

Towards Semantic Interoperability Between IFC and CityGML Using Semantic Web Technologies

G.R. Sirigireddy  and J. Krischler 

Chair of Intelligent Technical Design, Bauhaus-Universität Weimar Coudraystraße 13A, 99423 Weimar, Germany

E-mail(s): ganeshwar.reddy.sirigireddy@uni-weimar.de, judith.krischler@uni-weimar.de

Abstract: This research investigates the semantic interoperability between two prominent data models: Industry Foundation Classes (IFC), predominantly used in BIM, and City Geography Markup Language (CityGML), commonly employed in GIS. Utilizing Semantic Web technologies and domain-specific ontologies such as ifcOWL, CityGML ontology, and the Building Topology Ontology (BOT), this study proposes a systematic methodology for maintaining semantic consistency across data transformations. An ETL-based approach is employed, converting IFC and CityGML data into RDF representations. A practical use case involving the integration of an IFC-modeled city sewer pipe segment with a CityGML-based building model demonstrates the efficacy and limitations of semantic preservation during data integration. Initial results indicate that semantic content can be preserved, providing essential connections between detailed building information and urban-scale spatial data. This work contributes to the advancement of urban information modeling by facilitating the semantic integration of parametric building models with city-level spatial data.

Keywords: Building information modeling (BIM), Geographic Information Systems (GIS), Extract-Transform-Load (ETL), Resource Description Framework (RDF), Semantic Web Technologies



DOI: 10.18154/RWTH-CONV-254897. Published in the conference proceedings of the 36. Forum Bauinformatik 2025, Aachen, Germany, © 2025 The copyright for this article lies with the authors. This publication, except for quotations and otherwise indicated parts, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1 Introduction

The integration of BIM and GIS has gained importance in managing construction, infrastructure, and urban planning workflows. BIM provides structured, object-based representations of buildings [1], while GIS supports spatial analysis and data management at broader geographic scales [2]. However, differences in semantic models, scales, and data structures create challenges for direct interoperability between the two systems [3].

Two standards commonly associated with these domains are the IFC and the CityGML. IFC is widely used in the AEC sector to describe detailed geometric and semantic properties of building elements [4]. CityGML provides a framework for modeling and exchanging urban features in 3D, supporting multiple levels of detail for different application needs [5]. The need to combine these arises when detailed building-scale information must be analyzed alongside city-scale infrastructure data. For example, an urban planning or municipal engineering office may need to check whether newly planned

sewer pipes conflict with existing building foundations or to coordinate utility networks such as gas, water, or power lines with surrounding structures. In infrastructure projects, a transportation authority may need to evaluate how a planned bridge interacts with surrounding buildings, underground utilities, and road networks, ensuring compliance with clearance, safety, and maintenance requirements. Such analyses require both datasets to be represented in a shared, semantically consistent model that can be queried comprehensively. One of the main challenges in aligning IFC and CityGML is preserving semantic meaning during data exchange, as conventional methods often lead to information loss [6]. Semantic Web technologies, such as RDF and domain ontologies, enable the representation and linking of heterogeneous datasets while supporting querying and reasoning [7].

This research addresses semantic interoperability between IFC and CityGML through an ontology-driven approach using ifcOWL, the CityGML ontology, and BOT. An ETL process transforms data into RDF format, guided by these domain-specific ontologies to ensure semantic preservation and meaningful integration.

The approach is demonstrated through a case study integrating an IFC-modeled city sewer pipe segment with a CityGML-based building model. Results reveal both the potential and the challenges of preserving semantics in cross-domain integration, validating the proposed ETL-based RDF framework for linking infrastructure element and building data.

2 Background

Building on the foundational goal of semantic interoperability between BIM and GIS, ontology driven methods provide a promising framework to represent, integrate, and query heterogeneous datasets while preserving their semantic richness. Ontologies, which are formal models of domain concepts and their interrelationships, are central to this approach. They provide a shared, machine-readable vocabulary to resolve semantic ambiguities and enable automated reasoning. Complementing these, the RDF establishes a universal data model that encodes information as subject predicate object triples using Uniform Resource Identifiers (URIs), thereby decoupling data from original source schemas and enabling integration.

Most of the integration processes typically begin with data instantiation, wherein diverse data sources, including BIM models and GIS features, are converted into RDF triples to form a unified semantic knowledge graph [8]. For BIM, IFC is translated into ontological representations such as ifcOWL, which capture detailed semantic properties of building elements in compliance with the IFC schema [9]. Similarly, the BOT abstracts fundamental structural concepts like spaces and zones, offering consistent semantic constructs that bridge various BIM datasets [8]. This step ensures that building data is represented in a form conducive to semantic linking and inference.

In parallel, GIS datasets such as CityGML and Shapefiles are mapped onto GeoSPARQL, an OGC standard ontology designed to represent geographic features and their spatial relationships semantically [8], [9]. Although these mappings incorporate spatial context, the focus remains on maintaining semantic coherence by classifying urban elements meaningfully, such as designating pedestrian pathways or land use zones [8]. This complementary semantic encoding of BIM and GIS data establishes the groundwork for integrated urban knowledge.

Further, to reconcile semantic heterogeneity across these domains, ontology mapping employs the Semantic Web Rule Language (SWRL) to create rules that define equivalences, conditional relationships, and property mappings [8], [9]. For example, SWRL rules can align an IFC building concept with its counterpart in BOT or GeoSPARQL, facilitating consistent interpretation across datasets. Moreover, declaring BOT classes as subclasses of GeoSPARQL features enables property inheritance, thereby strengthening semantic connections between building and geographic domains [8].

At the instance level, semantic enrichment continues through the use of GeoSPARQL's topological functions, such as spatial intersections or containment, which automatically infer spatial relationships based on existing geometry references [8]. These inferred relations are appended to the knowledge graph as new semantic assertions, effectively linking indoor and outdoor spatial contexts. Additionally, the `owl:sameAs` property resolves co reference by linking instances representing the same real-world entities across disparate datasets [8].

Reasoning engines such as Apache Jena and Protégé apply these SWRL mappings and ontological definitions to derive new knowledge and implicit relationships that are not explicitly stated in the source data [9], [10]. Ensuring compatibility with reasoning mechanisms sometimes requires redefining certain properties; for instance, converting annotation properties to datatype properties to facilitate effective inference [8].

Together, these stages culminate in the creation of an integrated, semantically rich RDF knowledge base stored in triplestores like Jena TDB or Ontotext GraphDB. This knowledge base supports complex querying and retrieval through SPARQL, empowering urban management and infrastructure applications with a consistent semantic foundation [9], [10].

3 Methodology

This study presents a structured methodology for the semantic integration of heterogeneous building and urban data models, specifically those based on IFC and CityGML. The integration framework is designed around an ontology-driven ETL pipeline that converts and aligns both datasets into a unified RDF graph representation. The overall process, depicted in Figure 1, comprises three major phases: *file processing*, *semantic conversion*, and *graph-based integration*. Each phase contributes to the transformation of domain-specific datasets into semantically aligned, queryable knowledge structures.

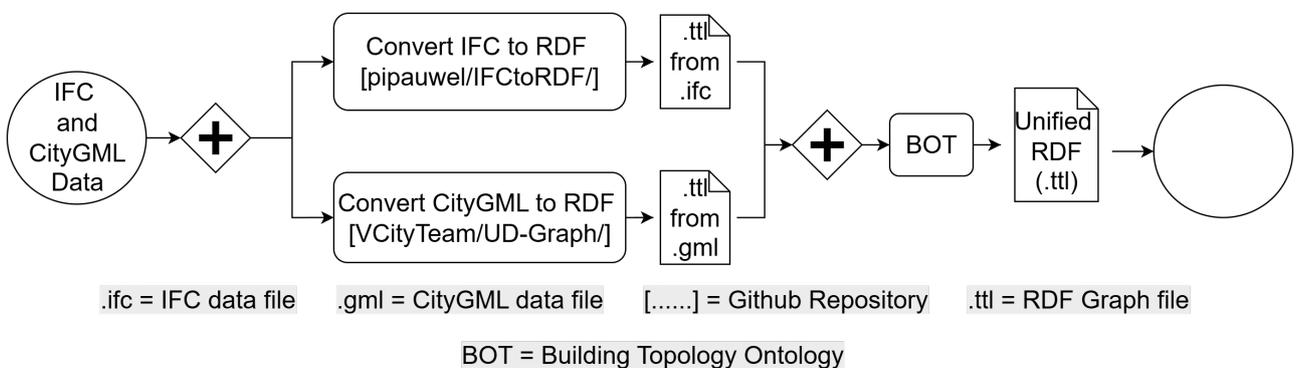


Figure 1: General architecture of the IFC-CityGML integration workflow.

3.1 File Processing and Input Validation

The workflow begins with the upload of two primary data sources: an IFC model capturing civil infrastructure components (e.g., sewer pipe segments) and a CityGML model containing above-ground building representations at Level of Detail 3 (LOD3). While the system was implemented and tested primarily with LoD3 datasets, because URIs and mapping rules were defined for this level, the same approach is applicable to LoD2. Minor adjustments to the RDF templates and mapping rules would allow LoD2 representations to be integrated without changing the overall framework. Upon upload, the system identifies the format and version of each file to ensure compatibility with supported specifications (e.g., IFC4x3, CityGML 2.0). Version metadata is parsed to route files through appropriate converters. Validation routines check for syntactic correctness, ontology consistency, and missing or malformed metadata. Files that fail validation are excluded, and detailed error reports are provided to facilitate correction.

3.2 Semantic Conversion into RDF

In the second phase, both validated datasets are independently transformed into RDF graphs. Each transformation process is aligned with relevant Semantic Web ontologies to ensure the preservation of semantic structures and relationships.

IFC to RDF Conversion: The IFC source is parsed into RDF using an IFC-specific transformation mechanism based on the *ifcOWL* ontology. This conversion preserves object hierarchies, geometrical parameters, and parametric attributes defined in the original IFC schema. Geometry-specific information and semantic descriptors are mapped to ontology terms referred to as A, through schema alignment and rule-based transformation logic.

CityGML to RDF Conversion: The CityGML model is converted to RDF using a process that applies both the CityGML ontology and GeoSPARQL vocabulary. This ensures that spatial constructs, such as surfaces and building footprints, are semantically represented using well-defined RDF triples. Thematic elements (e.g., Door, RoofSurface, BuildingPart) are mapped to their RDF class equivalents using JSON-based mapping rules referred to as B. For example, `<bldg:Door>` is consistently aligned to `building#Door`:

```
"{http://www.opengis.net/citygml/building/2.0}Door":  
"https://dataset-dl.liris.cnrs.fr/rdf-owl-urban-data-ontologies/  
Ontologies/CityGML/2.0/building#Door"
```

This ensures that semantically similar entities across the dataset use a uniform RDF vocabulary.

Each conversion results in a semantically enriched RDF graph that preserves the domain-specific semantics of the respective input models. These graphs serve as the input to the subsequent integration phase.

3.3 Integration and Semantic Alignment

The final phase of the methodology is a structured six-stage RDF integration process designed to generate a semantically interoperable knowledge graph. The integration logic is ontology-driven and

guided by a rule-based configuration that defines class mappings, property translations, spatial linking strategies, and preservation constraints.

Stage 1: All RDF graphs, including the reference ontology, are merged into a single working graph. During merging, all non-resolvable or local Internationalized Resource Identifiers (IRI) (e.g., `file:///`) are replaced with stable, ontology-aligned IRIs derived from class-level namespaces. This operation ensures long-term IRI consistency and supports linkability across external data sources.

Stage 2: Each instance in the merged graph is evaluated against declarative mapping configurations referred to as A and B to replace source-specific classes and attributes with ontology-compliant equivalents. The system applies comprehensive cross-domain alignments defined in JSON rules that map infrastructure elements such as `IfcPipeSegment` to BOT-based classes like `bot:UtilityNetwork`, and CityGML buildings to `bot:Building` with property preservation strategies. When the target ontology lacks a corresponding term, a placeholder OWL concept is generated, incorporating the missing key term from the source data as a proxy URI. This mitigates the semantic gap, albeit with inherent limitations.

A: IFC → BOT Mapping:

```
{
  "ifc:IfcPipeSegment": {
    "bot:UtilityNetwork": {
      "subclassOf": true, .....
    }
  }
}
```

B: CityGML → BOT Mapping:

```
{
  "bldg:Building": {
    "bot:Building": {
      "equivalentClass": true, .....
    }
  }
}
```

Stage 3: Semantic preservation rules are applied to maintain critical attributes and establish spatial relationships throughout the integration process. While Section 3.1 validates the structural correctness of the raw input files (syntax, schema compliance, and presence of basic metadata), Stage 3 ensures that all properties required for meaningful semantic integration are present in the RDF graph. It should be noted that Stage 3 relies on the input files passing Section 3.1 checks; any missing or inconsistent information in the source files can propagate into Stage 3, potentially causing incomplete semantic links or integration errors. Attributes such as `nominalDiameter`, `wallThickness`, and `measuredHeight` are used as examples; the system can be configured to validate any essential set of properties required for the analysis. Spatial relationships are established through utility validation rules that define buffer zones, depth constraints, and connectivity requirements between pipes and buildings. When pipe geometry falls within buffered regions around buildings, semantic links are created using the `servicesBuilding` predicate. Missing semantic properties are logged for review to ensure comprehensive semantic integrity.

Stage 4: The RDF graph is validated using SHACL shapes that define structural expectations and semantic constraints. Although the RDF graphs are generated beforehand, Stage 4 ensures that the integration process has not introduced inconsistencies, missing properties, or misaligned relationships. SHACL validation can fail in cases such as incomplete property propagation, unexpected null values, or violations of project-specific constraints that are not enforced during earlier stages. Violations are classified into three levels `Violation`, `Warning`, and `Info` to facilitate informed debugging and ensure compliance with the target ontology.

Stage 5: In the final step, the integrated RDF graph is cleaned, normalized, and serialized into Turtle format. A summary of integration metrics, such as triple counts, class/property replacements, and validation outcomes, is generated to support reproducibility and subsequent analysis.

This multiphase methodology ensures that heterogeneous spatial data from BIM and GIS domains are transformed into a unified semantic framework that maintains domain fidelity and supports integrative applications. Its modular design supports extensibility to additional ontologies and data types, offering a reusable foundation for smart city data integration scenarios.

4 Implementation

This section details the technical realization of the proposed integration framework using a representative use case. The implementation connects a city sewer pipe segment modeled in IFC with a residential building from a CityGML dataset. Building upon the ontology-driven ETL pipeline described in Section 3, the focus here is on the practical aspects of RDF transformation, rule-based alignment, and data serialization as applied to the selected models. The alignment of concepts is achieved by mapping both IFC and CityGML classes to a shared ontology layer based on the BOT and its extended modules. These mappings form the semantic backbone for reasoning across datasets. Mappings play the key schema-level alignments, covering both topological entities and descriptive attributes. For example, `IfcPipeSegment` is mapped to `bot-elements:SewerPipe`, while `data_measuredHeight` from CityGML is aligned to `bot-elements:height`. In addition to schema-level mappings, rule-based alignment strategies were implemented to establish dynamic semantic relationships between spatial entities. The system creates RDF graphs that preserve semantic properties from the original models. The CityGML building maintains its semantics of geometry and EPSG:25832 coordinate reference system. The IFC pipe segment preserves its property network, including geometric parameters and material specifications. The visualization displays the interconnected structure of the converted IFC data. One key rule establishes a connection between sewer pipes and nearby buildings based on spatial proximity. This rule states that if a pipe instance lies within 3 meters of a building, a relationship is instantiated using the property `bot:servicesBuilding`. The final RDF output is organized using named graphs for modularity and traceability. Metadata about entity origin, transformation steps, and version identifiers are preserved, allowing validation and provenance tracking. Semantic queries are supported via SPARQL endpoints, enabling cross-domain data retrieval, such as querying which building is served by which underground utility or identifying connected infrastructure based on semantic class types. This implementation provides a scalable foundation for linking heterogeneous spatial data sources at the semantic level. It allows interoperability without altering the data's structural and semantic integrity. However, the current implementation supports one-way transformation from IFC and CityGML into RDF; direct reconstruction back into the original IFC or CityGML formats is not implemented. While the RDF graphs preserve the necessary schema-level classes and properties, reverse conversion would require an additional rule set to serialize RDF triples back into the native syntactic structures of the source standards.

5 Discussion

The RDF output is compatible with triple stores and graph databases, enabling BIM and GIS data queries via SPARQL. Schema-level alignment supports class and property linking across heterogeneous datasets. The pipeline is transferable to other geographic or infrastructure contexts with minimal rule file changes, facilitating interoperability in urban information systems.

However, the framework has certain limitations. The mapping logic depends on predefined schema correspondences. If source models use non-standard property names or lack metadata, alignment accuracy is reduced. For example, when IFC models omit Pset attributes or CityGML datasets lack consistent address structures, the matching process fails or requires fallback strategies. In such cases, manual rule revision becomes necessary. Moreover, RDF conversion is currently a static process: any change in the source BIM or GIS datasets requires re-running the transformation to preserve consistency, which can be resource-intensive and error-prone. Incremental update strategies or change-detection mechanisms could mitigate this by reprocessing only affected elements.

Another limitation concerns spatial metadata, which affects accurate linking of objects across datasets. When coordinate reference systems or positional anchors are missing or inconsistent, spatial linking strategies such as buffer-based matching are invalidated. Although the integration avoids direct geometry transformation, it still depends on harmonized reference frames to compute spatial relations. Disparate coordinate systems place objects in unaligned spatial reference frames, so their semantic spatial relations cannot be established without coordinate transformation. As full CRS harmonization was not performed, differing reference systems may result in incomplete or inconsistent semantic links in the graph.

In addition to spatial constraints, performance issues arise with large datasets. Processing large datasets is resource-intensive. RDF conversion from nested IFC structures generates many triples and high memory usage, while post-conversion queries slow down with long property chains or reasoning. Without indexing or triple store optimization, execution time rises sharply with model size.

In its current form, the BOT ontology does not natively support domain concepts such as utility flow direction, network connectivity, or pipe-specific properties like gradient or service depth. To represent such attributes, project-specific extensions were introduced. These are not part of any external standard and may not be interoperable with other semantic tools unless formally published and adopted. This restricts portability unless the extensions are clearly documented and shared in a compatible form. Furthermore, while the current rule set demonstrates interoperability for a representative case, it does not cover the full attribute range of IFC and CityGML. Broader applicability would require expanding the mapping rules or incorporating adaptive techniques to handle diverse model structures.

6 Conclusion and Future Work

This study introduces a semantic integration framework linking BIM and GIS datasets via ontology-aligned RDF representations. An IFC-modeled sewer pipe segment and a CityGML-based residential building were converted to RDF using ifcOWL and CityGML2OWL. Schema-level alignment with BOT and its extensions ensured interoperability, while rule-based mappings enabled semantic links between

entities without geometric fusion. The resulting knowledge graph supports cross-domain queries and traceable metadata for urban infrastructure analysis.

The case shows how static BIM and GIS datasets can be integrated through declarative rules and modular ontology mappings. Transformation logic is externalized, enabling reuse across domains through schema-level correspondence, though without validation at all levels.

The framework relies on explicit metadata and consistent property naming; deviations from standards require rule adjustments. Large datasets introduce performance constraints in RDF generation and querying. BOT's generalized structure lacks native concepts for features like utility connectivity, flow direction, or temporal states, requiring extensions for full coverage.

Planned improvements include integrating ontologies such as SAREF, FIWARE, or CityGML 3.0 ADEs to expand utility and network modeling; implementing SHACL and SPARQL-based validation for data completeness and structural integrity; and incorporating live or sensor data streams to support temporal queries and dynamic infrastructure monitoring in RDF environments.

References

- [1] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*. John Wiley & Sons.
- [2] P. A. Longley, M. F. Goodchild, D. W. Maguire, and D. W. Rhind, *Geographic Information Systems and Science*. John Wiley & Sons, 2005.
- [3] S. Herle, R. Becker, R. Wollenberg, et al., "Gim and bim", *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 88, no. 1, pp. 33–42, 2020.
- [4] International Organization for Standardization, *ISO 16739-1:2018: IFC for data sharing in the construction and facility management industries*, 2018.
- [5] Open Geospatial Consortium, *OGC CityGML 3.0 Conceptual Model Standard*, 2021.
- [6] Y. Ji, Y. Wang, Y. Wei, et al., "An ontology-based rule mapping approach for integrating ifc and citygml", *Transactions in GIS*, vol. 28, no. 3, pp. 675–696, 2024.
- [7] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web", *Scientific American*, vol. 284, no. 5, pp. 28–37, 2001.
- [8] J. Dao, S. T. Ng, and C. Y. Kwok, "Interlinking bim and gis data for a semantic pedestrian network and applications in high-density cities", DOI: <https://doi.org/10.1016/j.dibe.2024.100367>
- [9] J. Shi, Z. Pan, L. Jiang, and X. Zhai, "An ontology-based methodology to establish city information model of digital twin city by merging bim, gis and iot", DOI: <https://doi.org/10.1016/j.aei.2023.102114>
- [10] S. Howell, Y. Rezgui, and T. Beach, "Integrating building and urban semantics to empower smart water solutions", DOI: <https://doi.org/10.1016/j.autcon.2017.02.004>