Structured Databases for Heritage BIM: Enhancing Digital Documentation and Restoration Workflows

Laura Marcela Valderrama Niño[1] and Sabrina Becker[1]
[1]Institute of Numerical Methods and Informatics in Civil Engineering (iib), TU Darmstadt, Darmstadt, Germany
E-mail(s): valderrama@iib.tu-darmstadt.de, becker@iib.tu-darmstadt.de

Abstract: Heritage Building Information Modeling (HBIM) plays a central role in the digital documentation and conservation of historic buildings. Yet, many current systems lack the flexibility and semantic depth required to manage evolving, multidisciplinary datasets. Rigid data structures and tool-dependent workflows hinder the representation of restoration processes, the integration of contextual knowledge, and long-term data reuse. This research introduces a hybrid data architecture that combines flexible document storage with semantic graph modeling. The system enables the integration and structured analysis of diverse heritage data—linking historical, material, and procedural information across the HBIM lifecycle. While this work is conceptual in scope, it outlines a pathway toward a modular, extensible, and open HBIM infrastructure. Aligned with ontologies such as CIDOC CRM and FAIR data principles, the proposed framework supports future integration with linked open data platforms and external heritage repositories. By advancing a semantically expressive and interoperable foundation, this approach promotes sustainable documentation, transparent conservation workflows, and deeper historical insight in digital heritage environments.

Keywords: Heritage Building Information Modeling (HBIM), NoSQL, Graph Database, CIDOC CRM



DOI: 10.18154/RWTH-CONV-254892. Published in the conference proceedings of the 36. Forum Bauinformatik 2025, Aachen, Germany,
© 2025 The copyright for this article lies with the authors. This publication, except for quotations and otherwise indicated parts, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1 Introduction

The use of digital technologies has become central to the documentation and preservation of cultural heritage. Heritage Building Information Modeling (HBIM) offers a structured framework to integrate geometric, historical, and restoration data into unified digital representations that support conservation planning and interdisciplinary collaboration [1], [2].

However, current HBIM implementations often rely on rigid relational databases or proprietary software, limiting their ability to manage heterogeneous, evolving datasets and semantic relationships [3], [4]. This results in poor interoperability, restricted data reuse, and difficulties in long-term access [5], [6].

Ontology-based approaches such as CIDOC CRM offer a solution by enabling structured, semantically rich modeling of heritage knowledge [7], [8]. Yet, technical barriers and missing infrastructure have hindered their broader adoption in HBIM contexts [9].

This paper proposes a hybrid NoSQL system architecture that combines flexible document storage (MongoDB) with semantic graph modeling (Neo4j). The aim is to create a scalable, modular platform for data integration, versioning, and semantic querying—aligned with FAIR principles and cultural heritage ontologies [10], [11]. Rather than replacing existing HBIM platforms, this approach offers a complementary foundation for semantically enriched and interoperable heritage data systems.

2 Related Work

2.1 From Geometric Models to Semantically Enriched HBIM

Historic Building Information Modeling (HBIM) has developed from a tool for geometric reconstruction into a semantically enriched infrastructure for documenting and analyzing cultural heritage. Early methods translated point clouds into parametric models [1], while more recent approaches integrate historical data, material attributes, and restoration processes into structured, queryable models. This shift enables HBIM to function not only as a visual model but as a knowledge system, facilitating reasoning, data reuse, and interoperability [3].

2.2 Structural and Semantic Limitations of Relational Data Models

Relational databases are still widely used in HBIM to store attributes of building elements. However, their rigid schemas hinder integration of diverse data types—such as archival documents, 3D geometries, or annotations—and complicate modeling of temporal or interpretive relationships. Representing changes across restoration phases or aligning heterogeneous metadata often requires workarounds that compromise semantic clarity [5]. Critically, relational systems are not natively compatible with semantic web standards like Resource Description Framework (RDF), limiting their interoperability and long-term usefulness.

2.3 NoSQL Data Models for Flexible HBIM Workflows

NoSQL databases address many of the structural limitations of relational systems. Document-oriented systems like MongoDB support schema-flexible storage of HBIM records—including geometries, pathologies, and intervention logs—within nested data structures. Graph databases such as Neo4j or ArangoDB, in turn, provide native capabilities for modeling historical, material, and actor-based relationships via subject–predicate–object semantics. This aligns closely with ontology-driven frameworks and allows for advanced querying across semantically linked data [9].

2.4 Semantic Enrichment and Ontology Integration in HBIM

The use of ontologies is increasingly central in HBIM to represent knowledge beyond geometric attributes. CIDOC CRM has become a widely accepted standard for modeling events, actors, materials, and sources in a semantically interoperable way. For example, elements like wall fragments (E22 Man-Made Object) can be linked to specific conservation actions (E11 Modification) and responsible agents (E39 Actor), facilitating semantic tracing of interventions across time [7].

2.5 Towards FAIR and Interoperable HBIM Infrastructures

The FAIR principles—Findable, Accessible, Interoperable, and Reusable—have become a guiding framework for managing cultural heritage data. In this context, RDF graphs, persistent URIs, and alignment with shared vocabularies (such as CIDOC CRM) enable HBIM data to be integrated with external repositories and semantic web platforms. While such approaches are gaining traction in the digital humanities, their systematic application within HBIM environments remains limited [8].

2.6 Research Gap in Semantic HBIM Systems

Despite increased awareness of semantic data needs, most HBIM systems continue to rely on proprietary platforms or relational backends, limiting interoperability and adaptability. Hybrid NoSQL architectures—combining flexible document storage with graph-based reasoning—have shown potential in related fields but are rarely operationalized in heritage practice. Coordinated use of MongoDB and Neo4j, aligned through CIDOC CRM, offers a promising path forward, which this research develops into a modular infrastructure [10].

3 System Architecture and Methodology

This section introduces a modular HBIM architecture that combines flexible document storage, semantic graph reasoning, and interoperable metadata standards. It addresses the limits of relational databases and fragmented NoSQL use by enabling structured ingestion and querying of heterogeneous data—from 3D scans to historical records. A prototype was tested using simulated restoration data and point cloud-based models.

3.1 Conceptual Framework

This research proposes a multilayer NoSQL architecture that addresses structural and semantic limitations in conventional HBIM workflows. It combines MongoDB for flexible data storage and Neo4j for semantic querying, enabling scalable integration of heterogeneous data such as IFC models, photogrammetry meshes, annotations, and restoration records.

The system functions as a dynamic information infrastructure—supporting the linkage of restoration data, semantic relationships, and external sources through ontology alignment. Its design is guided by three principles: schema flexibility, semantic expressiveness, and interoperability with established ontologies.

The architecture comprises:

- A document storage layer for ingesting and managing heterogeneous HBIM data
- A semantic graph layer for modeling complex relationships among heritage entities
- An integration layer for coordinating queries and maintaining data consistency between components

This layered approach supports complex queries and enrichment workflows. For example, an element such as `WALL_102` is linked to condition reports, restoration events, materials, and survey data in a coherent and queryable format.

3.2 MongoDB Layer: Flexible HBIM Data Storage

The document storage layer manages core HBIM data using a flexible, schema-less format capable of storing geometry, materials, conditions, and restoration records. Each architectural element is represented as a structured document with fields for identification, spatial context, and intervention history. For example, `WALL_102` includes IFC geometry, STL mesh, materials like limestone, pathologies such as cracking, and associated restoration actions. This structure supports extensibility, versioning, and links to external survey data, adapting to diverse heritage documentation needs.

3.3 Neo4j Layer: Semantic Reasoning and Historical Tracing

The semantic graph layer enables querying relationships such as restoration sequences, material usage across elements, and classification of interventions. Data is aligned with CIDOC CRM, which provides standardized classes for actions (e.g., E11 Modification), building components (E22), and responsible actors (E39), ensuring semantic clarity and interoperability across platforms.

3.4 Interoperability Layer and Query Engine

The architecture enables synchronization between document-based and semantic data by transforming storage-layer content into graph representations while preserving context. This bidirectional structure supports both structured filtering and semantic querying for tasks like restoration planning and documentation analysis.

Although no user interface is included, the modular design supports future integration of dashboards or web tools without altering core logic.

3.5 Standards and Ontologies

Semantic consistency is achieved through CIDOC CRM, which structures data on interventions, materials, spaces, and actors. Metadata is exportable as RDF or JSON-LD (Linked Data), supporting future integration with linked data platforms.

This combination of flexible storage and semantic modeling enhances HBIM workflows, enabling traceability, interoperability, and long-term conservation support.

4 Implementation and Use Case

4.1 Overview

To test the proposed architecture, a prototype was developed to simulate the documentation and analysis of restoration activities associated with historical architectural elements. A representative synthetic data sample was assembled based on common HBIM conditions, including material degradation, restoration actions, and survey data. The goal was to assess the multilayer approach in terms of data structuring, semantic linking, and query execution across both the document and graph-based layers.

4.2 Data Modeling Scope

Several architectural elements were modeled and stored in the document database using a flexible JSON structure. Each record includes geometry (e.g., IFC files, meshes), material data, location metadata, condition annotations, and restoration events with dates, techniques, and responsible actors. A representative example is `WALL_102`, which combines photogrammetric survey data, historical notes (e.g., surface erosion, staining), and restoration metadata. This synthetically constructed case reflects

typical HBIM scenarios and demonstrates the system's capacity for data integration and semantic linkage (see Figure 1).


```
_id: ObjectId('6849311ab74b62214d174dc7')
element_id : "WALL_102"
▾ annotations : Array (2)
    0: "surface erosion"
    1: "biological staining"
geometry_ifc : "wall_102.ifc"
▾ linked_survey : Object
    file : "wall_test.stl"
    absolute_path : "/absolute/path/to/models/wall_test.stl"
    survey_date : "2023-09-01"
    description : "Terrestrial photogrammetry mesh for detailed surface mapping"
location : "North aisle"
▾ materials : Array (2)
    0: "limestone"
    1: "stucco"
name : "North Nave Wall"
▾ restoration_events : Array (1)
  ▸ 0: Object
type : "IfcWall"
cidoc_class : "E22 Man-Made Object"
▾ responsible_actor : Object
    name : "Elena Rossi"
    role : "architect"
    cidoc_class : "E39 Actor"
```


Figure 1: `WALL_102` Element: Document Structure in MongoDB

In parallel, the graph database was populated with the corresponding semantic entities. These include nodes for elements, restoration events, materials, and techniques, along with relationships such as `RESTORED_ON`, `HAS_MATERIAL`, and `USED_TECHNIQUE`. The resulting structure enables reasoning over intervention histories, material usage, and patterns of deterioration (see Figure 2).

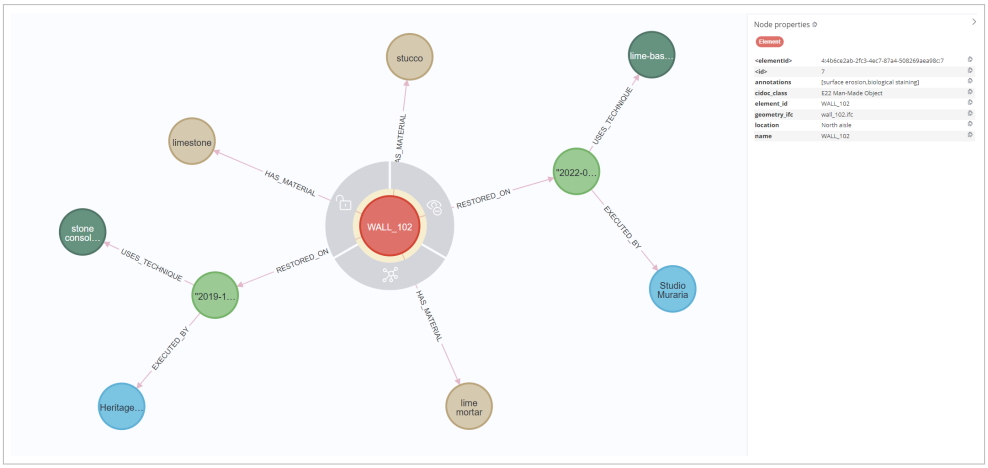


Figure 2: `WALL_102` : Semantic Graph of Restoration Relationships

In addition to individual elements, the system supports the semantic enrichment of full IFC-based digital models. These cases represent a different example from `WALL_102` and are used to demonstrate how restoration metadata, scan data, and geometry can be linked at the model level.

For instance, a file `WALL_test.ifc` includes synthetic metadata connecting the model to STL mesh files, scan techniques (e.g., terrestrial laser scanning), the executing contractor, and a list of linked IFC

elements, see Figure 3 . Each linked element, such as a wall or door, is separately documented with its own restoration history, location, and classification.


```
_id: ObjectId('6849311ab74b62214d174dc6')
file_name : "Wall_test.ifc"
contractor : "BauKultur GmbH"
created_on : "2025-06-11"
▾ linked_elements : Array (2)
    0: "1xLMsTjQr9N88Q3rX5Sot7"
    1: "1xLMsTjQr9N88Q3rX5So9i"
path_ifc : "C:\Users\valde\Documents\05_DenkmalBIM\05_Konferenz\fbi\hbim_prototype…"
path_stl : "C:\Users\valde\Documents\05_DenkmalBIM\05_Konferenz\fbi\hbim_prototype…"
scan_file : "Wall_test.stl"
scan_technique : "Terrestrial Laser Scanning"
source : "academic simulation"
type : "IFC"
```


Figure 3: Digital Model Record for `WALL_test.ifc` in MongoDB

This integration is further represented in the semantic graph layer. The `DigitalModel` node for `WALL_test.ifc` is connected to its STL mesh, scanning process, and the individual IFC components that are part of the model. Each of those components is also linked to restoration events, materials, and responsible actors, enabling complete semantic tracing from the model level down to individual interventions (see Figure 4).

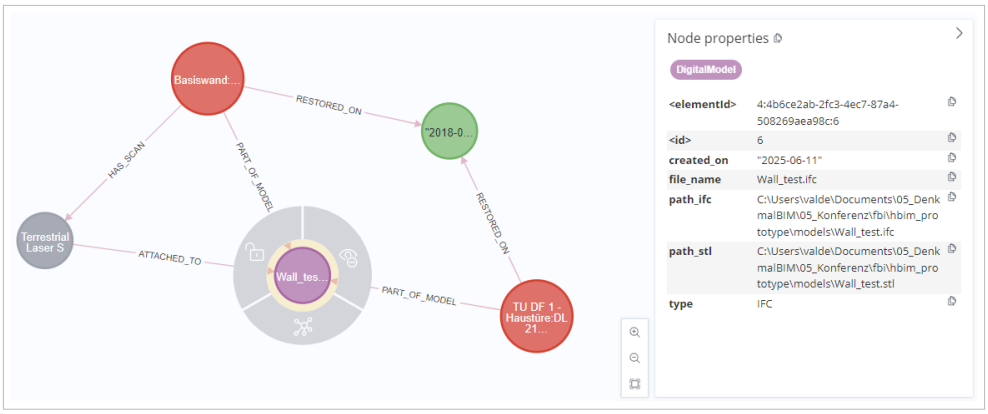


Figure 4: Digital Model Record for `WALL_test.ifc` in MongoDB

This structure allows digital models to act not only as geometric references, but also as information containers that connect technical documentation, historical interventions, and survey data. It supports advanced queries that bridge IFC elements, restoration metadata, and spatial models in an integrated and extensible way.

4.3 Evaluation and Limitations

The prototype validates the feasibility of a hybrid NoSQL architecture for HBIM, enabling structured data management and semantic enrichment of restoration information. It effectively links architectural elements with their historical and material context and offers a scalable foundation for future development. Limitations include the absence of version control, full ontology alignment, and a front-end interface. Coordination between document and graph layers was achieved via scripts rather than automated integration.

Despite these constraints, the system demonstrates clear practical value. Its combination of flexible storage and semantic reasoning exceeds the capabilities of conventional HBIM tools and supports advanced heritage conservation workflows.

5 Conclusions and Future Work

5.1 Conclusions

This research presented and evaluated a hybrid NoSQL architecture designed to support scalable and semantically enriched data management in HBIM. Addressing challenges such as fragmented sources, limited interoperability, and rigid relational models, the system combines document-based flexibility (MongoDB) with semantic graph reasoning (Neo4j) to improve restoration workflows and heritage documentation.

The prototype confirmed the feasibility of managing heterogeneous HBIM data across two layers. Documents support the ingestion of IFC files, point clouds, and restoration records, while the graph layer enables reasoning over events, materials, and actors using CIDOC CRM alignment.

The system addresses key gaps in current HBIM practices, particularly the integration of historical data into structured, queryable models. Though version control and synchronization were not implemented, the architecture anticipates these capabilities and supports future development.

By aligning with open standards and enabling RDF/JSON-LD export, the system offers a reusable and extensible foundation. Compared to conventional HBIM tools, it enhances data interpretation, interoperability, and long-term preservation.

5.2 Future Work

Future development will focus on transforming the prototype into a fully operational platform for real-world heritage documentation. Several key directions include:

- **Semantic and Temporal Enrichment:** Further integration of temporally annotated interventions and event-based modeling will build on the existing use of CIDOC CRM in Neo4j. Future work may include inferencing or historical reasoning features.
- **Interoperable Integration:** Development of standardized middleware will enable consistent connections between components, allowing integration with GIS platforms, IFC viewers, and heritage registries.
- **Linked Data Compatibility:** Connecting to external vocabularies (e.g., Wikidata, Getty AAT) through persistent URIs and SPARQL endpoints would enhance semantic interoperability and linked open data access.
- **User Interface Development:** A visual interface would improve accessibility, allowing heritage professionals to navigate the semantic graph, interact with data, and execute queries more intuitively.
- **Real-World Validation:** Applying the system in partnership with heritage institutions will be essential for assessing scalability, usability, and workflow integration.

In summary, this research establishes a foundation for a modular, semantically enriched HBIM infrastructure. It supports the long-term reuse, interpretation, and interoperability of restoration data—ensuring that interventions are not only documented but meaningfully connected across time and systems.

References

[1] M. Murphy, E. McGovern, and S. Pavia, "Historic building information modelling (hbim) - adding intelligence to laser and image based surveys of european classical architecture", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 76, no. 4, pp. 89–102, 2013. DOI: 10.1108/02630800910985108

[2] A. Baik and Y. Alshawabkeh, "Harnessing heritage bim for enhanced architectural documentation of ad deir in petra", *Applied Sciences*, vol. 14, no. 11, p. 4562, 2024. DOI: 10.3390/app14114562

[3] F. Giuliani, F. Gaglio, M. Martino, and A. de Falco, "A hbim pipeline for the conservation of large-scale architectural heritage: The city walls of pisa", *Heritage Science*, vol. 12, no. 1, 2024. DOI: 10.1186/s40494-024-01141-4

[4] T. Penjor, S. Banihashemi, A. Hajirasouli, and H. Golzad, "Heritage building information modeling (hbim) for heritage conservation: Framework of challenges, gaps, and existing limitations of hbim", *Digital Applications in Archaeology and Cultural Heritage*, vol. 35, pp. 1–15, 2024. DOI: 10.1016/j.daach.2024.e00366

[5] F. Matrone, E. Colucci, V. de Ruvo, A. Lingua, and A. Spanò, "Hbim in a semantic 3d gis database", *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W11, pp. 857–865, 2019. DOI: 10.5194/isprs-archives-XLII-2-W11-857-2019

[6] M. Schlögl, M. Durco, I. Börner, P. Andorfer, and K. Illmayer, *Connecting the dots: The case of omnipot*, 2020. DOI: 10.1007/978-3-319-63962-8

[7] M. Doerr, K. Schaller, and M. Theodoridou, "Integration of complementary archaeological sources", in *Beyond the Artifact: Digital Interpretation of the Past. Proceedings of CAA2004*, Archaeolingua, 2010, pp. 64–69.

[8] T. Tietz, O. Bruns, H. Fliegl, E. Posthumus, T. Schrade, and H. Sack, "Knowledge graph-basierte forschungsdatenintegration in nfdi4culture", in *DHd2023: Open Humanities, Open Culture: Zenodo*, A. Busch and P. Trilcke, editors, Zenodo, 2023. DOI: 10.5281/zenodo.7748740

[9] F. Michel, "Integrating heterogeneous data sources in the web of data", Ph.D. dissertation, Université Côte d'Azur, 2017.

[10] V. A. Cotella, "From 3d point clouds to hbim: Application of artificial intelligence in cultural heritage", *Automation in Construction*, vol. 152, p. 104 936, 2023. DOI: 10.1016/j.autcon.2023.104936

[11] F. Kräutli and M. Valleriani, "Corpustracer: A cidoc database for tracing knowledge networks", *Digital Scholarship in the Humanities*, vol. 33, no. 2, pp. 336–346, 2018. DOI: 10.1093/llc/fqx047