions should be added to GeoSPARQL to accommodate such recent developments. Suggestions: - 3D-aware functions of the RCC8 calculus - ST\_Distance3D - ST\_Length3D - ST\_Difference3D

### **6.1.13. Extension 13: Extending the GeoSPARQL ontology with support for M and T coordinates**

#### **Standards Tracker Change Request URI**

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/45>

## Category

Query language

## Description

Many geospatial libraries such as JTS (<https://github.com/locationtech/jts>) provide explicit support for geometries with measurement (M) coordinates. These are useful in a variety of applications e.g. when a road is simplified in a query statement but users still would like to query the correct amount of kilometres since its start. The time coordinates (T) are useful when working with GNSS tracks in order to track per-point when a user went to a particular place. While the latter can also be achieved by modeling every point of a GNSS track as its own point geometry, it is unnecessary if the points provide no further semantic information apart from the time point. While the support XYZM or XYZMT coordinates is not a matter of GeoSPARQL itself but more of the formats which are supported as literals in the query language, GeoSPARQL could provide definitions of functions which are aware of these extended coordinate concepts such as: - ST\_M/ST\_T - ST\_FilterByM/ST\_FilterByT - ST\_PartOfGeometryBefore - ST\_PartOfGeometryAfter - ST\_PartOfGeometryAt

### 6.1.14. Extension 14: Propose how authoring metadata should be included

#### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/46>

## Category

Semantic improvement

## Description

In AEC, authoring metadata (e.g. author, date, revision, etc.) is of high relevance for multiple aspects of collaboration, such as coordination and legal issues. Hence, the domains requires authoring metadata to be attached to any kind of information—and with geometry oftentimes being the core structure for non-geometric information—this is even more important for geometry descriptions. To avoid inconsistent and varying attachments of authoring metadata, we suggest to formulate a best practice for enhancing GeoSPARQL triples with authoring information, ideally by reusing existing concepts of provenance ontologies, e.g. PROV-O, DCTerms, etc.

### 6.1.15. Extension 15: Add concepts for accuracies and tolerances

#### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/47>

#### Category

Semantic improvement / Increased Scope

#### Description

The AEC domain struggles with geometric representations of planned objects and built objects and corresponding tolerances and inaccuracies, respectively. Planned objects are created on a scratchpad, while construction sites do not offer perfect conditions to recreate the planned geometry completely. Depending on the construction material, the geometry descriptions are commonly enriched with tolerance values, ranging from millimeters (steelwork) to centimeters (masonry). For the geometry of already built objects, the (measured) accuracy is also not perfect, as measuring techniques cannot provide flawless representations. Furthermore, by processing or simplifying geometry descriptions (e.g. from point cloud to mesh), inaccuracies can occur, which are also of interest to attach (represented accuracy). Hence, the possibility to attach accuracies (measured or calculated accuracy) and tolerances would be beneficial for building geometry.

### 6.1.16. Extension 16: Enable semantic descriptions of the applied geometry representation contexts

#### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/48>

#### Category

Semantic improvement

#### Description

Some geometry schemes can contain a variety of geometry representation contexts (BREP, CSG, NURBS, etc.) but not all applications can deal with every geometry representation context. To ease the integration of geometry descriptions into applications, metadata regarding the geometry representation context at hand could help to automatically retrieve suitable descriptions only. Example: DWG can contain a 2D drawing, a 3D mesh and/or 3D BREP geometry.

### 6.1.17. Extension 17: Allow multiple modelling levels for connection patterns between objects and geometry descriptions

#### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).



## GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/49>

## Category

Modelling patterns

## Description

Currently, GeoSPARQL requires one intermediate node to attach geometry descriptions to objects. However, in some cases, this might be too complicated and reduce querying performance, while other cases might require additional nodes (e.g. version control). In the Ontology for Managing Geometry (OMG, <https://w3id.org/omg>), multiple levels for connecting geometry descriptions and objects exist. OMG level 1 implements direct connections between objects and their geometry description (omg:hasSimpleGeometryDescription and omg:hasComplexGeometryDescription), level 2 adds one intermediate node (omg:Geometry), similar to GeoSPARQL 1.0, and the third level adds another node (omg:GeometryState) for versioning purposes. The selection which level is used in a situation depends on the required features (balance between simplicity and functionality).

In AEC, for example, versioning, and thus level 3, is needed during (1) design phase of buildings and (2) for modeling change over time of a building (e.g. changes to building elements, geometry from multiple surveys over time), while data exchange without need for additional metadata would be most performant with level 1 and storage of certain planning stages would be ideally implemented with level 2 to allow multiple geometry descriptions.

Also see the following publications:

- Wagner, A., Bonduel, M., Pauwels, P., & Uwe, R. (2019). Relating geometry descriptions to its derivatives on the web. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 304–313). Chania, Greece. <https://doi.org/10.35490/EC3.2019.146>
- Bonduel, M., Wagner, A., Pauwels, P., Vergauwen, M., & Klein, R. (2019). Including widespread geometry formats in semantic graphs using RDF literals. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 341–350). Chania, Greece. <https://doi.org/10.35490/EC3.2019.166>

## 6.1.18. Extension 18: Need for a universal linking method between objects and geometry descriptions in any existing schema

### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

## GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/50>

## Category

Increased scope

## Description

Geometry can be described in different ways than SFA/GML snippets in RDF literals. A uniform connector ontology with supplementary metadata would help collaboration across different domains and companies. Hence the different approaches to include geometry in a Semantic Web context should be allowed by the linking method: 1) RDF-based geometry following a dedicated ontology as the geometry schema (e.g. GEOM (<https://github.com/w3c-geom-cg/geom>), OntoBREP (<https://github.com/OntoBREP/ontobrep>), OCC (<http://w3id.org/occ>), etc.); 2) RDF literal embedding the content of a geometry file (~ GeoSPARQL 1.0 for 2D WKT and 2D GML); 3) RDF literal containing a reference to an external geometry file.

For each approach, a wide variety of geometry schemes exist for different use cases. All these schemes should be supported (i.e. binary or text-based, open or proprietary, 2D (vector or raster) or 3D (BREP, CSG, NURBS, point clouds, meshes), RDF-based or other (XML-based, JSON-based, SPFF-based, custom, etc.)). To achieve this support, we suggest the following adaptations:

- Binary geometry descriptions: GeoSPARQL 1.0 uses both the datatype property (e.g. `geosparql:asWKT`) and the datatype (e.g. `geosparql:wktLiteral`) to express the used geometry schema. Instead, the datatype could be used to add information about the used text encoding of the binary geometry descriptions (base64, hexadecimal, base32, base122, etc.) using XSD or custom datatypes
- RDF-based geometry descriptions: `geosparql:hasSerialization` cannot be used, as it is an `owl:DatatypeProperty`. We suggest to implement two properties for connecting geometry descriptions, as implemented in the Ontology for Managing Geometry (OMG, <https://w3id.org/omg>): `omg:hasSimpleGeometryDescription` a `owl:DatatypeProperty`—to link to RDF literals; `omg:hasComplexGeometryDescription` a `owl:ObjectProperty`—to link to the first node of RDF-based geometry descriptions
- Links to external files: Instead of making the file location an individual (node), we argue to add the URL/location as RDF literal with the datatype `xsd:anyURI`. Otherwise, if the location of the geometry file would change, the URI of the RDF node would have to be updated, which is against the COOL URIs best practice.

By widening the scope via the proposed adaptation, the serialisation of the geometry can no longer be defined in the datatype. Hence, a novel approach to identify the serialisation for automated processing by software applications is required. In OMG this issue is solved by creating a taxonomy, the File Ontology for Geometry formats (FOG, <https://w3id.org/fog>), that extends the OMG properties. Each FOG property corresponds with a geometry schema and is specialised further via subproperties to indicate the schema version and—if the geometry schema demands/allows multiple files (e.g. a separate material, texture, ... file) or serialisations—the individual serialisation.

- Remark: defining properties in FOG for every geometry schema, version, and serialisation that exist is impossible. The taxonomy can and should be extended by users (locally or by suggesting extensions to a centrally managed repository)
- Remark: if external files are integrated as proposed above, only one taxonomy has to be created for geometry schemes that are not RDF-based, in disregard of whether they are integrated as snippets (datatype properties) or external files (which would be object properties, if they would be represented by individuals instead of RDF literals)

Also see the following publications for OMG and FOG and their respective use cases:

- Wagner, A., Bonduel, M., Pauwels, P., & Uwe, R. (2019). Relating geometry descriptions to its derivatives on the web. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 304–313). Chania, Greece. <https://doi.org/10.35490/EC3.2019.146>
- Bonduel, M., Wagner, A., Pauwels, P., Vergauwen, M., & Klein, R. (2019). Including widespread geometry formats in semantic graphs using RDF literals. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 341–350). Chania, Greece. <https://doi.org/10.35490/EC3.2019.166>

#### NOTE

This extension proposal has some overlap with issues raised in extensions 8, 9 and 11

### 6.1.19. Extension 19: Terminology needed to express relations between geometry descriptions

#### Standards Tracker Change Request URI

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/51>

#### Category

Semantic improvement / Increased scope

#### Description

If multiple geometry descriptions can be attached to one object, or if objects are related to each other, it would be useful to also define relations between the geometry descriptions directly, that are of other nature than purely topological. We propose to extend the schema to also describe the following relations:

1. grouping of geometry descriptions, e.g. for use cases;
2. describing derivations of multiple geometry descriptions, i.e. to ease maintaining consistent data;
3. transforming geometry descriptions to avoid redundant geometry descriptions;
4. referencing parts of a large geometry description as geometry representation of a smaller object

These relations are currently implemented in the Ontology for Managing Geometry (OMG, <https://w3id.org/omg>). The individual use cases and current implementations are presented below.

1. Grouping: In the AEC industry, geometry can be necessary in multiple use cases, e.g. heating calculation software needs BREP geometry of internal and external space volumes and their connections or an architect wants to communicate the geometry of a specific version of his design to a client. If the according geometry representations (BREP, mesh, CSG, etc.) of all

relevant geometry descriptions could be extracted by a simple query that extracts the relevant group, these processes could be automated more swiftly. In OMG, this is currently implemented via a geometry context. Several `omg:Geometry` or `omg:GeometryState` (version) instances can be linked to an `omg:GeometryContext` instance via the `omg:hasGeometryContext` property.

2. **Deriving:** Derivation of geometry descriptions occur in two cases: either, a geometry is converted from one schema into another, or a geometry is processed for a certain use case. The first case is usually conducted due to software application interoperability, where one application outputs one schema and another application requires a different one. The second case has more diverse reasons, for example, a BREP/CSG/NURBS geometry could be modeled based on a point cloud or a mesh coming from a survey, a 3D BREP model is created from 2D CAD drawings (elevations, plans, sections), or certain parts of the geometry are filtered for simulations, e.g. only outside faces for raytracing simulations. This relation is currently implemented in OMG on both geometry description, as well as geometry version level via the `omg:isDerivedFromGeometry` and `omg:isDerivedFromGeometryState` properties, respectively. These properties can create links between two instances of `omg:Geometry` or `omg:GeometryState` to indicate the derivation or—if version control is used—potential derivations (e.g. a geometry in OBJ serialisation can be derived from geometries in STEP or DWG serialisations)
3. **Transforming:** A building model can contain a manifold of identical objects, such as doors, that share the same geometrical form, but have a different location. The object's geometry is supplied by the manufacturer, but the designer has to decide the location of each door in the building. If the geometry has to be copied for every instance of the object, this can immensely inflate the total size of geometry descriptions and also bears the danger of data inconsistency. For example, if the manufacturer changes a minor detail of the description, each copy needs to be adapted or replaced. Instead, the geometry instances could contain a transformation matrix and a link to the original geometry description, effectively reducing file size and risk for inconsistency. Currently, this relation is implemented in OMG on the `omg:Geometry` node level. The `omg:transformsGeometry` property can link between two instances of `omg:Geometry`, where the instantiated geometry (subject) has no individual geometry description (thus no `omg:hasSimpleGeometryDescription` / `omg:hasComplexGeometryDescription`), but only a transformation definition connected (matrix, vector, etc) while the origin geometry (object) contains the complete geometry description in its own custom coordinate system.
4. **Referencing:** During a modeling phase, it is easier to store the geometry in a single file using the native geometry schema of the modeling application. At the same time, it is relevant to know which parts of the larger geometry description correspond to an individual building object. Hence, subparts of the building model should be connected to individual building elements, if the applied geometry format allows this, e.g. by providing identifiers for parts of the geometry description. In OMG this relation is currently implemented via the `omg:isPartOfGeometry` property between two instances of `omg:Geometry`. The partial geometry (subject) is referenced using one or multiple `omg:hasReferencedGeometryId` properties (with subproperties per kind of identifier per geometry schema in File Ontology for Geometry formats (FOG, <https://w3id.org/fog>)) that can be applied to the main geometry (object) to extract the subgeometry by processing.

Also see the following publications for OMG and FOG:

- Wagner, A., Bonduel, M., Pauwels, P., & Uwe, R. (2019). Relating geometry descriptions to its

derivatives on the web. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 304–313). Chania, Greece. <https://doi.org/10.35490/EC3.2019.146>

- Bonduel, M., Wagner, A., Pauwels, P., Vergauwen, M., & Klein, R. (2019). Including widespread geometry formats in semantic graphs using RDF literals. In Proceedings of the European Conference on Computing in Construction (EC3 2019) (pp. 341–350). Chania, Greece. <https://doi.org/10.35490/EC3.2019.166>

## **6.1.20. Extension 20: Support for parametric geometry descriptions via links between a non-geometric property and geometric properties**

### **Standards Tracker Change Request URI**

This issue has not yet been added to the OGC issue tracker (<http://ogc.standardstracker.org/>).

### **GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/52>

### **Category**

Semantic improvement / Increased scope

### **Description**

During the design phase of a building, a series of geometric descriptions are made based on input parameters (lengths, size, location, orientation). These parameters are commonly also part of the non-geometric description, causing redundant information. In traditional descriptions, such as the Industry Foundation Classes (IFC), this can lead to the non-geometric description contradicting the geometric one, since they are not updated accordingly. For example, a wall can be defined in IFC, where the height of the wall is a non-geometric property, while the extrusion of the geometry is stored separately. The non-geometric property is neither connected to the extrusion nor automatically updated, when the geometry changes, resulting in the potential of inconsistent data. Thus, a link between related properties from geometric and non-geometric descriptions should be established, to ease the detection of contradicting data and the subsequent updating process.

In the Ontology for Managing Geometry (OMG, <https://w3id.org/omg>), a first implementation of such links is realised. For one, the `omg:isExplicitlyDerivedFrom` property can be applied in cases, a geometric property of RDF-based geometry (object) and a non-geometric (subject) property describe exactly the same situation (e.g. the height/extrusion), where a chain axiom can be used to automatically update the according values. On the other hand, the `omg:isImplicitlyDerivedFrom` property can be applied to indicate that a non-geometric property (subject) can be derived from a geometry description (object), as is the case for volumes or surface areas.

- Remark: adding metadata regarding the used coordinate system, units and other metadata as mentioned in extension 1 and 4 can also result in double information. Some geometry schemas such as OBJ are unitless so additional info on the length unit is useful, while others allow to define the used units internally (e.g. STEP) or have a fixed length unit defined in the schema (e.g. glTF uses metre). Adding metadata in RDF for OBJ geometry description enriches the knowledge over the geometry. If this is also added for STEP or glTF geometry, this results in

redundant data, but the metadata is now externalized in RDF can thus be queried directly (no interpretation needed).

## 6.2. Defects reported in GeoSPARQL 1.0

### 6.2.1. Bug 1: Corrections of example data and queries

**Standards Tracker Change Request URI**

[http://ogc.standardstracker.org/show\\_request.cgi?id=173](http://ogc.standardstracker.org/show_request.cgi?id=173)

**GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/4>

**Category**

Documentation bug

**Description**

Errors in example data and queries might lead to wrong implementations. In the given example data in B.1. (page 51) the LineString start with and ends with, although it must be single brackets.

In the third example query in B.2 (page 52) the subject of the fifth triple pattern is a variable called ?my:D although it should be no variable (my:D).

In the fourth example query in B.2 (page 54) the URL given in the prefix definition for geof is wrong. Instead of < 0 > it must be < 1 >

### 6.2.2. Bug 2: GeoSPARQL Schema v1.0.1: hasDefaultGeometry

**Standards Tracker Change Request URI**

[http://ogc.standardstracker.org/show\\_request.cgi?id=548](http://ogc.standardstracker.org/show_request.cgi?id=548)

**GitHub Issue URI**

<https://github.com/opengeospatial/geosemantics-dwg/issues/5>

**Category**

Documentation bug

**Description**

There is a mismatch between the published GeoSPARQL standard v1.0 (11-052r4) and the schema v1.0.1 ( 0 ).

On 8.3.12, page 13, of the standard is defined the property hasDefaultGeometry. The schema defines



the property instead as defaultGeometry but with an otherwise equivalent definition. The schema does not contain a hadDefaultGeometry property.

This mismatch prevents RDFS and OWL inferencing being performed correctly on a dataset written to comply with the standard.

Please can the schema be updated and new version issued.

### 6.2.3. Change 1: Decouple CRS and WKT

#### Standards Tracker Change Request URI

[http://ogc.standardstracker.org/show\\_request.cgi?id=595](http://ogc.standardstracker.org/show_request.cgi?id=595)

#### GitHub Issue URI

<https://github.com/opengeospatial/geosemantics-dwg/issues/10>

#### Category

Coordinate Reference System support

#### Description

WKT seems a good way to easily encode geographic geometry, but the datatype geo:wktLiteral makes it hard to work with GeoSPARQL. In a next version, there should be an option to use only a WKT literal and use a different way to express the CRS of a geometry. Reasons why concatenation of CRS URI and WKT can be considered bad design are:

- GeoSPARQL deviates from the WKT standard, resulting in poor software support.
- Allowing not to specify a CRS and defaulting to CRS84 may be useful in North America, but is of little value for serious usage in other parts of the world.
- The proper data type for expression of CRS is an IRI. Therefore it should be defined as such, not as part of a string literal.

Especially when non-geographical geometry is considered, CRS is not necessarily a known property. Therefore it should be possible to leave out CRS data in publications, without this resulting in wrong interpretations.

CRS can be considered an intrinsic or fundamental aspect of geometry, but so are other properties like dimensionality or accuracy. This does not mean all of this information should be lumped together in one literal.

It seems better to introduce a new property for CRS and to let WKT literals be just WKT literals. Should a new property for indicating CRS be introduced, it would be good to allow it to be applied not only to individual geometries, but to geometry collections too.

# Chapter 7. Other Related Activities

This section provides an overview of related activities identified by the authors of this white paper.

## 7.1. W3C Linked Building Data Community Group

### 7.1.1. Scope of Activity

This group brings together experts in the area of building information modelling (BIM) and Web of Data technologies to define existing and future use cases and requirements for linked data based applications across the life cycle of buildings. A list of recommended use cases will be produced by this community group. The envisioned target beneficiaries of this group are both industrial and governmental organisations who use data from building information modelling applications and other data related to the building life cycle (sensor data, GIS data, material data, geographical data, and so forth) to achieve their business processes and whom will benefit from greater integration of data and interoperability between their data sets and the wider linked data communities. [15]

### 7.1.2. Web Site(s)

- <https://www.w3.org/community/lbd/>

### 7.1.3. Source Repositories

- <https://github.com/w3c-lbd-cg/lbd>

### 7.1.4. Liaisons

- TBD

## 7.2. ISO/IEC JTC 1/SC 32/WG 3

### 7.2.1. Scope of Activity

SC 32 provides enabling technologies to promote harmonization of data management facilities across sector-specific areas. Specifically, SC 32 standards include:

reference models and frameworks for the coordination of existing and emerging standards; definition of data domains, data types, and data structures, and their associated semantics; languages, services, and protocols for persistent storage, concurrent access, concurrent update, and interchange of data; methods, languages, services, and protocols to structure, organize, and register metadata and other information resources associated with sharing and interoperability, including electronic commerce. [16]

### 7.2.2. Web Site(s)

- TBD



### **7.2.3. Source Repositories**

- TBD

### **7.2.4. Liaisons**

- TBD

# Annex A: Revision History

Date	Release	Editor	Primary clauses modified	Description
2019-10-28	0.1	J. Abhayaratna	all	initial version
2019-11-15	0.2	R. Atkinson	1	Editing
2019-11-28	0.3	T. Homburg	2,4	Editing and adding additional use cases and change requests
2019-11-28	0.4	F. Thierry	2	Additional Archeology use case
2019-12-09	0.5	M. Bonduel	4	Editing and adding additional change requests
2020-01-15	0.6	F. Knibbe	all	Editing
2020-02-29	0.7	J. Abhayaratna	2	Editing, cleaning up use cases
2020-02-29	0.8	K. McGlinn, M. Bonduel, A. Wagner, M. H. Rasmussen, J. Abhayaratna	2	Adding additional use cases from the Linked Building Data Community Group
2020-03-01	0.9	J. Abhayaratna	5	Adding Other Related Activities section

# Annex B: Bibliography

*Example Bibliography (Delete this note).*

The TC has approved Springer LNCS as the official document citation type.

Springer LNCS is widely used in technical and computer science journals and other publications

## NOTE

- For citations in the text please use square brackets and consecutive numbers: [1], [2], [3]

– Actual References:

[n] Journal: Author Surname, A.: Title. Publication Title. Volume number, Issue number, Pages Used (Year Published)

[n] Web: Author Surname, A.: Title, <http://Website-Url>

[1] Web: Wikipedia: Semantic technology, [https://en.wikipedia.org/wiki/Semantic\\_technology](https://en.wikipedia.org/wiki/Semantic_technology)

[2] Web: Wikipedia: Graph database, [https://en.wikipedia.org/wiki/Graph\\_database](https://en.wikipedia.org/wiki/Graph_database)

[3] Web: Data Intelligence: Virtual Knowledge Graphs: An Overview of Systems and Use Cases, <http://www.data-intelligence-journal.org/p/24/#Ch-S2>

[4] Web: Youtube: Introducing the Knowledge Graph, <https://www.youtube.com/watch?v=mmQl6VGvX-c>

[5] Web: OGC: OGC GeoSPARQL - A Geographic Query Language for RDF Data, <http://www.opengis.net/doc/IS/geosparql/1.0>

[[PROV DM]] [6] Web: W3C: PROV-DM The PROV Data Model, <https://www.w3.org/TR/prov-dm/>

[7] Web: Australian Government: <https://www.bioregionalassessments.gov.au>

[8] Web: Constitute Project: <https://www.constituteproject.org>

[9] Web: Bonduel M, Vergauwen M, Pauwels P, Klein R : <https://www.researchgate.net/project/Flemish-Cities-in-Transition-A-Framework-for-a-Web-based-Built-Heritage-Platform-Using-BIM-GIS-and-Linked-Data>

[10] Wagner A, Bonduel M, Pauwels P, Rüppel U: Relating geometry descriptions to its derivatives on the web. In Proceedings of the 2019 European Conference for Computing in Construction (O'Donnell J, Chassiakos A, Rovas D and Hall D, eds), University College Dublin, Chania, Greece, Pages 304–313 (2019) <https://doi.org/10.35490/EC3.2019.146>

[11] Bonduel M, Wagner A, Pauwels P, Vergauwen M, Klein R: Including widespread geometry formats in semantic graphs using RDF literals. In Proceedings of the 2019 European Conference for

Computing in Construction (O'Donnell J, Chassiakos A, Rovas D and Hall D, eds), University College Dublin, Chania, Greece, Pages 341–350 (2019), <https://doi.org/10.35490/EC3.2019.166>

[12] McGlinn K, Debruyne C, McNerney L, O'Sullivan D: Integrating Ireland's Geospatial Information to Provide Authoritative Building Information Models. In SEMANTiCS 2017, the 13th International Conference on Semantic Systems. Held in Amsterdam, Netherlands (2019)

[13] Rasmussen M H, Lefrançois M, Pauwels P, Hviid C A, Karlshøj J: Managing interrelated project information in AEC Knowledge Graphs. *Automation in Construction*, 108, 102956 (2019)

[14] Web: Building Topology Ontology: <http://www.semantic-web-journal.net/content/bot-building-topology-ontology-w3c-linked-building-data-group-0>

[15] Web: Linked Building Data Community Group: <https://www.w3.org/community/lbd/>

[16] Web: Data Management and Interchange: <https://www.iso.org/committee/45342.html>