

## EXTENSION AND CONTEXTUALISATION FOR LINKED SEMANTIC 3D GEODATA

Claudine Métral<sup>1</sup>, Gilles Falquet<sup>1</sup>

<sup>1</sup> Centre Universitaire d'informatique, University of Geneva,  
7 route de Drize, 1227 Carouge, Switzerland  
(claudine.metr@unige.ch, gilles.falquet@unige.ch)

**KEY WORDS:** 3D linked geodata, nD linked geodata, CityGML, RDF, SPARQL, OWL

### ABSTRACT:

The application of the linked data principles provide a simple and effective way to publish and share data on the Web. The publication of geodata on the Web has been studied and developed for a long time, leading to the creation of geo vocabularies and ontologies as well as geographic query languages. This is also true for 3D geodata, for which multiple RDF vocabularies and OWL ontologies exist (at the building, city, or territory level). In this paper we show that at least two issues remain to be addressed to reach a satisfactory publication framework for 3D linked geodata. First, the semantic model interconnection issue is not magically solved by the creation of ontologies, it requires the careful design of interconnection and extension schemes and their management. Second, the representation of additional dimensions, such as time, level of detail, confidence level, provenance, etc. cannot be directly handled in RDF. For these two issues we propose methodologies and design patterns and show how they can actually be used to create a semantically rich representation of 3D or nD linked geodata.

### 1. INTRODUCTION

Linked data is a set of principles defined to identify and publish data on the web. The main expected, and generally reached, benefits of this approach include:

- Sharing data without copying and manipulating (large) datasets, the desired information being selected by querying endpoints with the (Geo)SPARQL language;
- (Ideally) referring to the same entity with a worldwide identifier (URI).
- Using a formally defined logical model that supports logical inference;
- Obtaining a semi-structured graph model that does not depend on fixed predefined schemas;
- Augmenting each dataset by connecting it to other datasets that can provide additional information in specific domains, levels of details, etc.

In this paper we study the application of the linked data approach to 3D geographic data (geodata) with the aim to propose solutions to issues that are still hindering the publication of 3D geodata as linked data. The rest of this article is organized as follows: In the following section (2) we briefly present the linked data principles and their application to geodata; Section 3 presents 3D linked geodata and the issues that must still be addressed; Section 4 discusses the model interconnection and extension issue and proposes an approach for the management of 3D model extensions; Section 5 discusses the introduction of other dimensions (times, level of detail, provenance, ...); Section 6 provides conclusions and perspectives,

### 2. LINKED DATA PRINCIPLES

#### 2.1 Definitions

The term Linked Data is associated with the more general concept of Semantic Web and refers to a set of best practices for publishing and interlinking structured data on the Web using World Wide Web Consortium (W3C) standards (Wood et al. 2014, W3C 2015).

In a Web-published note, Tim Berners-Lee (2006) proposed the following four principles for producing Linked Data:

- “1. Use URIs (Uniform Resource Identifier) as names for things;
2. Use HTTP URIs so that people can look up those names;
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL);
4. Include links to other URIs so that they can discover more things.”

RDF is a standard graph-based data model for data interchange on the Web. An RDF graph is constituted by a set of subject-predicate-object triples, establishing an RDF statement (W3C 2014). The subject and the predicate in a triple are usually URIs, while the object can be either an URI or a literal value. RDF statements can also include blank nodes, representing resources or objects for which a URI or literal is not given. An RDF graph can be written using various syntaxes such as RDF/XML, or such as N3 (or Turtle) syntax more readable by humans.

Giving a meaning to RDF triples by accurately defining predicates and objects requires the use of vocabularies and ontologies. An ontology offers a formal description of a given knowledge domain including the definition of its classes, types,

properties and hierarchy. While the RDF Schema recommendation provides a data-modelling vocabulary, the Web Ontology Language (OWL), an extension of RDF Schema, offers a formal syntax for writing Web ontologies (Heflin 2007, W3C 2012).

While relatively small RDF datasets (i.e. several hundred RDF triples) can be simply served as static files, more important datasets should be stored in RDF stores, also known as triplestores. Triplestores can then be queried using the SPARQL query language through SPARQL endpoints (W3C 2013).

## 2.2 Linked geodata

Some large spatial datasets such as LinkedGeoData (<http://linkedgeo.org>), an RDF transcription of Open Street Map, are available with a free access.

Some National Mapping Agencies (NMA) publish linked geodata. For examples, the swiss federal agency produces linked geodata about the administrative units and national boundaries of Switzerland, and about public transport stops and other locations related to public transportation (<https://www.geo.admin.ch/en/geo-services/geo-services/linkeddata.html>).

There are also geodata available in dbpedia (<https://wiki.dbpedia.org/>), that makes available on the World Wide Web the data created in the Wikipedia project.

Various platforms and applications for producing and consuming geospatial Linked Data have also been developed recently (Garbis et al. 2013).

The rise of geospatial Linked Data tools has been mainly relying on the recent development of GeoSPARQL, an OGC standard (OGC 2012). GeoSPARQL is a spatial extension of the SPARQL query language including an RDF/OWL vocabulary which allows representing and querying geospatial data.

## 3. 3D LINKED GEODATA

### 3.1 Current state

Since the linked data approach relies on OWL ontologies or RDF Schemas (RDFS), we first listed the main 3D geodata already expressed as OWL ontologies or RDFS.

At the geometry and topology level:

- GML, the Geography Markup Language, exists in OWL (<https://bioportal.bioontology.org/ontologies/GML/>);
- the translation from GML/XML is not direct (because of srsDimension, etc.);
- geoSPARQL exists as an ontology (<http://www.opengis.net/ont/geosparql>);
- there exists an OWL ontology of spatial relations (Bucher et al, 2012).

At the building level, IFC/BIM exists in OWL with the ifcOWL ontology (<http://openbimstandards.org/standards/ifcowl/>).

At the city level, CityGML also exists in OWL (see for instance <http://cui.unige.ch/isi/onto/citygml2.0.owl>) and in JSON-LD with CityJSON ([www.cityjson.org](http://www.cityjson.org)).

Nevertheless, we identified two important issues that must be addressed if we want to use the linked data idea at its full potential for 3D geodata: How to interconnect 3D geographic models? How to represent supplementary dimensions (time, level of detail, etc.)?

### 3.2 Interconnection and extension requirements of 3D linked geodata

Extension mechanisms, such as the CityGML Application Domain Extensions (ADE), have showed their utility in various domains such as, for example, the energy ADE that relates to urban energy simulation (Agugiaro et al, 2018). The general objective of such an extension process is providing support to develop applications that deal with different aspects (subdomains) of 3D geodata, e.g. adding underground structures to city objects, or to geodata associated to non-geo data, e.g. buildings + noise, or buildings + transportation network.

In the linked data approach, we want a mechanism to either interconnect existing models or extend a model with new concepts. Moreover, we should have a global view of these model extensions and interconnections.

### 3.3 From 3D to nD (contextualization) requirements

A current trend is defining and using 4D GIS, where the tridimensional data are supplemented with time change data.

In CityGML, for example, time exists in a basic form :

- *creationDate* and *terminationDate* parameters for the *\_CityObject* class,
- *yearOfConstruction* and *yearOfDemolition* parameters for *\_AbstractBuilding*, *\_AbstractTunnel* or *\_AbstractBridge* classes.

Spatio-temporal data can also be supplemented with a fifth dimension associated to scale.

In CityGML, for example, scale is defined as Level Of Detail (LOD), which is an integer ranging from 0 to 4. Increasing LODs are associated to more detailed models. The geometry of the city objects differs with the LOD. Each different geometry is associated to a city object using a different property.

In the linked data approach we have to take in account that the Web is not a centrally controlled or curated environment. What is published on the Web may have various characteristics, such as:

- validity time, since the published data can evolve over time;
- level of detail or accuracy;
- confidence, generally depending on the provenance (how to evaluate quality from the provenance) and ranging from data with no accuracy mechanism to highly curated data.

Therefore it must be possible to add contextual information to any piece of information, leading to data with several dimensions.

## 4. INTERCONNECTION AND EXTENSION

The goal of a model interconnection or extension is to provide a unified view of two or more data models that represent different aspects or domains of the real world. Although they do not provide a magic solution to the difficult model-mapping

problem, the Semantic Web techniques can simplify the expression of model connections and model extensions.

#### 4.1 Model interconnection

We suppose that two different models M1 and M2 have been defined and represented as OWL or RDFS ontologies named O1 and O2.

In the linked data approach an interconnection means defining an interconnection ontology that contains:

- simple (direct) connection statements using *SubClassOf*, *EquivalentClass*, *sameAs*, or *DisjointClass* axioms;
- complex connection statements using logical expressions (axioms).

Thus the interconnection can be expressed without modifying the original models. However, as we will see in the following subsections, the interconnection ontology must be carefully designed to avoid some semantic problems.

#### 4.2 Model extension

Extending a model consists in adding new classes and attributes (properties) that are intended to represent a part of the reality or system that is not covered by the initial model. In the linked data approach a model is an OWL ontology or an RDF schema. Therefore extending a model amounts to:

- define additional classes and properties;
- define additional axioms to specify these classes and properties and their relationships with the initial model.

To enforce a modular approach, these new definitions can be grouped in a new ontology (that refers to the initial one).

For instance, to define a SolarPotential extension of the CityGML model one could define the following class and property:

```
sol:SolarPotential a owl:Class .
sol:ofBuilding a owl:ObjectProperty .
```

and the axiom (in Manchester syntax):

```
sol:SolarPotential
  rdfs:subClassOf
    (sol:ofBuilding exactly 1 gml:Building) .
```

which means that a solar potential relates to one building.

The computation of spatial relations such as externally connected (EC) from RCC8 should normally be based on the geometry of the considered city objects. However this may prove impracticable or imprecise, for several reasons such as :

- determining if two surface are tangent or if they are connected is numerically difficult ;
- the geometries may be known up to a given precision, leading to erroneous results ;
- the geometrical knowledge may be incomplete.

However, one may know that some relations exist, e.g. a road segment is connected to the next one. Moreover, one may be interested in spatial relations that do not directly correspond to the relations between the geometries of the objects. For

instance, one may consider that a X is in (proper part) a Y even though some part of X is outside of Y, provided this relation satisfies the same axioms as the usual PP (proper part) relation.

Therefore it can be interesting or necessary, for some applications, to explicitly state spatial relations between city objects even though they cannot be computed from their geometries.

#### 4.3 Extension management

In fact, this extension mechanism is extremely simple and expressive. However, its sheer power may lead to undesired effects on the initial model. For instance, suppose that the initial ontology O1 contains two classes A and B and that the extension ontology O2 defines a new class C and the axioms:

```
O2:C rdfs:subClassOf O1:B .
O1:A rdfs:subClassOf O2:C .
```

Since *rdfs:subClassOf* is transitive, a logical consequence of these axioms is:

```
O1:A rdfs:subClassOf O1:B .
```

This amounts to add a new axiom to O1 and therefore to modify O1.

**4.3.1 Conservative extension requirement:** As seen in the previous subsection, when O1 is used together with O2 it is different than the initial O1 ontology. Such an extension is called non-conservative. Unfortunately there is no automated technique to prove that an extension is conservative. One way to ensure that the extension has no effect on the initial ontology is to restrict the kind of axioms that can be employed in an extension ontology. This is exactly what the CityGML extension mechanism does. It corresponds to admit extension axioms of the form:

```
O2:Class rdfs:subClassOf O1:Class .
O2:Class rdfs:subClassOf <propertyRestriction> .
```

**4.3.2 Multiple extension consistency:** The management of extensions also includes validating the consistency of multiple extensions. If a model M (represented by an ontology O) possesses two extensions M1 and M2 (represented by O1 and O2), and an application must use both extensions, then one must check that the ontology  $O \cup O1 \cup O2 \cup O12$  is consistent (where O12 is an interconnection ontology for O1 and O2).

## 5. FROM 3D TO ND

### 5.1 A nD pattern

One way to fulfil the requirements expressed in 3.3 consists of adding annotations to the RDF triples representing facts that have a time validity, a level of detail, etc. Several patterns have been proposed to add annotations on RDF triples. Among them, an extension of the n-ary relation patterns seems to be the most appropriate for adding contextualisation dimensions to triples (Aljalbout & Falquet, 2017).

Figure 1 shows how to use this pattern to add a validity time, level of detail, and confidence dimension to the property that

defines the geometry of a building. In the running example, a building *b* has a geometry *g* with a level of detail 2 from 1977 to 2000, with a confidence value of 0.84.

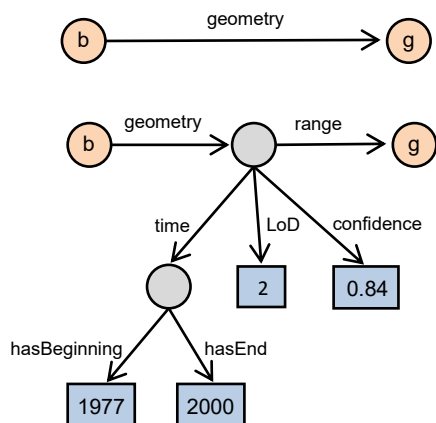


Figure 1. Proposed nD RDF pattern

This pattern can be seen as a generalisation of the time extension defined in ISO 19108:2002 Geographic Information - Temporal Schema, which has been partly implemented in OWL (<http://def.seegrid.csiro.au/isotc211/iso19108/2002/temporal>).

Relevant ontologies are used whenever publicly available. This concerns the ontologies related to 3D as quoted in section 3.1 as well as, for example, geospatial ontologies defined by the W3C consortium (W3C, 2017b) or basic ontologies, such as the time ontology, also defined by the W3C consortium (W3C, 2017a), which provides three basic categories of time: instants, intervals and durations. Allen's relations can be used, in particular with the *ProperInterval* class defined in the W3C time ontology.

In RDF/N3, the example of Figure 1 can be expressed (simplified) as:

```
:b :geometry [:range :g ;
              :time [ time:hasBeginning "1977" ;
                     time:hasEnd "2000" ] ;
              :LoD "2" ;
              :confidence "0.84"
            ] .
```

where the prefix *time:* refers to the W3C time ontology.

Any time representation can be used for the temporal dimension. It is thus possible to represent more complex notions of time, such as archeological time. Katsiani et al (2008) identified different temporal categories, such as:

- absolute time, which describes an absolute chronology, such as 3600 BC +/- 140 years ;
- stratigraphic time, which describes a relative temporal position between deposits. For example, layer L2 > layer L1 if layer L2 is later than layer L1.

Absolute time is usually obtained from scientific analysis, such as radio carbon, and is characterized by uncertainty (De Roo et al, 2013). Absolute time can be represented by a data range, (with minimal value 3740 BC and maximal value 3460 BC in the running example) and a most probable value (3600 BC in the example). The resulting pattern is shown in Figure 2.

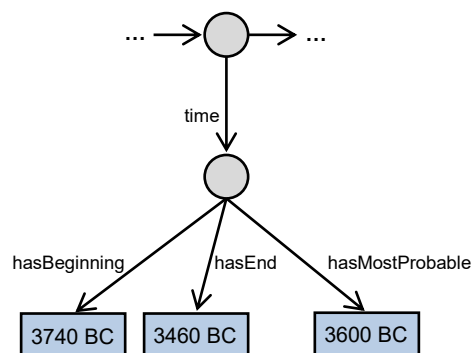


Figure 2. Pattern for absolute time

Stratigraphic time is in fact a topological temporal representation (De Roo et al, 2013). In the running example creation of deposit D2 (associated to layer L2) is later than creation of deposit D1 (associated to layer L1) and D2 is above D1. The resulting pattern is shown in Figure 3.

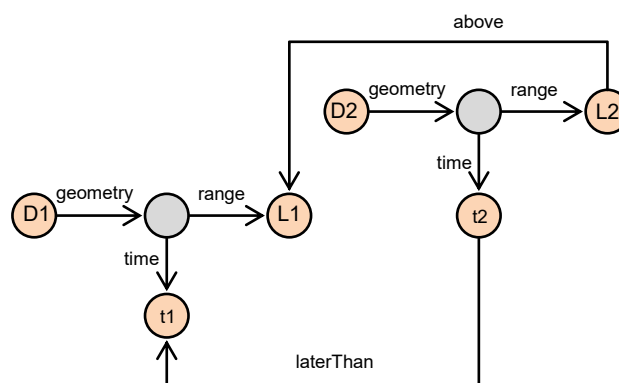


Figure 3. Pattern for stratigraphic time

Any kind of resource can be represented and, as illustrated by the previous example (Figure 3), transversal relations between resources can also be represented.

Recent works (Aljalbout & Falquet, 2017) have shown that the proposed pattern enjoys interesting properties. In particular it does not introduce a large number of new objects and one can still use the usual Semantic Web reasoning techniques and tools (with some adaptation) to perform multi-dimensional reasoning.

## 5.2 nD reasoning and operations

**5.2.1 Snapshot:** This is the basic dimension reduction operation. Given a value for a dimension domain it retains only those facts that are true in this "context". One can then work independently of the projected dimension(s).

Here is an example where one wants to :

- project on the value 1977 for the validity time dimension ;
- project on IGN for the provenance dimension ;
- project on LoD2 for the level of detail dimension.

Defining selections (using SPARQL queries) gives the right results even if n-dimensionality complexifies the queries. But it can be interesting to extract from the whole graph the part that is relevant for our purpose. Indeed this subgraph is easier to

query. The following SPARQL code shows how such a subgraph can be created by projection on the time dimension :

```
construct ?s ?p ?o.
where {?s ?p _:b. _:b :range ?o. _:b :time ?t}
filter (?t = 1977)

construct ?s ?p [ :range ?o; :lod ?l; :prov ?pr]
where {?s ?p _:b.
      _:b :range ?o.
      _:b :time ?t.
      _:b :lod ?l.
      _:b :prov ?pr}
filter (?t = 1977)
```

**5.2.2 Context-aware queries:** As seen before, adding dimensions complexifies SPARQL queries. The example below illustrates that some elements not directly appearing in the query have to be taken in account. For example, querying for two buildings that touch implies that these buildings must have a common period of existence.

The example below intends to obtain as results the buildings x whose function is school and that touch the buildings y whose function is factory :

```
function(x, school) and touches(x, y) and
function(y, factory) -> result(x)
```

Figure 4 illustrates the initial query (upper part) and the effective query taking in account time (lower part).

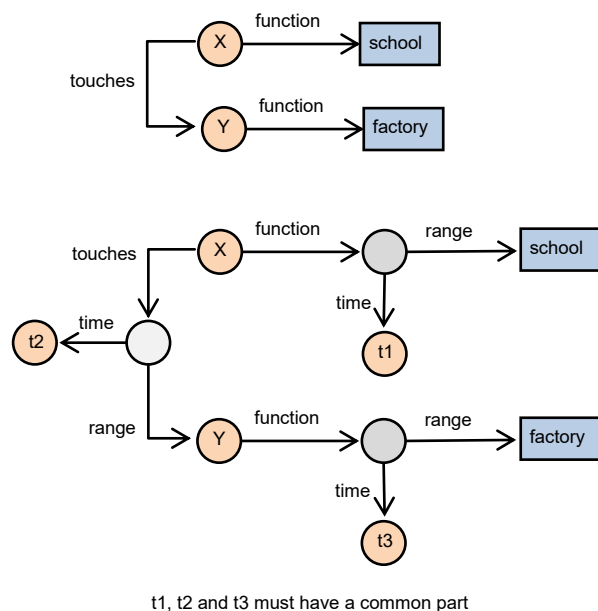


Figure 4. Context-aware pattern

The SPARQL query is the following :

```
select ?x
where
{?x gml:function [:range "school"; :time ?t1] .
 ?x srel:touches [:range ?y; :time ?t2] .
 ?y gml:function [:range "factory"; :time ?t3]
}
filter {overlaps(?t1, ?t2, ?t3)}
```

## 6. CONCLUSION

In this paper we have shown that the standard Semantic Web techniques are usable, and actually used, to create a Semantic Web of linked 3d geodata. Nevertheless, the direct application of the linked data principles, in particular the translation of geodata in RDF and OWL, is not sufficient in itself to solve two important issues : the management of 3D model extensions and the creation of 4D or nD models. For the management of model extensions we have proposed a modular solution based on ontology interconnection. This solution also provides a technique for validating the global consistency of multiple extensions. For the nD models we have proposed to use an RDF pattern that can support any number of dimensions and any type of dimension (linear validity time, archaeological time, provenance, level of detail, etc.). In addition we have shown how to implement nD reasoning on this pattern with the standard (geo)SPARQL language. We are currently developing tools to automate the proposed approaches and make them readily usable.

## REFERENCES

- Aguiaro, G., Benner, J., Cipriano, P., Nouvel R., 2018. The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations, In: *Open Geospatial Data, Software and Standards 3:2*
- Aljalbout, S., Falquet, G., 2017. Un modèle pour la représentation des connaissances temporelles dans les documents historiques : Applications sur les manuscrits de F.Saussure. In *Proc. 28es Journées francophones d'Ingénierie des Connaissances (IC 2017)*, Caen, July 2017.
- Berners-Lee T., 2006. Linked Data - Design Issues, <http://www.w3.org/DesignIssues/LinkedData.html> (24 May 2018)
- Bucher B., Falquet G., Clementini E., Sester M., 2012. Towards a typology of spatial relations and properties for urban applications, In: Leduc T., Moreau G., Billen R. (eds) *Usage, Usability, and Utility of 3D City Models*, 02010
- De Roo, B., Van de Weghe, N., Bourgeois, J., De Maeyer, P. (2013) The Temporal Dimension in a 4D Archaeological Data Model: Applicability of the Geoinformation Standard. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume II-2/W1, ISPRS 8th 3DGeoInfo Conference & WG II/2 Workshop, 27 – 29 November 2013, Istanbul, Turkey
- Garbis, G., Kyzirakos, K., & Koubarakis, M., 2013. Geographica: A benchmark for geospatial rdf stores (long version). In: *The Semantic Web–ISWC 2013*, Springer Berlin Heidelberg, pp. 343-359.
- Heflin, J., 2007. An Introduction to the OWL Web Ontology Language. Lehigh University. National Science Foundation (NSF), <http://www.cse.lehigh.edu/~heflin/IntroToOWL.pdf> (24 May 2018)
- Katsianis, M., Tsipidis, S., Kotsakis, K., Koussoulakou, A., 2008. A 3D digital workflow for archaeological intra-site

research using GIS. In : *Journal of Archaeological Science*,  
35(3), pp. 655–667.

OGC, 2012. OGC GeoSPARQL - A Geographic Query  
Language for RDF Data, OGC 11-052r4,  
<http://www.opengis.net/doc/IS/geosparql/1.0> (24 May 2018)

W3C, 2012. OWL 2 Web Ontology Language: Mapping to  
RDF Graphs (Second Edition), <https://www.w3.org/TR/owl2-mapping-to-rdf/> (24 May 2018)

W3C, 2013. SPARQL 1.1 Query Language,  
<https://www.w3.org/TR/sparql11-query/> (24 May 2018)

W3C. 2014. Resource Description Framework (RDF): Concepts  
and Abstract Syntax, <https://www.w3.org/TR/rdf-concepts/> (24  
May 2018)

W3C, 2015. Linked Data. <https://www.w3.org/standards/semanticweb/data> (24 May 2018)

W3C, 2017a. Time Ontology in OWL, W3C Recommendations,  
19 october 2017, <https://www.w3.org/TR/owl-time/> (5 July  
2018)

W3C, 2017b. W3C Geospatial Ontologies, W3C Incubator  
Group Report, 23 October 2017, <https://www.w3.org/2005/Incubator/geo/XGR-geo-ont/> (5 July 2018)

Wood D., Zaidman M., Ruth L. & Hausenblas M., 2014.  
Linked Data - Structured data on the Web. Manning, New York

*Revised July 2018*