AGILE: GIScience Series, 5, 3, 2024. https://doi.org/10.5194/agile-giss-5-3-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

AGILE Association of Geographic Information Laboratories in Europe

A model-driven methodology for integrating heterogeneous 3D geospatial urban entities

Clement Colin[1,4], Diego Vinasco-Alvarez[1], John Samuel[2], Sylvie Servigne[3], Christophe Bortolaso[4], and Gilles Gesquière[1]

[1]Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69676 Bron, France
[2]Univ Lyon, CPE Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69621 Villeurbanne, France
[3]Univ Lyon, INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69621 Villeurbanne, France
[4]Berger-Levrault, Limonest, France

Correspondence: Clement Colin (clement.colin@liris.cnrs.fr); Diego Vinasco-Alvarez (diego.vinasco-alvarez@liris.cnrs.fr)

Abstract. Data on geospatial entities is increasingly available from various fields, such as GIS, CIM (City Information Modeling), and BIM (Building Information Modeling). They are described using different data models, such as IFC and CityGML, and are composed of both semantic and geometric data. Integrating this heterogeneous data to create a unified view of geospatial entities requires the use of technologies and methods adapted to each type of data. In this work, we propose a data integration methodology that leverages geospatial and semantic web technologies to provide data views suited to facilitating spatial and semantic navigation of these data. Concerning semantic data, we propose a model-driven data transformation process for limiting data loss during integration and reducing semantic heterogeneity between data from different urban information domains. For geometric data, we propose data transformations towards data formats for efficiently sharing, storing and visualizing 2D and 3D data. An open-source application is presented to illustrate this methodology.

Keywords. BIM, GIS, CIM, 2D-3D city models, Integration, Methodology, Reproducibility

1 Introduction

As the production of urban geospatial data grows (Lei et al., 2023b; Batty, 2021), providing users (urban planners, maintenance workers/managers, citizens, etc.) with unified views of such data from different sources can improve access to enriched information and enhance understanding of urban objects and landscapes. Heterogeneous geospatial data can provide various information to describe **geospatial entities** (*things that have separate and distinct existences and objective or conceptual reality* (ISO, 2016)) physically, functionally and operationally. Recently, approaches such as Urban Digital Twins (Batty, 2018) have seen an increase in adoption as tools that integrate, visualize, and analysing the complex, evolving geospatial urban data.

These urban geospatial data are produced from heterogeneous information domains, such as GIS (Geospatial Information System), CIM (City Information Model) and BIM (Building Information Model) (Beck et al., 2021; Liu et al., 2021). These domains provide thematic and spatial (topology and location) information about urban entities using semantic and 2D or 3D geometric data. BIM includes information about a building's geometry, materials, and even operational data.

Navigating the context of an urban entity, i.e. any information that can be used to characterize the situation of an entity (Abowd et al., 1999), can help these users achieve a more complete understanding of the city. Additionally, integrating heterogeneous geospatial data would help to create more complete and informed views of an entity to help in its definition and comprehension by taking into account its context (Colin et al., 2022).

Being able to create views of all kinds of geospatial data (2D, 3D, and semantic data) can be essential for decision-making in urban development projects which may affect many citizens. Combining semantic web and computer graphics methodologies and technologies could contribute innovative solutions for the integration and comprehension of heterogeneous data.

Naturally, interoperability problems arise when trying to integrate data from different urban information domains, as they are produced according to various needs and are often structured according to different heterogeneous data models (Weil et al., 2023). This has been identified as a key barrier to the adoption of Urban Digital Twin applications (Lei et al., 2023a). For example, GIS is often used to describe objects at an urban scale, like roads and buildings, BIM was created to describe objects at a building scale, like walls and windows. This means that the same object can be represented by simple or complex geometries with varied details of description. For example, a building may be represented by a 3D model in a CIM context, while a more detailed representation is required in a BIM context. Integrating representations at different scales could help enrich and facilitate the resulting 3D visualization. For example by using the simplest 3D model when the entity is far away from the perspective of the user.

In order to represent urban entities and facilitate seamless interoperability, several data standards have been created depending on different urban information domains and their different point of view (Mcglinn et al., 2019; Guyo et al., 2021). These standards may provide data models and schemas to define the concepts and data structures that constitute the data they specify. For instance, the IFC (Industry Foundation Classes) standard (ISO, 2018) is employed in the BIM domain to depict physical infrastructure like roads, bridges, buildings, and their constituent elements such as walls, pipes, and windows. Buildings and their components can also be described by using DWG (AutoCAD Drawing file) or IndoorGML (OGC, 2018c) files, which encompass semantic and 2D or 3D geometries. To represent urban data in GIS, CityGML (OGC, 2018a) and CityJSON (OGC, 2018b) use 2D or 3D geometric objects, along with thematic data and interrelationships between objects (Kutzner et al., 2022). GeoJSON, another GIS standard, is employed to represent 2D geometries of geographic features, including their attributes and spatial extents (Butler et al., 2016).

Furthermore, as 3D models (or digital twins) of geospatial entities are also usually composed of semantic and geometric data, achieving complete integration of these models would facilitate and enhance both visualization and navigation of these types of data for users. In other words, this means giving access to relevant semantic data while being able to visualize the 2D or 3D geometry in a 3D scene. While approaches based on knowledge graphs have been proposed for storing semantic urban data in interoperable data formats, a generic model for supporting 3D geometry is currently in development (Wagner et al., 2020). **Fully integrating heterogeneous 3D models is a problem that requires handling both semantic data heterogeneity and distinct computer graphics problems.**

According to Kolbe et al. (2020), with many applications requiring data about the same object, the ability to integrate various data sources could help find the missing information. There may exist multiple different models of the same entity, which means different simplifications and descriptions of it (Batty, 2018, 2021). **Giving access to multiple representations of the same geospatial entity is a rising problem that needs to be solved to ensure its better comprehension.**

Our objective is to provide users with integrated and navigable views of available representation of urban objects through a data integration process (Tran et al., 2016). This process must support the different navigation needs or use cases. For instance, a user may only need to provide access to semantic urban data to analyze an entity; sometimes a user may only want to visualize the geometry of an entity; sometimes access, visualize, and navigate both semantic and geometric representations. Therefore, there is also a need to provide users or systems with a generically applicable integration methodology.

We propose a generic geospatial urban data integration methodology to provide more complete and effective spatial and thematic navigation between heterogeneous representations of a geospatial entity. We aim to make the integration more complete by allowing users or systems to navigate all original data. Therefore, data must not be lost (or data loss must be limited) during integration. Also, it should be possible to explore and visualize both semantic and geometric data. We aim to make the integration more effective by using technologies and data formats that are designed for storing, querying and visualizing geometric and semantic data.

We present the following contributions in this paper for solving these challenges:

- A generic geospatial urban data integration methodology:
  - A model-driven transformation of 3D geospatial semantic data models and data towards an efficient and common knowledge graph format
  - Transformation of geometric data towards an efficient and common 3D format
  - An approach for linking heterogeneous representations of geospatial entities
- An approach based on the reuse of standards and open-source tools
- The development of a reproducible data integration pipeline

A preliminary background and explanation of current works in geospatial data integration are presented in section 2. Section 3 presents the proposed methodology and the technical architecture of the platform created to demonstrate it. We also explain how the platform can be fully reproduced, using open-source tools and available data. Section 4 presents the results obtained and section 5 concludes the article and also presents the future course of action.

2 Related Works

Solutions to solve geospatial data integration problem, i.e, combining data from different sources to provide users a unified view of the data (Tran et al., 2016), can be classified into five categories (Fosu et al., 2015): conversion (Stouffs et al., 2018; Chen et al., 2018; Donkers, 2013), creation of a unified data model to represent all objects of multiple data models (El-Mekawy et al., 2012; Yan et al., 2021; Kumar et al.), data integration using semantic web and linked data (Karan et al., 2016; Hor et al., 2016; Huang et al., 2020; Pauwels et al.; Radulovic et al.). The last category is integration for web visualization in a 3D geospatial context (La Guardia et al.; Gaillard et al., 2015; Colin et al., 2022). To ensure that our proposed methodology is generic and can be applied to any data, both semantic web integration methodology for semantic data and integration for web visualization for geometric data need to be used.

2.1 Semantic data integration

Transformation towards formal knowledge graph standards, such as GeoSPARQL, stRDF, and BimSPARQL, facilitate urban data integration for enriching semantic urban data and data models. These standards are useful for providing rich data structures for storing and combining heterogeneous semantic data (Malinverni et al., 2020) and can provide a basis for exchanging data across urban information domains in industry and academia (Claramunt, 2020), both of which are essential for heterogeneous data integration. Achieving such a transformation ensures that the semantic data is described using a common, interoperable language, making it easier to store, share, and query.

Recent works have proposed using model-driven transformations as an approach to integrate geospatial and urban data conformant to these standards (Kyzirakos et al., 2018; Alvarez et al., 2021; Hor et al., 2016). Where data models give data a standardized, shareable definition and structure which provides an essential basis for interoperability. These data models can be formalized at different levels of abstraction. Often, more abstract, technologically independent data models are defined in languages such as UML, while more concrete data models can take the form of physical data schema that take into consideration specific programing languages, data structures and data (transfer) formats. Common languages for defining physical data schema include data description languages such as SQL for defining the structure of relational databases, and JSON Schema and XML Schema for defining semi-structured data. Model-driven transformation approaches use these models to guide the transformation of their conforming data towards other data formats or data models (Kutzner, 2016). Additionally, these transformations can be applied to data models themselves to transform them to knowledge graph formats such as RDF/OWL (Jetlund et al., 2019). Here, OWL is a language used to define highly formal, but machine-readable data models as computational ontologies.

The use of a model-driven transformation approaches towards knowledge graphs formats also allows users to reuse existing geospatial urban data knowledge graphs. Approaches such as ontology alignment (sometimes referred to as ontology matching) are powerful methods for integrating data models and data instances (Euzenat and Shvaiko; Usmani et al., 2021). This involves proposing links or *correspondences* between the concepts, relationships, or data instances of two ontologies. The set of correspondences between the concepts of two ontologies is an *alignment* between the ontologies (figure 1). Several works (Hbeich et al., 2020; Usmani et al., 2021) are underway to align BIM and GIS data through the use of these integration approaches.

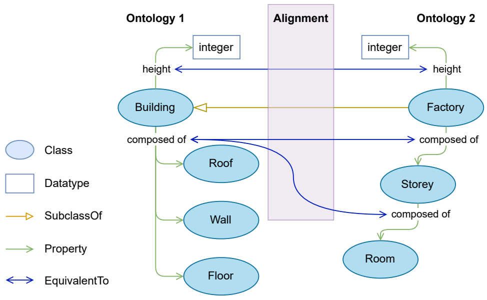


Figure 1. An illustration of an alignment (highlighted in red) between two example ontologies. The alignment is composed of several correspondences (in blue) between the equivalent classes, properties, and datatypes of each ontology.

2.2 Geometric data integration

Services and standards emerged during the last decade to make 2D and 3D data integration easier on the web (Guyo et al., 2021; Mcglinn et al., 2019). The Open Geospatial Consortium (OGC) has defined several standards to access geospatial data. The two main services, the Web Feature Services (WFS) (OGC, 2014) and the Web Map Services (WMS) (OGC, 2006) are useful for storing and sharing 2D geometry of features. Features, defined as an abstraction of a real-world phenomenon (ISO, 2014) are fundamental in both these services. To deliver 3D geospatial data, the OGC also adopted the 3DTiles (OGC, 2019) and the I3S (OGC, 2017) standards. They share similarities as they allow sharing, visualizing, and interacting with massive heterogeneous 3D geospatial content as 3D polygonal models or as 3D point clouds linked to semantic data. More specifically, 3DTiles uses the GLTF format (Khronos, 2021) to store and share efficiently 2D and 3D geometry on the web. A conceptual model, Gen3DCity (Jaillot et al., 2021) generalizes these standards. Those standards were used by several works (Gaillard et al., 2015; Chen et al., 2018; Zhang et al.; La Guardia et al.) to integrate geometric data of one data model, either CityGML or IFC. Works like (Marnat et al., 2022) allow transforming the geometric data of heterogeneous data models to 3DTiles. The OGC also proposes a standardized API, called 3D GeoVolumes (Miller et al., 2020b, a), to navigate in several representations on the same entity. A GeoVolume represents an entity with a distinct bounding volume, containing 3D model datasets that are relevant to that volume (items, content). It can include or reference other GeoVolume children whose bounding volumes are fully contained by the parent container's bounding volume.

However, these standards are dedicated to geometric data, with comparatively limited support for semantic data access and analysis. Improving this support may require answering various questions such as: do all semantic data need to be stored with a geometric format? If yes, how can they be stored? With what kind of structuration? Is querying the semantic data simple and efficient?

2.3 Hybrid data integration

In a previous paper, Colin et al. (2022) proposed a generic methodology to integrate heterogeneous geometry of 2D/3D models on the web while keeping a link to the sources to gather additional semantic data when required. Hor and Sohn (2021) described an approach to integrate BIM semantic data in parallel with geometric data using BimServer, which also focuses on respecting the data models of the data to be integrated. Another solution (Hijazi et al., 2020) allows navigating between BIM and GIS representation by linking instances of the same real-world object between two databases, 3DCityDB and BimServer. Another example of linking heterogeneous representations is the use of a MultimediaDB (Jaillot, 2020) to link CityGML representations and multimedia.

However, the few hybrid data approaches that exist either:

- do not take into account geometric visualization, sharing and storing problems
- are model specific, thus they are not generic as they can not be applied to other data models
- do not provide an efficient solution to link the heterogeneous representation of the same entity.

Wagner et al. (2020) studied four existing approaches to describe the geometric representation of BIM model in a Semantic Web context. The first two can be implemented using only Semantic Web standards, the second being more restricted as the storage and structure of the geometry description are compliant with both RDF and JSON. The last two approaches rely on other technologies than with semantic web to structure and store the geometry. The third uses a Semantic Web approach for linking and storing the geometry descriptions while the fourth only uses RDF for linking to the geometry descriptions and depends on other technologies for storage and structure of the content. Wagner et al. (2020) concludes that storing geometric data with adapted data models and linking them to the semantic data using semantic web technologies is:

- flexible: how flexible users are in choosing their most suitable geometry schema or linking method
- concise: the verbosity of an approach, the triple count or file size of geometry description
- supported: how well an approach is already supported by software applications
- less portable: how convenient it is to share geometry descriptions with other RDF data
- less semantically expressive: Semantic Web Technologies can not be used for spatial querying.

The fourth method is particularly well-suited to our problem, as storing heterogeneous geometric data using a homogeneous data model greatly simplifies their reuse and support. It also increases the conciseness of the data, which is necessary in our context as the geometry of BIM or CIM models is generally heavy, given the number and detail of the object to be represented.

However, progress is still required to achieve an integrated view of heterogeneous representations of the same entity in the same context that allows efficient navigation in both the 2D and 3D geometric data and the semantic data. The following section details our methodological contributions to these works.

3 Geospatial data integration methodology

This section presents the proposed methodology for integrating geospatial data (section 3.1), different urban data integration processes supported by our methodology (section 3.2), and a platform created using the proposed methodology for integrating and navigating multiple representations of an entity (section 3.3).

3.1 Methodology

Our proposed methodology for geospatial data integration, in figure 2, is based on the following approach: separating the semantic data of 2D-3D city models from the geometric data and store these data in formats specialized to make full use of their unique characteristics. Transformation and translation towards efficient data models and formats for each of these categories of data more easily supports the application of data storage, access, and integration methods from the semantic web and computer graphics domains. I.e., recognizing multiple representations of the same object would be possible using Semantic Web entity linking and Geospatial Entity Resolution methods as discussed in section 2. Additionally, we propose to remove redundant data in formats that are not designed for a type of data. More specifically, we propose to remove semantic data from geometric data stored in 3D formats and remove 3D data from data stored in graph formats.

As the semantic data is translated using a model-driven approach (to ensure the validity and completion of the data) into a single knowledge graph format, **the first step** in the methodology involves integrating the relevant data models by translating them into a unique and common formal graph language, also known as an ontological language, which is then stored in a knowledge base. The aim is to increase interoperability between data models and enable efficient use across various applications and BIM/CIM information domains. This step is only necessary for data models that do not already have an official representation in the identified ontological language.

The second step involves transforming and translating the data themselves into formats that are efficient for each specific type of data (spatial or semantic). The initial instantiation of these data is usually done by different users or stakeholders to create 2D/3D city models as open urban data, which we can utilize to test our proposed approach. Semantic and geometric data are separated and transformed into a common standardized graph format and a common standardized 2D/3D data format, respectively. Here, model-driven transformations and the data models integrated in step 1 are used to effectuate the transformation of semantic data. We propose to extract and transform all geometric data into 2D and 3D data model formats that are efficient for sharing, visualizing and interacting with it. During this process, a link between the semantic data and the geometric data must be created to ensure that navigation between them is possible.

Once these steps are achieved, we have at our disposal two datastores: a semantic datastore, a knowledge base containing the data models and semantic data instances, and a geometric datastore, containing all geometries in a common format, made for a more optimal integration in terms of data volume and access.

Our methodology can be categorized under the fourth type of approach of BIM Model geometry integration approaches described by (Wagner et al., 2020), as the geometry is stored in an adapted format while the semantic and the link with the geometry are stored separately using semantic web technology. However, a notable distinction is our methodology is generically applicable to geometry from larger urban scales such as CIM and GIS data.

Our methodology allows for easy access to the data and enables users to analyze, explore, and visualize the data in various ways, depending on their requirements. Overall, this methodology enables efficient and standardized use of geospatial data by users and systems while limiting data loss. Also, by using technologies and data formats that are designed for storing, querying and visualizing geometric and semantic data, we ensure a more optimal integration.

3.2 Applying geospatial and semantic data integration processes

Once the geospatial and semantic data are separated into their relative formats and storage solutions, additional steps are taken to make use of this integration methodology to take geospatial data integration a step further as shown in figure 3. The first step of this activity diagram involves identifying the existing representations of the same unique entity in heterogeneous data which can be done by using semantic entity linking or geospatial entity linking techniques as introduced in section 2.

Next, the transformed data models can be aligned using the approaches discussed in section 2 to minimize semantic heterogeneity between BIM and CIM information domains. After entity linking and ontology alignment, the formal nature of ontologies permits the use of tools such as reasoners to verify that the concepts and relationships defined by these translated data models are logically consistent, and logically infer new relations between concepts in the integrated data models and between representations of entities.

3.3 Experimentation and reproducibility

This section presents the technical architecture of the platform created for the experiment to test the aforementioned methodology and how it can be reproduced.

Figure 4 shows the sequences of activities at the technical level of our methodology of geospatial data integration. In the experiment, we used two standards of geospatial urban data: IFC and CityGML from the BIM and CIM information domains respectively. For the first step of data model translation into a graph, we used the ShapeChange tool to transform the CityGML 2.0 and 3.0 data model to RDF, and store it in a graph database, Blazegraph. The IFC standard ontology is publicly available[1], therefore it does not need any transformation before being added to the graph database.

Following step two, the geometric data are homogenized and transformed as 3DTiles (OGC, 2019), a standard that allows sharing, visualization, and interaction with massive heterogeneous 3D geospatial content. More specifically, 3DTiles uses the GLTF format (Khronos, 2021) to store and share efficiently 2D and 3D geometry on the web. This transformation was handled by a tool named py3dtilers (Marnat et al., 2022). It is an open-source tool to convert and manipulate 3D Tiles from the most common 3D geospatial data models: CityGML, IFC, OBJ, and GeoJSON. Next, the semantic data transformation to graph was made possible using the *UD-Graph* (Alvarez et al., 2021) and IFCtoRDF tools, which produce RDF files that can be uploaded to the graph database. The link between the semantic and the geometric representation is ensured

---

[1] https://technical.buildingsmart.org/standards/IFC/IFC-formats/IFCowl/

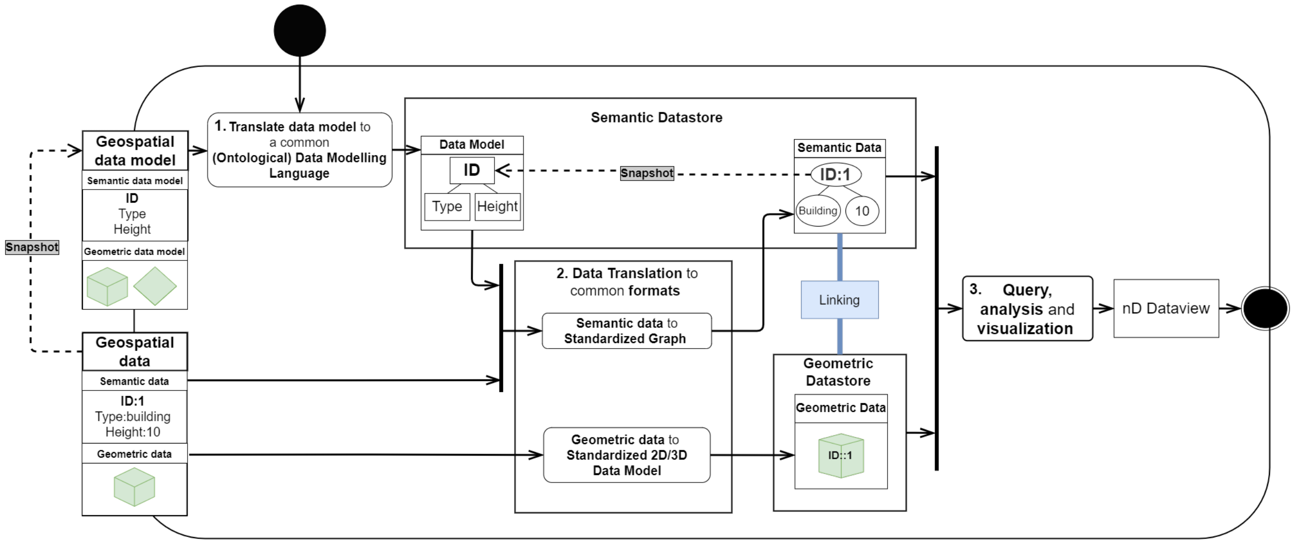


Figure 2. UL Activity diagram of the proposed methodology at the macro level

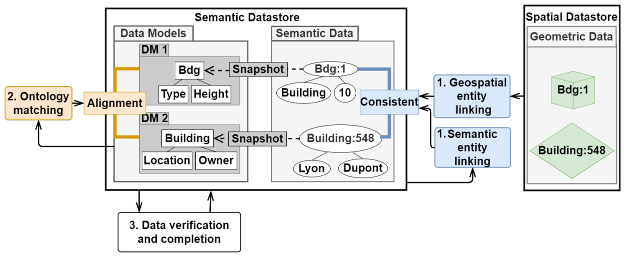


Figure 3. Example of the applicable geospatial data integration processes where the geospatial entities ‘Bdg:1’ and ‘Building:548’ are two identifiers referring to different representations of the same real-world entity.

using the GeoVolume model, by linking those representations to the entity they represent. This step was ensured by hand in this experiment.

Finally, the navigation and visualization of the geospatial data are ensured by *UD-Viz*, which allows the creation of web applications for visualizing and interacting with geospatial urban data.

All tools used to create this platform are open-source and accessible online. They are listed in table 2 in section 6. Some (*UD-Viz, Py3DTilers, UD-Graph*) were developed as part of the urban data services and visualization framework (Samuel et al., 2023), an open-source framework for multidisciplinary research to handle complex processing, analysis, and visualization of urban data. One major advantage of using multiple components is that, if needed, each component may be replaced by another that will fulfill the same role in this methodology.

The presented platform is fully reproducible, a link to the repository is available in table 2. The repository holds information to reproduce the results shown, the data created and how to recreate it. It is also accessible online[2].

4 Results

This section shows some results obtained by applying the presented methodology in the aforementioned platform as a geospatial urban data web application, using the technical architecture 3.3 presented above. The navigation in both geometric and semantic representations of a same entity is presented, followed by a study of the efficiency of the methodology.

4.1 Navigation

The demonstration makes it possible to navigate in a 3D geospatial environment created from geometric and semantic data of heterogeneous sources described in part 4.2. The geometry was extracted and then stored as 3DTiles and semantic data was transformed and then stored as an RDF graph. As all representations of the same entity are linked using the GeoVolume model, it is possible to navigate between them. As shown in figure 5, all representations linked to an entity (here, a building in the Doua district) are queried using SPARQL and listed on the left menu. The bounding box of the entity is shown in the geospatial environment to help locate it. The visualization is adapted depending on the type of data: geometric data can be visualized in the geospatial environment, as shown in figure 6, and semantic data can be explored through a graph visualization of it, as shown in 7.

---

[2] https://geodatanavigation.vcityliris.data.alpha.grandlyon.com/

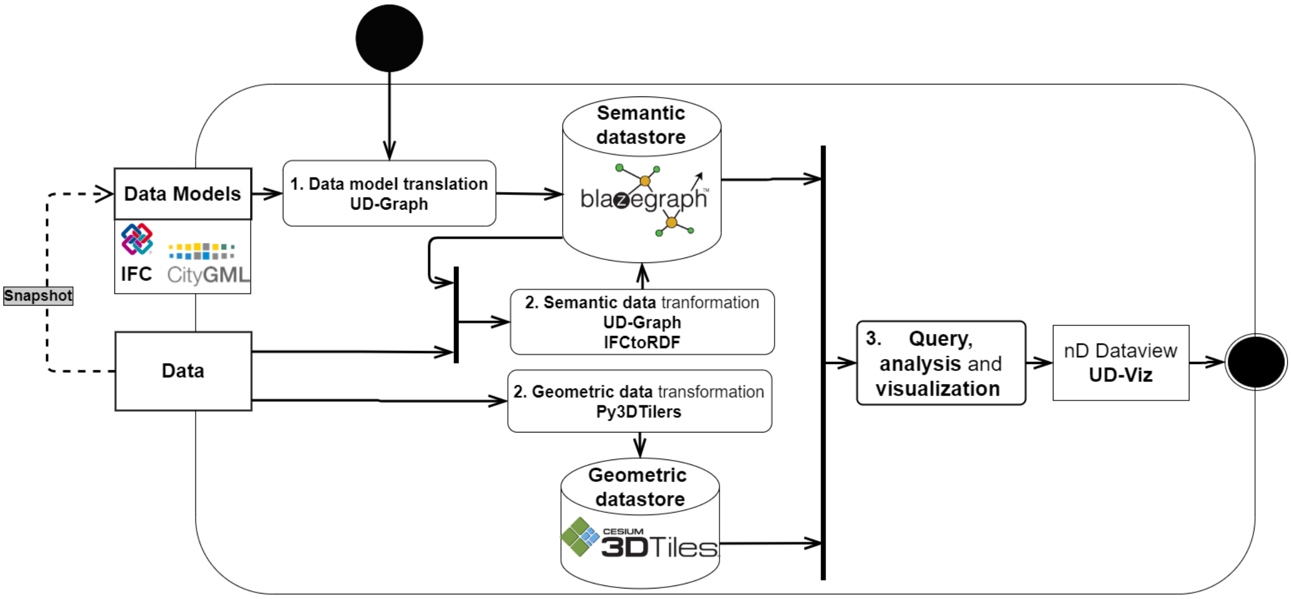
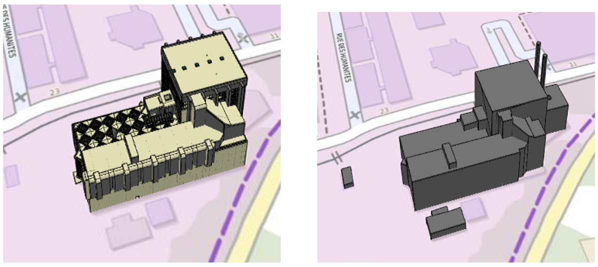


Figure 4. UML Activity diagram of our proposed methodology at the technical level, using open-source tools and standards.

Figure 5. List of the available representations (on the left menu) of a building in the Doua district with its bounding box used to locate it on a geospatial environment (on the right)

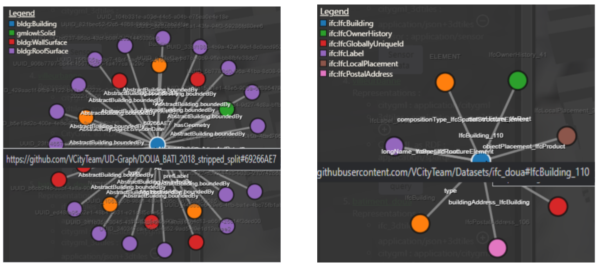
Thanks to the use of the homogeneous 3D format, 3DTiles, it is possible to visualize geometric data from a wide variety of 2D and 3D data standards. In figure 6, two representations of the same building are shown. They were transformed from a CityGML and an IFC file, respectively.



(a) (b)

Figure 6. Visualization of two geometric representations in 3DTiles from an IFC and a CityGML file of the same building

Semantic data that have been transformed into a graph can also be browsed (figure 7). The parent node that corresponds to the entity is shown in blue. Each edge represents a specific relationship with said entity. Additional information regarding each node and the relationships it can have with other nodes is described by their integrated data model (i.e., an instance of a `bldg:Building` in CityGML is linked to the class of buildings and its documentation). Through these graph interfaces, semantic data can be explored freely by the user.



(a) (b)

Figure 7. Visualization of two semantic representations as graph extracted from an IFC and a CityGML file

Figure 8 illustrates how simultaneous navigation of both geometric and semantic representations of an entity can be effectuated. On the right is a navigable graph that shows the semantic data issued from an IFC file of this building. In the middle, a 3D geometric representation issued from a CityGML file of this building. By using links created in the semantic datastore between these representations, upon interaction with a node that has a geometric representation in the geospatial environment, its geometry is highlighted in blue.

4.2 Data

Both geometric and semantic data are available by accessing the relevant service: the 3DTiles files in the geome-

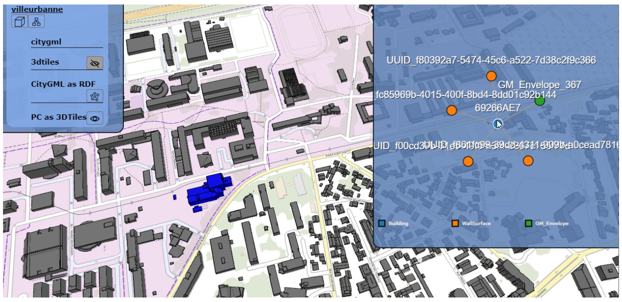


Figure 8. Visualization of a geometric (highlighted in blue) and a semantic representation (transparent blue frame on the right) of the same building extracted from a CityGML file. The geometry of the building was highlighted thanks to the link created between those representations.

try datastore using HTTP requests and the semantic RDF graph using the semantic datastore and SPARQL. As we used an adaptable model-driven data integration method and officially supported transformation tools, we ensured that no semantic data was lost during the transformation. The geometry is visualized and spatially analyzed using 3DTiles, which RDF and SPARQL have comparatively limited support for, especially concerning 3D models.

By using a semantic datastore, it is possible to link heterogeneous representations of the same entity, without altering the original data and data models. In this demonstration, the geometric and semantic data issued from a geospatial representation are linked using the GeoVolume model, which allows linking all representations of the same entity while geolocating it.

Table 1 shows the size of the integrated IFC and CityGML files before and after transformation. We can see that removing the geometry from the graphs greatly reduces their size. We can also note that the geometry, stored as 3DTiles, is still lighter than a graph, and is usable and optimized for this kind of data. Moreover, as the format of the geometry is homogenized, all geometric representations can be visualized, whatever the original data model they conformed to. By using optimized solutions for each kind of data, the solution is more efficient as the geometry is not transformed as a graph.

Another major advantage of this methodology is its flexibility: it is possible to only use the semantic or the geometric data, depending on the use case. A user may require an application that only uses heterogeneous semantic data for querying and analyzing, another may only need to visualize the geometric representation in a 3D environment.

The geometry from any data model may be stored in a homogeneous and optimized format meant to visualize or analyze geometry. It helps to gain portability (Wagner et al., 2020), i.e., how convenient it is to share geometry descriptions.

File	Format	Original	TTL	TTL without geometry	3DTiles
FZK-LOD4	CGML	15.9 mb	7.4 mb	39 kb	39.3 kb
Doua district	CGML	23.2 mb	20.7 mb	8.1 mb	5.5 mb
Limonest	CGML	56.6 mb	59.2 mb	15 mb	8.8 mb
FZK	IFC	2.4 mb	20.4 mb	15.9 mb	2.07 mb
AC20 Institute	IFC	10.4 mb	27.9 mb	24.4 mb	5.30 mb
Doua building	IFC	17.5mb	133.7 mb	71 mb	24 mb
Schependomlaan	IFC	47 mb	216 mb	78 mb	26.2 mb

Table 1. Size of files before and after transformation of semantic and geometric data towards RDF and 3D homogeneous formats.

5 Conclusion

The problem of integrating heterogeneous geospatial data lacks a generic methodology that can be applied to urban data models to visualize both semantic and geometric data that are commonly found in such data. We proposed a methodology to handle both semantic and geometric data by using the most efficient technology for each. They are also stored in a homogeneous adapted format that allows efficient semantic querying and efficient sharing and visualizing of geometric data.

This methodology allows a more complete integration of urban data as we propose to use a model-driven data integration process to avoid loss of data and make it possible to study other data integration problems such as ontology alignment. Additionally, the integrated data is easily navigable, independent of their original data model. The proposed methodology is also efficient as no unnecessary data is stored in an unadapted format, i.e., semantic data are not stored in geometric formats or vice-versa. As demonstrated this makes the semantic graph much lighter.

The navigation of heterogeneous data integrated through this methodology was exemplified using a fully reproducible demonstration based on standards and open-source tools, presented in section 3.3.

Another advantage of this methodology is the possibility of adding additional geospatial data and data models conforming to different data standards. Links between various representations of the same entity can be ensured in the semantic datastore, without altering original data or data models. Navigating all the data is made possible by only using two data stores, whatever the number of data model inputs.

The technical architecture presented in section 3 allows further study of other geospatial data integration approaches such as ontology matching to align data models in future works. As many representations of the same entity can be stored in the same place, semantic, geospatial or hybrid solutions for entity linking can be explored as well. By using those links, it would also be possible to enrich a representation using other representations of the same entity. It would also be possible to study data verification by comparing and analyzing various existing representations. One key problem not handled in this article is data synchronization between the distributed semantic and geometric data stores. Modification of data on one of the databases may imply modification of data in the other. This challenge still needs further study to be handled generically.

6 Data and Software Availability

Research data and code supporting this publication are available in the repositories listed in table 1 and are accessible via the corresponding links in table 1. More particularly, they are issued from the Open Datasets of the Lyon Metropole[3], Building Smart IFC examples[4] and from the Institute for Automation and Applied Computer Science (IAI) / Karlsruhe Institute of Technology (KIT)[5][6] (accessed on Dec 19, 2023). The computational workflow supporting this publication is provided as a container published at 10.5281/zenodo.10810329 with instructions included in the README.md file in the repository.

S.No.	Name	Repository
1.	UD-Viz	https://github.com/VCityTeam/UD-Viz
2.	Py3DTilers	https://github.com/VCityTeam/py3dtilers
3.	BlazeGraph	https://blazegraph.com/
4.	ShapeChange	https://shapechange.net/
5.	UD-Graph	https://github.com/VCityTeam/UD-Graph
6.	IFCToRDF	https://github.com/pipauwel/IFCtoRDF
7.	Platform	https://zenodo.org/doi/10.5281/zenodo.10810329

Table 2. Repositories and links of the open-source components used

7 Acknowledgment

The authors would like to thank both the Université Lumière Lyon 2 and Berger-Levrault for the funding of this project. This work was carried out within the VCity project[7] of the LIRIS laboratory.

References

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P.: Towards a better understanding of context and context-awareness, in: Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1, pp. 304–307, Springer, 1999.

Alvarez, D. V., Samuel, J., Servigne, S., and 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, pp.37–44, https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-37-2021, 2021.

Batty, M.: Digital twins, Environment and Planning B: Urban Analytics and City Science, 45, pp.817–820, https://doi.org/10.1177/2399808318796416, 2018.

Batty, M.: Multiple models, Environment and Planning B: Urban Analytics and City Science, 48, pp.2129–2132, https://doi.org/10.1177/23998083211051139, 2021.

Beck, S. F., Abualdenien, J., Hijazi, I. H., Borrmann, A., and Kolbe, T. H.: Analyzing Contextual Linking of Heterogeneous Information Models from the Domains BIM and UIM, ISPRS International Journal of Geo-Information, 10, 807, https://doi.org/10.3390/ijgi10120807, number: 12 Publisher: Multidisciplinary Digital Publishing Institute, 2021.

Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., and Hagen, S.: The GeoJSON Format, RFC 7946, https://doi.org/10.17487/RFC7946, 2016.

Chen, Y., Shooraj, E., Rajabifard, A., and Sabri, S.: From IFC to 3D Tiles: An integrated Open-Source solution for visualising BIMs on Cesium, ISPRS International Journal of Geo-Information, 7, pp.393, https://doi.org/10.3390/ijgi7100393, 2018.

Claramunt, C.: Ontologies for geospatial information: Progress and challenges ahead, https://doi.org/10.5311/JOSIS.2020.20.666, 2020.

Colin, C., Samuel, J., Servigne, S., Bortolaso, C., and Gesquière, G.: Creating contextual view of CMMS assets using geospatial 2D/3D data, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022, pp.37–44, https://doi.org/10.5194/isprs-annals-X-4-W2-2022-37-2022, 2022.

Donkers, S.: Automatic generation of CityGML LoD3 building models from IFC models, 2013.

El-Mekawy, M., Östman, A., and Hijazi, I.: A Unified Building Model for 3D Urban GIS, ISPRS Inter-

---

[3] https://data.grandlyon.com

[4] https://github.com/buildingSMART/Sample-Test-Files

[5] https://www.citygmlwiki.org/index.php?title=KIT_CityGML_Examples

[6] https://www.IFCwiki.org/index.php?title=KIT_IFC_Examples

[7] https://projet.liris.cnrs.fr/vcity/; online, accessed March 2024

national Journal of Geo-Information, 1, pp.120–145, https://doi.org/10.3390/ijgi1020120, 2012.

Euzenat, J. and Shvaiko, P.: Ontology Matching, Springer Berlin Heidelberg : Imprint: Springer, 2nd ed. 2013 edn.

Fosu, R., Suprabhas, K., Rathore, Z., and Cory, C.: Integration of Building Information Modeling (BIM) and Geographic Information Systems (GIS)–a literature review and future needs, pp. pp.27–29, 2015.

Gaillard, J., Vienne, A., Baume, R., Pedrinis, F., Peytavie, A., and Gesquière, G.: Urban data visualisation in a web browser, in: Proceedings of the 20th International Conference on 3D Web Technology, Web3D '15, pp. pp.81–88, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2775292.2775302, 2015.

Guyo, E., Hartmann, T., and Ungureanu, L.: Interoperability between BIM and GIS through open data standards: An overview of current literature, sign, 3, pp.5–9, 2021.

Hbeich, E., Roxin, A., and Bus, N.: Previous BIM-GIS Integration Approaches: Analytic Review and Discussion, 2020.

Hijazi, I. H., Krauth, T., Donaubauer, A., and Kolbe, T.: 3DCITYDB4BIM: A system architecture for linking Bim Server and 3D CityDB for BIM-GIS integration, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, V-4-2020, pp.195–202, https://doi.org/10.5194/isprs-annals-V-4-2020-195-2020, 2020.

Hor, A.-H. and 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, https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-175-2021, iSSN: 2194-9042, 2021.

Hor, A.-H., Jadidi, A., and Sohn, G.: BIM-GIS Integrated geospatial information model using semantic web and RDF graphs, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, III-4, pp.73–79, https://doi.org/10.5194/isprsannals-III-4-73-2016, 2016.

Huang, W., Olsson, P.-O., Kanters, J., and Harrie, L.: RECONCILING CITY MODELS WITH BIM IN KNOWLEDGE GRAPHS: A FEASIBILITY STUDY OF DATA INTEGRATION FOR SOLAR ENERGY SIMULATION, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. VI-4-W1-2020, pp. 93–99, Copernicus GmbH, https://doi.org/10.5194/isprs-annals-VI-4-W1-2020-93-2020, iSSN: 2194-9042, 2020.

ISO: ISO19101 Geographic information — Reference Model - Part 1: Fundamentals, https://www.iso.org/standard/59164.html, [Online; accessed 23-May-2023], 2014.

ISO: ISO19119 Geographic information — Services, https://www.iso.org/standard/59221.html, [Online; accessed 23-May-2023], 2016.

ISO: ISO70303 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries, https://www.iso.org/standard/70303.html, [Online; accessed 23-May-2023], 2018.

Jaillot, V.: 3D, temporal and documented cities : formalization, visualization and navigation, Theses, Université de Lyon, issue: 2020LYSE2026, 2020.

Jaillot, V., Rigolle, V., Servigne, S., Samuel, J., and Gesquière, G.: Integrating multimedia documents and time-evolving 3D city models for web visualization and navigation, Transactions in GIS, 25, pp.1419–1438, https://doi.org/10.1111/tgis.12734, publisher: Wiley, 2021.

Jetlund, K., Onstein, E., and Huang, L.: Adapted Rules for UML Modelling of Geospatial Information for Model-Driven Implementation as OWL Ontologies, ISPRS International Journal of Geo-Information, 8, https://doi.org/10.3390/ijgi8090365, 2019.

Karan, E. P., Irizarry, J., and Haymaker, J.: BIM and GIS Integration and Interoperability Based on Semantic Web Technology, Journal of Computing in Civil Engineering, 30, 04015 043, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000519, 2016.

Khronos: GLTF Specification, https://www.khronos.org/registry/glTF/specs/2.0/glTF-2.0.pdf, [Online; accessed 23-May-2023], 2021.

Kolbe, T. H., Moshrefzadeh, M., Donaubauer, A., and Chaturvedi, K.: The Data Integration Challenge in Smart City Projects, https://doi.org/10.13140/RG.2.2.21093.22241, publisher: Technische Universität München, 2020.

Kumar, K., Labetski, A., Ohori, K. A., Ledoux, H., and Stoter, J.: The LandInfra standard and its role in solving the BIM-GIS quagmire, Open Geospatial Data, Software and Standards, 4, pp.5, https://doi.org/10.1186/s40965-019-0065-z.

Kutzner, T.: Geospatial data modelling and model-driven transformation of geospatial data based on uml profiles, PhD Thesis, Technische Universität München, https://mediatum.ub.tum.de/1280909?id=1280909&change_language=en, 2016.

Kutzner, T., Chaturvedi, K., and Kolbe, T. H.: CityGML 3.0: New Functions Open Up New Applications, PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88, pp.43–61, https://doi.org/10.1007/s41064-020-00095-z, 2022.

Kyzirakos, K., Savva, D., Vlachopoulos, I., Vasileiou, A., Karalis, N., Koubarakis, M., and Manegold, S.: GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings, Journal of Web Semantics, 52-53, 16–32, https://doi.org/10.1016/j.websem.2018.08.003, 2018.

La Guardia, M., Koeva, M., D'ippolito, F., and Karam, S.: 3d data integration for web based open source webgl interactive visualisation, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W4-2022, pp.89–94, https://doi.org/10.5194/isprs-archives-XLVIII-4-W4-2022-89-2022.

Lei, B., Janssen, P., Stoter, J., and Biljecki, F.: Challenges of urban digital twins: A systematic review and a Delphi expert survey, Automation in Construction, 147, 104 716, https://doi.org/10.1016/j.autcon.2022.104716, 2023a.

Lei, B., Stouffs, R., and Biljecki, F.: Assessing and benchmarking 3D city models, International Journal of Geographical Information Science, pp. pp.788–809, https://doi.org/10.1080/13658816.2022.2140808, 2023b.

Liu, A. H., Ellul, C., and Swiderska, M.: Decision Making in the 4th Dimension—Exploring Use Cases and Technical Options for the Integration of 4D BIM and GIS during Construction, ISPRS International Journal of Geo-Information, 10, 203, https://doi.org/10.3390/ijgi10040203, 2021.

Malinverni, E. S., Naticchia, B., Lerma Garcia, J. L., Gorreja, A., Lopez Uriarte, J., and Di Stefano, F.: A semantic graph database for the interoperability of 3D GIS data, https://doi.org/10.1007/s12518-020-00334-3, 2020.

Marnat, L., Gautier, C., Colin, C., and Gesquière, G.: Py3DTilers: An open source toolkit for creating and managing 2D/3D geospatial data, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W3-2022, 165–172, https://doi.org/10.5194/isprs-annals-X-4-W3-2022-165-2022, 2022.

Mcglinn, D. K., Wagner, A., Pauwels, P., Bonsma, P., Kelly, P., and O'sullivan, D.: Interlinking geospatial and building geometry with existing and developing standards on the web, Automation in Construction, pp. pp.235–250, https://doi.org/10.1016/j.autcon.2018.12.026, 2019.

Miller, T., Trenum, G., and Lieberman, J.: 3D Data Container Engineering Report, https://docs.ogc.org/per/20-029.html, [Online; accessed 23-May-2023], 2020a.

Miller, T., Trenum, G., and Simonis, I.: 3D Data Container and Tiles API Pilot Summary Engineering Report, https://docs.ogc.org/per/20-031.html, [Online; accessed 23-May-2023], 2020b.

OGC: WMS Specification, https://www.ogc.org/standards/wms, [Online; accessed 23-May-2023], 2006.

OGC: WFS Specification, https://www.ogc.org/standards/wfs, [Online; accessed 23-May-2023], 2014.

OGC: I3S Specification, https://www.ogc.org/standards/i3s, [Online; accessed 23-May-2023], 2017.

OGC: CityGML Specification, https://www.ogc.org/standard/citygml/, [Online; accessed 23-May-2023], 2018a.

OGC: CityJSON Specification, https://docs.ogc.org/cs/20-072r2/20-072r2.html, [Online; accessed 23-May-2023], 2018b.

OGC: IndoorGML Specification, https://www.ogc.org/standard/indoorgml/, [Online; accessed 23-May-2023], 2018c.

OGC: 3DTiles Specification, https://www.ogc.org/standards/3DTiles, [Online; accessed 23-May-2023], 2019.

Pauwels, P., Zhang, S., and Lee, Y.-C.: Semantic web technologies in AEC industry: A literature overview, Automation in Construction, 73, pp.145–165, https://doi.org/10.1016/j.autcon.2016.10.003.

Radulovic, F., Poveda-Villalón, M., Vila-Suero, D., Rodríguez-Doncel, V., García-Castro, R., and Gómez-Pérez, A.: Guidelines for Linked Data generation and publication: An example in building energy consumption, Automation in Construction, 57, pp.178–187, https://doi.org/10.1016/j.autcon.2015.04.002.

Samuel, J., Jaillot, V., Colin, C., Alvarez, D. V., Boix, E., Servigne, S., and Gesquière, G.: UD-SV: Urban data services and visualization framework for sharing multidisciplinary research, Transactions in GIS, n/a, https://doi.org/https://doi.org/10.1111/tgis.13049, 2023.

Stouffs, R., Tauscher, H., and Biljecki, F.: Achieving Complete and Near-Lossless Conversion from IFC to CityGML, ISPRS International Journal of Geo-Information, 7, pp.355, https://doi.org/10.3390/ijgi7090355, 2018.

Tran, B.-H., Plumejeaud, C., Bouju, A., and Bretagnolle, V.: Towards a semantic framework for exploiting heterogeneous environmental data, International Journal of Metadata, Semantics and Ontologies, 11, https://doi.org/10.1504/IJMSO.2016.081586, 2016.

Usmani, A. U., Jadidi, M., and 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, https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-183-2021, iSSN: 2194-9042, 2021.

Wagner, A., Bonduel, M., Pauwels, P., and Rüppel, U.: Representing construction-related geometry in a semantic web context: A review of approaches, Automation in Construction, 115, 103 130, https://doi.org/10.1016/j.autcon.2020.103130, 2020.

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., and Alahi, A.: Urban Digital Twin Challenges: A Systematic Review and Perspectives for Sustainable Smart Cities, Sustainable Cities and Society, 99, 104 862, https://doi.org/10.1016/j.scs.2023.104862, 2023.

Yan, J., Zlatanova, S., and Diakité, A.: A unified 3D space-based navigation model for seamless navigation in indoor and outdoor, International Journal of Digital Earth, https://doi.org/10.1080/17538947.2021.1913522, 2021.

Zhang, H., Cheng, C., and Miao, S.: A Precise Urban Component Management Method Based on the GeoSOT Grid Code and BIM, 8, 159, https://doi.org/10.3390/ijgi8030159.