Leveraging Standards in Model-Centric Geospatial Knowledge Graph Creation

Diego Vinasco-Alvarez[1][0000−0002−7171−5754]

Université Lumière Lyon 2, LIRIS UMR-CNRS 5205, Université de Lyon, Lyon, France; diego.vinasco-alvarez@univ-lyon2.fr;
https://liris.cnrs.fr/page-membre/diego-vinasco-alvarez

Abstract. Understanding the complex urban landscapes of cities and their evolution is becoming an ever more essential area of research for urbanists, city planners, historians, and industry leaders. Toward this endeavor, data-driven 3D semantic city models can be implemented to create tools for understanding, simulating, and modeling these urbanization processes and many other urban phenomena. These implementations often require integrating multidimensional (2D/3D, temporal, and thematic), heterogeneous, and multisource urban data to provide users with more complete views of the changing urban landscape. In recent years, researchers have turned toward Semantic Web technologies such as knowledge graphs as common platforms for integrating these data and their underlying data models. However, simple transformation or conversion of urban data towards these formats is prone to data loss, and integration of urban data model standards lacking interoperability poses its own challenges. This work proposes a model-centric urban data transformation approach towards Semantic Web data formats, based on international standards for facilitating the integration of these urban data and promoting their interoperability in the context of multidimensional city modeling.

Keywords: Ontologies · Data integration · Urban data · Multidimensional data · Conceptual Modeling.

1 Introduction and Motivation

Urbanization is a complex and increasingly relevant change process. In 2018 the United Nations reported that 55% of the world's inhabitants live in urban areas and predicted that number to increase to 68% by 2050 [23]. Today, the utilization of 3D and 4D semantic city models [2,27], urban digital twins, and smart city applications [20] is a growing area of research for understanding these urban processes. These approaches often go beyond purely spatial 2D and 3D city models and incorporate more complex multidimensional (nD) structured city data including temporal data such as 4D (3D+Time) [1] and thematic or categorical data defined by urban data vocabularies and data model standards such as

CityGML[1] and BIM-IFC[2]. In addition, these data are produced from a variety of data sources and data capturing techniques such as LIDAR, photogrammetry, Internet of Things (IoT) sensors, etc. leading to a large heterogeneity in data formats.

To provide users a more complete view of the evolving urban landscape, multidimensional city models often require the combination of these heterogeneous urban data [22]. Consequently, they must overcome the data integration challenges that arise from a lack of interoperability between these structured and semi-structured urban data formats and data models. To this end, approaches relying on Semantic Web technologies as a platform for urban data integration are being proposed [8,9,16] where both these data and their underlying data models can be represented, formalized, combined, and shared as machine processible open-linked-data through the use of Resource Description Framework (RDF) knowledge graphs and Ontology Web Language (OWL) ontologies.

These approaches often take either a bottom-up or a top-down method of integration [10]. Bottom-up approaches such as Ontology-Based Data Access (OBDA) can bridge the interoperability gap between heterogeneous urban data by mapping the schema of existing relational databases to a common ontological model [14,19]. Conversely, top-down (or model-centric) approaches such as Model-Driven-Development (MDD) or Model-Driven-Architecture (MDA) [13] rely on data models and data transformations to provide effective and reproducible methods for creating OWL ontologies from existing urban data models and datasets. These approaches may provide the possibility of reusing existing urban data model standards with existing structured data as opposed to creating new ontologies manually.

This evolving **middle-stage** work proposes a model-centric standards-based data transformation approach for improving nD urban data quality and interoperability during urban data integration as RDF knowledge graphs.

This paper presents the work as follows: Section 2 describes the state of the art and related works in nD urban data modeling, transformations to semantic web formats, and heterogeneous urban data integration; Section 3 presents the problem statement and each contribution; Section 4 presents the research methodology and the implementation of the proposed approach; Section 5 details how the approach will be evaluated; Section 6 will present the intermediate results achieved thus far; and Section 7 will present the lessons learned and conclude.

2 State of the Art

The representation and querying of geospatial and spatio-temporal urban data with Semantic Web technologies have seen much advancement in recent years

[1] https://www.ogc.org/standards/citygml
[2] https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/

with proposals such as GeoSPARQL, stRDF, Building Topology Ontology, Ontology for Managing Geometry, and BimSPARQL [3,29]. In the context of using transformation towards these technologies to facilitate nD urban data integration for enriching 3D semantic city models, there are several existing works in the literature to take into consideration.

2.1 nD Urban Data Model Transformation and Alignment

When creating ontologies with transformation approaches, it is important to consider the modeling language with which the input data model is formalized in, as the same data model formalized in different languages can produce OWL ontologies with varying structure and vocabulary.

In general, there are two primary kinds of urban data models used as input in these transformations:

- Conceptual data models (such as UML models) [6,11]
- Physical data models (such as XML schema) [24]

Top-down approaches using conceptual data model formats such as UML can exploit the similar modeling concepts and relationships to OWL (e.g. classes, attributes, properties, cardinality, etc.) to propose more straightforward mappings compared to XML Schema-based transformations [4]. To this end, geospatial model transformation standards such as ISO 19150-2[3] have been defined by international organizations such as ISO[4] and the Open Geospatial Consortium[5] for defining transformation mapping rules between geospatial UML models to OWL. In the case of both UML and XML Schema-based approaches - which use more constrained, closed-world formalizations of data models compared to OWL's open-world interpretation - a choice exists in how constrained the generated ontology should be. Which method of transformation should be implemented may vary based on the motivation for integration or use-case [4].

After a data model is expressed in OWL, it can be integrated with other OWL ontologies through alignment and ontology matching techniques [21]. In nD urban data integration, much research has been proposed with commonly implemented data models in Geographic Information Systems (GIS) and Building Information Modeling (BIM) [7,18,24], 3D Cultural Heritage Applications [17], and 4D urban data modeling [22].

2.2 Model-centric nD Urban Data Transformation

Model-centric transformation approaches such as Extract-Transform-Load (ETL) data pipelines can use data models or schema to define or guide the transformation of heterogeneous urban data towards RDF and OWL [4,12] with some works proposing scalable data-lake integration solutions [15]. Similar to the data

[3] https://www.iso.org/standard/57466.html
[4] https://www.iso.org/home.html
[5] https://www.ogc.org/

model transformations noted in section 2.1, the data format used to represent the underlying data models of the data to be transformed affects the structure of the output data. Here, bottom-up approaches are more commonly used as relying on schemas that directly define the structure of the data to be transformed may have the benefit of better preserving the geospatial semantics of these data [4].

3 Problem Statement and Contributions

Limiting semantic data loss and maximizing data interoperability are crucial to improving heterogeneous multisource nD urban data integration for enriching 3D semantic city models, Smart City Applications, and Urban digital Twins. This work hypothesizes that *a standards-based model-centric data transformation approach guided by conceptual, as opposed to physical, data models can improve the quality and interoperability of generated nD urban data models and data during integration as linked-open-data.*

To evaluate this hypothesis, three research questions arise:
Q1: *How can geospatial model transformation standards such as ISO 19150-2 be used to create more concise and useful nD urban data ontologies for nD city modeling while facilitating the integration of existing spatio-temporal ontologies?*
Q2: *What data quality issues arise (e.g. semantic data loss, inconsistencies) when using either conceptual or physical data models to guide data transformation pipelines and how can they be mitigated?*
Q3: *How can these transformations be structured to improve scalability when integrating large datasets covering the district and city-sized urban areas?*

Towards this end, several contributions have been realized [25,26]:

- A state of the art study in creating OWL ontologies from nD urban data UML and XML Schema data models
- A reproducible nD urban data model transformation pipeline based on the ISO standard 19150-2:2015 UML to OWL mappings
- An nD urban data ontology network based on urban data standards
- A proposed approach for integrating heterogeneous multisource nD urban data as linked-open-data with a model-centric data transformations

4 Research Methodology

To confront the aforementioned problem statement, this work utilizes the following research methodology: Firstly, several state-of-the-art studies must be conducted to better understand the domain of nD urban data integration. Two studies are conducted in the creation of OWL ontologies from nD urban data models to compare the use of conceptual and physical data models. In addition, a study of data integration approaches of nD urban data as RDF knowledge graphs will be done to better understand how to limit data loss during these processes and produce interoperable linked-open-data. These studies will also take into consideration which data models are being commonly implemented in these applications as reusing vocabularies is key to improving interoperability in OWL [4]; Secondly, a data model transformations will be implemented using the information gleaned from the state of the art to create ontologies from widely used urban data model standards. These ontologies will be aligned with existing spatio-temporal ontologies to create an nD urban data ontology network and ensure their interoperability as linked-open-data vocabularies; Thirdly, a data integration methodology (see figure 1) is proposed based on model-centric data transformations that exploit the nD urban ontology network to define their transformation mappings to RDF; Finally, this approach (and its resulting data and data models) will be tested according to an evaluation plan (detailed in section 5) to validate if it promotes the data quality of nD urban data standards during transformation and to determine its viability for data integration in enriching nD semantic city models, urban digital twins, and smart city applications.

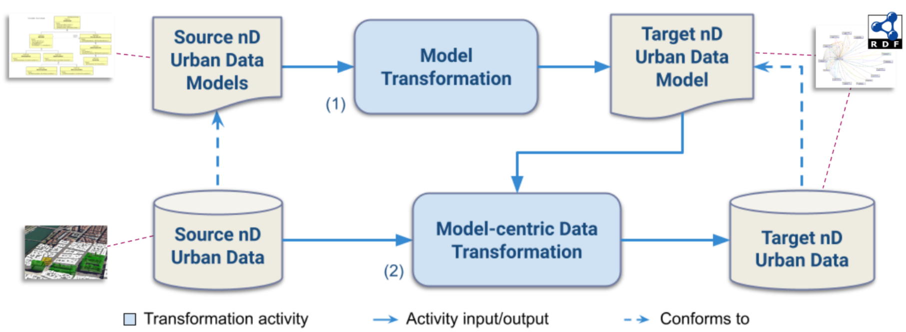

Fig. 1. Proposed integration methodology

5 Evaluation Plan

To test the hypothesis stated in section 3, this work will measure and profile several criteria of the proposed integration approach and resulting knowledge graphs. In regards to **Q1**, two OWL ontologies representing the same data model standard will be automatically generated (one from a UML model and another from an XML Schema) and then compared on criteria such as consistency and conciseness in the context of nD urban data applications using requirements and tests developed using the ROMEO [28] ontology evaluation methodology. In response to **Q2** and **Q3**, several knowledge graphs will be created from real-world open urban data repositories using model-centric transformations guided by the generated ontologies and their respective physical schemas. Profiling will be implemented to re-evaluate these graphs when combined with their previously generated ontologies. A method for determining if, and how much data loss has occurred during transformation is also in development. Additionally, the SPARQL query execution time of oft-asked nD questions from historians and urban planners (such as how city objects and urban phenomena evolve over time [5]). A synthesis of these tests will be done to measure the possibility of real-time analysis and scalability of these data in urban digital twins and smart city applications.

6 Intermediate Results

At this point, this work has achieved several initial results. A reproducible urban data transformation tool, UD-Graph, was implemented to generate urban data RDF knowledge graphs. This tool can currently be used to implement two types of transformations for generating OWL ontologies from either:

- Geospatial XML Schema [25] extending the mappings based on [24] to take advantage of geospatial standards in the Semantic Web
- Geospatial OWL ontologies created from UML models using the ISO 19150-2 standard [26]

Thus far, an initial comparison of the resulting ontologies and datasets has been performed using the CityGML 2.0 and 3.0 conceptual models [26] within the context of improving the integration of 3D city model snapshots to model spatio-temporal building evolution. These data models were chosen for their rich vocabularies and widespread use in nD urban data research and industry. In the UML-based approach, the generated ontology is queried using to determine how XML (where nodes and attributes are represented as a tree) should be converted into RDF triples. Since a conceptual model does not directly define the structure of the data instances, some human intervention is required in defining mappings in the case where the structure of the data diverges from the UML model. For example, in the case where an element in XML uses a different name or namespace in its XML schema compared to the UML model. Defining custom mappings has the added benefit of being able to reuse existing OWL vocabularies such as SKOS, GeoSPARQL, OWL-Time, and ISO standard ontologies similarly generated from geospatial UML models[6].

Alongside these contributions, some work has been done in conjunction with the OGC's CityGML 3.0 XML Encoding Sub-working Group[7] to produce standardized 4D urban data.

[6] https://def.isotc211.org/ontologies/

[7] https://github.com/opengeospatial/CityGML-3.0Encodings/tree/master/CityGML/Examples/Versioning

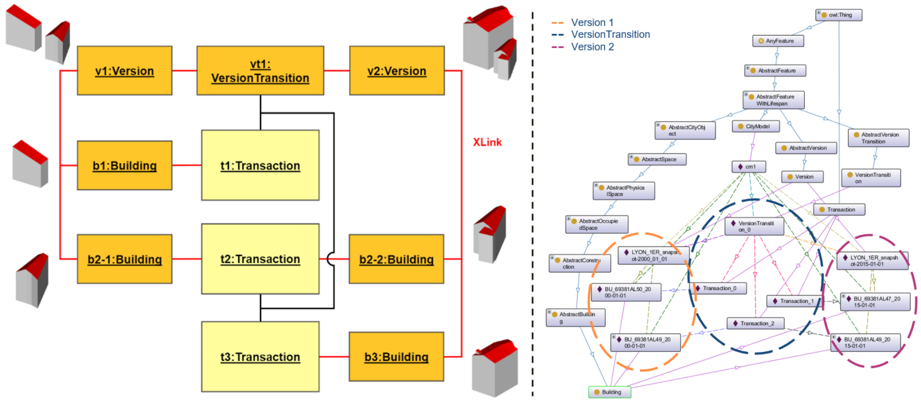

Fig. 2. 3 building historical successions based on the CityGML 3.0 Versioning module; Version graph overview[7] (left); Ontograf visualization of generated RDF graph produced by UD-Graph [26] (right)

Reproducibility

This work makes use of open data, especially city models of Lyon in CityGML format[8]. We also use SoftwareHeritage persistent identifiers[9] for sharing code and data presented in table 1.

S.No.	Name	SWHID
1.	(Transformation tool) UD-Graph	https://github.com/VCityTeam/UD-Graph
2.	Proposed model	https://github.com/VCityTeam/UD-Graph/tree/master/Ontologies/CityGML/3.0
3.	Generated datasets	https://github.com/VCityTeam/UD-Graph/tree/master/Transformations/test-data/RDF

Table 1. Reproducibility: URLs of code, proposed data model, and produced data

7 Conclusions and Lessons Learned

This work explores a method for model-centric data integration with semantic web technologies based on standards to create interoperable data models and improve data interoperability of nD urban linked-open-data. By using semi-automated mapping transformations, as proposed in UD-Graph [25,26], existing

[8] https://data.grandlyon.com/
[9] https://archive.softwareheritage.org/swh:1:dir:222696a70325b62af1b728b20714a5a0e0d415d9

nD urban data standard data models and data can be combined with existing spatio-temporal semantic web standards.

Currently, using UML models appears to be a favorable approach for creating OWL ontologies [26] due to the similarity in modeling concepts between UML and OWL, but a more formal evaluation of this approach is needed to measure how much semantic data may be lost in the case of urban data standards which present a large divergence between their conceptual and physical data models, and thus more human intervention is required to converge on a complete mapping. Additionally, this approach requires testing with additional data models aside from CityGML, such as IndoorGML[10], or BIM-IFC and profiling the data volume and query execution time of the datasets generated.

Once this work is concluded, the proposed approach can be implemented to integrate heterogeneous urban data sources for enriching nD city models, urban digital twins, and smart city applications.

Acknowledgments

This research is funded by the Université Lumière Lyon 2, as a part of the Virtual City Project[11] within the LiRiS[12] laboratory. I would like to thank my thesis advisors Gilles Gesquière, Sylvie Servigne, and John Samuel for their invaluable guidance and support, in addition to Clément Colin and Eric Boix for their insights.

References

1. Beck, S.F., Abualdenien, J., Hijazi, I.H., Borrmann, A., Kolbe, T.H.: Analyzing Contextual Linking of Heterogeneous Information Models from the Domains BIM and UIM. ISPRS International Journal of Geo-Information **10**(12), 807 (Dec 2021). https://doi.org/10.3390/ijgi10120807, number: 12 Publisher: Multidisciplinary Digital Publishing Institute
2. Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A.: Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information **4**(4), 2842–2889 (Dec 2015). https://doi.org/10.3390/ijgi4042842, number: 4 Publisher: Multidisciplinary Digital Publishing Institute
3. 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. pp. 341–350. European Council on Computing in Construction (2019). https://doi.org/10.35490/ec3.2019.166, iSSN: 2684-1150
4. Brink, L., Janssen, P., Quak, W., Stoter, J.: Linking spatial data: automated conversion of geo-information models and GML data to RDF. International Journal of Spatial Data Infrastructures Research **9**, 59–85 (Oct 2014). https://doi.org/10.2902/1725-0463.2014.09.art3

[10] http://indoorgml.net/
[11] https://projet.liris.cnrs.fr/vcity/
[12] https://liris.cnrs.fr/

5. Chaturvedi, K., Kolbe, T.: A Requirement Analysis on Extending Semantic 3D City Models for Supporting Time-Dependent Properties. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-4/W9**, 19–26 (Sep 2019). https://doi.org/10.5194/isprs-annals-IV-4-W9-19-2019
6. Chávez-Feria, S., García-Castro, R., Poveda-Villalón, M.: Converting uml-based ontology conceptualizations to owl with chowlk. In: Verborgh, R., Dimou, A., Hogan, A., d'Amato, C., Tiddi, I., Bröring, A., Mayer, S., Ongenae, F., Tommasini, R., Alam, M. (eds.) The Semantic Web: ESWC 2021 Satellite Events. pp. 44–48. Springer International Publishing, Cham (2021)
7. Chiang, Y.H., Huang, C.Y., Fuse, M.: The integration of ogc sensorthings api and ogc citygml via semantic web technology. In: Di Martino, S., Fang, Z., Li, K.J. (eds.) Web and Wireless Geographical Information Systems. pp. 55–67. Springer International Publishing, Cham (2020)
8. Claramunt, C.: Ontologies for geospatial information: Progress and challenges ahead. Journal of Spatial Information Science (Jun 2020). https://doi.org/10.5311/JOSIS.2020.20.666
9. Daniil, R., Wohlgenannt, G., Pavlov, D., Emelyanov, Y., Mouromtsev, D.: A new tool for linked data visualization and exploration in 3d/vr space. In: Hitzler, P., Kirrane, S., Hartig, O., de Boer, V., Vidal, M.E., Maleshkova, M., Schlobach, S., Hammar, K., Lasierra, N., Stadtmüller, S., Hose, K., Verborgh, R. (eds.) The Semantic Web: ESWC 2019 Satellite Events. pp. 167–171. Springer International Publishing, Cham (2019)
10. Daraio, C., Glänzel, W.: Grand challenges in data integration—state of the art and future perspectives: an introduction. Scientometrics **108**(1), 391–400 (Jul 2016). https://doi.org/10.1007/s11192-016-1914-5
11. De Paepe, D., Thijs, G., Buyle, R., Verborgh, R., Mannens, E.: Automated UML-Based Ontology Generation in OSLO2. In: Lecture Notes in Computer Science. pp. 93–97 (Nov 2017). https://doi.org/10.1007/978-3-319-70407-4_18
12. Hor, A.H., Sohn, G.: Design and Evaluation of a BIM-GIS Integrated information Model Using RDF Graph Database. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. VIII-4-W2-2021, pp. 175–182. Copernicus GmbH (Oct 2021). https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-175-2021, iSSN: 2194-9042
13. Kutzner, T.: Geospatial data modelling and model-driven transformation of geospatial data based on uml profiles. Ph.D. thesis, Technische Universität München (2016)
14. Lefrançois, M., Zimmermann, A., Bakerally, N.: A SPARQL extension for generating RDF from heterogeneous formats. In: Proc. Extended Semantic Web Conference (ESWC'17). Portoroz, Slovenia (May 2017)
15. Mandilaras, G., Koubarakis, M.: Scalable transformation of big geospatial data into linked data. In: International Semantic Web Conference. pp. 480–495. Springer (2021)
16. Métral, C., Falquet, G.: Extension and Contextualization for Linked Semantic 3D Geodata. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. XLII-4-W10, pp. 113–118. Copernicus GmbH (Sep 2018). https://doi.org/10.5194/isprs-archives-XLII-4-W10-113-2018, iSSN: 1682-1750
17. Nishanbaev, I., Champion, E., McMeekin, D.A.: A survey of geospatial semantic web for cultural heritage. Heritage **2**(2), 1471–1498 (2019). https://doi.org/10.3390/heritage2020093

18. Psyllidis, A.: Ontology-Based Data Integration from Heterogeneous Urban Systems: A Knowledge Representation Framework for Smart Cities. In: 14th International Conference on Urban Planning and Urban Management (Jul 2015)
19. Rodriguez-Muro, M., Kontchakov, R., Zakharyaschev, M.: Ontology-based data access: Ontop of databases. In: International Semantic Web Conference. pp. 558–573. Springer (2013)
20. Santos, H., Dantas, V., Furtado, V., Pinheiro, P., McGuinness, D.L.: From data to city indicators: A knowledge graph for supporting automatic generation of dashboards. In: Blomqvist, E., Maynard, D., Gangemi, A., Hoekstra, R., Hitzler, P., Hartig, O. (eds.) The Semantic Web. pp. 94–108. Springer International Publishing, Cham (2017)
21. Shvaiko, P., Euzenat, J.: Ontology matching: State of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering **25**(1), 158–176 (2013). https://doi.org/10.1109/TKDE.2011.253
22. Tran, B.H., Aussenac-Gilles, N., Comparot, C., Trojahn, C.: Semantic Integration of Raster Data for Earth Observation: An RDF Dataset of Territorial Unit Versions with their Land Cover. ISPRS International Journal of Geo-Information **9**(9), 503 (Sep 2020). https://doi.org/10.3390/ijgi9090503, number: 9 Publisher: Multidisciplinary Digital Publishing Institute
23. United Nations, Department of Economic and Social Affairs, P.D.: World urbanization prospects: the 2018 revision. Tech. Rep. ST/ESA/SER.A/420, United Nations, New York (2019), oCLC: 1120698127
24. Usmani, A.U., Jadidi, M., Sohn, G.: Towards the Automatic Ontology Generation and Alignment of BIM and GIS Data Formats. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. VIII-4-W2-2021, pp. 183–188. Copernicus GmbH (Oct 2021). https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-183-2021, iSSN: 2194-9042
25. Vinasco-Alvarez, D., Samuel, J.S., Servigne, S., Gesquière, G.: Towards a semantic web representation from a 3D geospatial urban data model. In: SAGEO 2021, 16ème Conférence Internationale de la Géomatique, de l'Analyse Spatiale et des Sciences de l'Information Géographique. pp. 227–238. Actes de la Conférence SAGEO 2021, La Rochelle [Online Event], France (May 2021)
26. Vinasco-Alvarez, D., Samuel, J.S., Servigne, S., Gesquière, G.: Towards Limiting Semantic Data Loss In 4D Urban Data Semantic Graph Generation. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **VIII-4/W2-2021**, 37–44 (Oct 2021). https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-37-2021
27. Virtanen, J.P., Jaalama, K., Puustinen, T., Julin, A., Hyyppä, J., Hyyppä, H.: Near real-time semantic view analysis of 3d city models in web browser. ISPRS International Journal of Geo-Information **10**(3) (2021). https://doi.org/10.3390/ijgi10030138
28. Yu, J., Thom, J.A., Tam, A.: Requirements-oriented methodology for evaluating ontologies. Information Systems **34**(8), 766–791 (2009)
29. Zhang, C., Beetz, J., Vries, d.: BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data. Semantic Web pp. 1–27 (Aug 2018). https://doi.org/10.3233/SW-180297