KNOWLEDGE GRAPH CONSTRUCTION FOR SUBSURFACE OBJECTS INCLUDING UNCERTAINTY AND TIME VARIATION

A. Caselli, G. Falquet, C. Métal
Centre Universitaire d’Informatique (CUI), University of Geneva, Switzerland
{ashley.caselli, gilles.falquet, claudine.metral} @unige.ch

KEY WORDS: Subsurface objects, Knowledge graphs, Uncertain geometry, Temporal evolution, Collision detection

ABSTRACT:
In the recent years the concept of knowledge graph has emerged as a way to aggregate information from various sources without imposing too strict data modelling constraints. Several graph models have been proposed during the years, ranging from the “standard” RDF to more expressive ones, such as Neo4J and RDF-star. The adoption of knowledge graph has become established in several domains. It is for instance the case of the 3D geoinformation domain, where the adoption of semantic web technologies has led to several works in data integration and publishing. However, yet there is not a well-defined model or technique to represent 3D geoinformation including uncertainty and time variation in knowledge graphs. In this paper we propose a model to represent parameterized geometries of subsurface objects. The vocabulary of the model has been defined as an OWL ontology and it extends existing ontologies by adding classes and properties to represent the uncertainty and the spatio-temporal behaviour of a geometry, as well as additional attributes, such as the data provenance. The model has been validated on significant use cases showing different types of uncertainties on 3D subsurface objects. A possible implementation is also presented, using RDF-star for the data representation.

1. INTRODUCTION

1.1 Background

Cities contain many diverse subsurface objects, either manmade or natural, such as utility networks, building basements, tunnels, geothermal drills, tree roots, groundwater, or archaeological remains. Knowing the potentially occupied underground space is mandatory to effectively plan new underground settlements.

Knowing as precisely as possible the shape of the underground objects, namely their geometry and their positioning, is essential to know which spaces are already occupied by the existing objects. However, geometry and positioning have not always to be considered valid during the object lifetime. For instance, they might be considered valid only within a certain context (i.e., for a given period of time). Indeed, tree roots generally grow during the years, etc.

A major problem with underground objects is that actual geometries may differ from those expected: for example, geothermal elements may not be exactly vertical, pipes may not be exactly aligned with adjacent ones, or their actual depth may differ from the one defined in standards or plans. In addition, some objects may move over the time. Due to this reason, the geometry (including the positioning) associated with underground objects must represent this uncertainty along with its variation over the time.

Furthermore, it may occur that the geometry of the object might not be represented as one single geometry. While one geometry is generally sufficient for manmade objects, it may be useful to associate several geometries to objects of other kind, such as the natural objects. For instance, a tree root can have one geometry representing its extent and a second geometry representing a critical root zone. Such a zone defines the minimum area that must be left intact in order to maintain the tree healthy and increase its chances of survival (Guerrero Iñiguez, 2017).

Another need relates to the data source integration where metadata, such as provenance (and trust), must be represented. In the recent years the concept of knowledge graph has emerged as a way to aggregate information from various sources without imposing too strict data modelling constraints. Several graph models have been proposed for knowledge graphs, such as RDF (the “standard” semantic web model), Neo4J (a similar attributed graph model), and more recently RDF-star (a kind of meta-RDF in which the subject or the object of a statement can be a statement) (Hartig, 2017). Large knowledge graphs have been developed by private companies, such as Google and Facebook, to enhance their products with new functionalities. Other knowledge graphs, such as Wikidata, are collaboratively developed by volunteers.

In the (3D) geoinformation domain, semantic web technologies have led to several works in data integration, data publishing, and geo-service provision. One can also find a large corpus of geoinformation in Wikidata and other knowledge graphs, however, there is currently no well-defined model or technique to represent 3D geoinformation in knowledge graphs. Our goal is to propose such a model for the specific, but representative case, of 3D subsurface data.

1.2 Problem statement

The question addressed in this paper is how to build knowledge graphs that:

- represent subsurface objects and their geometry (including their positioning)
- support different ways of representing geometric uncertainty
can represent changes in geometry and position over the
time (validity time for geometries)
- comprise metadata such as provenance, quality, reliability

In addition, the implementation of the knowledge graph must
be compatible with the semantic web technologies such as RDF
or RDF-star in order to be exploitable by standard 3D geodata
tools (e.g., ArcGIS). The information stored in the knowledge
graph must also be suitable to perform typical tasks such as
collision detection or rule compliance checking.

2. RELATED WORK

2.1 Ontologies and knowledge graphs for subsurface 3D
geodata

There is no explicit notion of schema in knowledge graphs.
Nevertheless, knowledge graph models offer a way to define
classes, instances, and properties (binary relations). Therefore,
an ontology language, such as OWL, may be used to define
a conceptual model that may be implemented in a knowl-
dge graph. In such knowledge graphs data are represented as
RDF triples (subject, predicate, object) that can be accessed by
SPARQL queries.

Métral et al. (2020) defined a set of interconnected ontologies
for representing subsurface objects with their semantic prop-
eties and their geometries. As quoted by the authors, the
ontologies provide an integration schema to collect geospatial
data from different sources and merge them into a single RDF
graph. The defined ontologies also provide a way to represent
the vocabulary with which completion or compliance rules can
be expressed.

However, such ontologies, usually represented in OWL lan-
guage do not allow to qualify properties. Representing that a
triple is only valid under specific conditions, such as within a
time interval, or representing the provenance of the data means
adding contextualisation dimensions to it. This can be per-
formed using different techniques. Métral and Falquet (2018)
propose to add intermediate nodes to the initial RDF triples to
represent the link annotations. The RDF and SPARQL stand-
ards have recently been extended, giving rise to RDF-star and
SPARQL-star that provide a more convenient way to annotate
RDF statements and query those annotations (Hartig, 2017).

2.2 Provenance/Quality

When integrating data from different sources it may be impor-
tant to keep the information of where data is sourced from and
how it has been transformed. We refer to such information with
the term “data provenance”.

A model with which the provenance information can be repres-
ented is the PROV Data Model. Such model, written in the
OWL2 language, is defined by the PROV Ontology (PROV-
O) (Sahoo et al., 2013). It provides a set of elements that can
be used to represent and interchange the data provenance in-
formation. It may also be extended to model the provenance
information for different applications and domains.

Additionally, the provenance supports the data quality. Indeed,
by enriching it with supplementary information, for instance,
it is possible to assess the quality of the data. To this end it
would be required to assign a confidence value to each data
source. A similar approach is presented by olde Scholtenhuis
et al. (2018) where they assign different quality levels (weights)
to a utility line when combining multiple existing geometric
representations fetched from multiple datasets (see Section 2.4).

2.3 Temporal aspects

Temporal changes are common in an urban environment for dif-
f erent aspects. For example, a building may change over the
time in its form (geometry), its function, or proprietary (se-
mantic). Furthermore, changes involving buildings may also be
sudden (e.g., its demolition) or gradual (e.g., its degradation).

Morel and Gesquière (2014) propose a CityGML extension to
take into account the temporal aspects and the state modifica-
tions of city objects for changes related to semantic, geometry,
topology, or appearance. This approach is based on tags and
flags. A tag is a temporal information (e.g., a date) linked to a
city object. A static flag is defined for a given tag and used to
describe the behaviour of an object (e.g., its demolition).
A dynamic flag contains a data source associated with a given
CityGML attribute, this source being a set of possible values
with a date for each one (e.g., an object color that changes in
accordance with information from a sensor). As quoted by the
authors the addition of tags and flags helps manage the life-
cycle of the city objects.

Temporal aspects have also been taken into account in RDF
graphs, not only for representing data but also for querying
them. Tappolet and Bernstein (2009) proposed a syntax and
storage format based on named graphs to express temporal RDF
by allowing the annotation of RDF triples with temporal valid-
ity intervals. Although they defined τ-SPARQL (an extension
of SPARQL) to query this temporal RDF, they argue that this
extension can be directly mapped to standard SPARQL.

2.4 Geometry uncertainty

Joskowicz et al. (2010) have defined the linear parametric geo-
metric uncertainty model (LPGUM) for modelling part shape
and position uncertainty of parametrically defined geometric
entities.

A CityGML ADE extension have been proposed by olde Schol-
tenhuis et al. (2018) to represent geographical uncertainties
of 3D utility data. With their model one can capture multiple
potential coordinates to calculate and visualise the utility’s un-
certainty. Indeed, they extend the CityGML ADE extension by
adding a FuzzyUtility class and several subclasses. The model
stores four different potential coordinates (standard, estimated,
surveyed, and unknown) and attributes a quality value to each of
them (weight). Thus, the centerline of the uncertain geometry
is computed as a weighted arithmetic mean.

3. PROPOSED MODEL

The proposed model the extends the model of Métral et al.
(2020) by adding new object classes and relations to repres-
ent uncertain and time-dependent geometries, and to associate
a provenance and a temporal validity to geometric values.
3.1 Geometric uncertainty and Values

Since actual geometries may differ from expected ones, the geometry uncertainty (which may vary over the time) must be taken into account. Probability density functions (PDFs) are used for representing the probability that a subsurface object is present at a certain location. To this aim the Value class has been extended with two subclasses: MeasuredValue and EstimatedValue (see Figure 1). It is thus possible to represent either known values (MeasuredValue), with their accuracy and precision, or uncertain values (EstimatedValue) defined by a probability distribution. The distribution is represented by a density function type (Triangular, PERT, Normal, etc) together with its parameters. The probability density function (PDF) and its parameters can in general be determined by empirical studies on surveyed data.

![Figure 1. Proposed extension of the Value class.](image)

3.2 Provenance/Quality

In the proposed model each property can be qualified with provenance information. Following the PROV-O model, the provenance of an entity is either an entity (dataset, database, document, etc.) or an activity such as an external process (surveying, standards, estimation) or a computational process (value aggregation, inference rules for missing values, etc.). For geometries, provenance information propagates down the hierarchy. For instance, if the geometry property of a tree root has provenance $p$, this means that all its component properties (center, depth, radius) have provenance $p$, unless another provenance is specified for some component (see the example shown in Figure 2).

![Figure 2. Propagation of the provenance information ($p$) to the radius component property of a tree root’s geometry.](image)

3.3 Temporal validity

As with provenance, each property can be qualified by a validity time. This indicates that the qualified object-property-value association is known to be true only during the specified validity time (which may be a time interval or a point in time). This works well for values that remain the same over the time interval. However, there are values that change continually over time (e.g., tree roots grow). In this case the value must be associated with a function that provides its temporal evolution during the validity period. If the value is uncertain, and thus represented by a probability density function (PDF), the temporal evolution function must provide a value for each parameter of the density function at each point in time.

![Figure 3. Representation of the temporal validity of a cylinder’s depth.](image)

A comprehensive example of usage including temporal validity, uncertainty, and provenance is provided by Figure 4. The example shows how the extent of a tree root object is represented. Its geometry is approximated by a vertical cylinder whose center is known, since it is the center of the tree, while the radius and the height (tree root depth) of the cylinder are estimated values, each being represented as an EstimatedValue. For the sake of space, only the radius property of the cylinder is represented.

4. IMPLEMENTATION

The vocabulary of the proposed model has been defined as an OWL ontology. It extends the ontologies defined by Métal et al. (2020) by adding classes and properties to represent uncertainty, data provenance, and time dependent geometries.

In the proposed model, information such as the data provenance and the temporal validity are associated with properties (geometry, center, radius, height, etc.). In a conceptual language such as UML this is represented by association classes that have attributes. In RDF there is no way to qualify a (subject, predicate, object) statement with some attributes. Several design patterns have been proposed for this purpose (n-ary relation pattern, 4D-fluents, etc.) but none of them is entirely
Our proposed model has been validated on typical use cases showing different types of uncertainties. In this section we present three representative examples, showing how it is possible to represent the parameterized geometries using our model. All the provided examples are expressed using RDF-star and they exploit the subsurface ontology. This ontology is linked to several ontologies, such as the geometry ontology and the time ontology, to cite a few.

In the following use cases we exploit the mentioned ontologies through their prefixes. This is quite common in the semantic web field. In order to ease the understanding of the examples we provide a detailed description of the prefixes in Table 1.

### 5.1 Use cases

**Tree root.** A tree root is represented (simplified) as a vertical cylinder defined by a center, a radius, and a depth. The center corresponds to the center of the tree. It is known and represented by a set of MeasuredValue with each value associated to a dimension. The radius and the depth are estimated from the size of the tree. Each of them is represented by an EstimatedValue and associated with a PDF. Since the extent of a tree root usually grows over the time, the parameters of the PDF (e.g., the minimum, mode, and maximum for a triangular PDF) are associated to a temporal evolution function. An example that uses RDF-star is provided by Listing 1. It implements the example shown in Figure 4, extending the geometry with additional properties.

![Figure 4. Example of a tree root object with cylindrical geometry. It has an estimated radius defined by a PERT distribution and only valid within the defined time interval.](image)

#### Table 1. Prefixes of the ontologies used in the examples.

| PREFIX | URI |
|--------|-----|
| geom   | https://purl.org/onto/geometry# |
| sub    | https://purl.org/onto/subsurface# |
| rdf    | http://www.w3.org/1999/02/22-rdf-syntax-ns# |

#### Listing 1. RDF-star representation of a tree root object and its cylindrical geometry.

**Geothermal drill.** Such an object should be positioned vertically but, in reality, this is not always the case. The expected geometry is a vertical cylinder, whilst the actual geometry is inclined. Thus, if the center, the radius, and the depth remain the same, the orientation according to a coordinate system \((x, y, z)\) is no longer \((0, 0, −1)\) but \((x_0, y_0, −1)\) associated to a vertical angle and a horizontal angle. Each of these angles is defined as an EstimatedValue and associated with a PDF. Since the center is known, it is represented by a set of MeasuredValue, one value per dimension. The radius and the depth are also known, thus represented as a MeasuredValue. An example written in RDF-star is provided by Listing 2. For the sake of space, the representation of all the geometry properties is omitted. We only show how the angles and the center properties are represented. In addition, the provenance of the center values is also specified.

#### Listing 2. RDF-star representation of a geothermal drill object and its cylindrical geometry.

This contribution has been peer-reviewed.
Pipe. A pipe is represented as a set of network link objects, each one connected to its previous and following one. We consider that all the network links have the same shape as the whole pipe, which is known (and does not change over the time). However, its positioning and/or orientation may differ from what is expected. The geometry of a network link can be defined as a tube, with radius and length. In addition, the network link has two nodes that represent the endpoints that are connected to the endpoints of the adjacent network links. The radius and the length are both represented as a measured value. The coordinates \((x, y, z)\) of the center of each endpoint are estimated and associated with a PDF. That is not the case for objects such as manholes (and similar elements) for which the position is known and can be represented as a measured value.

Listing 2. RDF-star representation of a geothermal drill object and its cylindrical geometry.

```
<<:center geom:x :point_x>>
geon:provenance "SIG" .
<<:center geom:y :point_y>>
geon:provenance "SIG" .
<<:center geom:y :point_y>>
geon:provenance "SIG" .
:geothermalDrill rdf:type sub:GeothermalDrill;
sub:geometry [ rdf:type geom:Cylinder ;
geon:center :center ;
geon:radius .... ;
geon:depth .... ; ] ;
sub:verticalAngle :Angle ;
sub:horizontalAngle :Angle .
:hAngle rdf:type geom:EstimatedValue ;
geon:pdf [ rdf:type geom:PERT ;
geon:a 0.13 ; geom:b 0.64 ; geom:c -0.4 ; ] .
:vAngle rdf:type geom:EstimatedValue ;
geon:pdf [ rdf:type geom:PERT ;
geon:a 0.25 ; geom:b 1.28 ; geom:c -0.4 ; ] .
```

Listing 3. RDF-star representation of a utility network link object and its geometry/positioning with time constraints.

```
:networkLink rdf:type sub:UtilityNetworkLink ;
sub:geometry :tube ;
sub:start :startNode ;
sub:end .... .
:tube rdf:type geom:Tube .
:startNode rdf:type geom:Node .

<<:startNode geom:position :startPoint>>
geon:hasBeginning "2020-01-01"^^xsd:dateTime ;
geon:hasEnd "2021-01-01"^^xsd:dateTime .
:startPoint rdf:type geom:Point ;
geon:x :point_x ;
geon:y .... ;
geon:z .... .
:point_x rdf:type geom:EstimatedValue ;
geon:pdf [ rdf:type geom:PERT ;
geon:a 0.63 ; geom:b 0.30 ; geom:c -0.8 ; ] .
```

Listing 4. Example of a GeoSPARQL-star query that retrieves the tree roots with a valid geometry (in terms of time) contained in a certain area.

```
SELECT DISTINCT *
WHERE {
    ?treeRoot rdf:type sub:TreeRoot .
    ??treeRoot geom:geometry ?cylinder .
    ?cylinder geom:center ?center .
    ?center geom:aWKT3D ?geom .
    ?geom geom:isX3DWellKnown ?geometry .
    FILTER (?geom :aWKT3D "POLYGON ((2499379.9 1117990.5 , 2500612.9 1119388.3 , 2499376.1 1120157.9 , 2499379.7 1117990.5))"^^geo:wktLiteral) .
    FILTER (xsd:dateTime (?endTime) > NOW())
}
```

5.2 Usage

We here discuss the usage of the model, proposing two possible scenarios. Before presenting them, we show how it is possible to retrieve RDF-star geodata from a knowledge graph. In the previous examples we showed how to represent parameterized geometries of underground objects using RDF-star. Such data are only queryable using SPARQL-star, the -star extension of the SPARQL language. Listing 4 provides an example combining SPARQL-star and the geographic query language for RDF data, namely GeoSPARQL. The example shows how to retrieve the tree roots object that are contained within a specified area and have a valid geometry at the time of querying.

Subsurface data integration. The proposed model may be used as data integration model. A successful use case has been carried out in the SUBSURFACE project, where the model was used to represent and integrate heterogeneous subsurface data of the Geneva canton (Switzerland). Those data have been provided by several different sources, such as, to mention a few,
Le système d’information du territoire à Genève (SITG), Services industriels de Genève (SIG), et Système d’information du Patrimoine Vert (SIPV).

The data integration led to the creation of a knowledge graph with nearly 1.4 million triples, some of which are RDF-star. In addition, since the imported data were incomplete, the data import process included completion operations to infer missing data. Such data are represented as uncertain data whose provenance is defined as “completion_process”.

**Probabilistic collision detection.** The collision probability of a test object (e.g., a planned borehole) with a subsurface object with uncertain geometry can be computed by a Monte Carlo-like technique. It consists in randomly generating a large number of possible geometries of the subsurface object (following the PDFs of the uncertain values). The percentage of the geometries that intersect the test object yields an approximation of the collision probability.

If instead of a test object we consider a single point and compute its collision probability with a subsurface object, we obtain a value between 0 and 1 for each point of the 3D space. These values can be taken to define the membership function of the fuzzy geometry (a fuzzy subset of \( \mathbb{R}^3 \)) of the subsurface object. Figure 5 shows a cross section of the fuzzy geometry of a pipe with uncertain vertical and horizontal positions that have a PERT distribution.

![Figure 5](image)

**Figure 5.** Section of a fuzzy pipe. The contour lines correspond to membership levels 0.001, 0.01, 0.1, 0.25, 0.5.

### 6. CONCLUSION AND FUTURE WORK

We have proposed a model to represent parameterized geometries by extending the ontologies defined by Métral et al. (2020). It defines additional classes and properties to represent geometrical uncertainty and geometrical variations over time. Thanks to this model it is possible to create a knowledge graph that supports different ways of representing geometric uncertainty and changes in geometry (including position) over time. In addition, metadata such as data provenance can also be represented.

As future goals, we aim to extend the proposed model to represent uncertain geometries of objects that are not physically well-defined, such as contaminated sites or archaeological remains.

### ACKNOWLEDGEMENTS

This research has been partially supported by Innosuisse in the framework of the Innovation project 35265.1 IP-ICT, “Impulse-Subsurface: Efficient data exploitation in urban subsurface planning”.

### REFERENCES

Guerrero Iñiguez, J. I., 2017. Geometric Modelling of Tree Roots with Different Levels of Detail. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W3(4W3), 29–35. https://doi.org/10.5194/isprs-annals-IV-4-W3-29-2017.

Hartig, O., 2017. Foundations of RDF* and SPARQL*: (An alternative approach to statement-level metadata in RDF). *CEUR Workshop Proceedings*, 1912, CEUR-WS.

Joskowicz, L., Ostrovsky-Berman, Y., Myers, Y., 2010. Efficient representation and computation of geometric uncertainty: The linear parametric model. *Precision Engineering*, 34(1), 2–6. https://doi.org/10.1016/j.precisioneng.2009.05.002.

Métral, C., Daponte, V., Caselli, A., Di Marzo, G., Falquet, G., 2020. Ontology-based Rule Compliance Checking for Subsurface Objects. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-, 91–94. https://doi.org/10.5194/isprs-archives-XLIV-4-W1-2020-91-2020.

Métral, C., Falquet, G., 2018. Extension And Contextualisation For Linked Semantic 3D Geodata. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W10, 113–118. https://doi.org/10.5194/isprs-archives-XLII-4-W10-113-2018.

Morel, M., Gesquière, G., 2014. Managing temporal change of cities with CityGML. *Eurographics Workshop on Urban Data Modelling and Visualisation, UDMV 2014 - Proceedings*, 37–42.

olde Scholtenhuis, L. L., den Duijn, X., Zlatanova, S., 2018. Representing geographical uncertainties of utility location data in 3D. *Automation in Construction*, 96, 483–493. https://doi.org/10.1016/j.autcon.2018.09.012.

Sahoo, S., Lebo, T., McGuinness, D., 2013. PROV-o: The PROV ontology. W3C recommendation, W3C. https://www.w3.org/TR/2013/REC-prov-o-20130340/.

Tappolet, J., Bernstein, A., 2009. Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL. *Lect Notes Compu Sci (LNCS)*, 5554, 308–322.