

Article

Not peer-reviewed version

BIM2BI: An ETL Architecture Based on openBIM Standards for Integrating BIM Data into Business Intelligence Environments

[Diego Jesús Sánchez García](#)* and [Rafael Vicente Lozano Díez](#)

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1489.v1

Keywords: Building Information Modeling (BIM); business intelligence; Industry Foundation Classes (IFC); Information Delivery Specification (IDS); ETL pipeline; data lineage



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

BIM2BI: An ETL Architecture Based on openBIM Standards for Integrating BIM Data into Business Intelligence Environments

Diego Jesús Sánchez García * and Rafael Vicente Lozano Díez

School of Building Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

* Correspondence: diegojsanchez@gmail.com

Abstract

The construction (AEC) industry has consolidated Building Information Modelling (BIM) as the standard for producing and managing project information, yet its analytical exploitation in Business Intelligence (BI) environments remains manual, ad hoc and dependent on proprietary platforms. Existing literature addresses partial aspects of the problem — IFC extraction, dashboards, semantic approaches, data quality— without articulating a coherent architecture that integrates dimensional modelling, open-standard-based ETL, granular lineage and pre-ingestion validation. This work proposes BIM2BI, a BIM-BI integration architecture organised into four functional layers (data sources, transformation and orchestration, analytical storage and exploitation) that formalises a separation of responsibilities, explicit data contracts between layers and an extensibility-without-redesign principle. The architecture is grounded in the openBIM standards IFC, IDS and BCF, adopts the IfcGlobalId as a technical key for end-to-end lineage and uses IDS as a pre-ETL quality gate with a staged three-level validation strategy. The proposal is validated empirically through an open-source reference implementation (MIT licence) applied to ten representative use cases grouped in three complexity profiles. The results demonstrate the viability of the architecture in terms of data quality, traceability, reproducibility and scalability, and document empirical findings on the real behaviour of openBIM standards within automated analytical workflows.

Keywords: Building Information Modeling (BIM); business intelligence; Industry Foundation Classes (IFC); Information Delivery Specification (IDS); ETL pipeline; data lineage

1. Introduction

Building Information Modelling (BIM) has consolidated its role as the standard for producing and coordinating project information in the architecture, engineering and construction (AEC) sector. However, BIM was not originally conceived for the systematic analysis of project information: the data stored in the models are heterogeneous, scattered across disciplines and semantically weak for structuring consistent analytical workflows [1]. The result is a widely documented paradox in the recent literature: a growing number of AEC organisations accumulate BIM models in their Common Data Environments, yet only a minority exploit that information via advanced analytics, big data integration or Business Intelligence tools [2].

This underutilisation manifests itself in daily sector practice. Updating budgets from the model still requires significant manual effort for exporting, cleansing and reconciliation, preventing real-time tracking of cost deviations [3]. Historical analysis of changes across projects is practically non-existent, with the consequent loss of institutional learning [4]. And risk and cost-overrun forecasting lacks the analytical support that is routine in other sectors [5]. The causes of this situation extend beyond purely technological limitations and are structured around four technical barriers: structural heterogeneity without automated reconciliation mechanisms, semantic limitations inherent to the practical use of the IFC standard, absence of traceability and lineage from BIM models to analytical

indicators, and dependence on manual integration that prevents scalability. To these a transversal aggravating factor is added: the widespread lack of data governance frameworks at the organisational level [6,7].

The contrast with digitally mature sectors is notable. In manufacturing, retail, banking and logistics, Business Intelligence has consolidated as a core capability, with automated ETL architectures, advanced analytics and data-governance frameworks that sustain substantial improvements in productivity and operational efficiency [8,9]. The AEC sector, by contrast, maintains a disproportionate dependence on personal experience and intuition compared with industries where decision-making is systematically supported by BI platforms [10]. Information produced during design and construction phases is seldom reused continuously for asset management or for optimising decisions across the life cycle [11], and the native analytical capabilities being incorporated by commercial BIM platforms operate on pre-configured metrics, premium licensing models with high total cost of ownership, perpetuating dependence on closed proprietary ecosystems [12]. The industrial relevance of the problem is further evidenced by the direction the main software vendors are taking: Autodesk released in 2025 a public roadmap for its AEC Data Model, with a public-beta API for granular access to model data within Autodesk Construction Cloud [13], and in March 2026 consolidated its cloud offering under the unified Autodesk Forma brand –which integrates the former Autodesk Construction Cloud and its associated modules (Docs, BIM Collaborate Pro, Build, Takeoff)– as confirmation of a strategic shift towards an end-to-end AEC data ecosystem [14]. The fact that the market-leading vendor has made this agenda a public part of its roadmap shows that the extraction and analytical exploitation of BIM data outside the authoring environment is a recognised market need, although these moves remain anchored to the proprietary ecosystem and therefore do not resolve the need for open, reproducible and platform-independent architectures. This structural gap between the BI paradigm consolidated in other sectors and its effective application in AEC, together with the limitations of the proprietary approaches that have begun to emerge, motivates the present research.

The research is guided by the following question: can an ETL architecture based on openBIM standards reduce the manual effort of BIM data preparation and, at the same time, improve the traceability and quality of the analytical datasets of AEC projects? From this question the working hypothesis is formulated that it is possible to define and validate a BIM-BI integration architecture that significantly reduces the manual intervention required in project data preparation, while at the same time preserving the traceability and quality of the information throughout the full extract, transform and load flow. The architecture relies on the openBIM standards Industry Foundation Classes (IFC), Information Delivery Specification (IDS) and BIM Collaboration Format (BCF), and on consolidated data-engineering practices, with the goal of generating coherent, verifiable datasets fully compatible with Business Intelligence tools.

The overall aim of the work is to design and validate a BIM-BI integration architecture that combines an ETL pipeline with openBIM standards and that significantly reduces the time and manual effort of data preparation, while increasing the traceability and quality of analytical datasets in AEC projects. This aim is specified in six specific objectives: (i) to analyse the current state of BIM-BI integration through a systematic review, identifying gaps in interoperability, data quality and traceability; (ii) to design an analytical data model that structures BIM information in schemas compatible with BI tools; (iii) to develop an automated ETL pipeline that extracts, validates and transforms data from IFC files by applying openBIM standards to guarantee quality and traceability; (iv) to implement the architecture in representative sector use cases; (v) to evaluate the proposal quantitatively and qualitatively against traditional workflows; and (vi) to formulate a reproducible methodological framework and recommendations for its transfer to professional practice.

The contributions of the work are fourfold. First, a scientific contribution: the BIM2BI architecture in four functional layers (sources, transformation and orchestration, analytical storage and exploitation) that formalises the separation of responsibilities, data contracts between layers and an extensibility-without-redesign principle absent from the previous literature. Second, a

methodological contribution: the reinterpretation of the IDS standard as a pre-ETL quality gate within automated analytical workflows, and the development of a three-level staged validation strategy applicable to high-volume scenarios. Third, a practical contribution: an open-source reference implementation published under the MIT licence that materialises the architecture in a portable, reproducible system. And fourth, a contribution to the state of the art: empirical evidence on the real behaviour of openBIM standards in analytical workflows, including findings not previously documented on the reference implementation of the IDS 1.0 schema.

The remainder of the article is organised as follows. Section 2 places the work in context by characterising the digital evolution of the AEC sector, the openBIM standards as technical enablers, the current data-governance frameworks and prior work on BIM-BI integration, with explicit identification of the gaps that the proposed architecture addresses. Section 3 describes the research methodology based on Design Science Research, the phases of the work and the evaluation criteria. Section 4 presents the BIM2BI architecture layer by layer and introduces the reference implementation. Section 5 documents the empirical validation on ten use cases grouped in three complexity profiles and summarises the most relevant findings. Section 6 discusses the results against the state of the art, and Section 7 synthesises the conclusions and future lines of work.

2. Related Work and Background

2.1. Digital Evolution of the AEC Sector: From BIM to the Data Economy

Over the past four decades the architecture, engineering and construction (AEC) sector has undergone a progressive digitalisation trajectory, moving from CAD systems to parametric BIM models and Common Data Environments (CDEs). Whereas CAD digitalised graphical documentation without integrating semantic information, the consolidation of BIM meant a conceptual leap: the model ceased to be a graphical container and became a structured repository of multidimensional information in which each construction element carries semantic, geometric, relational and temporal attributes [15,16].

The subsequent emergence of CDEs (Autodesk Construction Cloud, Trimble Connect, BIM 360, Procore and similar platforms) professionalised collaborative documentation management and paved the way for standards such as PAS 1192 and ISO 19650. However, digital centralisation has not eliminated information silos: every discipline still generates data in proprietary formats, and converting that documentary mass into analytical information useful for decision-making remains one of the most complex challenges in the sector. The recent literature agrees that the genuine leap towards a data economy only occurs when repositories are able to integrate and transform heterogeneous sources into structured, traceable and auditable flows [17,18].

Business Intelligence (BI) constitutes the consolidated paradigm for this transformation [19]. In sectors such as retail, banking, logistics and manufacturing, BI architectures have reached a maturity that enables demand forecasting, inventory optimisation, fraud detection and service personalisation [20]. By contrast, BI adoption in AEC remains predominantly descriptive, occasional and fragmented, relying on isolated sources and on limited integration of the information generated across the project life cycle [2]. The native analytical capabilities of commercial BIM platforms partially mitigate this shortfall but at the same time intensify it: they are limited to pre-configured metrics, are commercialised as premium modules with a high total cost of ownership, and perpetuate dependence on closed proprietary ecosystems [12]. The need for BIM-BI architectures grounded in open standards and independent of platform constitutes the central motivation of the work presented in this article.

2.2. openBIM Standards as Enablers: IFC, IDS and BCF

The openBIM ecosystem is articulated around three complementary standards. Industry Foundation Classes (IFC) [21] provides the structured data model representing geometry, properties, spatial relationships and metadata of the project's construction elements, and is the primary data

source for any analytical architecture. The Information Delivery Specification (IDS) [22] specifies, in a machine-readable form, the informational requirements that an IFC model must meet to be suitable for a specific use case, enabling automated conformance validation. The BIM Collaboration Format (BCF) [23] records incidents, comments and decisions from the BIM coordination process, contributing causal traceability of changes made to IFC models between versions.

Despite their fundamental contributions, the openBIM standards leave at least five relevant gaps unresolved for the analytical exploitation of BIM data. The first is post-import semantic normalisation: IFC does not guarantee that the heterogeneous Property Sets of different projects use consistent naming conventions, so any analytical architecture requires an explicit mapping layer [24]. The second is analytical transformations: the standards do not define how to compute derived indicators, aggregations or KPIs. The third is the granular lineage of transformed data: although the `IfcGlobalId` allows entities to be tracked across versions, the standards do not document the ETL transformations applied [25]. The fourth is the organisational governance of data: roles, responsibilities and quality policies are not covered by technical standards. And the fifth is analytical dimensional modelling: IFC is an object-oriented model, not a dimensional one, so efficient exploitation in BI tools requires an explicit transformation to star schemas [26].

2.3. Data Governance: ISO 19650 and DAMA-DMBOK

Information management in BIM projects is regulated mainly by the ISO 19650 standard [27], derived from the UK BIM Level 2 initiative, which establishes the CDE as the centralised source of project information and defines a hierarchy of information requirements (OIR, AIR, PIR, EIR) together with a BIM Execution Plan (BEP) and a set of key roles (appointing party, appointed party, information manager). However, the standard focuses predominantly on documentary management and collaborative coordination, without explicitly addressing the systematic analytical exploitation of the information contained in the models nor the transformation towards enterprise analytical schemas [28,29].

By contrast, the DAMA-DMBOK framework [30] provides a comprehensive approach to enterprise-level data governance and articulates eleven knowledge areas, among which stand out, for this work, Metadata —with data lineage as the capability of tracking origin, transformations and destination— and Data Quality, with dimensions of integrity, consistency, accuracy and timeliness. These mechanisms, consolidated in enterprise BI environments through tools such as Apache Atlas, Marquez or Collibra, remain absent from the BIM-BI implementations documented in the literature, where traceability is limited to version management of IFC models without tracking the ETL transformations or the propagation of the `IfcGlobalId` through to the final BI indicators [17].

The comparative analysis between ISO 19650 and DAMA-DMBOK evidences an explicit conceptual gap: the roles defined in the BIM framework do not contemplate figures equivalent to Data Steward, Data Engineer or BI Analyst, and the normative traceability in ISO 19650 —version management and BCF— does not cover the attribute- and field-level data lineage that DAMA-DMBOK requires in mature analytical environments. The explicit articulation of both frameworks is a methodological requirement for any BIM-BI architecture that aspires to end-to-end auditability.

2.4. Prior Work on BIM-BI Integration and Gap Analysis

The literature on BIM-BI integration can be grouped into four complementary but not integrated lines of work. The first addresses extraction via direct querying of IFC files, with works such as Barzegar et al. [31] that use PostGIS and FME as ETL tooling to load IFC models into relational databases with spatial capabilities. This line enables the systematic extraction of information but replicates the object-oriented structure of IFC without transforming it into a dimensional model suitable for OLAP analysis and does not integrate enterprise BI tools.

The second line explores the semantic web and ontologies. Pauwels and Terkaj [32] developed `ifcOWL` as an OWL representation of the complete IFC schema; Rasmussen et al. [33] proposed the Building Topology Ontology (BOT) as a lightweight alternative; and Kone and Mahesh [34] presented

a bidirectional ontology-driven workflow to integrate project-management data with IFC. Subsequent work has also explored the validation of ifcOWL instances through constraint languages such as SHACL (Shapes Constraint Language), converting EXPRESS schema rules into constraints able to detect inconsistencies in RDF graphs, although the approach operates exclusively in the RDF domain without connecting to dimensional schemas or BI tooling. These semantic approaches contribute advanced reasoning and rich semantics but introduce technical complexity and a learning overhead that hinders their transfer to productive environments, and do not connect with conventional BI tooling or dimensional schemas.

The third line develops BIM-BI dashboards for project management: Rodrigues et al. [12] integrate Revit with Power BI for construction-site monitoring; Apellániz et al. [35] present 3D dashboards for visualising life-cycle assessment (LCA) results; Di Giuda et al. [36] articulate BIM-GIS-BI for the management of university assets; and Abdulaziz and Ahmed [37] propose a conceptual framework for BIM 5D with Power BI. These works demonstrate the value of integration in specific use cases, but the extraction depends on manual exports or vendor-specific connectors, the schemas do not follow a systematic dimensional design and none of them documents lineage mechanisms that allow BI indicators to be tracked back to specific IFC entities.

The fourth line addresses data quality and governance. Zhang et al. [38] propose quality metrics for IFC models oriented to geometric and semantic consistency; the IDS standard contributes the automated validation of informational requirements over IFC models [22]. Nevertheless, these metrics are oriented to pre-coordination validation, not to the preparation of analytical data, and IDS validation operates on native IFC data, not on the dimensional schemas resulting from an ETL process.

The comparative synthesis of these four groups (Table 1) identifies four transversal gaps that no prior work addresses jointly: (i) the absence of systematic dimensional modelling that transforms IFC data into star schemas compatible with consolidated BI tools [39]; (ii) the lack of granular data lineage that allows BI indicators to be tracked back to IFC entities through persistent identifiers (IfcGlobalId); (iii) the absence of automated pre-ingestion validation integrated as a quality gate in an automated BIM-BI ETL pipeline; and (iv) dependence on proprietary platforms, derived from the use of exports from specific authoring software instead of operating on the open IFC standard as the input format. These gaps configure the research space that the BIM2BI architecture addresses jointly in the following sections.

Table 1. Comparative synthesis of prior BIM-BI integration approaches in the academic literature, grouped by methodological focus. None of the four groups addresses jointly the four transversal gaps identified in Section 2.4 (systematic dimensional modelling, granular data lineage, automated pre-ingestion validation, and independence from proprietary platforms).

Approach	Representative works	Strengths	Limitations for BIM-BI
Direct IFC query	Barzegar et al. [31]	Systematic extraction; open-source tools; low technical overhead	No dimensional modelling; no data warehouse; no enterprise BI integration
Semantic web and ontologies	Pauwels & Terkaj [32]; Rasmussen et al. [33]; Kone & Mahesh [34]	Advanced reasoning; rich semantics; cross-domain interoperability	Limited scalability; high technical overhead; no conventional BI; massive graphs
BIM-BI dashboards	Rodrigues et al. [12]; Apellániz et al. [35]; Di Giuda et al. [36]	Real use cases; demonstrated value; interactive visualisation	Ad-hoc pipelines; no dimensional schema; no lineage; vendor dependency

Quality and governance	Zhang et al. [38]; IDS [22]	IFC quality metrics; automated validation of informational requirements	No post-ETL governance; no BIM-to-BI lineage; coordination-oriented, not analytics-oriented
------------------------	-----------------------------	---	---

3. Materials and Methods

3.1. Research Paradigm and Design Science Research Approach

The research is framed within a pragmatic paradigm oriented to the generation of knowledge useful for professional practice [40] and adopts Design Science Research (DSR) as its methodological framework [41]. This choice responds to the prescriptive nature of the problem: the aim is not to describe an existing phenomenon but to design and evaluate an artefact that solves a concrete need of the AEC sector. The BIM2BI architecture is posed as a DSR artefact simultaneously of model and method type —a conceptual model for organising BIM data and a reproducible method for transforming IFC models into analytical datasets— and is built by articulating the three classical DSR cycles [42]: the relevance cycle, which identifies and characterises the BIM-BI gap and the needs for data quality and traceability in AEC projects; the rigour cycle, which relies on the systematic review of the state of the art and on the adoption of consolidated standards and frameworks (IFC, IDS, ISO 19650, DAMA-DMBOK); and the design cycle, which encompasses the iterative construction of the architecture, its implementation and its evaluation in controlled scenarios. This structure is complemented by a dual reasoning logic: deductive, to contrast the architecture against manual workflows in delimited scenarios, and inductive, to identify emerging patterns, limitations and improvement opportunities during repeated implementation on different cases.

3.2. Research Phases

The work is organised into three phases that adapt the DSR process model proposed by Peffers et al. [43] to the BIM-BI integration problem. Phase 1 —problem determination and scope delimitation— establishes the conceptual foundations through systematic literature review, analysis of the BIM-BI gap, identification of relevant data typologies (entities, properties, classifications) and formulation of information and quality requirements. Phase 2 —design and implementation— materialises the artefact: it formalises the data-engineering framework, specifies the analytical data model, designs the ETL pipeline with automated IDS-based validation, articulates the governance and lineage mechanisms and develops the functional prototype. Phase 3 —verification and review— applies the architecture to the use cases, evaluates the results against the defined criteria and derives recommendations and future lines of work.

3.3. Evaluation Criteria and Validation Scenarios

Artefact evaluation combines technical and practical-utility dimensions aligned with the recommendations of DSR [41] and with consolidated data-management frameworks such as DAMA-DMBOK [30]. Five criteria are adopted: manageable complexity level (the pipeline's ability to adapt to models of different scale and structure without complete redesign), level of automation (proportion of tasks that can be executed without manual intervention), execution time, data quality and traceability (integrity, consistency and lineage down to specific model versions) and interoperability and scalability (the ability to integrate new sources and use cases without in-depth reconfiguration of the architecture).

Validation is carried out in controlled but representative scenarios constructed to cover three sources of variability: different levels of geometric and semantic complexity (from simple architectural models to federations integrating architecture, structure and MEP systems), different data-quality conditions (clean models versus models with typical gaps in properties and classifications) and different analytical use cases (element quantification, cost indicators, space

management and functional-distribution analysis). The concrete materialisation of these scenarios in ten use cases grouped in three complexity profiles is described in Section 5.

3.4. Technology Stack and Experimental Environment

The experimental environment is executed on a Windows 11 local-development machine, without connection to external servers or network dependencies, so that the results are reproducible by any researcher or professional without additional infrastructure. The complete technology stack –Python 3.11 with the libraries ifcopenshell, ifctester, pandas and SQLAlchemy; Streamlit interface; SQLite 3.45 repository; and Microsoft Power BI Desktop connected through a Python connector– is detailed in Section 4.6 as part of the description of the reference implementation. Full deployment of the system in a clean environment requires three steps (cloning the repository, installing Python dependencies and executing the launcher INICIAR.bat) with a total time below five minutes. Each use case is configured through a delivery.meta.json file generated by the interface itself during the delivery-preparation flow, and the IDS validation profiles are elaborated specifically for each case from the informational requirements of the reference project. IFC models are exported directly from Autodesk Revit when the native file is available, ensuring coherence between authoring model and interchange file.

4. The BIM2BI Architecture

4.1. Overview and Design Principles

The BIM2BI architecture is conceived as an intermediate data-integration layer between BIM models and Business Intelligence platforms. Its purpose is to transform the artefacts contained in openBIM deliveries –IFC models [21], IDS specifications [22], BCF incidents [23] and FIEBDC-3 budgets– into a stable, traceable and reusable analytical model that serves as a generic basis for different types of analysis (cost, space, informational quality, operation) without the need to redesign the integration pipeline for each use case.

The design of the architecture is articulated around five structural principles: (i) explicit separation of responsibilities between orchestration, processing logic, configuration and data; (ii) isolation by project and delivery, so that each execution is managed independently and enables the comparative analysis of a project over time; (iii) process reproducibility, understood as the ability to reconstruct a specific execution from its inputs and configuration; (iv) data traceability and lineage, so that every analytical observation can be tracked back to its originating IFC entity; and (v) controlled extensibility, which allows the incorporation of new sources, requirements or dimensions without redesigning the fundamental structure of the system.

Effective operation of the architecture requires the formalisation of three specific roles without consolidated referents in the AEC literature: the BIM Data Steward, responsible for guaranteeing the quality and governance of the data; the BIM Data Engineer, responsible for the design and maintenance of the ETL pipeline; and the BIM/BI Analyst, who transforms data into actionable information for decision-making. The technical materialisation of the architecture is articulated in four functional layers (Figure 1), whose operational responsibilities are described in the following subsections.

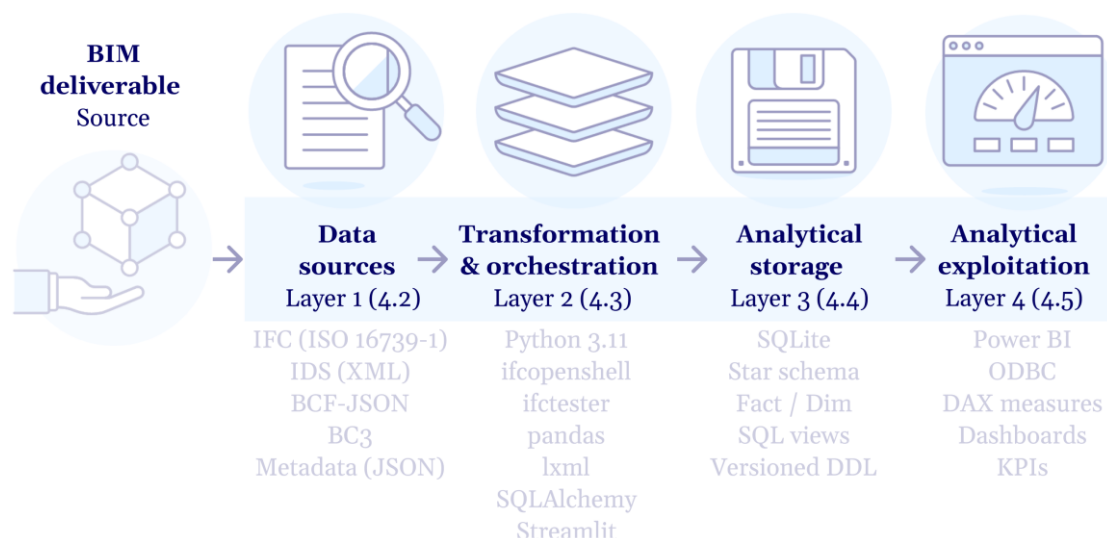


Figure 1. BIM2BI four-layer architecture. The data flow moves from the source deliverables –IFC, IDS, BCF and FIEBDC-3 budgets– through an ETL pipeline and an analytical repository to the BI exploitation layer. Each layer assigns explicit and non-overlapping responsibilities and is supported by an open-source technology stack.

4.2. Layer 1: Data Sources

Layer 1 is the entry point of the system and is conceived as a passive repository of raw information: it receives, versions and contextualises the data without applying any transformation or validation process. This passivity is a deliberate design decision that preserves data traceability from its origin and prevents early contamination of the pipeline with unaudited transformations. The layer also acts as the formal informational-contract point between the project’s production environment and the analytical system.

Two typologies of sources are distinguished. The sources associated with the delivery (inbox/) represent the informational state of the project at a given moment –IFC models, BC3 budget files, BCF incidents and applicable IDS files– and are incorporated into the system in a versioned manner under the hierarchy `inbox/{project_id}/{delivery_id}/`. The reference and requirement sources (reference/), external to any specific delivery, define the normative frameworks, reusable IDS profiles, classification dictionaries and validation rules that govern the interpretation and transformation of the data.

Each delivery is accompanied by a `delivery.meta.json` file that describes its identity (project, delivery, author, date, milestone, applied IDS profile) and a `_READY.flag` marker that triggers the pipeline execution. This organisation materialises the principle of isolation by project and delivery and enables the management of the data life cycle –retention, correction or selective deletion– to be applied at individual-delivery granularity, covering an operational gap of existing BIM governance frameworks.

4.3. Layer 2: Transformation and Orchestration

Layer 2 is the operational core of the architecture. It is articulated as an ETL pipeline orchestrated under the principles of modularity and lineage control, whose objectives are: to isolate transformation processes from the original sources, to preserve the `IfcGlobalId` identifiers as the axis of traceability, to systematically validate compliance with the informational requirements through IDS, to normalise the extracted information into structures compatible with analytical models and to record the intermediate pipeline state as support for auditing.

The reference implementation materialises this layer through twelve decoupled Python scripts that independently execute the different phases of the process: IFC model extraction with the

ifcopenshell library (extract_ifc.py), extraction and parsing of FIEBDC-3 budgets (extract_bc3.py, parse_bc3.py), normalisation of BCF incidents in the four active variants of the standard (parse_bcf.py), IDS validation with ifctester (validate_ids.py), transformation to the star schema (transform_star.py), load into the analytical repository (load_dw.py), DDL schema management and versioned migrations (init_dwh.py), generation of delivery audit records (generate_report.py), structured logging of executions (pipeline_logger.py) and complete orchestration of the flow (run_pipeline.py). This separation favours reuse, debugging and the progressive evolution of the system without compromising the global architecture.

The orchestration of the phases is performed through a Streamlit interface that sequentially invokes the scripts from the user's confirmation, ensuring that the state of each delivery is visible at all times. The operational flow relies on a folder structure that separates raw data (inbox/), intermediate processing artefacts (staging/), output documentation (reports/) and automatic backups of the analytical repository (backups/).

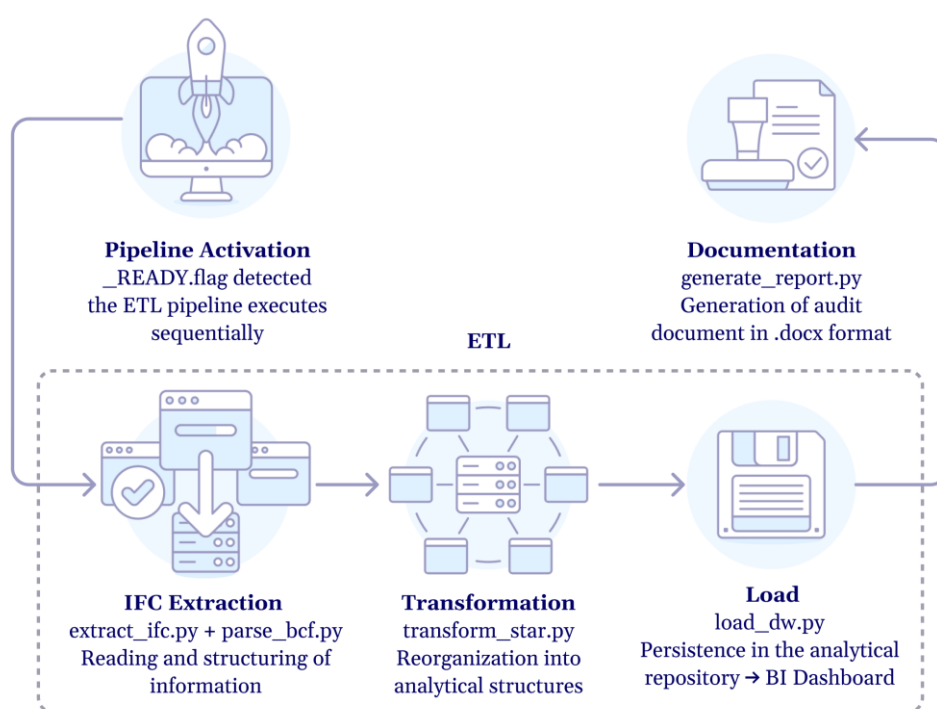


Figure 2. Operational flow of the BIM2BI ETL pipeline. The pipeline is triggered by the _READY.flag marker and executes sequentially five phases: IFC extraction and BCF parsing (extract_ifc.py, parse_bcf.py), transformation to the star schema (transform_star.py), load into the analytical repository (load_dw.py) and generation of the delivery audit document in .docx format (generate_report.py). The IDS quality gate, described in Section 4.3.1, is invoked between extraction and transformation and conditions the continuation of the flow.

4.3.1. Pre-ETL IDS Quality Gate

The pipeline incorporates a quality gate based on the IDS standard [22] that is executed before any analytical transformation. Its function is to guarantee that only IFC models meeting the minimum informational requirements of the use case advance towards the analytical repository, preventing the propagation of incomplete or inconsistent data to subsequent layers. This reinterpretation of the IDS standard – originally conceived as a validation mechanism for interchange between BIM authoring and coordination tools – extends its scope of application to data-quality governance within automated analytical pipelines.

The implementation provides two execution levels. When the ifctester library is available a full validation is run against the IDS file based on the official XSD schema of buildingSMART; otherwise

the pipeline applies a set of basic checks on the intermediate extracted data, preserving system operability in environments with limited dependencies. Detected deviations are recorded in BCF-JSON format and subsequently loaded into the fact_incident table, turning data quality into an analytical metric comparable over time.

The quality-gate result classifies each delivery as CONFORMING, NON-CONFORMING or PENDING. NON-CONFORMING deliveries only register the dim_delivery dimension with the attempt and its incidents, without contaminating the analytical fact tables; in this way the longitudinal traceability of the project is preserved without degrading the aggregated indicators. This differentiated treatment makes it possible to document the rejection of a delivery as an auditable fact, not as a valid analytical record.

4.3.2. Staged IDS Validation Strategy (L1/L2/L3)

The application of the monolithic quality gate on high-volume deliveries revealed an operational limitation: in a linear-infrastructure case with twenty-five IFC models and an approximate volume of two gigabytes, a single profile with thirty-eight rules generates an unmanageable number of incidents that prevents effective prioritisation and correction by the BIM team. In response, a staged IDS validation strategy was developed that distributes the rules across three independent profiles according to their impact on the BIM2BI pipeline.

The stratification criterion is grounded in three functional questions derived from the layered architecture. Level 1 (critical gate) groups the rules whose failure prevents the correct load into the analytical repository (GlobalId, asset identification, discipline); its failure blocks model ingestion. Level 2 (warnings) groups the rules that degrade dashboard indicators or compromise the integrity of the analytical dimensions (complete identification, minimum location, classification, main budget item); its failure generates high-priority BCF incidents but allows conditional ingestion. Level 3 (informative) groups the remaining rules, associated with model informational-maturity requirements whose failure does not affect the immediate indicators but documents the state of the model for phase closure.

The execution flow is sequential: the L1 profile is evaluated first on all delivery models; models that do not pass it are rejected and notified to the BIM team through BCF; those that pass it are processed by Level 2, which generates high-priority incidents without blocking ingestion; Level 3 is then executed asynchronously to produce the informational-maturity report. This stratification establishes an explicit correspondence between validation levels and architectural layers: L1 protects the integrity of Layer 3, L2 protects the analytical quality of Layer 4, and L3 documents maturity for long-term data governance (Table 2).

Table 2. Rule distribution in the staged IDS validation strategy applied to the infrastructure use case (CDU_004). Thirty-eight rules are distributed across three IDS profiles according to their impact on the BIM2BI pipeline: Level 1 protects the integrity of the analytical repository (Layer 3), Level 2 protects the analytical quality of the dashboards (Layer 4), and Level 3 documents informational maturity for long-term data governance.

Level	Name	Rules	Behaviour on failure	Affects
L1	Critical gate	3	Blocks ingestion	GlobalId, DGC_GEN asset code, DGC_IDE discipline
L2	Warnings	12	Generates high-priority BCF; ingestion conditional	Complete identification, minimum location, RCE classification, main budget item
L3	Informative	23	Generates low-priority BCF; does not block	Extended RCE, location PKs, plan references, complete measurement and additional items

4.4. Layer 3: Analytical Storage

Layer 3 is the analytical repository of the system and acts as the pipeline's convergence point: it is the final destination of the processed data and the starting point of the analytical exploitation. Unlike the preceding layers, oriented to processing and validation, this layer is conceived as a stable persistence layer. The data that reach it have passed the IDS quality gate and have been normalised in accordance with the informational requirements defined for each project. Its design directly conditions the analytical capabilities available in Layer 4, determining which metrics can be computed, with what granularity and with what capacity for comparison between projects and deliveries.

The analytical data model is grounded in a star schema [39], widely adopted in Business Intelligence architectures for its direct compatibility with consolidated BI tools, its explicit separation between metrics and context, and the controlled extensibility that allows new fact or dimension tables to be incorporated without altering the existing ones.

4.4.1. Star Schema Dimensional Model

The analytical schema of the reference implementation v1.0.0 includes four main fact tables. `fact_element` records the instances of construction elements of the model with their associated metrics (IFC type, net area, volume, discipline), identified by their `IfcGlobalId`. `fact_space` records the instances of spaces and rooms (surface, volume, level, intended use) and enables functional-programme and spatial-efficiency analyses. `fact_cost` records the budget items from BC3 files linked to the model elements when the budget references the `IfcGlobalId` in its `COMENTARIO` field, closing the cycle between the BIM model and its associated cost. `fact_incident` records the coordination incidents from BCF files, likewise linked to the model element, and makes it possible to analyse the evolution of the project's informational quality across deliveries.

These tables are contextualised through six dimensions: `dim_model` describes the model and the delivery (IFC version, discipline, authoring tool, applied IDS profile); `dim_element`, `dim_space`, `dim_cost_item` and `dim_incident_type` provide the descriptive attributes of elements, spaces, budget items and incidents respectively; and `dim_delivery` records the attributes of each delivery (date, author, quality-gate result, processing metrics), enabling longitudinal analysis and monitoring of pipeline performance (Figure 3).

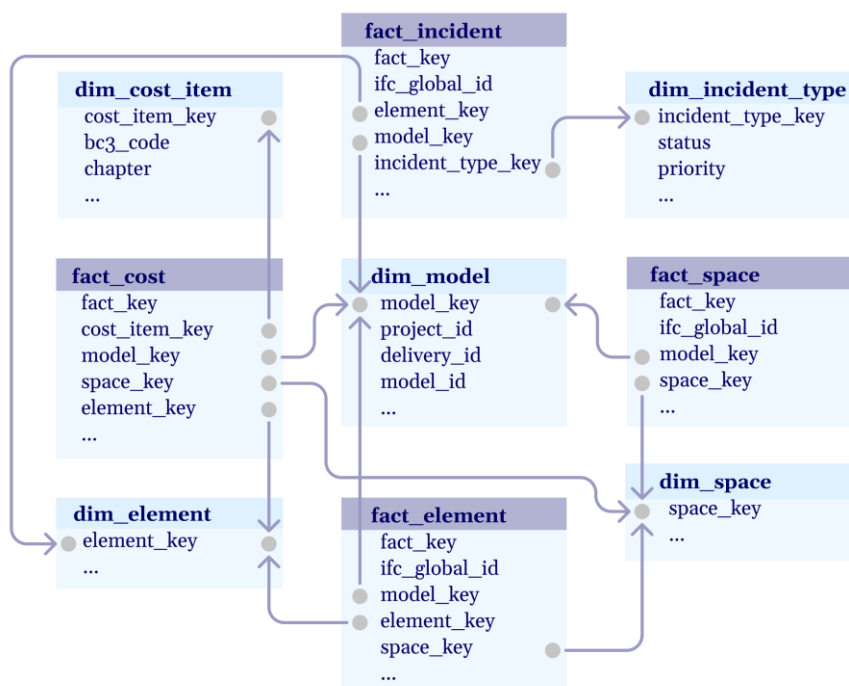


Figure 3. Star schema of the BIM2BI analytical repository (reference implementation v1.0.0). Four fact tables — `fact_element`, `fact_space`, `fact_cost` and `fact_incident`— capture the quantitative observations extracted from IFC, BC3 and BCF sources, and are contextualised by five dimensions — `dim_model`, `dim_element`, `dim_space`, `dim_cost_item` and `dim_incident_type`. The `IfcGlobalId` is preserved as a technical key in every fact table to guarantee end-to-end lineage between any analytical indicator and the originating IFC entity.

4.4.2. Data Lineage through `IfcGlobalId`

End-to-end lineage is articulated through the preservation of the `IfcGlobalId` as a technical key in every fact table, alongside surrogate keys that link to the dimensions. This dual level of identification combines the efficiency proper to dimensional models with the capability of reconstructing the link between any analytical indicator and the concrete IFC entity of the source model from which it originates, thereby guaranteeing result auditability regardless of the number of deliveries or models accumulated in the repository (Figure 4).

The mechanism is complemented by two additional resources. The `pipeline_logger.py` module generates a unique execution identifier and a cumulative historical CSV record that enables the reconstruction of the complete sequence of steps applied to the data. The `init_dwh.py` module performs an automatic backup of the analytical repository before each load, named with `{timestamp}_{project_id}_{delivery_id}`, which supports the selective cascade deletion of a specific delivery (`fact_*` → `dim_model` → `dim_delivery`) without affecting the history of the remaining deliveries. The combination of these three mechanisms materialises a traceable lineage at entity, execution and delivery level, covering a gap repeatedly documented in the BIM-BI literature.

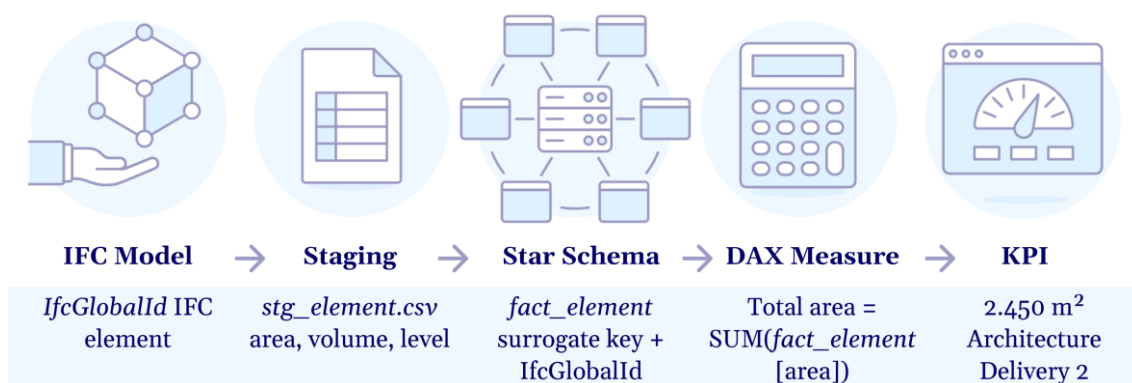


Figure 4. End-to-end data lineage in BIM2BI illustrated with a worked example. The `IfcGlobalId` of an IFC element (Layer 1) is preserved through the staging files (`stg_element.csv`), the analytical fact table (`fact_element` with both surrogate key and `IfcGlobalId`), the DAX measure (Total area = `SUM(fact_element[area])`) and the final BI KPI displayed on the dashboard, enabling any analytical indicator to be traced back to the originating IFC entity, project and delivery.

4.5. Layer 4: Analytical Exploitation

Layer 4 transforms the data persisted in Layer 3 into actionable knowledge for project stakeholders (technical management, BIM Manager, client, asset manager) through interactive indicators, metrics and visualisations. The layer is deliberately agnostic with respect to the visualisation tool employed: the star schema of the repository ensures its compatibility with any BI platform that supports standard relational connections, preserving the technological neutrality that articulates the whole architecture.

The reference implementation uses Microsoft Power BI for its consolidated adoption in the AEC sector, connected to the SQLite repository through a Python connector that removes the dependence on ODBC drivers on the user's machine. The indicators are defined as DAX measures on the semantic model and are organised in four categories: (i) model-composition indicators (number of elements

by IFC type, surface by discipline, volume by level); (ii) informational-quality indicators (ratio of compliant elements by IFC type, incidents per delivery, evolution of compliance across versions); (iii) cost indicators (construction budget by zone, cost by typology, variance across deliveries); and (iv) traceability indicators (surface variation between versions, elements added or removed, evolution of the information level).

These indicators are articulated in three complementary dashboards: a model-composition view aimed at the BIM Manager, an informational-quality view aimed at the BIM Quality Manager and a cross-delivery traceability view aimed at the project director and the client. The combination of the star schema of Layer 3 with the DAX measures of Layer 4 materialises the central goal of the architecture: converting IFC models into comparable, traceable and reproducible analytical knowledge, independently of the BIM authoring tool or the visualisation platform employed.

4.6. Open-Source Reference Implementation

Version 1.0.0 of the reference implementation materialises the architecture in a functional, documented and reproducible system, publicly available on GitHub under the MIT licence. The technology stack was selected under the criterion of maximum portability: the system must run on any Windows machine without server installation, without network dependencies and at zero licence cost. The technological decisions of each layer consolidate into a single open-source stack: local file system (Layer 1); Python 3.11 with ifcopenshell, ifctester, pandas, SQLAlchemy and a Streamlit interface (Layer 2); embedded SQLite 3.45 (Layer 3); and Power BI Desktop with a Python connector (Layer 4).

Deployment of the system in a clean environment requires three steps — cloning the repository, installing Python dependencies through pip and running the INICIAR.bat launcher— with a total time below five minutes. This portability reduces the adoption barrier for AEC organisations without a specialised technical profile and enables the full reproduction of the use cases of the validation section by researchers and professionals external to the development team. The published repository bundles the twelve ETL scripts, the Streamlit interface, the DDL schema of the analytical repository, the Power BI dashboard templates and the IDS profiles corresponding to the ten use cases documented in the validation.

5. Validation

5.1. Validation Strategy and Use Case Selection

Validation combines two complementary dimensions characteristic of the DSR framework. Technical validation evaluates compliance with functional requirements (correct extraction of IFC entities, operability of the IDS quality gate, transformation to the star schema without loss of traceability, and load with referential integrity) through inspection of the artefacts produced by the pipeline (staging CSVs, SQLite repository, delivery audit records) and the structured execution logs. Empirical validation evaluates the non-functional requirements (performance, scalability, maintainability and vendor independence) from the execution of the system on real data and documents the emerging findings that motivated architectural refinements not foreseen in the initial design.

The validation set was built through active collection of real BIM deliveries from projects executed in Spain and Chile between 2024 and 2025, applying three combined criteria: availability of complete real data (excluding synthetic models), diversity of delivery configuration (covering extreme ranges in volume, number of IFC files, presence or absence of BC3, IDS and BCF, and IFC schema version) and confidentiality of the originating project's data. The resulting set comprises ten use cases (CDU_001 to CDU_010) with an approximate total volume of 5 GB distributed over 69 IFC files, grouped in three profiles of increasing complexity.

Profile A (single-discipline deliveries) brings together four cases with a single architectural or structural IFC file accompanied by a budget and an IDS profile, and represents the baseline case of

the pipeline in everyday professional working conditions. Profile B (standard multidisciplinary deliveries) brings together four cases with between four and six IFC files processed in a single execution, introducing the consolidation of shared dimensions across models and, in two cases, BCF coordination incidents. Profile C (massive infrastructure deliveries) brings together two cases with more than twenty IFC files, volumes above 500 MB and, in CDU_004, the only delivery of the set with three staged IDS profiles applied to 25 IFC4X3 models approximately 2 GB in size. The diversity of IFC versions present in the set (IFC2X3 in CDU_007, IFC4 in the majority and IFC4X3 in CDU_004) is in itself a robustness requirement empirically verified: the pipeline processed the three versions without changes to the extraction code (Table 3).

Table 3. Characterisation of the ten use cases used in the empirical validation, grouped by complexity profile (A, B, C).

Code	Profile	IFC files	BC3	IDS	BCF	Size	IFC version
CDU_001_ARQ	A	1	1	1	—	177 MB	IFC4
CDU_002_EST	A	1	1	1	—	83 MB	IFC4
CDU_003_COR	B	6	1	2	—	85 MB	IFC4
CDU_004_COR	C	23	1	3	—	1,970 MB	IFC4X3
CDU_005_ARQ	A	1	1	1	—	49 MB	IFC4
CDU_006_COR	C	22	—	1	—	500 MB	IFC4
CDU_007_COR	B	6	—	1	6	388 MB	IFC2X3
CDU_008_COR	B	4	1	1	4	267 MB	IFC4
CDU_009_INS	A	1	2	1	—	898 MB	IFC4
CDU_010_COR	B	4	1	1	—	85 MB	IFC4
Total	—	69	9	13	10	4,502 MB	—

5.2. Evaluation of Functional Requirements

The extraction and load into the star schema executed correctly in all ten cases, with verifiable preservation of the `IfcGlobalId` from the input IFC file to the Power BI dashboard indicators. The systematic inspection of foreign keys in the SQLite repository confirmed the referential integrity between fact and dimension tables, and the end-to-end traceability was maintained independently of the delivery profile, the number of models processed and the IFC schema version. This result validates the principle of traceability as an architectural guarantee of the system and not as a property dependent on the configuration of each delivery.

The operability of the IDS quality gate was evaluated across the ten cases, with differentiated behaviours depending on the profile. In the Profile A cases, conformance rates were consistently high, with localised incidents traceable to specific model elements (for instance, 17 `IfcMember` and `IfcPlate` elements lacking the required discipline attribute in CDU_001). In the Profile C cases conformance rates were lower due to the greater complexity of the informational requirements associated with linear-infrastructure projects—a greater number of specialised disciplines, a wider variety of IFC classes and greater heterogeneity in naming conventions between subcontractors—, a pattern coherent with the literature on BIM data quality in infrastructure [44]. System reproducibility was verified through clean deployment of the repository and complete execution of the ten cases in a secondary environment in less than five minutes of installation, validating the portability criterion described in Section 4.6.

5.3. Evaluation of Non-Functional Requirements

The performance evaluation showed a stronger correlation between processing time and the number of IFC files than between time and total volume in megabytes. CDU_004 (23 IFC, 1.97 GB) exceeds CDU_006 (22 IFC, 500 MB) by a proportion lower than what the difference in volume would suggest, which evidences that the phase of opening and parsing IFC files with `ifcopenshell` dominates the processing time with respect to transformation and load operations in SQLite. The Profile A cases

were processed consistently below 90 seconds, whereas the most demanding case (CDU_004) set the upper bound of system performance around 15 minutes in the reference configuration with SQLite as analytical repository.

Multiproject scalability was materialised through a dual-write pattern: the pipeline maintains a central repository as the source of truth for transversal analyses and simultaneously writes into a per-project configurable repository that feeds the project-specific dashboards. This pattern — the operational materialisation of the classical distinction between corporate data warehouse and project data mart [39] adapted to the lightweight deployment of the system — avoids the access conflicts that arise when different project managers require independent Power BI connections on the same file. Operational maintainability was consolidated through the introduction of an explicit project registry in settings.json with formal metadata and automatic backward compatibility with pre-existing folders, a transition from implicit management by naming convention to explicit management with control metadata that ISO 19650 applies to BIM information containers and that DAMA-DMBOK extends to analytical repositories.

Vendor independence was verified by construction: no pipeline component depends on proprietary software beyond the optional IFC export from Autodesk Revit in those cases where the native file was available. Replacing Power BI with any other BI tool compatible with standard relational connections reduces to reconfiguring the connection, without changes to the analytical repository or to the ETL pipeline. This result operationalises the principle of technological neutrality that articulates the whole architecture.

5.4. Empirical Findings on openBIM Standards

Empirical validation generated three findings not previously documented in the BIM-BI architectures literature, constituting a concrete contribution to the knowledge of the real behaviour of openBIM standards within analytical workflows. The first is the incompatibility between the IFC4X3 identifier and the official XSD schema of IDS 1.0: during the implementation of the quality gate on the IFC4X3 models of CDU_004 it was detected that the value IFC4X3 is not valid in the ifcVersion attribute of the IDS 1.0 XSD schema, which only accepts IFC2X3, IFC4 or IFC4X3_ADD2. The discrepancy originates in a change introduced in the final version of the standard — the IDS 1.0 release notes document the update of the denomination from IFC 4.3 to the official value IFC4X3_ADD2 [45] —, and the same problem has been reported by other implementers in the official ifcopenshell repository [46]. The finding illustrates a real limitation of the maturity of IDS 1.0 that only emerges in practical implementations, not in the theoretical reading of the specification, and reinforces the value of the use cases documented in this work as empirical evidence of the real behaviour of the openBIM ecosystem.

The second finding affects the loading strategy of the conformed dimensions of the analytical repository [39]. The initial replace strategy on dimension tables removed records from dim_cost_item and dim_element that had dependent rows in the fact tables coming from prior deliveries, generating 6,109 orphan foreign keys during incremental-load testing. This phenomenon — deferred referential-integrity violation — does not produce an immediate error in SQLite, which does not enforce foreign-key constraints by default, but manifests silently as fact records without associated dimension and degrades the analytical indicators without any visible error signal. The implemented solution establishes that dimensions shared between deliveries are always loaded with an additive strategy (append + INSERT OR IGNORE) preserving the existing surrogate keys and adding only the new records. The finding confirms that in a BIM data warehouse where deliveries accumulate across the project's life cycle, the loading strategy of dimensions must be additive rather than substitutive, unlike transactional data warehouses where periodic replacement is usual.

The third is the concurrency limitation inherent to SQLite as analytical repository in an architecture where storage and exploitation layers share the same physical file. When Power BI maintains an active connection on bim2bi.db the write operations of the pipeline fail with the error “database is locked”, behaviour inherent to the single-writer model of SQLite that cannot be resolved

by configuration. The reference implementation handles this limitation through an explicit retry protocol (BEGIN IMMEDIATE with visual feedback in the interface, up to sixty seconds before an actionable error) and prevents the problem architecturally by means of the dual-write pattern described in the previous subsection. The detailed characterisation of the limit and its solution, transferable to engines with ACID concurrency (MariaDB, PostgreSQL), constitutes a practical contribution for future lightweight BIM-BI implementations. The optional activation of an alternative engine is envisaged in the implementation through the environment variable `DW_ENGINE`, without modification of the ETL code.

6. Discussion

The BIM2BI architecture presented in this article addresses jointly the four transversal gaps identified in Section 2 that no prior work covers together. Systematic dimensional modelling is materialised in the star schema of Layer 3, with four fact tables (`fact_element`, `fact_space`, `fact_cost`, `fact_incident`) and six contextual dimensions that transform object-oriented IFC data into structures compatible with consolidated BI tools, aligning with the practice established since Kimball [39] but explicitly adapting it to the AEC domain. Granular lineage is solved by preserving the `IfcGlobalId` as a technical key in every fact table, complemented by per-execution structured logging and per-delivery backups, which enables any analytical indicator to be tracked back to the concrete IFC entity of the source model. Automated pre-ingestion validation is materialised through the IDS quality gate and the three-level staged validation strategy, which extend the application scope of the IDS standard beyond interchange between authoring tools to the governance of data quality in analytical pipelines. And dependence on proprietary platforms is eliminated by construction by operating the pipeline directly on the open IFC standard and integrating with any BI tool compatible with standard relational connections.

The positioning of BIM2BI with respect to prior approaches can be synthesised by grouping them into their four lines of work. With respect to proposals of direct querying over IFC [31], BIM2BI incorporates the dimensional transformation missing from those and the granular lineage mechanism they do not contemplate. With respect to semantic-web approaches [32–34], BIM2BI avoids the learning overhead and the performance overhead of the RDF/OWL/SPARQL stack, maintaining native compatibility with BI tools consolidated in the sector. With respect to project-specific BIM-BI dashboards [12,35–37], BIM2BI contributes an automated ETL pipeline and a systematic dimensional schema that avoid dependence on manual exports from authoring software and allow cross-project analytical consolidation. And with respect to current commercial initiatives from the major vendors –singularly the AEC Data Model API consolidated in March 2026 under the unified Autodesk Forma brand [14]–, BIM2BI maintains platform independence as a foundational principle, offering an alternative reproducible in any organisation regardless of its technological stack.

The acknowledged limitations do not invalidate the proposed architecture but define the perimeter within which the results are valid and guide future work. Validation was carried out in controlled and representative scenarios –ten real use cases from Spain and Chile between 2024 and 2025– but does not equate to an exhaustive evaluation in production environments at industrial scale with multiple projects and heterogeneous teams operating simultaneously. Additionally, the adoption of Design Science Research as the methodological framework implies that the conclusions are inherently dependent on the designed artefact: the generalisation to BIM-BI architectures with different design decisions (alternative database engine, non-sequential orchestrator, analytical schema different from the star schema) would require further research. The reference implementation operates without authentication, authorisation or data-privacy mechanisms, which conditions its direct deployment in multi-user environments, and the quality of the IFC models used directly conditions the IDS conformance rates and the completeness of the generated dimensional schemas, an external variability that the architecture cannot control but can characterise and record as an analytical metric.

The implications of the work are distributed across three planes. On the academic plane, BIM2BI provides the unifying BIM-BI framework that the prior literature did not articulate and demonstrates the viability of applying Design Science Research to the design of data architectures in a technical domain with a strong normative component (openBIM, ISO 19650), furthermore generating quantitative and qualitative evidence on the real behaviour of the standards in analytical workflows that can serve as a reference point for future research on scalability or extension to other organisational contexts. On the industrial plane, the architecture offers to client organisations a reproducible framework to convert the BIM deliveries from their supply chain into longitudinally consolidable analytical assets; to consulting and construction firms, a way to reduce dependence on individual technician judgement in data preparation and to satisfy ISO 19650 requirements as a natural by-product of the pipeline; and to platform vendors, confirmation –reinforced by recent commercial moves– that the demand for granular and decoupled access to BIM data is a market need, not an exclusively academic subject.

On the standardisation plane, the empirical findings of Section 5 contribute three concrete directions to the buildingSMART agenda. The use of IDS as a pre-ETL quality gate demonstrates that the standard is technically viable in automated analytical workflows but reveals expressive limitations for requirements proper to BI environments (consistency of numerical values, admissible ranges, referential integrity between Property Sets) and discrepancies between the documented specification and its reference implementation (the IFC4X3 vs. IFC4X3_ADD2 case) that could guide future extensions of the schema. The propagation of the `IfcGlobalId` as a lineage key along the pipeline evidences that IFC already provides the traceability mechanisms required but that their effective exploitation requires implementation conventions that the current standards do not prescribe explicitly. And the integration of BCF-JSON as a structured data source for `fact_incident` evidences the potential of the standard beyond model coordination, suggesting that its analytical profile could be enriched with fields oriented to data-quality metrics. Together, BIM2BI contributes to the openBIM agenda not as normative development but as a documented analytical use case, in line with the iterative development model that buildingSMART applies to its standards and with the rigour cycle of Design Science Research.

7. Conclusions

This article has presented BIM2BI, a BIM-BI integration architecture in four functional layers that formalises the separation of responsibilities, data contracts between layers and an extensibility-without-redesign principle absent from the previous literature. The proposal relies on the openBIM standards IFC, IDS and BCF and contributes four differentiated contributions: a scientific contribution, which articulates for the first time a unifying framework for BIM-BI integration; a methodological contribution, which reinterprets the IDS standard as a pre-ETL quality gate within automated analytical workflows and proposes a three-level staged validation strategy applicable to high-volume deliveries; a practical contribution, materialised in an open-source reference implementation (MIT licence, portable deployment in less than five minutes) that reduces the adoption barrier for AEC organisations without specialised technical profile; and a contribution to the state of the art, consisting of empirical evidence not previously documented on the real behaviour of openBIM standards in analytical workflows (incompatibility IFC4X3 / IFC4X3_ADD2 in the IDS 1.0 XSD schema, need for additive loading of conformed dimensions and concurrency limits of the SQLite repository).

The empirical validation across ten real use cases grouped in three profiles of complexity makes it possible to answer the research question in a condensed form. The transformation of IFC data into dimensional schemas is solved through a three-phase ETL pipeline (extraction with `ifcopenshell` preserving the `IfcGlobalId`, transformation into the star schema and load with additive strategy for conformed dimensions). End-to-end lineage is guaranteed by the preservation of the `IfcGlobalId` as a technical key in the fact tables, complemented by per-execution structured logging and per-delivery automatic backups. Semantic validation is articulated through the two-level IDS quality gate, which

turns data quality into an analytical metric comparable over time. And reproducibility relies on the portable deployment of the system and on automatic DDL-schema migrations with prior backup, with scalability beyond the limits of SQLite architecturally foreseen through the switching of the analytical engine to MariaDB or PostgreSQL without modifying the ETL code.

Four lines of future work derive directly from the results and limitations identified. The first is the extension of the analytical repository towards new dimensions of BIM data (`fact_schedule`, `fact_energy`, `fact_sustainability`) to cover areas with high regulatory and market demand such as energy efficiency (EPBD) and sustainability (CSRD), exploiting the schema's extensibility without redesign. The second is the closure of the definition-validation cycle through LOINXML as a formal layer of informational requirements and integration with the buildingSMART Data Dictionary (bSDD) as a source of normalised definitions, which would enable every validation rule to be tracked back to the informational need that originates it. The third is the evolution of the deployment towards multi-user environments, combining authentication and access-control policies with the replacement of SQLite by an ACID-concurrency engine (MariaDB or PostgreSQL) and orchestration through Apache Airflow, already anticipated by the environment variable `DW_ENGINE` and the compatibility of the pipeline with both execution modes. And the fourth is the incorporation of Layer 1 connectors oriented to proprietary APIs for BIM data access in the cloud — singularly the AEC Data Model API operated under the Autodesk Forma brand — as a complementary source to the IFC file, keeping the same data contract towards the transformation and load layers and preserving vendor independence as a foundational principle. Taken together, these lines consolidate the enduring value of BIM2BI not only as a concrete implementation but as a transferable methodological framework that connects the BIM world with the BI world in a reproducible, auditable and extensible manner.

Author Contributions: Conceptualisation, D.J.S.G. and R.V.L.D.; methodology, D.J.S.G.; software, D.J.S.G.; validation, D.J.S.G.; formal analysis, D.J.S.G.; investigation, D.J.S.G.; resources, R.V.L.D.; data curation, D.J.S.G.; writing—original draft preparation, D.J.S.G.; writing—review and editing, D.J.S.G. and R.V.L.D.; visualisation, D.J.S.G.; supervision, R.V.L.D.; project administration, R.V.L.D.; funding acquisition, R.V.L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The reference implementation of the BIM2BI architecture (v1.0.0) is publicly available at <https://github.com/diegojsanchez/BIM2BI> under the MIT License. The repository is currently private and will be made public upon publication of the doctoral thesis from which this article is derived. It includes the twelve Python ETL scripts, the Streamlit interface, the SQLite DDL schema of the analytical repository, the Power BI dashboard templates and the Information Delivery Specification (IDS) profiles of the ten use cases documented in Section 5. The IFC, BC3 and BCF datasets used in the validation are not publicly redistributable due to confidentiality agreements with the originating projects; synthetic equivalents can be provided by the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the professionals and organizations that provided access to the real-world BIM deliveries used in the empirical validation of the architecture, as well as the anonymous reviewers for their constructive feedback. During the preparation of this manuscript, the authors used Claude (Anthropic) as a generative AI assistant for language editing and structural revision of material previously developed in the doctoral thesis from which this article is derived. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AEC: Architecture, Engineering and Construction; BCF: BIM Collaboration Format; BI: Business Intelligence; BIM: Building Information Modeling; CDE: Common Data Environment; DSR: Design Science

Research; ETL: Extract, Transform, Load; IDS: Information Delivery Specification; IFC: Industry Foundation Classes; KPI: Key Performance Indicator; LOIN: Level of Information Need; RQ: Research Question.

References

1. Poirier, E.; Forgues, D.; Staub-French, S. Understanding the impact of BIM on collaboration: A Canadian case study. *Build. Res. Inf.* 2017, 45, 1–15. <https://doi.org/10.1080/09613218.2017.1324724>.
2. Lopes, A.; Boscaroli, C. Business intelligence and analytics to support management in construction: A systematic literature review. *Rev. Bras. Comput. Apl.* 2020. <https://doi.org/10.5335/rbca.v13i1.11891>.
3. Iqbal, F.; Ahmed, S.; Tariq, M.A.B.; Waqas, H.; Al-Ammar, E.; Wabaidur, S.; Fawad, M. BIM-IoT integration for remote real-time concrete compressive strength monitoring. *Ain Shams Eng. J.* 2024. <https://doi.org/10.1016/j.asej.2024.102906>.
4. Chiponde, D.; Gledson, B.; Greenwood, D. The institutional field of learning from project-related failures—opportunities and challenges. *Constr. Econ. Build.* 2024. <https://doi.org/10.5130/ajceb.v24i1/2.8394>.
5. Canesi, R.; Gabrielli, L.; Marella, G.; Ruggeri, A. Probabilistic risk assessment framework for cost overruns predictions in infrastructure projects using randomized simulations. *Comput.-Aided Civ. Infrastruct. Eng.* 2025, 40, 4774–4796. <https://doi.org/10.1111/mice.70100>.
6. De Gaetani, C.; Mert, M.; Migliaccio, F. Interoperability analyses of BIM platforms for construction management. *Appl. Sci.* 2020. <https://doi.org/10.3390/app10134437>.
7. Kładź, M.; Borkowski, A. IDS standard and bSDD service as tools for automating information exchange and verification in projects implemented in the BIM methodology. *Buildings* 2025. <https://doi.org/10.3390/buildings15010126>.
8. Abdul-Azeez, O.; Ihechere, A.O.; Idemudia, C. Enhancing business performance: The role of data-driven analytics in strategic decision-making. *Int. J. Manag. Entrep. Res.* 2024. <https://doi.org/10.51594/ijmer.v6i7.1257>.
9. Udeh, C.A.; Orieno, O.H.; Daraojimba, O.D.; Ndubuisi, N.L.; Oriekhoe, O.I. Big data analytics: A review of its transformative role in modern business intelligence. *Comput. Sci. IT Res. J.* 2024. <https://doi.org/10.51594/csitj.v5i1.716>.
10. Brozovsky, J.; Labonnote, N.; Vigren, O. Digital technologies in architecture, engineering, and construction. *Autom. Constr.* 2024. <https://doi.org/10.1016/j.autcon.2023.105212>.
11. Rafsanjani, H.; Nabizadeh, A. Towards digital architecture, engineering, and construction (AEC) industry through virtual design and construction (VDC) and digital twin. *Energy Built Environ.* 2021. <https://doi.org/10.1016/j.enbenv.2021.10.004>.
12. Rodrigues, F.; Alves, A.; Matos, R. Construction management supported by BIM and a business intelligence tool. *Energies* 2022. <https://doi.org/10.3390/en15093412>.
13. Autodesk Platform Services. AEC Data Model Roadmap; Autodesk: 2025. Available online: <https://aps.autodesk.com/aec-data-model-roadmap> (accessed on 19 April 2026).
14. Autodesk. Autodesk Construction Cloud is Now Autodesk Forma; Autodesk News, 25 March 2026. Available online: <https://adsknews.autodesk.com/en/news/autodesk-construction-cloud-is-now-autodesk-forma/> (accessed on 19 April 2026).
15. Eastman, C.; Teicholz, P.; Sacks, R.; Liston, K. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2011.
16. Azhar, S. Building Information Modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry. *Leadersh. Manag. Eng.* 2011, 11, 241–252. [https://doi.org/10.1061/\(ASCE\)LM.1943-5630.0000127](https://doi.org/10.1061/(ASCE)LM.1943-5630.0000127).
17. Zawada, K.; Rybak-Niedziółka, K.; Donderewicz, M.; Starzyk, A. Digitization of AEC industries based on BIM and 4.0 technologies. *Buildings* 2024. <https://doi.org/10.3390/buildings14051350>.
18. Kovacs, A.T.; Micsik, A. BIM quality control based on requirement linked data. *Int. J. Archit. Comput.* 2021, 19, 431–448. <https://doi.org/10.1177/14780771211012175>.
19. Turban, E.; Sharda, R.; Delen, D.; King, D. *Business Intelligence: A Managerial Approach*, 3rd ed.; Pearson: Harlow, UK, 2014.

20. Eboigbe, E.O.; Farayola, O.A.; Olatoye, F.O.; Nnabugwu, O.C.; Daraojimba, C. Business intelligence transformation through AI and data analytics. *Eng. Sci. Technol. J.* 2023. <https://doi.org/10.51594/estj.v4i5.616>.
21. International Organization for Standardization. ISO 16739-1:2024 Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries—Part 1: Data Schema; ISO: Geneva, Switzerland, 2024.
22. buildingSMART International. Information Delivery Specification (IDS), Version 1.0; buildingSMART: 2024. Available online: <https://technical.buildingsmart.org/projects/information-delivery-specification-ids/> (accessed on 19 April 2026).
23. buildingSMART International. BIM Collaboration Format (BCF); buildingSMART: 2025. Available online: <https://technical.buildingsmart.org/standards/bcf/> (accessed on 19 April 2026).
24. Noardo, F.; Ohori, A.; Krijnen, T.; Stoter, J. An inspection of IFC models from practice. *Appl. Sci.* 2021. <https://doi.org/10.3390/app11052232>.
25. Lee, M.; Lee, U.-K. A framework for evaluating an integrated BIM ROI based on preventing rework in the construction phase. *J. Civ. Eng. Manag.* 2020. <https://doi.org/10.3846/jcem.2020.12185>.
26. Du, S.; Hou, L.; Zhang, G.; Tan, Y.; Mao, P. BIM and IFC data readiness for AI integration in the construction industry: A review approach. *Buildings* 2024. <https://doi.org/10.3390/buildings14103305>.
27. International Organization for Standardization. ISO 19650-1:2018 Organization and Digitization of Information about Buildings and Civil Engineering Works, Including Building Information Modelling (BIM)—Information Management Using Building Information Modelling—Part 1: Concepts and Principles; ISO: Geneva, Switzerland, 2018.
28. Patacas, J.; Dawood, N.; Kassem, M. BIM for facilities management: A framework and a common data environment using open standards. *Autom. Constr.* 2020. <https://doi.org/10.1016/j.autcon.2020.103366>.
29. Tomczak, A.; Benghi, C.; Van Berlo, L.; Hjelseth, E. Requiring circularity data in BIM with information delivery specification. *Circ. Econ.* 2024. <https://doi.org/10.55845/rejy5239>.
30. DAMA International. DAMA-DMBOK: Data Management Body of Knowledge, 2nd ed.; Technics Publications: Basking Ridge, NJ, USA, 2017.
31. Barzegar, M.; Rajabifard, A.; Kalantari, M.; Atazadeh, B. An IFC-based database schema for mapping BIM data into a 3D spatially enabled land administration database. *Int. J. Digit. Earth* 2021, 14, 736–765. <https://doi.org/10.1080/17538947.2021.1875062>.
32. Pauwels, P.; Terkaj, W. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Autom. Constr.* 2016, 63, 100–133. <https://doi.org/10.1016/j.autcon.2015.12.003>.
33. Rasmussen, M.H.; Lefrançois, M.; Schneider, G.F.; Pauwels, P. BOT: The building topology ontology of the W3C linked building data group. *Semant. Web* 2020, 12, 143–161. <https://doi.org/10.3233/SW-200385>.
34. Kone, V.; Mahesh, G. An ontology-driven bi-directional workflow for integrating project management data into the IFC standard. *J. Inf. Technol. Constr.* 2025, 30, 1768. <https://doi.org/10.36680/j.itcon.2025.089>.
35. Apellániz, D.; Alkewitz, T.; Gengnagel, C. Visualisation of building life cycle assessment results using 3D business intelligence dashboards. *Int. J. Life Cycle Assess.* 2024, 29, 1–12. <https://doi.org/10.1007/s11367-024-02309-7>.
36. Di Giuda, G.M.; Accardo, D.; Gasbarri, P.; Meschini, S.; Tagliabue, L.C.; Scomparin, L. BIM-GIS and BI integration for facility and occupancy management of university assets: The UNITO pilot case. In *Proceedings of the CONVR 2023—23rd International Conference on Construction Applications of Virtual Reality*, Florence, Italy, 2023.
37. Abdulaziz, H.N.; Ahmed, H.M. Proposed framework for intelligent dashboard that utilizes business intelligence (BI) and building information modeling (BIM) for better decision making. *Int. J. Eng. Res. Technol.* 2025, 14, IJERTV14IS100069. <https://doi.org/10.5281/zenodo.18081051>.
38. Zhang, C.; Beetz, J.; de Vries, B. BIM compliance checking based on a graph representation. *Autom. Constr.* 2020, 118, 103196. <https://doi.org/10.1016/j.autcon.2020.103196>.
39. Kimball, R.; Ross, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed.; John Wiley & Sons: Indianapolis, IN, USA, 2013.

40. Goldkuhl, G. Design research in search for a paradigm: Pragmatism is the answer. In *Practical Aspects of Design Science*; Helfert, M., Donnellan, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2012.
41. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *MIS Q.* 2004, 28, 75–106. <https://doi.org/10.2307/25148625>.
42. Hevner, A. The three cycle view of design science. *Scand. J. Inf. Syst.* 2007, 19, 4.
43. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* 2007, 24, 45–77. <https://doi.org/10.2753/MIS0742-1222240302>.
44. Cerovsek, T.; Omar, M. Advancing semantic enrichment compliance in BIM: An ontology-based framework and IDS evaluation. *Buildings* 2025. <https://doi.org/10.3390/buildings15152621>.
45. buildingSMART International. IDS v1.0 Final buildingSMART Standard—Release Notes; buildingSMART: 2024. Available online: <https://technical.buildingsmart.org/projects/information-delivery-specification-ids/> (accessed on 19 April 2026).
46. buildingSMART Developer Forums. FILE_SCHEMA IFC4X3_ADD2 or IFC4X3—buildingSMART Developer Forums, 2024. Available online: <https://forums.buildingsmart.org/> (accessed on 19 April 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.