

Article

Direct Translation Versus Semantic Reinterpretation in IFC-to-CityGML 3.0 Conversion: Assessing Traceability and Quantity Preservation Through RDF Knowledge Graphs

Jiyoung Kim 

Social Eco Tech Institute, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Republic of Korea; elliekim@konkuk.ac.kr

Abstract

Converting building information models in the Industry Foundation Classes (IFC) into CityGML 3.0 brings building interiors into urban digital twins and knowledge graphs, yet converted models are often assessed only for schema validity. Converters either translate IFC elements one by one (direct translation) or rebuild thematic surfaces, rooms and levels of detail through interpretive decisions (semantic reinterpretation). This study makes those decisions explicit and measurable. We implemented a reinterpretation converter with switchable decisions, an area-preserving polygon reconstruction, and a converter-agnostic GML-to-RDF transformation based on the unmodified LIRIS CityGML 3.0 ontology, GeoSPARQL and PROV-O. On a railway-station model (IFC2x3, 23,861 products), we compared our converter with an open-source direct-translation converter against reference quantities from IFC geometry. For the 3649 objects not simplified in the default run, both converters preserved surface area equally well (95.4% and 95.2% within tolerance); prism simplification multiplied the summed volume of the simplified objects by 7.2. Every IFC product was either converted one-to-one or excluded by a recorded decision, whereas the direct translation's output lacked 728 products. With curtain walls excluded and simplification applied, the reinterpreted GML was 26 times smaller and answered queries on room area in the standard vocabulary and on wall surfaces by standard class with auxiliary storey and area properties, which the direct translation could not. Controlled runs varying the two decisions independently attributed the volume loss and most of the size reduction to simplification; with it off, reinterpretation matched direct translation on all 23,133 common objects (98.7% versus 98.6% within tolerance).

Keywords: IFC; CityGML 3.0; BIM–GIS integration; knowledge graph; GeoSPARQL; provenance; information loss; model conversion



Academic Editors: Yebin Chen, Zhigang Zhao, Ding Ma, Yaxing Li, Luigi Barazzetti, Philippe De Maeyer and Wolfgang Kainz

Received: 21 September 2026

Revised: 19 October 2026

Accepted: 22 October 2026

Published: 25 October 2026

Copyright: © 2026 by the author. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Urban digital twins increasingly need to reach inside buildings. Emergency response, facility management, indoor navigation and asset-level planning all require room-scale information that city models have traditionally omitted. Building information models (BIM) exchanged in the Industry Foundation Classes (IFC) hold that information, but in a form tailored to design and construction rather than to spatial analysis [1,2]. CityGML 3.0 narrows the gap: its revised space concept, the new Construction module and the refined Building module can represent rooms, storeys, constructive elements, fillings and thematic surfaces within one urban data model [3,4]. Converting IFC to CityGML 3.0 is therefore

a practical route for integrating BIM into city-scale platforms, and representing the result as an RDF knowledge graph allows it to be queried and linked with other authoritative datasets [5].

Conversion from IFC to CityGML has a long research record. Earlier work established how to derive geometrically valid, semantically enriched LoD3 building shells [6], how to carry IFC content into CityGML with little loss through extensions [7,8], and how available tools perform on common benchmark data [9]. In 2026 an open-source converter targeting CityGML 3.0 directly was released [10]. Converting a model is thus increasingly well supported; less attention has been paid to knowing what a converted model still contains.

Two conversion paradigms coexist. Direct translation maps each IFC element to a corresponding CityGML object and keeps its triangulated geometry. The source geometry is preserved, but the result mirrors IFC rather than the city model: there are no thematic surfaces, no room areas and no coarser levels of detail. Semantic reinterpretation rebuilds the model in CityGML terms, deriving wall, floor and ceiling surfaces, merging duplicated storeys, classifying generic proxies and simplifying detailed families. The result is directly usable for spatial analysis, but every step is an interpretive decision, and the decisions are often hard-coded and not documented in the output. Schema validation does not reveal them: a converted file can be valid while silently misrepresenting areas or omitting components.

This paper argues that the choice between the paradigms should be informed by measurement, and that reinterpretation can be made accountable if each decision is explicit, switchable and traceable. We address four research questions:

- RQ1 (traceability): To what extent can each converted feature be traced back to its IFC source, and how do conversion decisions affect this?
- RQ2 (quantity preservation): How well are surface areas and volumes preserved through IFC → CityGML 3.0 → RDF, and which decisions cause the deviations?
- RQ3 (queryability): Which analytical questions can each converted knowledge graph answer with standard CityGML vocabulary, and at what cost?
- RQ4 (geometry encoding): How do a fully structural and a literal (WKT) encoding of geometry in RDF compare in size, query time and fidelity?

The main contribution is an integrated and reproducible workflow, together with the evaluation that it makes possible; two of its parts are also methodological advances. First, we present a reinterpretation converter in which every interpretive decision is an option that can be switched off, so that its effect can be examined. Second, we describe a planar polygon reconstruction that turns kernel triangles into CityGML polygons while preserving area within stated tolerances, including rings that touch themselves. Third, we provide a converter-agnostic GML-to-RDF transformation that reuses the unmodified LIRIS CityGML 3.0 ontology and adds only provenance, quantities and storey membership. Fourth, we define an evaluation framework built on reference values derived from IFC geometry, and apply it to a large railway-station model and a public validation model, comparing reinterpretation with direct translation. What distinguishes this from earlier conversion studies is not the individual steps, most of which reuse established techniques. It is that every deviation is attributed to a named decision at the level of the single IFC product, so that omissions, quantity changes and differing query answers can be explained rather than only observed. The reconstruction of self-touching rings and the per-plane area check are new in this context, and so is the combination of provenance links to IFC GlobalIds with recomputed quantities in one converter-agnostic graph. Within the knowledge graph alone a reader can resolve any converted feature to its IFC GlobalId and class, and recompute its quantities. Products removed by a decision are, in the present version,

recorded only in the conversion log; recovering them requires that log together with the reference table (Section 6.2).

Section 2 reviews related work. Section 3 presents the conversion pipeline, Section 4 the evaluation methods and Section 5 the experimental design. Section 6 reports the results for each research question, Section 7 discusses trade-offs and limitations, and Section 8 concludes.

2. Related Work

2.1. IFC-to-CityGML Conversion

IFC describes a building as a set of typed elements with rich semantics and solid geometry expressed by extrusions, Boolean operations and boundary representations [1]. CityGML, in contrast, describes spaces and their bounding surfaces at several levels of detail. Deng et al. [11] mapped IFC to CityGML through schema mediation across levels of detail. Donkers et al. [6] derived LoD3 exteriors by Boolean union of element solids and extraction of outer shells. Stouffs et al. [7] and Biljecki et al. [8] pursued near-lossless conversion by extending CityGML with IFC-derived content through application domain extensions. The GeoBIM benchmark compared georeferencing and conversion tools on shared datasets and found that the results of the tools differed considerably and that information was often lost or altered in conversion [9]. Reviews of BIM–GIS integration [12–15] consistently locate the loss of semantics and geometry at this conversion step and call for its quantitative assessment. Since the benchmark, conversion has been approached through configurable graph-transformation rule sets [16,17], through semantics-based simplification of IFC models before their use in GIS [18], and through reviews of the role that CityGML can play in integration [19]; heritage models have been converted from IFC to CityGML for multiscale GIS as well [20]. Arroyo Ohori et al. [2] reported practical difficulties in processing BIM and GIS models together, including georeferencing and model quality issues. The open-source converter by Kolbe [10] targets CityGML 3.0 and follows direct translation: each IFC element becomes a constructive element, installation or furniture object with triangulated geometry and an external reference to its IFC GlobalId.

2.2. CityGML 3.0 Building and Construction Modules

CityGML 3.0 restructured the conceptual model around spaces and space boundaries [3,4]. Buildings contain rooms, storeys, constructive elements, installations and furniture; constructive elements may contain fillings such as doors and windows; and thematic surfaces such as WallSurface, FloorSurface and CeilingSurface bound the spaces. Areas and volumes can be recorded as qualified quantities on spaces. Levels of detail were decoupled from the thematic structure, so that one building can carry LoD0 footprints, LoD1 blocks and LoD3 detail at the same time [21]. These features make CityGML 3.0 a natural target for reinterpretation, but the standard does not prescribe how IFC content should populate them.

2.3. Semantic Web Representations of Buildings and City Models

Semantic web technologies have been widely applied to buildings [5]. ifcOWL expresses the IFC schema in OWL [22], and the Building Topology Ontology captures building topology with a minimal vocabulary [23]. For city models, the LIRIS group derived OWL ontologies directly from the CityGML 3.0 conceptual model [24,25], which keeps the RDF representation aligned with the standard. Beyond these ontologies, CityGML content has been brought into knowledge graphs through the Semantic 3D City Database [26], through graph databases that integrate BIM and GIS for urban applications [27], and through ontologies generated automatically from the IFC and CityGML schemas and

aligned with each other [28]; graph databases have also been used to compare versions of large city models [29]. GeoSPARQL provides a vocabulary and functions for geometry in RDF [30]; its version 1.1 [31] comes with a compliance benchmark [32], and stores such as Strabon [33] and Ontop-spatial [34] implement its function set, whose topological predicates are evaluated on two-dimensional geometries even where the literals carry a third coordinate. PROV-O models provenance [35]. What remains open is how geometry should be encoded when a converted city model becomes a knowledge graph, and how its lineage to the original BIM can be retained for later checking.

2.4. Assessing Information Loss in Model Conversion

Quality checks for 3D city models mostly concern geometric and schema validity [36,37]. They detect invalid solids and malformed polygons, but not whether a valid output still represents the same building as its source. Conversion studies have typically reported loss qualitatively or through visual comparison [9,19], and the reviews cited above [12–15] note the absence of agreed measures for what a conversion preserves. Quantitative, object-level comparison against the source model, and attribution of any deviation to a specific conversion decision, has received little attention. This study addresses that gap.

3. Traceable Conversion from IFC to CityGML 3.0 and RDF

3.1. Two Conversion Paradigms

Figure 1 summarizes the study design. In part A, the source IFC model is evaluated by the geometry kernel, and reference values are derived from the resulting meshes (Sections 4.1 and 5.1). In part B, the same model is converted by two converters that represent the two paradigms: an existing direct-translation converter and the semantic reinterpretation of this study, whose switchable decisions and polygon reconstruction are described in Sections 3.2 and 3.3. In part C, both CityGML outputs are processed by one RDF transformation (Section 3.4), so that differences can be attributed to the conversion rather than to the RDF step, and the two knowledge graphs are evaluated against the reference values for RQ1–RQ4 (Sections 4.2 and 4.3, with results in Section 6). Table 1 contrasts the two converters.

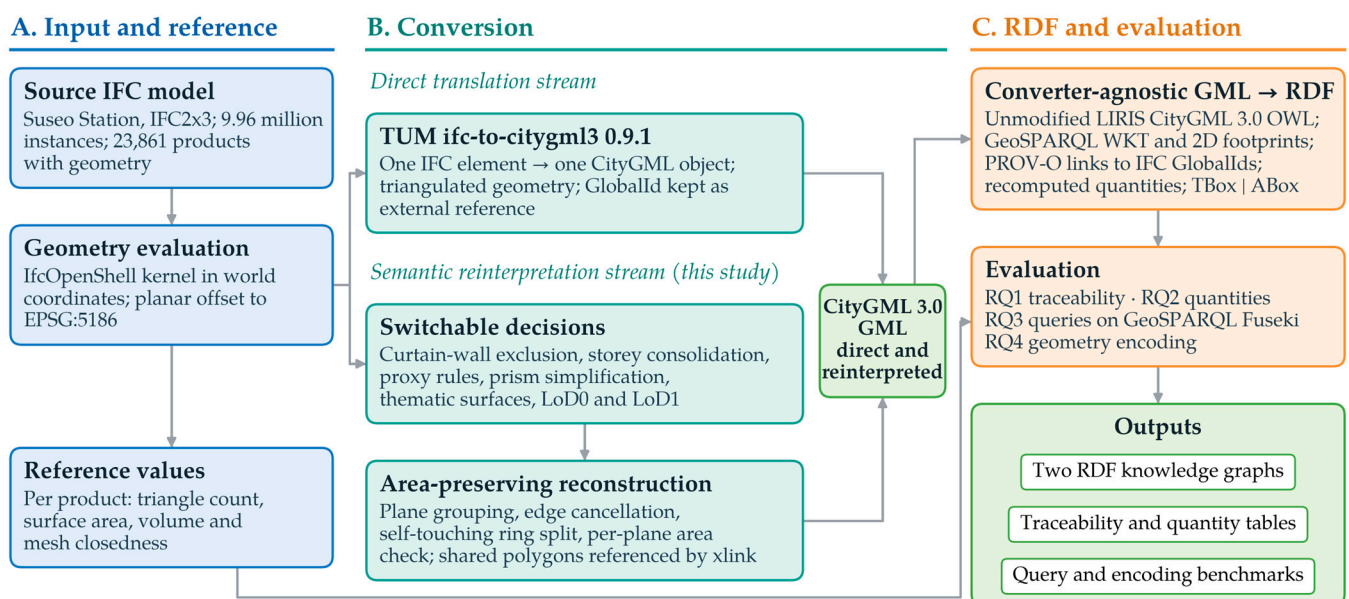


Figure 1. Overview of the study: (A) the source IFC model and the reference values derived from its geometry; (B) conversion of the same model by direct translation and by semantic reinterpretation to CityGML 3.0; (C) the converter-agnostic transformation of both outputs to RDF and their evaluation.

Table 1. Comparison of the two conversion paradigms as implemented in this study.

Aspect	Direct Translation (TUM Ifc-to-Citygml3 0.9.1)	Semantic Reinterpretation (This Study)
Mapping	One IFC element to one CityGML object	IFC elements reinterpreted as spaces, constructive elements and thematic surfaces
Thematic surfaces	Not generated	Wall, floor, ceiling, roof, ground, interior wall, door and window surfaces
Levels of detail	LoD3	LoD0 footprint, LoD1 block and LoD3 in one building
Geometry	Triangulated meshes	Planar polygons with holes; solids reference shared polygons by xlink
Storeys	As in IFC	Paired slab/finish levels consolidated (switchable)
Traceability	External reference to IFC GlobalId	External reference to IFC GlobalId
Quantities	Not exported	Room net floor area and volume as qualified quantities
Decisions	Implicit in code	Explicit and switchable (Table 2)

Table 2. Interpretive decisions of the reinterpretation converter and their options.

Decision	Rule	Switch	Default
Curtain-wall exclusion	Omit IfcCurtainWall and its decomposed members and plates	--no-curtainwall	off (on in the station configuration)
Storey consolidation	Merge storeys whose names differ only by slab/finish-level suffixes; keep the finish level as representative and the lower elevation	--no-merge-storeys	on
Proxy classification	Assign IfcBuildingElementProxy to constructive element, installation or furniture by ordered name rules (narrow terms first)	--no-proxy-rules	on
Prism simplification	Replace an object by a prism on its convex hull (at most 12 vertices) if it has more than 60 planar faces and either a bounding diagonal of at most 6 m or a mean face area below 0.05 m ² ; always for distribution and furnishing elements	--no-simplify	on
Thematic surfaces	Classify faces by the vertical component of the outward normal ($ n_z > 0.7$) and element type; coverings by predefined type	--no-surfaces	on
Derived levels of detail	LoD0 as the union of horizontal faces; LoD1 as its extrusion between minimum and maximum height	--no-lod01	on

3.2. Semantic Reinterpretation as Switchable Decisions

The reinterpretation converter implements each interpretive decision as a separate option (Table 2). Running the converter with one option switched off and comparing the result with the full configuration makes it possible to isolate the effect of that decision; in this study, the effects are attributed object by object in one configuration, which we call the station configuration throughout: all decisions of Table 2 are active, including curtain-wall

exclusion, which the command-line interface leaves inactive unless requested (Section 5.2). All decisions are applied after geometry has been evaluated by the IfcOpenShell geometry kernel [38] in world coordinates, so that they operate on the same triangulated geometry that a direct translation would export.

Several structural choices are fixed rather than switchable because they follow from the CityGML 3.0 model. Rooms become BuildingRoom objects; walls, slabs, roofs, coverings and curtain-wall components become BuildingConstructiveElement objects; columns, beams, stairs, railings and distribution elements become BuildingInstallation objects; and furnishing elements become BuildingFurniture objects. The treatment of beams and columns is a modeling choice of this study, not a reading of the standard: CityGML 3.0 defines a BuildingConstructiveElement as “an element of a Building which is essential from a structural point of view”, with walls, slabs, staircases and beams as examples (clause 8.17.1.4), and a BuildingInstallation as “a permanent part of a Building (inside and/or outside) which has not the significance of a BuildingPart”, with stairs, antennas, balconies and small roofs as examples (clause 8.17.1.6); beams could therefore be mapped to BuildingConstructiveElement, stairs appear among the examples of both classes, and columns are not named. We map them to BuildingInstallation for two reasons. The thematic surfaces of Section 3.2 are derived from the elements that bound rooms, and reserving BuildingConstructiveElement for walls, slabs, roofs and coverings keeps the boundary model free of members that bound no room. In the station model, moreover, many IfcBeam and IfcColumn instances are secondary framing of finishes rather than primary structure: 705 of the 1423 beams are small, detailed members that meet the simplification criteria of Table 2. Because every feature keeps its IFC class in the provenance link, the mapping can be reversed in the knowledge graph, and Section 6.4 notes where query results depend on it. Doors and windows are placed as fillings of the element they are embedded in, which is recovered from the IfcRelFillsElement and IfcRelVoidsElement relationships. The building owns all objects; storeys reference them by xlink, as in the aggregation defined by the conceptual model. Every object carries a core:ExternalReference to its IFC GlobalId, and rooms carry their net floor area and volume as core:QualifiedArea and core:QualifiedVolume.

The name rules for building element proxies are listed in full in Table A1 (Appendix A). They are applied to the concatenation of the Name, ObjectType and Tag attributes of each IfcBuildingElementProxy as case-insensitive regular expressions, in the fixed order of the table; the first matching rule decides. Narrow terms (furniture, sanitary fittings, signage, lighting, ducts, lifts, stairs) precede broad ones (ceilings, roofs, slabs, openings, walls), so that, for example, a ceiling molding is classified as a ceiling finish rather than as a wall. The keywords are listed in Korean and English because the station model follows Korean naming practice; a model in another language or naming convention requires its own keyword list, which is the limitation noted in Section 7.3.

Thematic surfaces are classified per face. For rooms, faces whose outward normal points downward become floor surfaces, faces pointing upward become ceiling surfaces and the rest become interior wall surfaces. Walls and curtain-wall components yield wall surfaces. Roof elements yield roof, ceiling or wall surfaces depending on normal direction. Slabs yield floor surfaces on top, ceiling surfaces below and ground surfaces below the lowest slab level. Coverings follow their IFC predefined type: the downward faces of ceiling coverings become ceiling surfaces and the upward faces of flooring become floor surfaces. The geometry of a constructive element is stored only in its thematic surfaces, whereas rooms, doors and windows additionally carry a closed LoD3 solid whose surface members reference the thematic-surface polygons by xlink, so that each polygon is written once. A face counts as facing upward or downward when the vertical component n_z of its unit outward normal satisfies $n_z > 0.7$ or $n_z < -0.7$; for the net floor area of rooms,

only faces with $n_z < -0.99$ are counted. The two thresholds serve different purposes. The classification threshold of 0.7 (about 45° from the vertical) partitions all faces of an object exhaustively into wall-like and floor- or ceiling-like surfaces, so that inclined faces are still assigned to one thematic class. The net floor area, in contrast, is meant to count only the horizontal floor of a room, so it sums the downward faces that lie within 8° of horizontal; inclined faces such as ramps or stair soffits, which the 45° rule would count as floor, are excluded. For the station model the choice has no effect: Section 6.3 reports that relaxing the floor-area threshold to the classification threshold changes the net floor area of none of the 199 rooms.

3.3. Area-Preserving Planar Polygon Reconstruction

The geometry kernel returns every object as a triangle mesh. Writing these triangles directly contributes to the size of direct translation and makes surface-level analysis harder. We reconstruct planar polygons from each mesh in six steps (Figure 2).

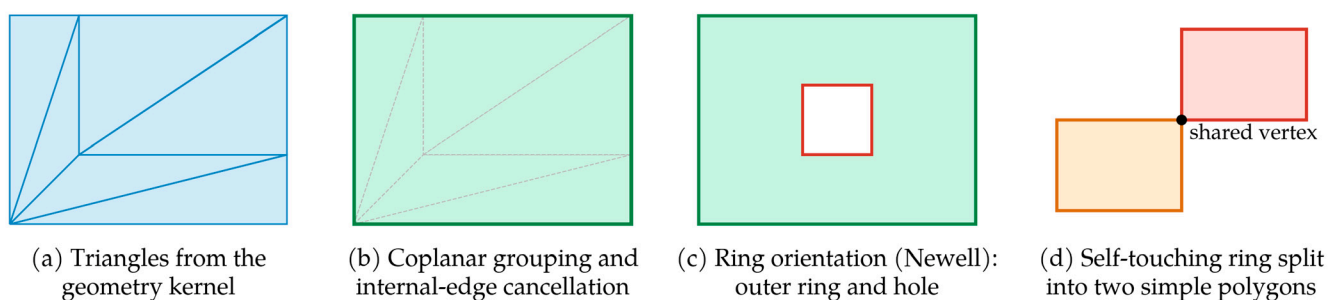


Figure 2. Planar polygon reconstruction: (a) kernel triangles; (b) grouping by supporting plane and cancelation of internal edges; (c) classification of outer rings and holes by the Newell area vector; (d) splitting a self-touching ring at its shared vertex instead of discarding it. Solid lines are triangle edges in (a) and ring boundaries in (b–d); dashed grey lines in (b) are internal edges that cancel; in (c), the green ring is an outer ring and the red ring a hole; in (d), the two colors mark the two simple rings obtained by splitting the self-touching ring.

- (i) Vertices closer than 10^{-6} m are welded, and degenerate triangles are removed.
- (ii) Triangles are grouped by their supporting plane, using the unit normal rounded to 10^{-4} and the plane offset rounded to 10^{-3} m as the key.
- (iii) Within each plane group, every directed edge is counted against its reverse. Internal edges cancel; the remaining boundary edges are chained into rings.
- (iv) The surviving directed edges of a plane group are chained into rings by walking from a start vertex along outgoing edges, each edge being consumed once. A vertex that is reached a second time while the current ring is still open identifies a self-touching ring. Instead of discarding the ring, the portion of the walk from the earlier occurrence of that vertex to the current position is closed off as a ring of its own and removed from the walk, which then continues from the touching vertex; a ring that touches itself at k vertices therefore yields $k + 1$ simple rings, and the partition is determined by the walk order. Because every edge is used exactly once, the areas of the simple rings sum to the area of the original region, which the area check of step (vi) verifies. Collinear vertices are removed from each ring, and rings whose Newell vector vanishes are dropped. The simple rings are then classified as outer rings or holes by the sign of their Newell vector against the plane normal, and each hole is assigned to the smallest outer ring of the group that contains a representative point of it.
- (v) Each ring is oriented with the Newell area vector. Rings whose area vector agrees with the plane normal are outer rings; the others are holes. A hole is assigned to the smallest outer ring that contains an interior point of the hole and is not smaller than it.

(vi) For every plane group, the area of the resulting polygons is checked against the summed area of its triangles; a mismatch is logged and triggers a second pass in which boundary edges are split at vertices lying on them (T-junctions).

Let a planar polygon consist of an outer ring R_0 and holes $R_1, \dots, R_k$, and let ring R_i have the vertices $p_{i,1}, \dots, p_{i,n_i}$ in order, closed by $p_{i,n_i+1} = p_{i,1}$. The area vector of the polygon is the sum of the Newell vectors of all its rings, and its length is the area. Because holes are oriented opposite to the outer ring, their contribution is subtracted automatically:

$$\mathbf{a} = \frac{1}{2} \sum_{i=0}^k \sum_{j=1}^{n_i} \mathbf{p}_{i,j} \times \mathbf{p}_{i,j+1}, \quad A = \|\mathbf{a}\| \quad (1)$$

where $\mathbf{a}$ is the area vector, whose direction is the unit normal of the polygon, A is the area, k is the number of holes, n_i is the number of vertices of ring R_i , $\mathbf{p}_{i,j}$ is the position vector of its j -th vertex, $\times$ is the cross product and $\|\cdot\|$ is the Euclidean norm. The volume enclosed by a closed shell of F polygons follows from the divergence theorem:

$$V = \frac{1}{3} \sum_{f=1}^F \mathbf{a}_f \cdot \mathbf{p}_f \quad (2)$$

where V is the enclosed volume, $\mathbf{a}_f$ is the area vector of polygon f from Equation (1), $\mathbf{p}_f$ is any vertex of polygon f and $\cdot$ is the scalar product. The outer rings must be oriented so that $\mathbf{a}_f$ points out of the shell.

The same expressions are applied to the kernel triangles, which provides the reference values of Section 4.1. Apart from welding vertices closer than 1 μm and grouping triangles within the stated plane tolerances, reconstruction only removes internal edges and collinear vertices; the per-plane area check of step (vi) verifies that the reconstructed polygons cover the same area as the triangles of each plane group. The split of self-touching rings matters in practice: a ring that touches itself is a valid boundary of a single planar region, and discarding it silently removes that region. In the station model, 9530 self-touching rings were split, and the area check of step (vi) reported no mismatch in any plane group (Section 6.3).

3.4. A Converter-Agnostic GML-to-RDF Transformation with Provenance and Quantities

Both CityGML outputs are transformed to RDF by one transformation that reads any CityGML 3.0 GML document. Classes and properties come from the LIRIS CityGML 3.0 ontology, used without modification [24]. Because the ontology names properties by their declaring class (for example `core:AbstractSpace.boundary`), the transformation resolves each XML property by walking up the class hierarchy of the element until a declared property is found. Nested objects become resources, xlink references become object properties, and literal values are typed by the range of the property.

Geometry is attached as GeoSPARQL geometries [30]. Each geometry element becomes a resource typed both by the corresponding GML class and by `geo:Geometry`, and carries a three-dimensional WKT literal with an explicit coordinate reference system: a solid becomes a `POLYHEDRALSURFACE Z` and a multi-surface a `MULTIPOLYGON Z`. Shared polygons referenced by xlink are resolved. In addition, every spatial object receives a two-dimensional footprint as its default geometry, which makes planimetric spatial functions available, together with its minimum and maximum height.

A small auxiliary vocabulary adds what the LIRIS ontology does not express: (i) `prov:wasDerivedFrom` links from each feature to a resource representing its IFC object, which carries the `GlobalId` and IFC class [35]; (ii) surface area and volume recomputed from the GML geometry with Equations (1) and (2), with units declared through QUDT [39];

(iii) explicit storey membership obtained from the xlinks of each storey; and (iv) a record of the producing converter and the transformation. Terminology (TBox) and instances (ABox) are written to separate files, and the instance data are split into a semantic file and a geometry file. The transformation streams the input object by object, so that multi-gigabyte GML documents can be processed with bounded memory.

Volumes are computed after translating each object so that one of its vertices, $\mathbf{o}$, becomes the origin:

$$V = \frac{1}{3} \sum_{f=1}^F \mathbf{a}_f \cdot (\mathbf{p}_f - \mathbf{o}) \quad (3)$$

where the symbols are as in Equation (2). For an exactly closed shell, $\sum_{f=1}^F \mathbf{a}_f = \mathbf{0}$ and Equations (2) and (3) agree for any $\mathbf{o}$. This holds even for non-planar polygons, because the area vectors cancel along shared edges whenever adjacent polygons use identical edge coordinates. In practice, the removal of collinear vertices, T-junctions inherited from the source meshes and millimeter rounding can leave small closure gaps, so that the sum of the area vectors deviates slightly from zero and Equation (2) becomes dependent on the origin. In a projected coordinate system with coordinates of several hundred kilometers, such gaps were amplified into volume errors of thousands of cubic meters. Evaluating the volume about a local origin with Equation (3) keeps this amplification at the scale of the object, although it does not remove the gaps themselves. The triangulated output of the direct translation keeps shared edge coordinates identical and was not affected.

4. Evaluation Methods

This section defines how the four research questions are evaluated: the reference values against which both conversions are measured (Section 4.1), the metrics for traceability, quantity preservation and queryability (Section 4.2) and the geometry encoding experiment (Section 4.3). The data, software and environment to which these methods are applied are described in Section 5.

4.1. Reference Values Derived from IFC Geometry

Because the main case study model (Suseo Station, Section 5.1) carries no quantities, reference values were computed from the kernel triangle meshes before any conversion decision was applied. For each IFC product we recorded the number of triangles, the surface area (sum of triangle areas), the volume by Equation (2), and whether the mesh is closed, i.e., whether every edge is used equally often in both directions. Two criteria decide whether a reference value is valid. A reference surface area is valid for every mesh, whereas a reference volume is valid only if the mesh is closed and its volume by Equation (2) is positive: a closed mesh with inconsistently oriented faces or duplicated bodies returns a non-positive value that is not a real volume. Reference volumes that fail either criterion are flagged in the evaluation output and excluded from the volume comparison; how many objects of the case study are affected is reported in Section 6.3. The reference table includes curtain-wall components so that the direct translation, which keeps them, can be evaluated as well. The same closedness test is applied to the output of the reinterpretation: for each feature, the polygons written to the GML document, including those of its thematic surfaces, are read back and an output is closed when every directed edge is matched by the opposite edge the same number of times. This test is strict, since a T-junction inherited from the source mesh leaves an unmatched edge even where the slit between the faces is of millimeter width; the outcome is recorded per object in the supplementary tables and reported in Section 6.3.

4.2. Metrics for Traceability, Quantity Preservation and Queryability

Traceability (RQ1). For every IFC product in the reference table we count the RDF features linked to it by `prov:wasDerivedFrom` and classify the correspondence as one-to-one, one-to-many or missing. Missing products are attributed to an explicit decision (for example curtain-wall exclusion) or to the absence of geometry. We also count features without an IFC counterpart, and doors and windows that could not be placed as fillings.

Quantity preservation (RQ2). For each product we compare the surface area and volume stored in the RDF with the reference values and report the relative deviation:

$$\delta = \frac{q - q_{\text{ref}}}{q_{\text{ref}}} \times 100\% \quad (4)$$

where q is the surface area or volume of an object in a knowledge graph, q_{ref} is its reference value from Section 4.1 and δ is the relative deviation in percent. Deviations are summarized by IFC class and by whether a decision (in particular prism simplification) was applied to the object. An object counts as preserved if its deviation is at most 0.1% or at most 0.01 m^2 (0.001 m^3 for volume), i.e., if $|\delta| \leq 0.1$ or $|q - q_{\text{ref}}| \leq 0.01 \text{ m}^2$ (0.001 m^3); the absolute bound prevents millimeter rounding from dominating the result for very small objects. The two criteria are reported separately in the supplementary evaluation output, because they can differ considerably for classes dominated by small objects (Section 6.3). As a consistency check for the room shells of this model, room floor areas are additionally compared with ceiling areas.

Queryability (RQ3). The same SPARQL queries are posed to both knowledge graphs; the full query set is listed together with the results in Section 6.4. They cover semantic navigation (rooms per storey, fillings per storey), standard quantities (room net floor area through `core:QualifiedArea`), thematic surfaces (wall surface area per storey), traceability (resolving an IFC `GlobalId`) and derived quantities. For each query we record whether it is answerable, the number of result rows and the median execution time over three runs. Spatial queries use GeoSPARQL functions (intersection with and containment in a room footprint). All queries are run on a GeoSPARQL-enabled triple store into which the semantic data and the two-dimensional footprints of each knowledge graph are loaded in turn, without RDFS inference; each query is sent three times over HTTP and the median time is reported. The store is specified in Section 5.2.

4.3. Geometry Encoding Experiment

To assess RQ4, one storey of the reinterpreted model is transformed twice: once with WKT literals as described in Section 3.4, and once with a structural encoding in which every polygon, ring and vertex becomes a resource with coordinate literals. Both encodings write coordinates with millimeter resolution, which is the resolution of the GML written by both converters, so that the values are carried over unchanged for these inputs; inputs finer than a millimeter would be rounded by the transformation, and the comparison below is therefore a statement about the two encodings at that resolution. We compare the number of triples, file size, loading time and query times, and check geometry by geometry that the coordinates recovered from the two encodings agree. Times are measured in-process with `rdflib` and exclude any HTTP transfer.

5. Experimental Design

5.1. Case Study Data

The main case is the BIM model of Suseo Station, a railway station in Seoul, Republic of Korea (Table 3; Figure 3). The model was exported from Autodesk Revit in IFC2x3 and contains 9.96 million entity instances in a file of about 510 MB. Its coordi-

dates are in Korea 2000/Central Belt 2010 (EPSG:5186) [40], although the file declares no coordinate reference system; a planar offset of (−0.978 m, 18.313 m) was applied to both converters after aligning the model with the national indoor spatial information dataset of the station and the building layer of the national address database. The model has three properties that shape the evaluation: it contains no element quantities (IfcElementQuantity), no space boundaries (IfcRelSpaceBoundary) and no map conversion. Thematic surfaces therefore have to be derived from geometry, and reference quantities have to be computed from geometry as well. The public AC20-FZK-Haus model (IFC4), distributed with the direct-translation converter, served as a small validation case whose outputs could be inspected completely.

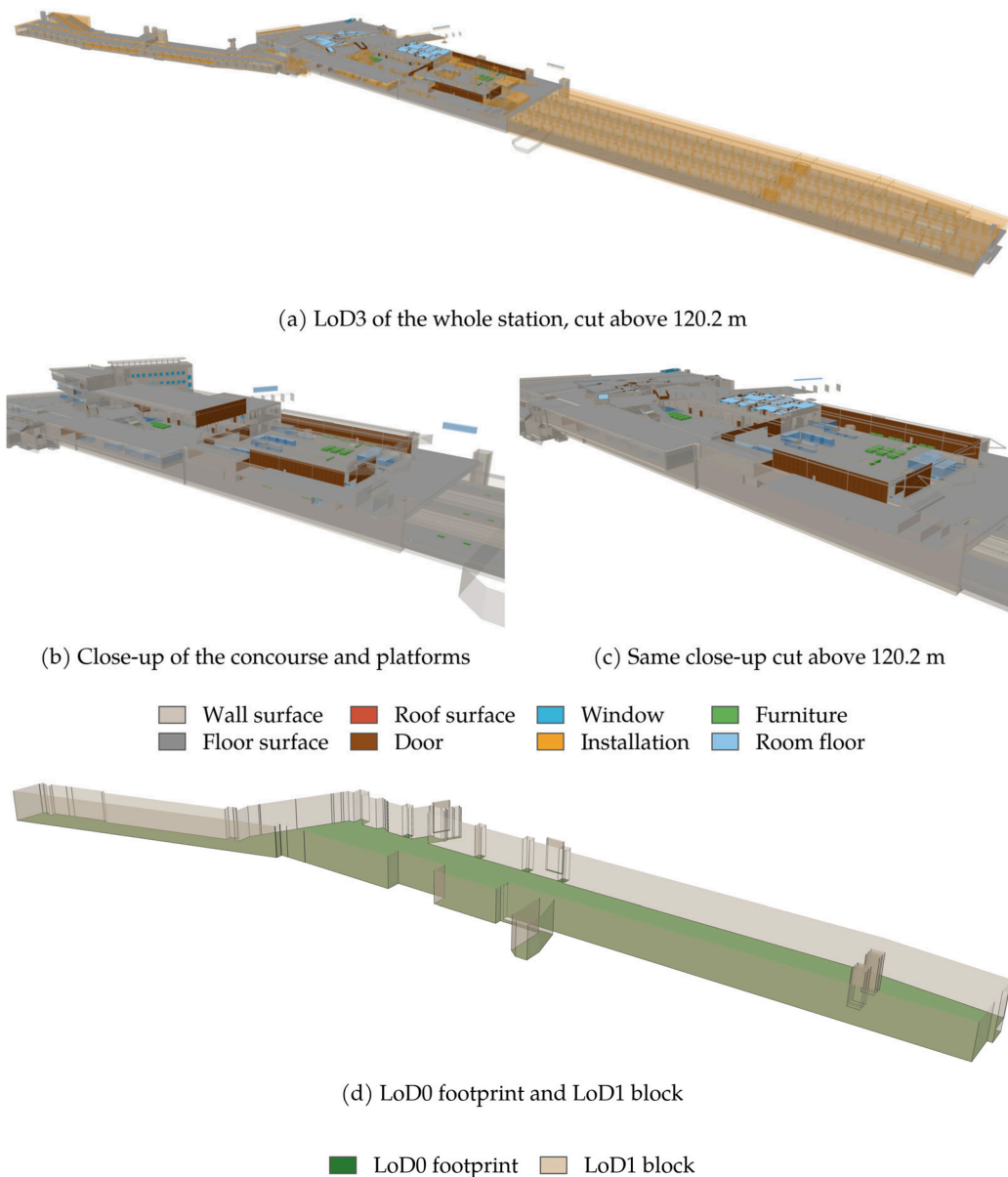


Figure 3. Suseo Station converted by the reinterpretation converter: (a) LoD3 model of the whole station cut above 120.2 m, with wall and roof surfaces rendered semi-transparent and installations at 15% opacity so that floors, doors, furniture and rooms remain visible; (b) close-up of the concourse and platform area; (c) the same close-up cut above 120.2 m, exposing room floors, doors and furniture; (d) LoD0 footprint and LoD1 block derived from the same model, shown from the same viewpoint and at the same scale as (a).

Table 3. Characteristics of the two case study models.

Item	Suseo Station	AC20-FZK-Haus
IFC schema/size	IFC2x3/510 MB	IFC4/2.6 MB
Entity instances	9,955,258	44,249
Products with geometry	23,861 (6852 excluding curtain-wall components)	89
Curtain-wall components	17,009 (11,832 members, 5177 plates)	0
Spaces/doors/windows	199/343/96	7/5/11
Storeys	13 (8 after consolidation)	2
Element quantities/space boundaries	None/none	Present/81
Coordinate reference system	EPSG:5186	Local

5.2. Implementation and Environment

The reinterpretation converter, the RDF transformation and the evaluation scripts were implemented in Python 3.12 with IfcOpenShell 0.8.5 [38], NumPy, Shapely 2.1, lxml and rdflib 7. The direct-translation converter was TUM ifc-to-citygml3 version 0.9.1 (commit 0368abb) [10], run with the same coordinate offset, without appearances and with property sets written as prefixed generic attributes, so that attribute handling was comparable. The station configuration of the reinterpretation converter was invoked as `--eps 5186 --shift = -0.978,18.313,0 --no-curtainwall`, with all other decisions of Table 2 at their default values; the controlled runs of Section 6.6 drop `--no-curtainwall`, add `--no-simplify`, or both; the full command of every run is recorded in the run log distributed with the Supplementary Materials. All GML outputs were validated against the official CityGML 3.0 XML schemas. Experiments ran on a laptop (LG Electronics, Seoul, Republic of Korea) with an Intel Core Ultra 5 125H processor (Intel Corporation, Santa Clara, CA, USA; 14 cores, 18 threads, 3.6 GHz), 32 GB of RAM and 64-bit Windows; the reinterpretation converter used 17 threads for geometry evaluation, and the triple store, Apache Jena GeoSPARQL Fuseki 6.2.0 running on Java 21 on the same machine, was given a maximum heap of 8 GB.

6. Results

6.1. Conversion Outputs

Table 4 summarizes the two conversions of the station model. The reinterpretation converter produced 6852 objects: 199 rooms, 2052 constructive elements, 3893 installations, 269 furniture objects, 343 doors and 96 windows, bounded by 3907 thematic surfaces and represented by 302,824 planar polygons. Of the 1114 building element proxies, 663 were assigned by the name rules of Table A1 (439 signage objects, 79 ceiling finishes, 50 furniture items, 35 stairs and railings, 34 roof elements, 14 lifts and escalators, 9 walls, one slab, one duct and one lighting fixture) and 451 matched no rule and remained installations. Prism simplification was applied to 3200 objects, including all 719 flow terminals and 219 furnishing elements. The GML document of 114.1 MB passed schema validation. The conversion took 19.4 min, of which 14.8 min were spent in the geometry kernel and 3.2 min in polygon reconstruction and classification, with a peak memory of 9.9 GB.

The direct translation produced 23,133 objects in a GML document of 2978 MB, 26 times larger, represented by 5,878,646 triangles and 23.5 million coordinate points against 302,824 polygons and 1.5 million points, i.e., 19 and 16 times as many. These ratios combine two effects that this comparison cannot separate: triangulation, and the retention of the 17,009 curtain-wall components, which the reinterpretation excludes. The two documents therefore differ in content as well as in representation, and the ratio should not be read as the compression achieved by reinterpretation as such. The direct translation took 32.7 min with a peak memory of 11.8 GB, against 19.4 min and 9.9 GB for the reinterpretation. The run times are not comparable as a measure of the two paradigms, because the

reinterpretation evaluates geometry on 17 threads while the direct-translation converter processes objects sequentially. Its RDF representation contained 1,073,085 triples, 2.2 times as many as the reinterpreted graph (481,214), although the difference in size was larger because each geometry literal lists triangles rather than polygons. The RDF transformation mapped every GML element of both documents; no element remained unmapped. Written as N-Triples, the direct translation occupied 1000 MB and took 30 min to transform, against 165 MB and 2.3 min for the reinterpreted model. Section 6.6 separates the two effects by converting the station again with each decision switched independently.

Table 4. Conversion outputs for Suseo Station.

Measure	Direct Translation	Reinterpretation
Objects	23,133	6852
Rooms/constructive/installations/furniture	199/22,216/69/219	199/2052/3893/269
Rooms/storeys	199/13	199/8
Doors/windows (fillings/top-level)	334/96 (430/0; 9 omitted)	343/96 (416/23)
Thematic surfaces	0	3907
Polygons	5,878,646 (all triangles)	302,824
Coordinate points (incl. ring closure)	23,514,584	1,503,682
GML size	2978 MB	114.1 MB
Conversion time/peak memory	32.7 min/11.8 GB	19.4 min/9.9 GB
RDF triples (semantic/geometry)	881,409/191,676	418,341/62,873
RDF size, N-Triples (semantic/geometry)	205.9 MB/794.6 MB	95.8 MB/69.5 MB
RDF transformation time	29.6 min	2.3 min

6.2. Traceability (RQ1)

Table 5 relates the 23,861 IFC products with geometry to the features of both knowledge graphs. In the reinterpreted graph, 6852 products correspond one-to-one to a feature, and the remaining 17,009 are curtain-wall members and plates excluded by an explicit decision. No product is unaccounted for, no product maps to several features, and no feature derived from an IFC product lacks an IFC counterpart; the 3907 thematic surfaces are parts of such features and are not counted separately. The exclusions are recorded per class in the conversion log and were attributed to individual products through the reference table; they are not written to the knowledge graph. In the direct translation, 23,133 products correspond one-to-one, but 728 are missing from its output: all 719 flow terminals, whose class the converter does not process, and nine doors that have no host element in IFC, which the converter reported as a warning during conversion but did not export under the default options. Direct translation is thus not automatically complete; the omissions only became visible by comparing the knowledge graph with the reference table through the IFC GlobalIds.

Traceability also exposes how a decision propagates. The IFC model relates 430 of its 439 doors and windows to a host element. The direct translation placed all 430 as fillings. The reinterpretation placed 416, because 14 doors are hosted by curtain-wall plates and the curtain wall had been excluded; these 14 and the nine doors without a host remained top-level objects, so that all 439 are present.

Table 5. IFC-product-level correspondence between IFC products with geometry and RDF features (derived thematic surfaces are not counted).

IFC Class	Products	Direct: 1:1	Direct: Missing	Reinterp.: 1:1	Reinterp.: Excluded
IfcWallStandardCase/IfcWall	1374/112	1374/112	0	1374/112	0
IfcSlab/IfcRoof/IfcCovering	372/2/69	372/2/69	0	372/2/69	0
IfcBeam/IfcColumn	1423/806	1423/806	0	1423/806	0
IfcBuildingElementProxy	1114	1114	0	1114	0
IfcDoor/IfcWindow	343/96	334/96	9/0	343/96	0
IfcFlowTerminal	719	0	719	719	0
IfcFurnishingElement/IfcStairFlight	219/4	219/4	0	219/4	0
IfcSpace	199	199	0	199	0
IfcMember/IfcPlate (curtain wall)	11,832/5177	11,832/5177	0	0	11,832/5177
Total	23,861	23,133	728	6852	17,009

6.3. Quantity Preservation (RQ2)

Two properties of the reference values and criteria of Section 4 bear on the figures of this section. Of the 23,445 reference meshes that pass the closedness test, 3963 (16.9%) return a non-positive volume from Equation (2), which indicates inconsistently oriented faces or duplicated bodies in the source geometry rather than a real volume; most of them, 3677, are curtain-wall members and plates. Within the objects used for the volume statistics below only 14 (0.4%) are affected, and excluding them changes the reported summed deviations by less than 0.02 m³. The relative and the combined criterion of Section 4.2 also differ considerably for classes dominated by small objects: among the 285 non-simplified building element proxies, for instance, 77.2% meet the relative criterion alone and 97.9% meet the combined criterion used here.

For the 3649 non-simplified objects converted by both converters (three doors without a host are missing from the direct translation), reconstruction preserved surface area to the same degree as direct translation (Figure 4a). Measured against the reference, 95.4% of these objects kept their area within tolerance in the reinterpreted graph and 95.2% in the direct translation; the summed absolute area deviation was 67.9 m² and 73.4 m², respectively, on a reference area of 566,310 m². For the 3557 closed objects among them with a positive reference volume (14 further closed objects were excluded by the rule of Section 4.1), the summed absolute volume deviation was 21.8 m³ and 19.6 m³ on a reference volume of 135,536 m³; because millimeter rounding exceeds the 0.001 m³ bound for small objects, only 56.2% and 57.2% of them met the volume tolerance in either graph. The residual deviations are shared by both converters and stem from writing coordinates with millimeter precision. The area check of the reconstruction reported no mismatch in any plane group, and, as observed for this model, floor and ceiling areas agreed for all 199 rooms (9682.4 m² of net floor area); relaxing the floor-area threshold of Section 3.2 from 8° to the 45° classification threshold leaves the net floor area of every room unchanged, because the room solids of this model contain no inclined downward faces. The shares of 95.4% and 95.2% refer to the combined criterion of Section 4.2; under the relative criterion alone they are 91.6% and 91.4%. Closedness was tested on the output as well as on the reference meshes (Section 4.1). In the station configuration, all 3200 prism-simplified objects are closed by construction, and 3353 of the 3574 non-simplified objects with a closed reference mesh (93.8%) also pass the test on their reconstructed output. The 221 that fail have a median of eight unmatched edges, typically where a T-junction inherited from the source mesh leaves a slit between two faces; they are mostly beams (65), columns (38), spaces (27), slabs (24) and walls (40). Their volumes deviate from the reference by 3.2 m³ in total on 26,683 m³, and their median relative deviation (0.12%) is the same as that of the closed outputs (0.14%), so the slits are too narrow to affect the volume. In the run without simpli-

fication (Section 6.6), where every object is reconstructed, 21,606 of the 23,445 objects with a closed reference (92.2%) have a closed output; the 1839 open outputs are concentrated in detailed objects (317 of the 343 doors, 498 of the 719 flow terminals, 552 proxies, 246 beams, 184 furnishing elements), and at the object level they meet the volume tolerance less often than closed outputs (23% against 68%), although their summed deviation remains small (18.4 m^3 on $28,298 \text{ m}^3$). Agreement of summed volumes is therefore not used as evidence of closedness; the per-object test is, and its outcome is available for every object. Surface areas are unaffected by the slits, since they are summed per polygon. The 416 objects whose reference mesh is already open, mostly ceiling coverings modeled as surfaces, are open in the output as well.

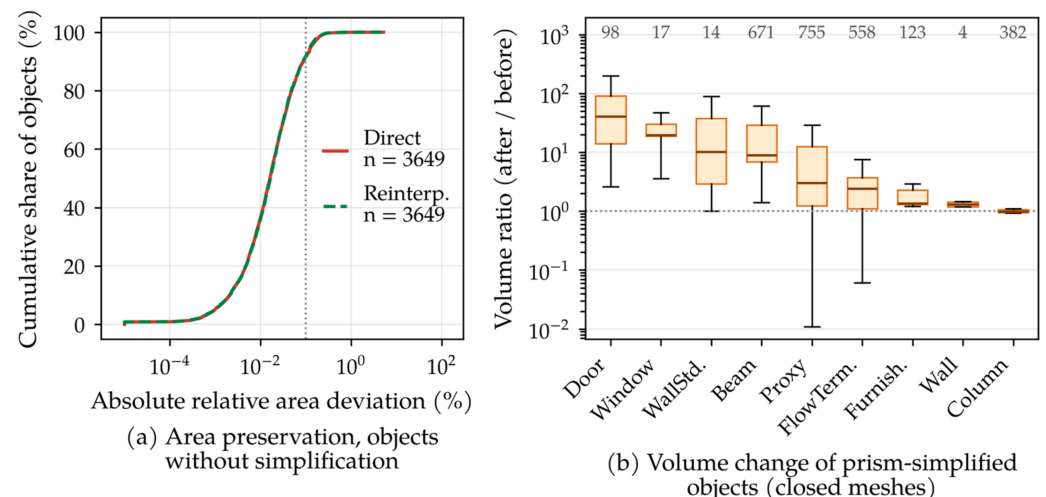


Figure 4. Quantity preservation for Suseo Station: (a) cumulative distribution of the absolute relative area deviation from the reference for objects that were not simplified, in both knowledge graphs; (b) ratio of the volume after prism simplification to the reference volume, by IFC class (closed meshes with a usable reference volume; whiskers at 1.5 IQR; the figure above each box is the number of objects). In (a), the solid red line shows the direct translation, the dashed green line the reinterpretation, and the dotted vertical line the relative tolerance of 0.1%; in (b), the dotted horizontal line marks a volume ratio of one (no change).

Losses were concentrated in prism simplification (Figure 4b; Table 6). A prism on the convex hull of a thin, detailed object fills the space around it, so that the median volume grew 40-fold for doors, 20-fold for windows, 10-fold for walls of standard case, 9-fold for beams and 3-fold for proxies, while round columns, approximated by a hull of at most 12 vertices, lost about 2%. Over the 2622 simplified objects with a usable reference volume the volume rose from 2641 m^3 to $19,035 \text{ m}^3$, which increased the summed usable volume of the evaluated objects by 11.9% ($154,569$ against $138,178 \text{ m}^3$); the sum counts overlapping elements and spaces separately and is not the physical volume of the building. Surface area behaves differently, and the difference matters for interpretation. Individual areas change substantially under simplification: the median deviation is -12.4% , and 78% of the simplified objects deviate by more than 10%. In the aggregate, however, the gains on small objects ($+10,368 \text{ m}^2$) almost cancel the losses on large ones ($-10,500 \text{ m}^2$), so that the summed surface area of the 3200 simplified objects is nearly unchanged ($39,845 \text{ m}^2$ against $39,713 \text{ m}^2$, -0.3%). Aggregate preservation must not be mistaken for object-level preservation: the area of an individual simplified object is not reliable, and the sum is preserved only because gains and losses cancel across the set, which need not hold for a spatial subset of it. The volume error, by contrast, is systematic and one-directional, because a prism on the convex hull can only enclose more space than the object it replaces. Ratios far below one may indicate reference meshes that contain overlapping or duplicated

bodies, whose volume the divergence theorem counts more than once; the reduction in the hull to 12 vertices can also lower the volume, and the two causes were not separated in this study.

Table 6. Quantity preservation for Suseo Station by IFC class. Share of objects whose area is within tolerance, and median volume ratio of simplified objects (closed meshes). Shares refer to the combined tolerance of Section 4.2; the column for the reinterpretation lists all 3652 non-simplified objects of that conversion, whereas the common-object figures quoted in Section 6.3 rest on the 3649 of them that the direct translation also converted. n is the number of simplified objects; the median volume ratio is computed over the closed meshes among them with a usable reference volume, whose per-class sizes are given in Figure 4b.

IFC Class	Direct: Area Within Tol.	Reinterp. Not Simplified: n /Area Within Tol.	Reinterp. Simplified: n /Median Volume Ratio
IfcWallStandardCase	99.4%	1359/99.4%	15/10.2
IfcWall	99.1%	108/99.1%	4/1.31
IfcSlab	97.8%	372/97.8%	–
IfcBeam	90.1%	718/85.2%	705/8.89
IfcColumn	93.8%	422/91.7%	384/0.98
IfcBuildingElementProxy	94.8%	285/97.9%	829/3.00
IfcDoor	87.7%	36/91.7%	307/40.5
IfcWindow	87.5%	78/100%	18/19.5
IfcFlowTerminal	not converted	–	719/2.41
IfcFurnishingElement	100%	–	219/1.35
IfcStairFlight	100%	4/100%	–
IfcSpace/IfcCovering/IfcRoof	100%	270/100%	–

6.4. Queryability (RQ3)

Table 7 lists the query set and the number of rows returned by each knowledge graph. Both graphs answered the queries on semantic navigation and traceability. The numbers of rows differ where a decision differs: the reinterpreted graph reports fillings on six consolidated storeys instead of ten, and maps IFC classes to more CityGML classes because proxies are classified by name. Only the reinterpreted graph answered the query on room net floor area expressed with standard CityGML vocabulary (Q2) and the query on wall surface area per storey (Q3), because the direct translation contains neither qualified quantities nor thematic surfaces. Q3 reaches the wall surfaces through the standard class `con:WallSurface`, but their areas and storey membership come from the auxiliary vocabulary of Section 3.4; it is therefore a query on standard thematic surfaces with auxiliary quantities rather than a fully standard-vocabulary query. The footprint query (Q2b), which relies on the auxiliary vocabulary, was answered by both, showing that part of the gap can be closed after conversion, but only outside the standard vocabulary.

The volume query (Q7) shows how decisions shape answers. Asked for installations and furniture of at least 20 m³, the reinterpreted graph returned 82 objects and the direct translation none. Two decisions explain the difference. First, the direct translation classifies beams, columns and proxies as constructive elements, so they fall outside the question; the reinterpretation classifies them as installations, which is why 21 beams and 29 proxies with genuine volumes above 20 m³ appear. Second, 32 of the 82 are simplified objects, and 19 of these have a reference volume below the threshold: they appear only because the prism inflated them. The same question thus receives different answers depending on class mapping and simplification, which illustrates why decisions should be switchable and recorded. The part of the difference that stems from the class mapping, on Q7 as on Q8 below, depends on the modeling choice of Section 3.2 rather than on the conversion paradigm: a reinterpretation that mapped beams and columns to constructive elements,

as the standard's examples permit for beams, would return fewer rows on Q7 and more on Q8.

Table 7. Query set for Suseo Station: rows returned and median execution time on the triple store (ms, three runs). Q8b and Q9b were run in a later session on the same machine and store (single run), so their times are indicative only and not comparable with the rest of the table.

ID	Question	Direct: Rows	Direct: ms	Reinterp.: Rows	Reinterp.: ms
Q1	Number of rooms per storey	6	64	6	49
Q2	Ten largest rooms by net floor area (core:QualifiedArea)	0	14	10	22
Q2b	Ten largest rooms by footprint area (auxiliary vocabulary)	10	35	10	15
Q3	Wall surface area per storey (con:WallSurface, auxiliary area and storey membership)	0	38	8	65
Q4	Features derived from a given IFC GlobalId	1	13	1	13
Q5	Doors and windows per storey (fillings)	10	80	6	54
Q6	IFC class to CityGML class correspondence	13	219	16	89
Q7	Installations and furniture with volume $\geq 20 \text{ m}^3$	0	30	82	47
Q8	Constructive elements of the same storey intersecting the footprint of a given room (geof:sfIntersects)	78	34	28	41
Q8b	Q8 with an additional overlap of height intervals (2.5D)	30	78	20	74
Q9	Objects of the same storey within the footprint of a given room (geof:sfWithin)	29	117	31	137
Q9b	Q9 with an additional overlap of height intervals (2.5D)	10	48	12	86

Both graphs also answered the spatial queries evaluated with GeoSPARQL functions on the room footprints. These two queries are restricted to the storey of the anchor room through the explicit storey membership of Section 3.4, so that elements on other levels are already excluded; within a storey, however, the test remains planimetric, and objects at a different height in the same storey, such as services above a suspended ceiling, are still returned. To measure this effect, Q8b and Q9b repeat the two queries with an additional condition that the height interval of the candidate object, taken from its stored minimum and maximum height, overlaps that of the anchor room. In the reinterpreted graph the 28 intersecting constructive elements reduce to 20 and the 31 contained objects to 12: eight elements and nineteen objects of the same storey lie entirely above or below the room, and the containment query is affected far more than the intersection query, because the objects that fall inside a room footprint without sharing its height are mostly ducts, cable trays and fixtures routed above its ceiling. In the direct translation the 78 intersecting constructive elements reduce to 30 and the 29 contained objects to 10. The containment query loses about two thirds of its answers in both graphs, whereas the intersection query loses 62% in the direct translation against 29% in the reinterpreted graph, because the beams, columns and proxies that the direct translation counts as constructive elements include many members that cross the footprint above the ceiling of the room. The height condition thus changes not only the counts but the ranking of the two graphs on Q8, and the figures of Table 7 for the planimetric queries should be read with this in mind. For the machine room used as the anchor, 28 constructive elements of the reinterpreted graph and 78 of the direct translation intersect its footprint (Q8); the larger number again follows from the direct translation classifying columns, beams and proxies as constructive elements. Thirty-one and 29 objects, respectively, lie within the footprint (Q9). On the triple store, every query of the set was answered within 0.22 s in both graphs (Table 7). The direct translation was not slower despite its larger size, because the queries touch semantic triples and two-dimensional footprints rather than the triangulated geometry; the cost of the larger graph lies in storage and loading, not in these queries.

6.5. Geometry Encoding in RDF (RQ4)

Table 8 compares the two encodings for the ground floor of the station. Both encodings carried identical coordinates: for all 1348 geometries, the coordinates recovered from the WKT literals and from the structural triples agreed exactly with each other. The costs differed by factors of 4.5 to 162. The retrieval times are not measured on equivalent outputs: the WKT query returns 1348 geometry literals, whereas the structural query returns 213,624 vertex records that a client must reassemble, so the ratio describes these two retrieval tasks rather than the cost of reading a single geometry. The structural encoding needed 1,283,551 geometry triples instead of 9721 (132 times as many), which made the whole storey graph 14.7 times larger in triples and 4.5 times larger on disk; the size ratio is smaller than the triple ratio because a single WKT literal holds all coordinates of a geometry. Loading took 55.1 s instead of 3.4 s, and retrieving all geometry of the storey took 23.0 s instead of 0.142 s. Semantic queries that do not touch geometry, such as counting rooms or fillings, ran in about 4 ms with either encoding. The validation case showed the same pattern (412 against 38,110 geometry triples for its ground floor).

Table 8. Geometry encoding of the ground floor of Suseo Station. Geometry retrieval returns 1348 WKT literals in the first row and 213,624 vertex records in the second, so the ratio compares two retrieval tasks rather than the cost per geometry.

Encoding	Geometry Triples	All Triples	Size	Load Time	Geometry Retrieval	Semantic Query
WKT literal	9721	92,738	14.8 MB	3.4 s	0.142 s	3–4 ms
Structural	1,283,551	1,366,568	66.5 MB	55.1 s	23.0 s	4 ms
Ratio	132	14.7	4.5	16	162	≈1

6.6. Controlled Runs: Separating the Two Decisions

The comparison of Section 6.1 combines two decisions whose effects a single run cannot separate: the exclusion of curtain walls, which changes the set of objects, and prism simplification, which changes the geometry of the objects kept. To attribute the observed differences, the station was converted three more times with the two decisions varied independently (Table 9): curtain walls retained with simplification on, curtain walls excluded with simplification off, and both switched off, which targets the same set of 23,861 IFC products as the direct translation; the direct translation, however, exports 23,133 of them, omitting the 719 flow terminals and nine unhosted doors noted in Section 6.2. All other decisions of Table 2 kept their default values, and every output was transformed to RDF and evaluated with the procedure of Section 4.

Retaining curtain walls restores the 17,009 components and raises the one-to-one correspondence to all 23,861 products, 728 more than the direct translation converts; it also hosts 430 of the 439 door and window fillings instead of 416, because 14 fillings belong to curtain-wall panels. Its cost in size is modest: the GML document grows from 114 MB to 183 MB and the semantic triples threefold, because every component carries its own wall surface, external reference and provenance link, while the geometry triples grow in proportion to the object count. Switching simplification off has the opposite profile. The object set is unchanged, but the polygon count rises sixteenfold and the document to 1484 MB, about half the size of the direct translation, and the geometry literals in RDF grow to 709 MB, close to the 795 MB of the direct translation. With both decisions off, the reinterpreted document is 1554 MB, 52% of the direct translation, and its RDF geometry 754 MB against 795 MB. Of the 26-fold size ratio of Section 6.1, simplification thus accounts for a factor of 13 (114 to 1484 MB); the curtain-wall exclusion accounts for a factor of 1.6 when simplification is on but only 1.05 when it is off, because the components are simple extrusions

that simplification does not touch; and the remaining factor of 1.9 (1554 against 2978 MB) is the difference observed between the two implementations on this model. That factor combines the representation of geometry as planar polygons rather than triangles with the content that the two outputs do not share: the 728 objects absent from the direct translation account for 18.1% of the polygons of the reinterpreted output, so that on matched inventories the factor would be about 2.3, and the attributes each converter exports also differ. It should therefore not be read as the effect of the polygon representation alone.

Table 9. Controlled runs of the reinterpretation converter on Suseo Station with curtain-wall exclusion and prism simplification varied independently, against the direct translation. Shares of objects within tolerance use the combined criterion of Section 4.2; common objects are those converted by both the run and the direct translation and not simplified; usable volumes exclude the suspect reference volumes of Section 4.1. The default column repeats Tables 4–7. Times of the reinterpretation runs were measured under uncontrolled background load and are indicative only.

Measure	Default (Station)	Curtain Walls Retained	Simplification Off	Both Off	Direct Translation
Curtain-wall exclusion/prism simplification	on/on	off/on	on/off	off/off	–
Objects converted (simplified)	6852 (3200)	23,861 (3202)	6852 (0)	23,861 (0)	23,133 (–)
Traceable one-to-one/excluded or missing	6852/17,009 excluded	23,861/0	6852/17,009 excluded	23,861/0	23,133/728 missing
Polygons	302,824	411,739	4,877,082	4,987,403	5,878,646 triangles
GML size (MB)	114.1	183.0	1484.2	1553.6	2978
RDF triples, semantic/geometry	418,341/62,873	1,327,723/198,945	403,925/61,903	1,313,299/197,975	881,409/191,676
RDF size, semantic/geometry (MB)	95.8/69.5	298.5/114.6	93.3/708.5	298.3/754.0	205.9/794.6
Conversion/RDF transformation time (min)	19.4/2.3	22.5/4.1	65.8/31.1	53.0/40.4	32.7/29.6
Peak memory of conversion (GB)	9.9	10.3	10.2	11.9	11.8
Area within tolerance, all converted objects	51.0%	85.9%	95.0%	98.6%	98.6%
Summed usable volume relative to reference	+11.9%	+11.8%	0.0%	0.0%	0.0%
Common objects with the direct translation (<i>n</i>)	3649	20,656	6124	23,133	–
Area within tolerance, common objects (run/direct)	95.4%/95.2%	99.2%/99.2%	95.1%/94.8%	98.7%/98.6%	–
$\Sigma \Delta V $ of common usable objects, m ³ (run/direct)	21.8/19.6	31.5/29.2	41.1/38.9	50.8/48.5	–
Q7 rows (installations and furniture ≥ 20 m ³)	82	82	63	63	0

Quantity preservation follows the simplification switch alone. With simplification off, the 3202 objects that the switch affects keep their area within tolerance at 94.6% (direct translation: 94.1% of the 2477 it converts) and their summed volume at 2638 m³ against

a reference of 2641 m³, so that the sevenfold inflation of Section 6.3 disappears and the summed volume of all usable objects matches the reference: 138,172 against 138,178 m³ with curtain walls excluded and 138,629 against 138,637 m³ with them retained. Over the 23,133 objects that both converters then share, 98.7% of the reinterpreted and 98.6% of the directly translated objects kept their area within tolerance, with summed absolute volume deviations of 50.8 and 48.5 m³ on 138,598 m³, and the per-class shares differ by at most 1.3 percentage points (beams, 91.4% against 90.1%). The curtain-wall components themselves, being planar extrusions, are preserved identically by both converters (all 17,009 within tolerance; 457 against 459 m³), so retaining them raises the share over all objects to 98.6% in both graphs. The share over all converted objects therefore depends on the composition of the object set, which is why Table 9 reports it alongside the share over common objects. The volume query Q7 returns 63 rows instead of 82 once simplification is off, in both curtain-wall settings: exactly the 19 objects that Section 6.4 identified as lifted above the threshold by the prism disappear, while the direct translation still returns none because of its class mapping.

The run times in Table 9 locate the time saved by simplification in polygon reconstruction and GML writing rather than in the geometry kernel, which evaluates the same products in every run: with both decisions off, reconstruction took 30 min instead of 3.2 min, and the RDF transformation 40 min instead of 2.3 min, because the geometry literals are ten times longer. These times were measured on the same laptop under uncontrolled background load and are indicative only. Taken together, the controlled runs show that the losses measured in Section 6.3 and the size reduction in Section 6.1 belong to one decision, prism simplification, and not to semantic reinterpretation as such; the curtain-wall exclusion changes completeness and size but not the preservation of what is kept.

7. Discussion

7.1. Trade-Offs Between the Two Paradigms

The results qualify the view of direct translation as a faithful baseline. In this case study, it preserved the quantities of the objects it converted. It was not complete, however: 728 products, or 3% of the model and all of its flow terminals, were missing from its output. Reinterpretation, in turn, did not lose surface area through reconstruction itself beyond the level of coordinate rounding; its losses were confined to one decision, prism simplification, and to the deliberate exclusion of curtain walls, and both are recorded in the conversion log and can be traced through the reference table. In this case, the difference between the paradigms lay less in how much was lost than in whether the loss could be seen.

Size and analytical reach separate the paradigms more clearly. In the evaluated configuration, the direct translation was 26 times larger in GML and could not answer the question on room areas posed in the standard vocabulary, nor the question on wall surfaces, which uses the standard surface class with auxiliary storey and area properties. The reinterpretation answered them, but its simplified volumes changed the answer to a volume query, and its class mapping, a choice of this study rather than of the paradigm, changed the answers to the volume and intersection queries as well. The controlled runs of Section 6.6 attribute most of the size ratio to simplification: with simplification off and curtain walls retained, the reinterpreted GML is 1554 MB, about half of the direct translation, and its RDF geometry is of the same size. This suggests a tentative guideline. Applications that need a complete inventory or reliable element volumes should use reinterpretation with both simplification and exclusion rules such as curtain-wall exclusion switched off, which in Table 9 reproduced the quantity preservation of the direct translation with a complete inventory, or direct translation checked against the source. Applications that

need spaces, surfaces and room quantities benefit from reinterpretation, provided that the decisions and their measured effects are published with the data.

The measured effects suggest when prism simplification can be left on. Its cost falls on the quantities of the simplified objects: it multiplies their volume by 7.2 in sum and by 40 for the median door, and it changes their individual surface areas by a median of 12% even though the aggregate area changes by less than one percent; the position, extent, classification and storey membership of each object are retained; switching it off removed the volume inflation entirely (138,172 against a reference of 138,178 m³ in Table 9) at the price of a thirteenfold larger document. Simplification should therefore be left on only where neither the area nor the volume of the simplified objects is read, that is, where their position, extent, classification and storey membership suffice. Visualization, indoor routing and the coarse occupancy of space by services are potential uses of this kind; they were not evaluated in this study, and the room quantities that such uses may also need come from the rooms, which are never simplified. Wherever quantities of the simplified objects are consumed, including area statistics that count them, volume-based quantity take-off, energy and HVAC calculations, material estimation and asset inventories, simplification should be switched off, accepting the larger output. Because the decision is recorded in the output, a consumer of the data can tell which of the two situations applies.

7.2. Implications for CityGML 3.0 and Its Ontologies

Three observations concern the standard and its encodings. First, CityGML 3.0 offers qualified areas and volumes but no place to record how they were derived; provenance and derivation had to be added through an auxiliary vocabulary. Second, quantities recomputed from polygonal geometry are sensitive to small closure gaps introduced by vertex removal and coordinate rounding; storing quantities at conversion time, or recomputing them about a local origin, limits this. Third, the LIRIS ontology could be used unmodified, but it leaves geometry to the GML ontology without an encoding that supports spatial queries; combining it with GeoSPARQL literals was sufficient for the queries evaluated here. The encoding experiment shows why: expanding geometry into triples preserved no coordinate or ring information that the literal lacked, but multiplied the geometry triples by more than a hundred and slowed geometry retrieval by two orders of magnitude.

The planimetric evaluation of spatial relations deserves a remark of its own, because it is the main limitation of the query results reported here. Three routes toward genuinely three-dimensional queries fit the converter-agnostic design of the transformation to different degrees. The first keeps the footprint and adds the height interval that every spatial object already carries, so that a planimetric test is combined with an interval overlap in the same SPARQL query; it needs no change to the data and no extension of the store, and it is the route we would recommend first. The second precomputes three-dimensional topological relations at transformation time and stores them as triples, which makes queries simple and fast but fixes the relations at conversion time and enlarges the graph. The third delegates the test to an external geometry engine that reads the three-dimensional WKT literals already present, which is the most general option but gives up the ability to answer the question inside the triple store. Only the first and second remain within the standard GeoSPARQL vocabulary that the transformation targets; the third would rely on an engine outside the store, as Ontop-spatial does for relational geometry [34], and the three-dimensional predicates it would need are not part of GeoSPARQL 1.1 [31] or of the stores that implement it [32,33].

7.3. Limitations and Generalizability

The main case is a single, large model from one authoring tool, and the validation case is small. The name rules for proxies rely on the naming practice of the source model. The model has no element quantities, so reference values were derived from the kernel geometry; errors in the tessellation would affect both converters equally, and for a few families with overlapping bodies the reference volume itself is doubtful; Section 4.1 defines how they are handled and Section 6.3 quantifies how many objects are affected. Spatial relationships in GeoSPARQL are evaluated planimetrically, which is a simplification for multi-storey buildings; we therefore stored heights separately. The vertical datum of the model could not be confirmed, so absolute heights were not evaluated. Further public models from other authoring tools would strengthen the generalization, and the quantitative findings should be read as a detailed case study rather than as a tool-independent benchmark; the comparison with the direct-translation converter likewise applies to the implementation and version tested here. Two decisions, curtain-wall exclusion and prism simplification, were evaluated separately in Section 6.6. The remaining decisions of Table 2, storey consolidation, proxy classification, thematic surfaces and the derived levels of detail, were evaluated only in their default setting, and interactions between decisions, such as curtain-wall exclusion changing the hosts of 14 doors, were not quantified. The reconstructed output is not closed for every object: 6% of the non-simplified objects in the station configuration and 8% in the run without simplification fail the strict edge test of Section 4.1, mostly because of T-junctions inherited from the source mesh; the defect is reported per object rather than repaired, and a consumer who requires closed solids should use the test, or the prisms, to select objects.

Heritage BIMs would test the decisions of Table 2 more severely than the station does, and the converter has not been evaluated on them; the conversion of damaged heritage models from IFC to CityGML reported by Colucci et al. [20] shows the kind of model that would be required. Such models contain vaults, domes, arches, out-of-plumb walls and moldings, and they lean heavily on IfcBuildingElementProxy with domain-specific naming. Classifying thematic surfaces by the vertical component of the normal would divide a vault or a dome inconsistently between wall, ceiling and roof surfaces; the proxy name rules would have to be rewritten for heritage vocabularies; and prism simplification would be particularly damaging for decorated elements, whose surface and volume are often the quantities a conservation project cares about. The provenance side of the approach, on the other hand, seems well suited to that domain, and the auxiliary vocabulary could be extended to record conservation-relevant statements or aligned with a heritage ontology such as CIDOC-CRM. Testing this on a public heritage model is left for future work.

How the measured costs scale with model size was not studied systematically, and the figures of Section 6.1 should not be extrapolated. Three observations bound what can be said. The geometry kernel accounts for 82% of the conversion time and grows roughly with the number of products, so the total time is dominated by a stage that the conversion paradigm does not control. The transformation to RDF streams the input object by object, so its memory demand follows the size of the largest object rather than of the document, which is why a 2978 MB input could be processed on the same laptop. The peak memory of the conversion, in contrast, comes from evaluating geometry on several threads at once and is therefore set by the thread count; on small models this parallelism brings no benefit and its overhead may dominate, while on models much larger than the station it would have to be reduced to stay within memory. The controlled runs add one observation: without simplification, polygon reconstruction and the RDF transformation rather than the kernel dominate the time, because both are proportional to the number of polygons written.

8. Conclusions

This study compared direct translation and semantic reinterpretation in IFC-to-CityGML 3.0 conversion by measuring what each preserves. We implemented a reinterpretation converter with switchable decisions, an area-preserving polygon reconstruction and a converter-agnostic transformation to RDF with provenance and quantities, and evaluated both paradigms against reference values derived from IFC geometry on a railway-station model with 23,861 products. For the objects to which simplification was not applied in the station configuration, reinterpretation preserved surface area as well as direct translation, with summed volume deviations of the same order. Its losses were concentrated in explicit decisions, above all prism simplification, whereas 728 products were absent from the output of the direct translation. Controlled runs that varied the two decisions independently confirmed that the volume loss and most of the size reduction belong to simplification and not to reinterpretation as such: with simplification off, reinterpretation matched the quantity preservation of direct translation on all 23,133 common objects (98.7% against 98.6% within tolerance) while retaining the complete inventory. Under the evaluated configuration, which excluded curtain walls and applied prism simplification, the reinterpreted model was 26 times smaller and answered questions on room areas in the standard vocabulary and on wall surfaces by standard class with auxiliary storey and area properties, which the direct translation could not, with all queries, including spatial ones, answered within a quarter of a second. Making conversion decisions explicit, switchable and traceable can turn information loss from a hidden side effect into a quantity that can be reported alongside the data. These figures describe one large model, one station configuration and one version of the direct-translation converter; the size and time ratios in particular depend on the decisions taken, as Table 9 shows. Future work will extend the evaluation to further models and authoring tools, vary the remaining decisions of Table 2 in the same way, and propose a standard way of recording derivation in CityGML 3.0 knowledge graphs.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijgi15110483/s1>, Archive S1: the conversion, transformation, evaluation, query and geometry-check scripts (ifc2citygml, gml2rdf, gml2rdf_rq4, eval_rq123, eval_rq4, run_fuseki_queries, fp_extract, count_polygons, make_table6_fig4, output_closed, room_floor_threshold); the CityGML 3.0 outputs, RDF knowledge graphs (TBox and ABox), reference values and evaluation results for the public AC20-FZK-Haus model; and the aggregated evaluation tables, controlled-run logs, closedness results and query outputs for Suseo Station.

Funding: This research was funded by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143804).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to legal restrictions. The Suseo Station model and the data derived from it were produced in a research project funded by the Korea Agency for Infrastructure Technology Advancement (KAIA). Because they contain georeferenced spatial information at a scale finer than 1:25,000, whose transfer outside the Republic of Korea is restricted under Korean law, they are not publicly available; they can be provided by the corresponding author for the purposes of peer review upon reasonable request. The conversion, transformation and evaluation code and the complete results for the public AC20-FZK-Haus model are provided in the Supplementary Materials.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BIM	Building information modeling
GML	Geography Markup Language
IFC	Industry Foundation Classes
LoD	Level of detail
OWL	Web Ontology Language
PROV-O	PROV Ontology
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
WKT	Well-known text

Appendix A

Table A1 lists the name rules by which IfcBuildingElementProxy instances are assigned to CityGML classes (Section 3.2). The rules are applied in the given order to the concatenated Name, ObjectType and Tag attributes as case-insensitive regular expressions, and the first match decides; proxies that match no rule become building installations. The last column gives the number of the 1114 proxies of the station model matched by each rule. Keywords are given in the Korean form used in the model, with English equivalents.

Table A1. Name rules for the classification of IfcBuildingElementProxy, in order of application.

Order	Rule	Keywords (Korean English)	Target Class	Surface Rule	Matched
1	furniture	가구, 의자, 책상, 테이블, 벤치, 캐비닛, 사물함 furniture, bench, chair, desk, table	BuildingFurniture	–	50
2	landscape	조경, 수목, 식재, 화단 landscap-, tree, plant	BuildingInstallation	–	0
3	sanitary	위생, 변기, 세면, 싱크 toilet, urinal, lavator-, sink	BuildingInstallation	–	0
4	sign	사인, 표지, 안내, 간판, 키오스크 sign, kiosk	BuildingInstallation	–	439
5	electric	조명, 등기구, 분전반, 배전, 전기, 트레이 light, lamp, panel	BuildingInstallation	–	1
6	mep	덕트, 배관, 급기, 배기, 디퓨저, 밸브, 펌프, 공조 duct, pipe, diffuser, valve, pump	BuildingInstallation	–	1
7	lift	엘리베이터, 에스컬레이터, 승강기, 무빙워크 elevator, escalator, lift	BuildingInstallation	–	14
8	stair	계단, 난간, 램프 stair, handrail, railing, ramp	BuildingInstallation	–	35
9	ceiling	천장, 천정, 마감, 몰딩 ceiling, finish, molding	BuildingConstructiveElement	wall	79
10	roof	지붕, 캐노피, 차양, 루버 roof, canopy, louver	BuildingConstructiveElement	roof	34
11	slab	슬래브, 바닥, 데크, 기초 slab, floor, deck, footing	BuildingConstructiveElement	slab	1
12	opening	도어, 창호, 창문, 커튼월 door (not indoor/outdoor), window, curtain	BuildingConstructiveElement	wall	0
13	wall	벽체, 벽, 파티션, 칸막이, 방화구획 wall, partition	BuildingConstructiveElement	wall	9
–	no match	–	BuildingInstallation	–	451

References

1. ISO 16739-1:2018; Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries—Part 1: Data Schema. International Organization for Standardization: Geneva, Switzerland, 2018.
2. Arroyo Ohori, K.; Diakité, A.; Krijnen, T.; Ledoux, H.; Stoter, J. Processing BIM and GIS Models in Practice: Experiences and Recommendations from a GeoBIM Project in The Netherlands. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 311. [CrossRef]
3. OGC 20-010; OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, Version 3.0.0. Open Geospatial Consortium: Arlington, VA, USA, 2021.
4. Kutzner, T.; Chaturvedi, K.; Kolbe, T.H. CityGML 3.0: New Functions Open Up New Applications. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 43–61. [CrossRef]
5. Pauwels, P.; Zhang, S.; Lee, Y.-C. Semantic Web Technologies in AEC Industry: A Literature Overview. *Autom. Constr.* **2017**, *73*, 145–165. [CrossRef]
6. Donkers, S.; Ledoux, H.; Zhao, J.; Stoter, J. Automatic Conversion of IFC Datasets to Geometrically and Semantically Correct CityGML LOD3 Buildings. *Trans. GIS* **2016**, *20*, 547–569. [CrossRef]
7. Stouffs, R.; Tauscher, H.; Biljecki, F. Achieving Complete and Near-Lossless Conversion from IFC to CityGML. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 355. [CrossRef]
8. Biljecki, F.; Lim, J.; Crawford, J.; Moraru, D.; Tauscher, H.; Konde, A.; Adouane, K.; Lawrence, S.; Janssen, P.; Stouffs, R. Extending CityGML for IFC-Sourced 3D City Models. *Autom. Constr.* **2021**, *121*, 103440. [CrossRef]
9. Noardo, F.; Harrie, L.; Arroyo Ohori, K.; Biljecki, F.; Ellul, C.; Krijnen, T.; Eriksson, H.; Guler, D.; Hintz, D.; Jadidi, M.A.; et al. Tools for BIM-GIS Integration (IFC Georeferencing and Conversions): Results from the GeoBIM Benchmark 2019. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 502. [CrossRef]
10. Kolbe, T.H. IFC to CityGML 3.0 Converter, Version 0.9.1. Chair of Geoinformatics, Technical University of Munich, 2026. Available online: <https://github.com/tum-gis/ifc-to-citygml3> (accessed on 19 September 2026).
11. Deng, Y.; Cheng, J.C.P.; Anumba, C. Mapping between BIM and 3D GIS in Different Levels of Detail Using Schema Mediation and Instance Comparison. *Autom. Constr.* **2016**, *67*, 1–21. [CrossRef]
12. Liu, X.; Wang, X.; Wright, G.; Cheng, J.C.P.; Li, X.; Liu, R. A State-of-the-Art Review on the Integration of Building Information Modeling (BIM) and Geographic Information System (GIS). *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 53. [CrossRef]
13. Song, Y.; Wang, X.; Tan, Y.; Wu, P.; Sutrisna, M.; Cheng, J.C.P.; Hampson, K. Trends and Opportunities of BIM-GIS Integration in the Architecture, Engineering and Construction Industry: A Review from a Spatio-Temporal Statistical Perspective. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 397. [CrossRef]
14. Wang, H.; Pan, Y.; Luo, X. Integration of BIM and GIS in Sustainable Built Environment: A Review and Bibliometric Analysis. *Autom. Constr.* **2019**, *103*, 41–52. [CrossRef]
15. Zhu, J.; Wu, P. BIM/GIS Data Integration from the Perspective of Information Flow. *Autom. Constr.* **2022**, *136*, 104166. [CrossRef]
16. Tauscher, H. Creating and Maintaining IFC–CityGML Conversion Rules. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4*, 115–122. [CrossRef]
17. Tauscher, H.; Lim, J.; Stouffs, R. A Modular Graph Transformation Rule Set for IFC-to-CityGML Conversion. *Trans. GIS* **2021**, *25*, 261–290. [CrossRef]
18. Zhu, J.; Wu, P.; Anumba, C. A Semantics-Based Approach for Simplifying IFC Building Models to Facilitate the Use of BIM Models in GIS. *Remote Sens.* **2021**, *13*, 4727. [CrossRef]
19. Tan, Y.; Liang, Y.; Zhu, J. CityGML in the Integration of BIM and the GIS: Challenges and Opportunities. *Buildings* **2023**, *13*, 1758. [CrossRef]
20. Colucci, E.; De Ruvo, V.; Lingua, A.; Matrone, F.; Rizzo, G. HBIM-GIS Integration: From IFC to CityGML Standard for Damaged Cultural Heritage in a Multiscale 3D GIS. *Appl. Sci.* **2020**, *10*, 1356. [CrossRef]
21. Biljecki, F.; Ledoux, H.; Stoter, J. An Improved LOD Specification for 3D Building Models. *Comput. Environ. Urban Syst.* **2016**, *59*, 25–37. [CrossRef]
22. Pauwels, P.; Terkaj, W. EXPRESS to OWL for Construction Industry: Towards a Recommendable and Usable ifcOWL Ontology. *Autom. Constr.* **2016**, *63*, 100–133. [CrossRef]
23. Rasmussen, M.H.; Lefrançois, M.; Schneider, G.F.; Pauwels, P. BOT: The Building Topology Ontology of the W3C Linked Building Data Group. *Semant. Web* **2021**, *12*, 143–161. [CrossRef]
24. Vinasco-Alvarez, D.; Samuel, J.; Servigne, S.; Gesquière, G. Towards Limiting Semantic Data Loss in 4D Urban Data Semantic Graph Generation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *8*, 37–44. [CrossRef]
25. Vinasco-Alvarez, D.; Samuel, J.; Servigne, S.; Gesquière, G. Towards an Automated Transformation of an nD Urban Data Model to a Computational Ontology Network: From UML to OWL, from CityGML 3.0 to “CityOWL”. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *10*, 231–238. [CrossRef]
26. Chadzynski, A.; Krdzavac, N.; Farazi, F.; Lim, M.Q.; Li, S.; Grisiute, A.; Herthogs, P.; von Richthofen, A.; Cairns, S.; Kraft, M. Semantic 3D City Database—An Enabler for a Dynamic Geospatial Knowledge Graph. *Energy AI* **2021**, *6*, 100106. [CrossRef]

27. Hor, A.-H.; Sohn, G.; Claudio, P.; Jadidi, M.; Afnan, A. A Semantic Graph Database for BIM-GIS Integrated Information Model for an Intelligent Urban Mobility Web Application. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 89–96. [[CrossRef](#)]
28. Usmani, A.U.; Jadidi, M.; Sohn, G. Towards the Automatic Ontology Generation and Alignment of BIM and GIS Data Formats. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *8*, 183–188. [[CrossRef](#)]
29. Nguyen, S.H.; Yao, Z.; Kolbe, T.H. Spatio-Semantic Comparison of Large 3D City Models in CityGML Using a Graph Database. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 99–106. [[CrossRef](#)]
30. OGC 22-047r1; OGC GeoSPARQL—A Geographic Query Language for RDF Data, Version 1.1. Open Geospatial Consortium: Arlington, VA, USA, 2024.
31. Car, N.J.; Homburg, T. GeoSPARQL 1.1: Motivations, Details and Applications of the Decadal Update to the Most Important Geospatial LOD Standard. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 117. [[CrossRef](#)]
32. Jovanovik, M.; Homburg, T.; Spasić, M. A GeoSPARQL Compliance Benchmark. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 487. [[CrossRef](#)]
33. Kyzirakos, K.; Karpathiotakis, M.; Koubarakis, M. Strabon: A Semantic Geospatial DBMS. In *The Semantic Web—ISWC 2012; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7649, pp. 295–311.
34. Bereta, K.; Xiao, G.; Koubarakis, M. Ontop-Spatial: Ontop of Geospatial Databases. *J. Web Semant.* **2019**, *58*, 100514. [[CrossRef](#)]
35. Lebo, T.; Sahoo, S.; McGuinness, D. *PROV-O: The PROV Ontology*; W3C Recommendation; World Wide Web Consortium: Cambridge, MA, USA, 2013.
36. Biljecki, F.; Ledoux, H.; Du, X.; Stoter, J.; Soon, K.H.; Khoo, V.H.S. The Most Common Geometric and Semantic Errors in CityGML Datasets. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *4*, 13–22. [[CrossRef](#)]
37. Ledoux, H. val3dity: Validation of 3D GIS Primitives According to the International Standards. *Open Geospat. Data Softw. Stand.* **2018**, *3*, 1. [[CrossRef](#)]
38. IfcOpenShell: The Open Source IFC Toolkit and Geometry Engine, Version 0.8.5. Available online: <https://ifcopenshell.org> (accessed on 19 September 2026).
39. QUDT: Quantities, Units, Dimensions and Types. Available online: <https://qudt.org> (accessed on 19 September 2026).
40. EPSG:5186; Korea 2000/Central Belt 2010. In EPSG Geodetic Parameter Dataset. International Association of Oil & Gas Producers (IOGP): London, UK, 2010. Available online: <https://epsg.org> (accessed on 1 October 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.