Towards a semantic web representation from a 3D geospatial urban data model
Diego Vinasco-Alvarez, John Samuel, Sylvie Servigne, Gilles Gesquière

To cite this version:

HAL Id: hal-03240567
https://hal.archives-ouvertes.fr/hal-03240567
Submitted on 28 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Towards a semantic web representation from a 3D geospatial urban data model

Diego Vinasco-Alvarez¹, John Samuel², Sylvie Servigne³, Gilles Gesquière⁴

1. Univ. Lyon, Université Lumière Lyon 2, LIRIS
   5 av. Pierre Mendès-France, 69622, Villeurbanne, France
diego.vinasco-alvarez@univ-lyon2.fr

2. Univ. Lyon, CPE Lyon, LIRIS
   43, bd du 11 novembre 1918, 69616, Villeurbanne, France
john.samuel@cpe.fr

3. Univ. Lyon, INSA Lyon, LIRIS
   7 Avenue Capelle, 69621, Villeurbanne, France
sylvie.servigne@insa-lyon.fr

4. Univ. Lyon, Université Lumière Lyon 2, LIRIS
   5 av. Pierre Mendès-France, 69622, Villeurbanne, France
gilles.gesquiere@univ-lyon2.fr

ABSTRACT. Urbanization is a continuous evolution process that is currently studied by a number of researchers. Multi-source and multidimensional city information models are often used to

RÉSUMÉ. L’urbanisation est un processus d’évolution continue qui est actuellement étudié par nombre de chercheurs. Les modèles d’information de la ville, multi-sources et multidimensionnels, sont souvent utilisés pour comprendre le paysage urbain, qui est en constante évolution. Ces modèles peuvent cependant poser des problèmes d’interopérabilité et de perte de données lors de conversions, souvent nécessaires pour permettre leur intégration. Aujourd’hui, nous pouvons nous baser sur des modèles conceptuels qui peuvent aider à traiter les pertes de données lors de conversions, et peuvent également contribuer à préserver l’interopérabilité des données. Ce type d’approche modèle-centrée permet de définir une représentation commune. Récemment, un mouvement vers la représentation de données basée sur les graphes et la sémantique a gagné en popularité pour répondre aux problèmes d’interopérabilité. Dans un premier temps, pour illustrer notre approche, nous considérons CityGML, un standard de représentation des informations urbaines en 3D. Nous proposons une stratégie pour convertir la sémantique du modèle conceptuel de CityGML vers des ontologies et, ensuite, en formats web sémantiques pour faciliter l’intégration. Nous proposons également une méthode pour convertir et stocker les instances CityGML en individus RDF qui respectent l’ontologie générée. L’approche proposée permet de pallier le déficit d’information sémantique résultant de la traduction directe de différents types de données en graphe, tels que RDF.
understand the ever-changing urban landscape. These models may encounter issues with interoperability and data-loss if conversion is required for integration. Today we can base ourselves on conceptual models which can help deal with data losses during conversion and may also help preserve data interoperability. This kind of model-driven approach can be useful as common representation. Recently, a movement towards graph and semantic based data representations has also grown in popularity to respond to these issues. As a first step, we consider CityGML, a common standard that can be used to represent 3D urban information. We propose a strategy for converting the semantics of CityGML conceptual model into ontologies and later to semantic web formats to facilitate integration. In addition, we propose a method for converting and storing CityGML instances into RDF individuals that respect the generated ontology. This proposed approach overcomes the loss of semantic information resulting from the direct translation of different types of data into graphs such as RDF.

MOTS-CLES : Modèle conceptuel de données urbaines 3D, Web sémantique, Interopérabilité, CityGML, RDF, GeoSPARQL

KEYWORDS: 3D urban data conceptual models, Semantic Web, Interoperability, CityGML, RDF, GeoSPARQL

1. Introduction

The urban landscape is a complex and heterogeneous source of information. The anthropization and urbanization of modern cities have brought into the fold various actors, issues, and systems that all have complicated interconnections and dependencies both on local and global scales. The data-driven approaches used to represent and model these urban environments often depend on multi-source and multidimensional urban data, which consists of a variety of different information coming from multiple actors and organizations (Barbosa et al., 2014). For example, city governments have different departments for handling different subjects or issues like water and sewage, pollution, traffic, energy, etc. (Rochet and Pinzon, 2016). Sometimes these organizations release open datasets using their own internally defined business formats or using formats that conform to a specific use-case as opposed to more generic or widely used standards, thus making it difficult to get an integrated view of the data from multiple sources.

In response to the limits of using internal formats, we are now seeing the growing use of geospatial and 2D/3D urban data standards. Different international organizational bodies (OGC, ISO, W3C...) propose standards and release conceptual data models for their standards to ensure that the data is represented uniformly by the users, thereby giving open datasets an aspect of interoperability. Take for example, the CityGML standard1 released by the Open Geospatial Consortium (OGC), a city

1https://www.ogc.org/standards/citygml
A semantic web representation of urban data

information model used for 3D urban modeling to represent city objects at different levels of detail.

Another aspect of urban data is the underlying n-dimensional structure. For a long time, 2D maps, aerial views, and cadastral views were used for the management of the urban lifecycle. With the growing availability of 3D data, thanks to new data acquisition processes (LIDAR, photogrammetry...), numerous applications are now possible with 3D urban data (Biljecki 2015). Recently, some urban data models also take into account the changes of cities as they evolve over time in addition to spatial information (Chaturvedi and Kolbe, 2019; Jaillot et al., 2020; Samuel et al., 2020), thereby adding the temporal dimension (4D). These models help historians and city planners visualize and contextualize the impacts previous key projects have had on the development of a city. Such studies require an integration of the urban models with document corpus consisting of project plans, newspaper articles, archives etc. However, it is challenging to integrate data from different sources with different standards for providing an integrated urban view (Biljecki et al., 2018; Ohori et al., 2018).

Semantic web models are now being used increasingly to ensure interoperability among multiple data sources (Malinverni et al., 2020; Nuninger et al., 2020; Tran et al., 2016). Due to its increasing popularity, a number of tools are now available to transform data in legacy formats to semantic web formats but, as said by C. Claramunt in (Claramunt, 2020) “3D models and Building Information Models should be fully integrated. However, any transformation must conform to the original conceptual model used for developing the standard or risk inconsistent or semantically incomplete representations (Bohring and Auer, 2015).

In this work, we take into consideration a model-centric approach for transformation of urban data into ontological and graph formats. We demonstrate our approach with CityGML, particularly making use of open data from the metropole of Lyon. Our approach is detailed in this article. Section 2 presents the problem statement of model-centric urban data transformation in detail. Section 3 presents the existing state of the art. Taking the case study of CityGML, we detail our approach in Section 4. However, the focus of our work is not limited to transforming CityGML, but to link data from other sources, especially those based on open standards. These perspectives are discussed in section 5. Section 6 concludes the article.

2. Problem Statement

Managing and analyzing urban data is complex because different stakeholders and different entities produce these data, sometimes in silos. City administrators and urban planners often require an integrated view of these data for analyses and for obtaining meaningful insights for daily routine tasks and future planning. Take for example, urban planners who wish to construct a new high-rise building in a particular sector, need to have information on the other structures present around the proposed building. Such information may help them to study the impact on visibility of historical buildings of interest, analyze the reduced or increased sunshine or shadows because
of the new construction. All these studies require data from multiple sources, such as the terrain information, 3D structures of urban objects, position of celestial bodies at different points of times etc.

However, the task of integrating urban data from multiple sources remains very difficult. Additionally, the various standards used to model this data evolve based on the growing requirements of the domain users. New versions of standards need to be released. Sometimes, the data released under previous versions is incompatible with the newer versions. Hence, any semantic web representation solution for an integrated view of urban data must deal with the heterogeneous data from multiple sources as well as their evolution.

Concerning 3D geospatial urban data models, many approaches propose directly translating different data formats into graphs, such as RDF, without taking into account the underlying conceptual model, and consequently omit semantic information during translation, weakening the initial model. The question is therefore to propose an approach that allows the preservation of the semantics of a 3D urban model when converting into a semantic web representation.

3. State of the art

There are growing efforts to ensure semantic interoperability across heterogeneous data from diverse multiple sources. To obtain an integrated view, one commonly used approach is to transform the data from one data format to a common format by making use of data transformers. Stylesheets like XSLT, JSONT are commonly used for this purpose. However, some of these stylesheets are written on demand basis and may not be available on the internet for use. This means that if the users want to reuse and reproduce some of the results based on the transformed data, they may not be able to achieve it, especially if transformers follow different methodologies. In addition, information loss is another major problem during data transformation (Levina, 2012). It is important to take into consideration the initial conceptual models used for the conception of the modeling.

Several information modeling techniques exist in the literature (Bork et al., 2020). One possible approach is to ensure that a common modeling language is used across domains, like the semantic web ontology (OWL) language. However, UML models are often used by domain experts (De Paepe et al., 2017) to conceive and represent information models instead of semantic web technologies like OWL, RDFS etc. Another possible approach is to build ontologies from existing standards, i.e., transforming models conceived using different modeling languages to a semantic web ontology. Researchers are currently exploring how to make the most of the familiar modeling languages like UML (De Paepe et al., 2017) and automatically generate ontologies in OWL from these models. Works like (Bohring and Auer, 2015; Kramer et al., 2015; Usmani et al., 2020) use similar approaches based on XML Schema as a model towards OWL ontologies.

Another possible solution to this problem is if a standards body releases the ontologies for their conceived models. Take for example, the GML 3.2 geometries
ontology\textsuperscript{2} made available by the OGC. Currently, an official ontology for CityGML 2.0 is missing from the standards body. Nevertheless, it is important to state some works like (Brink et al., 2014; Métral and Falquet, 2018) proposed ontologies for CityGML 2.0. Additionally, there is a semantic web working group looking for ensuring the release of OWL for future releases of CityGML.

As discussed above, we are looking for integrating heterogeneous data from multiple sources by proposing an approach that preserves the semantics of a 3D urban model when converting it to semantic web representations.

Hence, it is important to have standard procedures or guidelines for ensuring the mapping between different models so that it can be easily reproduced by other users. Our approach is to take into consideration the conceptual model of information related to cities and transform the model using these standard procedures so that the data converted is conformance to the original model and can also be reproduced by other users following these guidelines.

4. Proposed approach illustrated with the CityGML 2.0 Conceptual Model

As previously stated, the information produced by cities and the urban lifecycle can be broken down into hierarchies of concrete and abstract subdomains of information - such as energy, transportation, infrastructure, and many others - and be represented and visualized by 3D urban data models. CityGML is an international OGC standard that is commonly used for this purpose. Taking the CityGML conceptual model as an example, we see frameworks for describing the spatial and thematic information of the more ‘material’ subdomains within CityGML’s modules, e.g. water, buildings, terrain, etc. More immaterial subdomains such as education or public administration are not currently represented by these modules but as a first step the existing conceptual model is more than sufficient to begin transforming open urban data into semantic web formats.

To efficiently take advantage of the benefits of semantic web formats, a direct transformation of CityGML instances is not enough, the CityGML conceptual model itself must also be transformed and be used to ‘guide’ instance transformations. This results in meaningful RDF graphs that respect the resulting urban data conceptual model. Resources for reproducing the approach proposed here is discussed in Section 6: Data and code availability statement.

4.1. Proposed Approach Design

In order to transform the CityGML conceptual model and CityGML instances into semantic web representations, our proposal uses several XSLT-based transformations based on previous works on XML to RDF, RDFS, and OWL conversion (Bedini et

\textsuperscript{2}http://schemas.opengis.net/gml/3.2.1/gml_32_geometries.rdf
al., 2011; Bohring and Auer, 2015; Brink et al., 2014; Kramer et al., 2015; Métral and Falquet, 2018; Vinasco-Alvarez et al., 2020). The general proposed pipeline is described in Figure 1, which can be broken down into 3 main activities:

- Creation of an urban data model ontology from the CityGML conceptual model (as described by its application schema)
- Generation of a CityGML instance to RDF transformation stylesheet from the CityGML conceptual model (as described by its application schema)
- Transformation of CityGML instances to RDF graphs using the stylesheet generated in step 2.

**FIGURE 1. Overview of proposed pipeline applied to the CityGML conceptual model and instances**

During each transformation, the resulting information must be logically consistent and maintain its interoperability from CityGML and GML. To ensure this, several challenges need to be overcome. For instance, the generated RDF graphs and types must conform to the model described by the ontology. How can mappings be created to ensure this? The CityGML application schema often implements elements that do not have a direct equivalent in OWL or RDFS; how should these elements be represented to best describe CityGML with semantic web formats? In addition, CityGML schema often draws from elements, types, functionality from external schema such as xLinks and xAL addresses. How should these imported elements be addressed to preserve their original functionality? Finally, how can we make use of existing semantic web standards to make geospatial queries? These issues are addressed in sections 4.2 and 4.3.

In addition, to create holistic transformations, the CityGML conceptual model must be assimilated or imported in its entirety into our pipeline. Because the model is divided into several modules which are sometimes interdependent and dependent on the GML standard, the complete model is defined by 43 different XML schema
documents: 29 to define the GML 3.1 standard\(^3\) and 14 to define the CityGML 2.0 standard\(^4\) (not including xAL). Schema cannot be processed in isolation or we risk certain loss of data in the transformation (for example in distinguishing when to generate an `owl:ObjectProperty` from an `owl:Datatype` property when the XML element type in question is declared outside the current schema).

Our proposed approach chooses to consolidate these components in a composite XML schema, created from the definitions within each relevant schema in the conceptual model. This prerequisite step also requires that all definitions in the composite schema use normalized namespace prefixes to differentiate between references to one module or another. Once this is complete, the composite schema can be passed into the transformation pipelines.

### 4.2. Creating a 3D Urban Data Ontology

As an initial approach, we consider that the conceptual model can be expressed by XML schema and in a general simplification, XML schema can use `xs:elements`, `xs:complexType`, and `xs:simpleType` to define the concepts and semantics of domain specific urban. In addition, when creating the transformation from the CityGML conceptual model into an OWL ontology, it is important to reuse as many existing concepts as possible in the semantic web to enrich and render the model interoperable.

Using the mapping transformations proposed in the previously mentioned works (Bedini et al., 2011; Bohring and Auer, 2015; Kramer et al., 2015; Vinasco-Alvarez et al., 2020), this can be efficiently achieved.

Once the XML schema is transformed into a domain ontology, we can integrate the GeoSPARQL standard\(^5\). In GeoSPARQL, the `geo:SpatialObject` class represents any real world object with thematic and geospatial properties. Geometric properties of these classes are represented by `geo:Geometry`, a subclass of `geo: SpatialObject`, and can be serialized by `geo:gmlLiterals`. It is disjoint with the `geo:Feature` class, which represents the thematic properties of a `geo: SpatialObject` and can be linked to related geometric classes via the `geo:hasGeometry` property. As suggested in (Battle and Kolas, 2012), to integrate the thematic classes of our generated ontology with GeoSPARQL, we must declare our classes as a subclass of the `geo:Feature` class. Since our ontology is partially generated from the GML schema, any class which is derived from the `gml:_Feature` type is implicitly declared an `rdfs:subClassOf geo:Feature`. This also means that all of the geometry class declarations defined from the GML schema can be removed from the ontology as they are already defined in the

---

\(^3\)http://schemas.opengis.net/gml/3.1.1/
\(^4\)http://schemas.opengis.net/citygml/
\(^5\)https://www.ogc.org/standards/geosparql
These proposed transformation strategies to create OWL ontologies from XML schema are an amalgam of previous transformation approaches with several specializations for the CityGML application schema. However, they are largely generalized and can be applied to schema outside of the main GML and CityGML schema, such as the external xAL addressing schema used by CityGML.

4.3. Creating Meaningful Transformations of Urban Data Instances

When generating the CityGML instance to RDF transformation, it is important that the patterns created are general enough to be reused and can take advantage of the OWL/RDFS and GeoSPARQL vocabulary whenever possible. Several techniques suggested in (Brink et al., 2014) are implemented to achieve this. In addition, this process uses the general structure of the approaches proposed in (Bohring and Auer, 2015; Kramer et al., 2015; Vinasco-Alvarez et al., 2020) with consideration taken so that these mappings work with the ontology to be generated alongside the RDF data.

In general, the following mappings in table 1 are applied to the composite schema to generate the CityGML to RDF transformation stylesheet. The generated CityGML to RDF template also uses `gml:id` attributes as `rdf:ID` or `rdf:about` whenever possible, as proposed in (Brink et al., 2014). If no `gml:id` is available, a unique id is generated and appended to the local name of the element. Also as suggested in (Brink et al., 2014), the `rdf:type` of each individual should be generated from its local name in order to reference the generated ontology from section 4.2, or the official GML ontology in the case of geometric elements.

Table 1. General XSD to ‘XML to RDF XSLT’ mapping transformations

<table>
<thead>
<tr>
<th>XML Schema Concept</th>
<th>Resulting XML to RDF XSLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>global xs:element elements based on a xs:complexType</td>
<td>xsl:template for generating owl:NamedIndividuals of type xs:element that calls the template for its xs:complexType</td>
</tr>
<tr>
<td>xs:elements which are children of xs:complexType or xs:groups</td>
<td>xsl:templates for owl:ObjectProperties or owl:DatatypeProperties based on the type of the xs:element</td>
</tr>
<tr>
<td>xs:attributes of xs:complexType or xs:attributeGroups</td>
<td>xsl:templates for owl:DatatypeProperties</td>
</tr>
<tr>
<td>xs:complexType, xs:attributeGroup, and xs:group</td>
<td>xsl:template that calls relevant templates for every possible property or text the element could have</td>
</tr>
</tbody>
</table>

https://def.isotc211.org/ontologies/iso19136/
A semantic web representation of urban data

<table>
<thead>
<tr>
<th>xs:simpleType with an xs:restriction or xs:extension of a native XML datatype</th>
<th>xsl:template that selects the text of an element</th>
</tr>
</thead>
<tbody>
<tr>
<td>xs:simpleType with an xs:restriction or xs:extension of another xs:simpleType</td>
<td>xsl:template that calls the relevant template for the xs:simpleType it is based on</td>
</tr>
</tbody>
</table>

These transformations also map the geometry of a CityGML feature to an RDF triple using GeoSPARQL’s `geo:asGML` datatype property. This is applied whenever a branch of GML elements is detected which has a CityGML parent and is composed of only GML nodes. Such a branch represents a valid `geo:gmLiteral` if all of its nodes are in the substitution group `gml:_Geometry`. This is ensured by placing the `geo:gmLiteral` transformation within the template for `gml:AbstractGeometryType`, as any element within the same substitution group as this type can be serialized as a GML literal.

This script also scans the transformed instances for malformed RDF triples and fully qualifies any RDF attributes that contain namespace prefixes such as `rdf:resources`, `rdf:type`, and `rdf:about`. These transformations follow the “garbage in, garbage out” concept that poorly formed data input into a program, will produce nonsensical results, and thus assume that the GML and CityGML instance documents provided are well structured and conform to their application schema. If this assumption is met, the resulting data should conform to the ontology transformation discussed in the previous section.

4.4. Leveraging the GeoSPARQL Standard

Once the dataset was generated, a GeoSPARQL endpoint was installed to perform geospatial queries. The Parliament triple-store\(^7\) was used for this purpose as proposed in (Battle and Kolas, 2012) since it contains a SPARQL endpoint with GeoSPARQL support based on the Apache-Jena libraries\(^8\). We noted that if a particular coordinate reference system (CRS) is used, it must be declared in the literal value as a `gml:srsName` attribute in order for Parliament to parse and index the coordinates and geometry as the default reference system is `<http://www.opengis.net/def/crs/OGC/1.3/CRS84>`. Thus any GML data that uses a different reference system must include it in the GML instances themselves during transformation or manually add it to the output `geo:gmLiterals` after transformation. This implementation adds these CRS declarations to the data through a Python script after the transformation since they are not declared in the source CityGML data. Through these practices, the geometry of CityGML instances can effectively be generated, stored, and queried through GeoSPARQL.

\(^{7}\)https://github.com/SemWebCentral/parliament

\(^{8}\)https://jena.apache.org
5. Discussion

The CityGML conceptual model is a useful starting point for creating a 3D urban data ontology as a city information model. Its domain specific, modular structure provides a solid base for describing several hard domains of the urban ecosystem, which can be easily extended and integrated once in ontological form (Métral and Falquet, 2018). Even ADE extensions of CityGML could be applied to this approach to describe existing sub-domains of information if an XML schema is provided - such as the NoiseADE9 provided by the OGC. Another benefit of this modular approach is that the ontology can easily be pruned into domain specific sub-ontologies as each domain can be referenced by a unique URI namespace. In addition, enriching the ontology after transformation with existing geospatial standards is straightforward.

However, a limitation of this approach is its reliance on the semantic limitations of XML Schema. For example, the recently declared ISO 19150-2 standardizes guidelines for UML to OWL transformation of geographic information standards. For GML, (Brink et al., 2014) proposes using Shapechange10, which takes advantage of this standard, to generate ontological models and acknowledges that ontology generation from UML and XML schema conceptual models are not semantically equivalent. When comparing the resulting ontologies between Shapechange and this proposed approach, still focusing on CityGML, initial explorations indicate that the ISO 19150-2:201511 ontology mappings are more expressive and direct than XML Schema to OWL. For example, an aggregation in UML can be converted into an OWL object property, while XML Schema expresses these entities as complex types which may be converted into extraneous OWL classes and properties to represent the same semantic information. Yet, Shapechange does not facilitate the transformation of CityGML or XML data into graph formats, and thus can only improve the city information model generated by our approach without supplemental transformations.

6. Conclusion

In this article, we proposed a model centric approach towards semantic 3D urban data representations, where we consider that any data transformation from one data format to another format must conform to the equivalent underlying conceptual model. We implemented this proposed approach to convert the CityGML 2.0 conceptual model into an OWL ontology and to convert CityGML 2.0 data to RDF based on the conceptual model. Subsequently, geospatial queries were made on the resulting data using a GeoSPARQL endpoint. This open-source solution is available

9http://schemas.opengis.net/citygml/examples/2.0/ade/noise-ade/CityGML-NoiseADE.xsd
10https://shapechange.net/
11Geographic information - Ontology - Part 2: Rules for developing ontologies in the Web Ontology Language (OWL), https://www.iso.org/standard/57466.html, Publication date: 2015-07
A semantic web representation of urban data

Online\textsuperscript{12} and has been tested with the data from the 1st borough of Lyon (see the data and code availability statement below). Our future works will explore integrating other information sources - taking into consideration their conceptual models - and integrating a UML based approach. This approach will lead us to facilitate integration of 3D data in city information models and building information models. We are working to test this approach on CityGML 3.0. This will also demonstrate the application of this approach across different versions of standards.

Acknowledgements

The authors would like to thank University Lumière Lyon 2 for funding the PhD Thesis of Diego Vinasco-Alvarez and the LIRIS UMR 5205 Laboratory for funding his internship as part of a transversal effort with the GOAL team. This work has been done within the Vcity project (https://projet.liris.cnrs.fr/vcity/). Special thanks to Eric Boix for fruitful discussions and kind help.

Data and code availability statement

Detailed notes and the code for reproducing the results of section 4.2 can be found at: https://github.com/VCityTeam/UD-Graph/tree/master/CityGML_Transformations/XSD-to-OWL. Detailed notes and the code for reproducing the results of sections 4.3 can be found at: https://github.com/VCityTeam/UD-Graph/tree/master/CityGML_Transformations/XML-to-RDF.

Bibliographie et références


\textsuperscript{12}https://archive.softwareheritage.org/swh:1:dir:7a2459481066a4f77a433ec9a220d8fe5af279;origin=https://github.com/VCityTeam/UD-Graph;visit=swh:1:snp:d95c3eaaa6a4f22e77d63f1531be901e71f7e6744b;anchor=swh:1:rev:22b63103c50d21234961d1854c03cb3ab4b2abc/


