

Article

A Semantic Backbone for Heritage Digital Twins: Ontology Development and Knowledge Graph Generation

Gozde Basak Ozturk ^{1,2,*} , Fatih Soygazi ³  and Busra Ozen ² ¹ Civil Engineering Division, Department of Engineering, University of Cambridge, Cambridge CB3 0FA, UK² Civil Engineering Department, Aydin Adnan Menderes University, 09100 Aydin, Türkiye³ Computer Engineering Department, Aydin Adnan Menderes University, 09100 Aydin, Türkiye

* Correspondence: gbozturk@adu.edu.tr

Abstract

This study develops an ontology-driven framework for the semantic enrichment of the Ephesus Ancient City within a Heritage Digital Twins (HDT) context. The proposed approach integrates heterogeneous cultural heritage data into a unified semantic structure, enabling consistent representation of complex relationships. A domain-specific ontology is developed, followed by the construction of a knowledge graph in Neo4j to support structured querying. The framework moves beyond visualization-oriented digital twins by introducing a knowledge-centered semantic layer intended to support future reasoning and decision-making capabilities. Rather than constituting a complete digital twins system, the proposed approach establishes a semantic backbone that can serve as a foundation for future Heritage Digital Twins implementations. The Ephesus case suggests the practical applicability of the framework on a complex CH site. The results suggest a scalable and semantic foundation for developing interoperable Heritage Digital Twins.

Keywords: digital twins; heritage digital twins; ontology; knowledge graph; semantic enrichment; cultural heritage

1. Introduction

Cultural heritage (CH) assets are constantly under threat from a range of negative external forces, including natural disasters, the passage of time, human activity, and environmental factors. In the meantime, these assets bring together knowledge from multiple domains such as history, architecture, archaeology, and social sciences [1,2]. These pressures give rise to two primary knowledge management problems, namely, the physical disappearance of the CH asset itself, and the consequent loss of the knowledge associated with it. Even when full disappearance is avoided, assets may suffer damage, leading to incomplete or missing CH knowledge. CH knowledge management solutions can be approached on two levels to address these challenges. The first solution is physical knowledge management, which deals with the preservation of the tangible asset, and the second is virtual knowledge management, which focuses on capturing and sustaining the knowledge digitally. Supporting this virtual layer is a broad toolkit of digital technologies that form an integrated ecosystem for virtual CH knowledge management, enabling the documentation, reconstruction, and long-term accessibility of cultural heritage assets even when the physical originals are lost or deteriorated (Figure 1). Recent advances in digital technologies have accelerated the collection of CH data from different sources, including 3D models, archival records, and textual content [3,4]. However, this data often remains



Academic Editor: Jose María Alvarez Rodríguez

Received: 25 June 2026

Revised: 10 July 2026

Accepted: 14 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

scattered across disconnected platforms. Differences in formats, standards, and terminology limit interoperability and reuse [5]. As a result, managing CH knowledge in an integrated and sustainable way is still a major challenge [6].

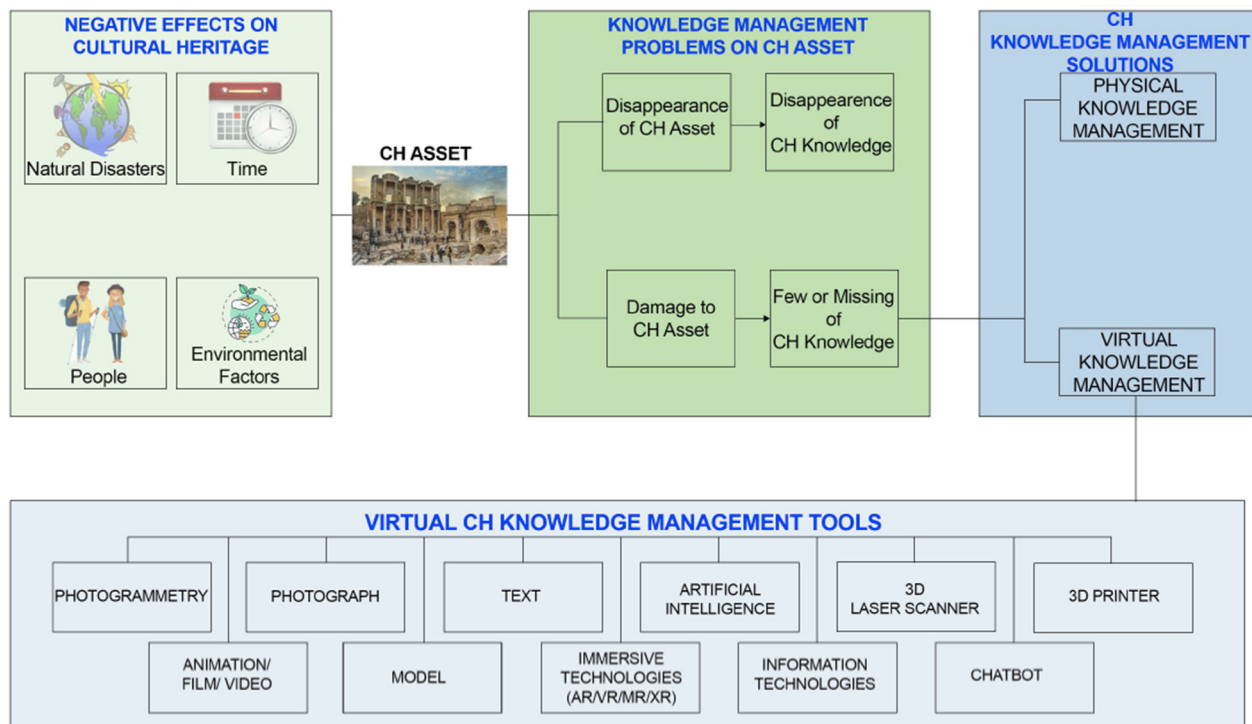


Figure 1. Problems, solutions and tools in cultural heritage knowledge management.

Ontologies, which can help to structure domain knowledge and define relationships between concepts in a machine-readable way, have been widely used to address these issues [7,8]. In CH research, ontology-based approaches support the integration of interdisciplinary data and improve semantic interoperability [9,10]. At the same time, Knowledge Graphs (KGs) extend this capability by connecting data through semantic relationships. This allows more advanced querying and knowledge sharing [11]. Existing KG applications often lack domain-specific depth or rely on loosely connected data sources [12].

Recent studies on Heritage Digital Twins (HDTs) highlight a similar limitation. Digital Twins connect physical assets with their digital representations through data and models [13] and public engagement [14,15]. However, many HDT applications do not include a strong semantic layer. This limits their ability to support interoperability, reasoning, and intelligent interaction [16]. A structured and semantically enriched knowledge base is still missing.

The literature shows that ontology development, semantic enrichment, and knowledge graph construction are often studied separately. Some studies focus on ontology design, while others explore enrichment techniques or KG creation [17,18]. However, there is limited work that connects these steps into a single workflow. In particular, the transformation from ontology to a fully enriched CH knowledge graph remains underexplored. This gap becomes more critical when considering applications such as HDTs, which require integrated and continuously evolving knowledge structures.

This study proposes an integrated semantic framework that connects ontology development, semantic enrichment, and knowledge graph construction within a unified workflow. The proposed approach does not aim to deliver a fully operational digital twins system. Instead, it provides a semantic foundation that enables the structured integration of knowledge, supporting the future development of Heritage Digital Twins. The framework

is evidenced to be a prototype through its development using Ephesus as a case and is positioned as a semantic foundation for Heritage Digital Twins [19,20].

The rest of the paper is organized as follows. Section 2 reviews the literature and defines the research gap. Section 3 presents the methodology. Section 4 reports the results. Section 5 discusses the findings. Section 6 concludes the paper.

2. Related Works

This section reviews the literature across four thematic areas that underpin the present study, namely cultural heritage ontologies, knowledge graphs in cultural heritage, semantic enrichment approaches, and digital twins in cultural heritage. The review identifies the current state of research, highlights emerging directions, and establishes the gaps that motivate this work.

2.1. Cultural Heritage Ontologies

Cultural heritage (CH) ontologies provide structured representations of heritage knowledge by defining entities, relationships, and constraints that support interoperability across heterogeneous datasets. Recent studies move beyond schema definition toward extensibility and context-awareness. These approaches attempt to capture layered historical interpretations, uncertain temporal boundaries, and incomplete archaeological records [21,22]. Domain-specific extensions are increasingly introduced to represent architectural components, conservation processes, and material properties. Despite these advances, a key limitation lies in reusability and cross-domain scalability. Many ontology implementations remain project-specific, making it difficult to transfer knowledge models across different CH sites or integrate them into broader digital systems. This highlights the need for ontology designs that balance domain specificity with interoperability and long-term adaptability.

2.2. Knowledge Graphs in Cultural Heritage

Knowledge graphs (KGs) extend ontologies by enabling flexible, graph-based data structures that support large-scale integration across diverse datasets. In the CH domain, KGs are increasingly used to connect archival records, 3D models, and historical narratives, improving data discoverability and semantic querying capabilities [23]. Graph databases such as Neo4j have become widely adopted due to their ability to efficiently model complex relationships. These systems enable multi-hop queries and richer contextual analysis compared to traditional data models [24]. However, a major challenge lies in semantic consistency and alignment across heterogeneous sources. When integrating multiple datasets, inconsistencies in terminology, structure, and granularity often emerge. Existing alignment techniques remain insufficient for resolving these discrepancies in complex cultural heritage contexts, limiting the reliability and interpretability of resulting knowledge graphs.

2.3. Semantic Enrichment Approaches

Semantic enrichment aims to transform raw or unstructured data into meaningful, machine-readable information. In cultural heritage, this process enables the integration of textual, geometric, and contextual data into a unified semantic framework. Early approaches relied on manual annotation, which ensured high accuracy but required significant expert effort. Recent research has shifted toward automated and semi-automated methods using natural language processing (NLP) and machine learning techniques. Named entity recognition and relation extraction are widely applied to process archival texts and historical documentation [25]. Another important research direction focuses on linking geometric data, such as HBIM models and point clouds, with semantic representations. However, a key limitation remains in handling domain-specific ambiguity and interpretive uncertainty.

Cultural heritage data often involves contested interpretations, incomplete records, and context-dependent meanings, which automated enrichment methods struggle to capture reliably without expert validation.

2.4. Digital Twins in Cultural Heritage

Digital twins (DTs) in cultural heritage aim to create dynamic, data-driven representations of physical assets. These systems integrate geometry, semantics, and real-time data streams. They support monitoring, conservation, and decision-making. Recent literature shows rapid growth in this area. Recent studies explore DT applications in buildings and infrastructure systems. These works highlight the potential of DTs to improve asset management and risk assessment. Despite this progress, many implementations remain at an early stage. Most systems focus on visualization and data integration. They do not fully exploit predictive analytics or autonomous decision-making. This limitation aligns with the lower maturity levels of digital twins. Another critical issue concerns semantic integration. Many DT frameworks lack a robust semantic backbone. Data from sensors, models, and archives often remain loosely connected. Without a strong semantic layer, interoperability and scalability suffer [26]. Recent studies begin to address this gap [27]. Researchers propose combining ontologies and knowledge graphs with DT architectures. This integration supports better data alignment and enables more advanced reasoning [28,29]. However, practical implementations are still limited. In addition, most DT models do not capture the complexity of heritage environments. They often simplify temporal dynamics and ignore uncertainties in historical data.

2.5. Research Gaps and Positioning

A comparative analysis of representative studies published between 2022 and 2026 reveals a consistent pattern across the literature (Figure 2). There are four principal components examined, namely: ontology development, knowledge graph construction, NLP-based semantic enrichment, and Heritage Digital Twins integration. Cultural heritage ontologies often struggle with scalability beyond project-specific implementations, knowledge graphs face challenges in maintaining semantic consistency across heterogeneous sources, and semantic enrichment approaches encounter difficulties in handling interpretive ambiguity inherent in multidisciplinary heritage data. Collectively, these limitations result in systems that remain fragmented, difficult to generalize, and limited in their ability to support integrated, continuously evolving digital environments such as Heritage Digital Twins.

There are three specific gaps identified in the literature. The first concerns technical integration. Studies that develop strong ontological models [5,30] achieve robust compliance with existing ontologies but do not connect their models to NLP-based enrichment pipelines or to knowledge graph databases. Conversely, studies that apply NLP to heritage texts operate independently of ontological standards and produce outputs that are not interoperable with existing systems. Studies that implement knowledge graphs in Neo4j or GraphDB rarely pair their graph structures with a formally specified domain ontology aligned to international standards. The result is a set of technically capable but disconnected contributions, none of which demonstrates an end-to-end workflow from ontology to enriched knowledge graph to digital twins foundation.

The second gap concerns CH application. The most technically advanced recent frameworks, including [29], remain at a conceptual level without CH-specific implementation and do not apply their methods to heritage contexts. Felicetti and Niccolucci [30] ground its work in the documentation of individual artworks, which does not expose the challenges of integrating heterogeneous, temporally stratified, and institutionally distributed data

that characterise large CH sites. As Huggett [31] has argued, automated enrichment tools applied without domain-expert validation risk standardising interpretive decisions.

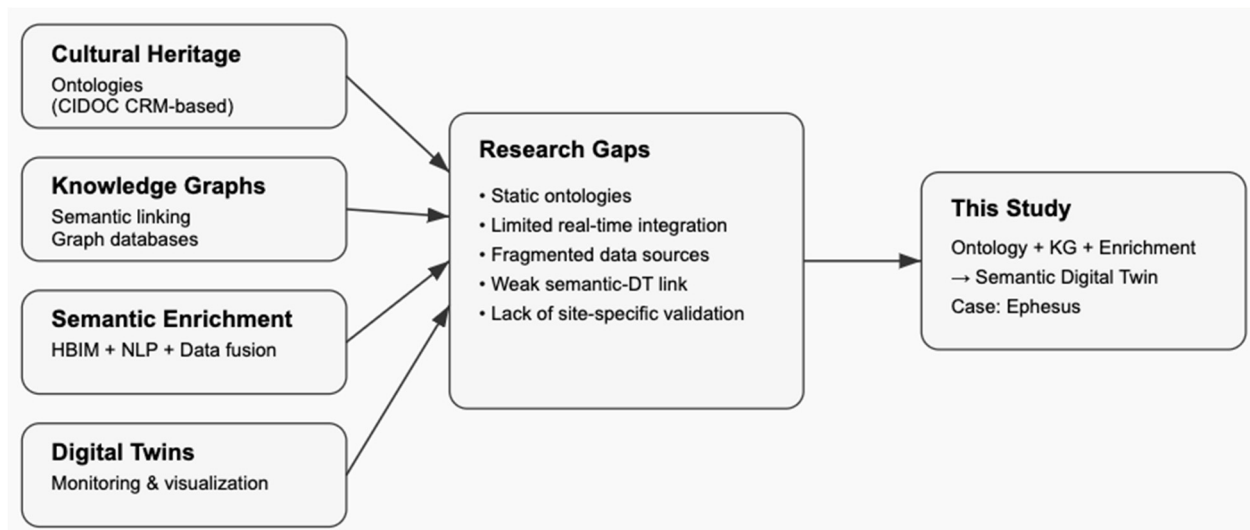


Figure 2. Positioning of the study within the existing literature.

The third gap concerns data source transparency and reliability. A recurring limitation in existing enrichment studies is the use of general web sources, encyclopaedic databases, or unspecified text corpora as input to NLP pipelines. Conflation of institutionally validated and unverified sources introduces factual uncertainty into the resulting knowledge graphs, particularly for CH where the archaeological record is complex, contested, or incompletely digitised. Heath [32] points to the need for more explicit data modelling decisions in heritage knowledge graph construction, noting that even well-structured field data requires transparent semantic choices during graph conversion. This study responds to this concern by restricting textual inputs to verified web-based sources and publications to ensure annotation consistency.

In the broader perspective of Heritage Digital Twins research, recent developments confirm the direction this study has taken. The work in [30] identifies the combination of ontologies and AI as the defining characteristic of a new generation of Heritage Digital Twins capable of both reasoning and knowledge accumulation. This study seeks to address this gap specifically in the CH domain. The present contribution positions itself at the convergence of these emerging directions. Therefore, this study aims to address these gaps by developing a domain-specific ontology for Ephesus, integrates an NLP-based semantic enrichment pipeline validated against expert-reviewed sources, constructs a knowledge graph in Neo4j representing the heritage data, and positions this integrated semantic layer as the foundation for Heritage Digital Twins.

Building on the identified gaps, this study makes four main contributions (Figure 2). First, it develops a domain-specific CH ontology, capable of representing the complex, multi-layered structure of a CH. Second, it introduces a semantic enrichment pipeline using an NLP-based approach. Third, it constructs a CHKG implemented in Neo4j from this enriched data, enabling structured querying and multi-hop semantic reasoning. The end output is a prototype to demonstrate the system as a platform to be evolutionarily improved. Finally, it positions the integrated ontology–knowledge graph framework as a semantic backbone for Heritage Digital Twins. Rather than delivering a complete digital twins implementation, the study focuses on establishing the underlying knowledge infrastructure required for future, more advanced HDT systems.

- It develops an Cultural Heritage Ontology that brings together knowledge from multiple domains;
- It introduces a semantic enrichment pipeline that integrates structured and unstructured data;
- It constructs a Cultural Heritage Knowledge Graph (CHKG) populated from this enriched data, that aims to support interoperability and enhance knowledge sharing through a structured semantic approach;
- It positions the ontology-KG framework as a semantic foundation for Heritage Digital Twins.

Current research streams in cultural heritage ontologies, knowledge graphs, semantic enrichment, and digital twins remain largely fragmented. This study addresses the identified gaps by integrating these domains into an integrated semantic framework that connects ontology development, semantic enrichment, and knowledge graph construction within a unified workflow (Figure 2). This in turn provides a semantic foundation that enables the structured integration of knowledge, supporting the future development of Heritage Digital Twins.

The following research questions are sought to be addressed in this study.

RQ1. How can a domain-specific ontology be structured to represent the complex knowledge of a CH?

RQ2. How can ontology-based knowledge be transformed into a knowledge graph to support scalable and flexible semantic relationships?

RQ3. What semantic enrichment strategies can effectively integrate CH data sources within a unified framework?

RQ4. How can the integration of ontology, knowledge graph, and semantic enrichment contribute to the development of semantically enriched digital twins for cultural heritage?

3. Methodology

This section presents the methodology developed to address the research questions defined in Section 2. The methodology is structured into four stages, described in this section. The proposed methodology follows a layered architecture to support Heritage Digital Twins development. At the core, a domain-specific ontology is designed to define the semantic schema of cultural heritage entities and relationships. This ontology serves as the foundational layer for structuring heterogeneous data. Building upon this layer, semantic enrichment processes are applied to link and augment data with external knowledge sources. Finally, a knowledge graph is constructed as the semantic knowledge population layer, instantiating the ontology with enriched data and enabling advanced querying and interoperability. This layered approach ensures that the ontology functions as a semantic backbone.

The process of ontology development encompassed distinct stages, including goal and scope definition, information gathering, initial structuring, formalization and evaluation. These sequential steps were executed to ensure the systematic construction of the CH Ontology (Figures 3 and 4).

3.1. Overall Framework

This study follows a structured workflow that combines ontology development, semantic enrichment, and knowledge graph construction within a unified framework. The process moves from domain understanding to system implementation. Each step directly addresses the research questions defined in Section 2.

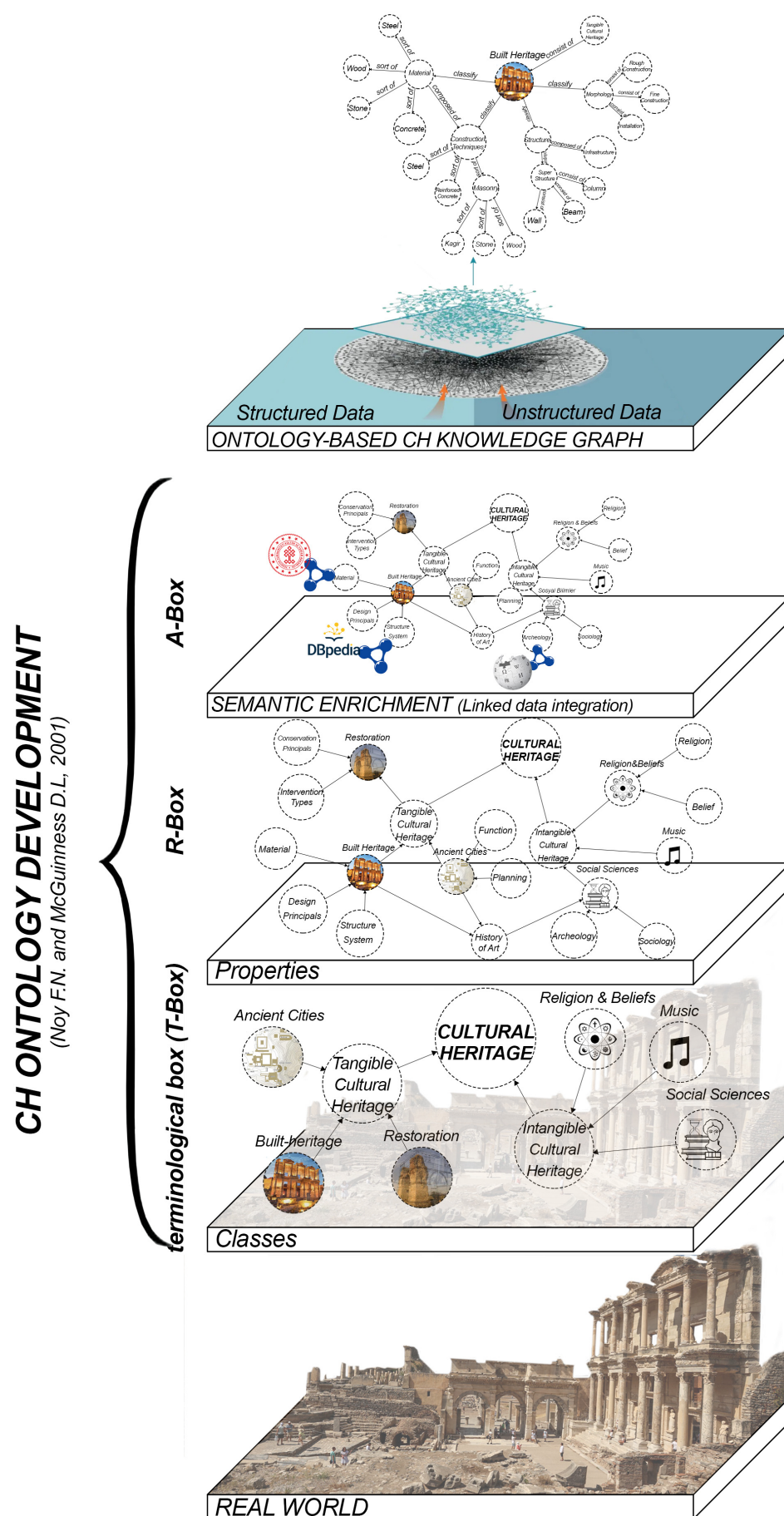


Figure 3. Ontology Development and Semantic Enrichment Process.

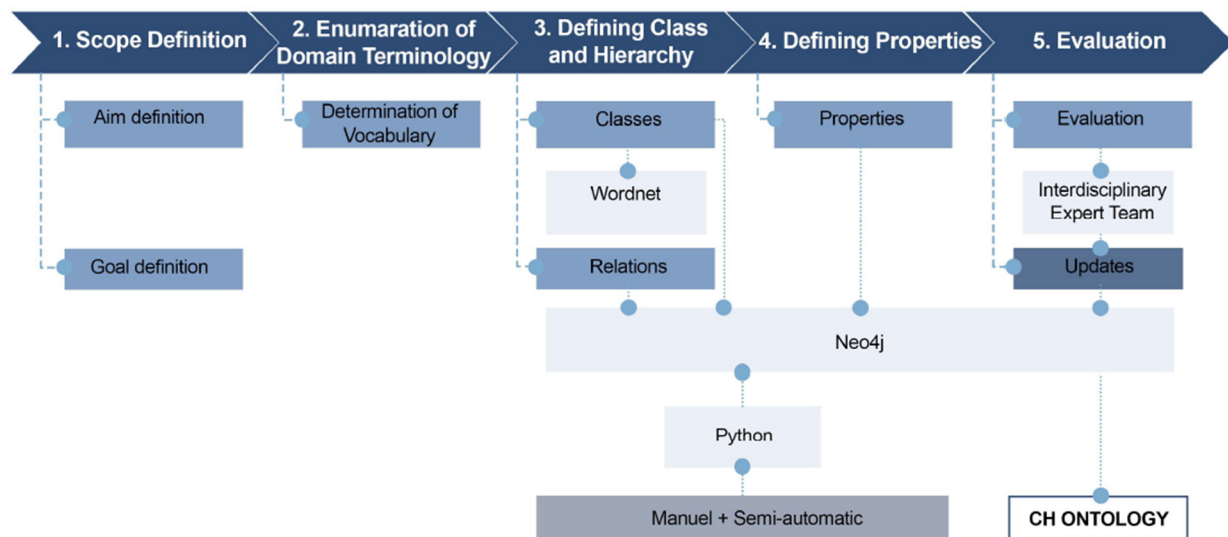


Figure 4. Ontology Development Process.

The framework includes five main stages. First, the scope of the cultural heritage domain is defined based on the cultural heritage asset. Second, domain knowledge is collected from multiple sources, including verified web-based sources and publications. Third, an ontology is designed and formalized. Fourth, semantic enrichment techniques are applied to integrate heterogeneous datasets. Finally, the enriched data is transformed into a knowledge graph and connected to the heritage digital twins structure. The workflow avoids isolated modelling steps. Instead, it treats ontology, enrichment, and graph construction as interdependent processes. This integrated approach reflects recent shifts in semantic digital twins research [29]. It also supports extensible data integration. The framework is designed as a semantic backbone. It does not implement a full digital twins system, but instead supports the structured representation and integration of knowledge required for future HDT development.

3.2. Cultural Heritage Ontology Development

The ontology development process follows a domain-driven approach. It builds on established semantic standards while adapting them to the specific characteristics of the CH asset. The process includes five steps: scope definition, knowledge acquisition, class design, property definition, and implementation. Consistent with established ontology engineering methodology [8], the process is treated as necessarily iterative: the evolving ontology is revisited and refined at each stage, with earlier decisions revised as new domain knowledge or modelling requirements emerge.

3.2.1. Scope Definition

The expert panel that guided the ontology development consisted of five domain specialists with complementary expertise. The panel included: (i) an archaeologist from Aydin Adnan Menderes University with fieldwork experience at Ionian period sites in western Turkey; (ii) an art and architecture historian; (iii) an architect; (iv) a civil engineer with expertise in Heritage Building Information Modelling (HBIM); and (v) a computer engineer specialized in semantic web technologies and ontology engineering. Expert consultations were conducted in two rounds, the first of which was an initial interview phase for concept elicitation, and the second as a subsequent review phase in which panelists evaluated the first version of the ontology. The scope focuses on representing the architectural, archaeological, and contextual elements of the CH asset. The model captures both tangible and intangible components. These include structures, artifacts, historical

events, and conservation activities. The ontology is developed in English, which serves as the working language for all class labels, property names, and definitions. English was selected to maximize international accessibility and to ensure compatibility with major heritage ontology standards, which are also expressed in English [22]. The scope remains intentionally bounded. Overly detailed classifications are avoided. This decision improves usability and reduces model complexity. Recent studies emphasize the importance of controlled scope in ontology design, especially for heritage systems with high data variability [21,22]. Following this approach, the ontology targets key entities and relationships. A set of competency questions was formulated to operationalize the scope [33]. These questions define the types of queries the knowledge base must be capable of answering. Competency questions are presented in detail in Section 4.

3.2.2. Domain Knowledge Acquisition

Domain knowledge is collected from multiple sources. The process combines manual review with structured data extraction. The primary data sources used for the ontology data fusion are verified as (i) scholarly publications; (ii) web-based sources. Key concepts and relationships are identified through iterative analysis. Terminology is aligned across sources to reduce semantic inconsistencies. This step ensures that the ontology reflects both academic knowledge and site-specific characteristics.

3.2.3. Class and Hierarchy Design

The class structure defines the core entities of the ontology. These include classes such as *Building*, *Artifact*, *Material*, *Historical Event*, and *Conservation Activity*. The hierarchy organises these classes into logical relationships. The hierarchy remains modular. This allows future extensions without restructuring the entire model. Clear class definitions improve consistency across datasets. They also support semantic queries and reasoning. Recent studies show that well-structured hierarchies play a key role in interoperability [24].

3.2.4. Property Definition

Properties define the relationships between classes. Both object properties and data properties are included. Object properties capture semantic relationships such as *located_in*, *constructed_by*, and *associated_with*. Data properties describe attributes such as material type, construction date, and condition state. Property definitions follow a clear naming convention. Each property has a defined domain and range. This structure reduces ambiguity and supports automated reasoning. Special attention is given to temporal and spatial properties. These aspects are critical in cultural heritage contexts. They allow the model to represent historical changes and spatial relationships more accurately.

3.2.5. Mapping with CIDOC CRM

CIDOC CRM (ISO 21127) [34] is the internationally accepted conceptual reference model for cultural heritage documentation. The proposed ontology was aligned with the CIDOC CRM (ISO 21127) to improve semantic interoperability and facilitate integration with existing cultural heritage knowledge bases. Rather than replacing CIDOC CRM, the proposed ontology extends and specializes its concepts to represent the domain-specific knowledge requirements of the Ephesus Ancient City. CIDOC CRM is deliberately generic, providing domain-agnostic classes for cross-institutional (museums-archives-archaeological sites) interoperability but no domain-specific vocabulary. This study adds the domain-specific classes while aligning its related core classes back to CIDOC CRM (Table 1), keeping the resulting knowledge graph interoperable rather than a closed schema. For Ephesus, this dual structure is what makes the ontology operational. CIDOC CRM alignment alone would classify the Library of Celsus as an “E22 Human-Made Object,”

but the domain extension lets the graph represent it as a building of marble, dated to the Roman period, with a recorded condition state and linked excavation activity for granularity. Broader domain vocabulary (styles, hazards, conservation values) is captured via E55 Type hierarchies, CIDOC CRM's own recommended mechanism for controlled sub-vocabularies. While developed for Ephesus, the core-class structure is deliberately generic to Hellenistic-Roman urban heritage rather than site-specific, and is intended to extend to other Ionian coast cities (e.g., Miletus, Priene, Didyma) sharing comparable building typologies, materials, and periods, without requiring changes to the core schema. Table 1 presents the alignment between the core classes of the proposed ontology and their closest CIDOC CRM equivalents. This mapping ensures semantic compatibility.

Table 1. CIDOC CRM Mapping.

Ontology Class	CIDOC CRM Mapping	Mapping Rationale
heritage_asset	E18 Physical Thing	Uncontested general superclass covering any tangible entity; parents cultural_heritage and its subtypes.
building	E22 Human-Made Object	Per CIDOC CRM's own reference examples (the Colosseum, the Forth Railway Bridge), large immovable structures with discrete physical boundaries are classified as E22.
architectural_component	E22 Human-Made Object	A structural part of a building (wall, roof, superstructure/substructure) is itself a discrete human-made physical object.
artifact	E22 Human-Made Object	Direct correspondence to E22's scope notes for movable human-made objects.
material	E57 Material	Dedicated, one-to-one CIDOC CRM class for material types.
historical_event	E5 Event	Matches E5's general, non-agentive spatiotemporal occurrence scope.
conservation_activity	E11 Modification	E11's scope notes explicitly names "the preventive treatment or restoration of an object for conservation."
excavation_activity	E7 Activity	Excavation is an intentional action carried out by an actor, matching E7's general definition.
historical_period	E4 Period	E4 is defined precisely as a set of phenomena limited in space and time exhibiting common characteristics—the standard CIDOC CRM class for historical eras.
place	E53 Place	Direct correspondence to E53's definition as an extent in space.
spatial_region	E53 Place	Maps to the same CIDOC CRM class as Place; internally modeled as a physiographic sub-extent (e.g., coastal, valley, upland type) related to Place via a part-whole relationship, since CIDOC CRM does not subdivide E53 further.
person	E21 Person	One-to-one correspondence for real-world individuals.
archaeologist	E21 Person, role-typed via E55 Type	A professional role, not a distinct actor class; the individual is E21 Person, the role is an E55 Type qualifier.

Table 1. *Cont.*

Ontology Class	CIDOC CRM Mapping	Mapping Rationale
organization	E74 Group	Museums, universities, and heritage authorities are social groupings, matching E74's scope exactly.
inscription	E34 Inscription	Dedicated, purpose-built CIDOC CRM class for textual inscriptions.
documentation	E31 Document	Reports, excavation records, and publications fall directly within E31's scope.
digital_model	E73 Information Object	Standard practice in digital heritage literature for representing digital/informational content objects (including HBIM models).
image	E38 Image	Dedicated, purpose-built CIDOC CRM class for visual documentation.
condition_state	E3 Condition State	Dedicated, purpose-built CIDOC CRM class; no ambiguity.
time_span	E52 Time-Span	Dedicated, purpose-built CIDOC CRM class for temporal extents.

3.2.6. Implementation

The ontology is implemented using a graph-based approach. A Neo4j database is used to represent classes as nodes and relationships as edges. This structure supports flexible data modelling and efficient querying. The implementation process includes schema translation, data mapping, and graph population. Ontology elements are converted into graph structures while preserving semantic meaning. Graph databases provide advantages over traditional relational systems. They handle complex relationships more effectively and support scalable data integration. Recent studies confirm the growing use of Neo4j in semantic heritage applications [35].

3.3. Ontology Evaluation

The ontology is evaluated using competency questions and consistency checks. Competency questions test whether the model can answer relevant domain queries. These queries reflect real-world use cases, such as identifying relationships between structures, materials, and historical events. Consistency checks ensure logical coherence within the ontology. They verify class hierarchies, property definitions, and relationship constraints. Evaluation also includes a qualitative comparison with existing CH ontologies. This step highlights improvements in contextual representation and integration capability. Similar evaluation strategies are widely used in recent ontology studies [22,23].

3.4. Semantic Enrichment Pipeline and Knowledge Graph Population

The semantic enrichment pipeline transforms raw textual input data into structured, ontology-aligned knowledge that can be directly populated into the Cultural Heritage Knowledge Graph (CHKG). The pipeline follows a four-step sequential architecture adapted from [36] and is depicted in Figure 5: (1) Coreference Resolution, (2) Named Entity Linking, (3) Relationship Extraction, and (4) Knowledge Graph population. Each step operates on the output of the previous stage, progressively increasing the semantic precision and structural consistency of the data. To ensure factual reliability throughout, all textual inputs were restricted to verified scholarly publications and institutionally validated web-based sources, including excavation reports and peer-reviewed studies pertaining to the Ephesus archaeological site.

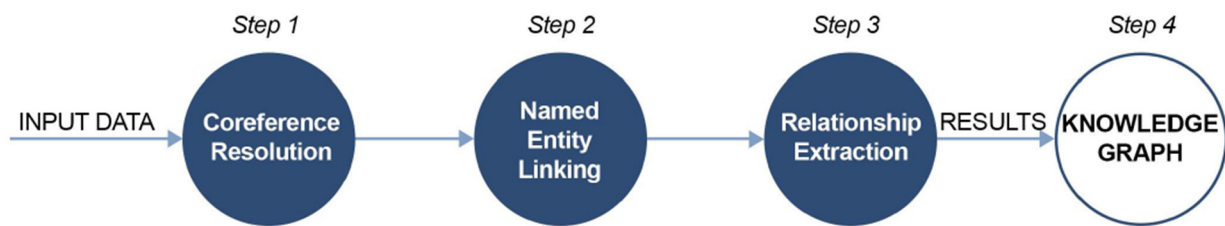


Figure 5. Semantic Enrichment of CH Ontology Steps for CH Knowledge Graph [36].

Step 1: Coreference Resolution. The first stage of the pipeline addresses the problem of co-referential expressions in heritage texts, where a single real-world entity is referred to by multiple surface forms across a document. In archaeological and historical sources, it is common for a monument such as the Library of Celsus to be referred to interchangeably as “the Library,” “the building,” “it,” or “the structure” within the same passage. Without resolving these references to a single canonical entity, downstream entity extraction would generate duplicate or fragmented nodes in the knowledge graph, undermining semantic coherence. Coreference resolution was applied to each input document prior to further processing, using a rule-based approach supplemented by contextual heuristics tailored to the domain-specific linguistic patterns observed in CH texts. Pronoun chains, nominal anaphors, and abbreviated references were all mapped to their antecedent entity mention. The output of this stage is a set of coreference-resolved documents in which each distinct real-world entity is consistently represented by a single, unambiguous surface form throughout the text. This normalization ensures that all subsequent processing stages operate on a consistent and deduplicated entity representation, reducing noise in both entity linking and relationship extraction.

Step 2: Named Entity Linking. The second stage performs named entity recognition (NER) and entity linking (NEL) on the coreference-resolved text. A domain-adapted NER model was applied to identify and classify entity mentions into the core types defined by the CH ontology: architectural structures (e.g., the Great Theatre, the State Agora), archaeological artefacts (e.g., inscriptions, statuary), historical periods (e.g., Hellenistic Period, Roman Imperial Period), spatial locations (e.g., Ionia, the Aegean coast), persons and actors (e.g., the Austrian Archaeological Institute), and heritage events (e.g., UNESCO inscription, excavation campaigns). NER alone produces surface-level entity spans but does not associate them with ontology classes or external knowledge bases. Entity linking resolves this by mapping each detected mention to a specific node in the ontology or, where applicable, to an external identifier [25,37]. Canonical entities were linked to DBpedia and Wikidata identifiers where available, providing external interoperability. These external links serve only as identifier references and do not substitute for the verified source texts used for factual content extraction in Step 1. Entities without external equivalents were assigned ontology-internal identifiers following the naming conventions established during ontology development (Section 3.2). The output of this stage is an annotated corpus in which each entity mention carries a class assignment and a unique identifier, forming the node set from which the knowledge graph will be constructed. Because these ontology classes are themselves aligned with CIDOC CRM (Section 3.2.5), each entity mention is implicitly assigned a CIDOC CRM-compatible type at this stage, rather than through a separate downstream mapping step. Manual validation by domain experts was applied at this stage to verify the correctness of entity–class assignments, particularly for ambiguous or polysemous terms common in heritage contexts.

Step 3: Relationship Extraction. The third stage identifies semantic relationships between the linked entity pairs produced by Step 2. Relationship extraction was performed using a rule-based approach grounded in syntactic dependency patterns, consistent with the

method described by [38]. Dependency parse trees were generated for each sentence, and predefined syntactic templates were matched against verb–argument structures to identify candidate relations between entity pairs. For example, the sentence “The Library of Celsus was constructed during the Roman period using marble” yields three candidate triples: (Library_of_Celsus, constructed_during, Roman_Period), (Library_of_Celsus, has_material, Marble), and, through ontology inference based on transitivity, (Library_of_Celsus, belongs_to, RomanArchitecture). Each candidate triple was evaluated against the property definitions established in the ontology (Section 3.2.4): the subject entity must fall within the domain of the candidate property, and the object entity must fall within its range. Triples that violated domain or range constraints were flagged for manual review rather than automatically discarded, preserving potentially valid relationships that may reflect ontology gaps. Approved triples were aligned with the corresponding ontology object or data properties, using the canonical property names defined during ontology development. This alignment step ensures that all extracted relationships are expressed in a vocabulary consistent with the ontology schema, enabling direct graph population without additional transformation. The output of this stage is a set of validated semantic triples of the form (subject entity, property, object entity), constituting the edge set of the knowledge graph.

Step 4: Knowledge Graph Population. The fourth and final stage of the pipeline transforms the enriched and validated triple set into a populated knowledge graph. Entities identified and linked in Step 2 are instantiated as typed nodes within the Neo4j graph database, with each node assigned the ontology class label established during entity linking and populated with data property values (e.g., construction date, condition state, material type) derived from the extraction process. Relationships validated in Step 3 are instantiated as directed, labelled edges connecting the corresponding node pairs, following the node-label and edge-type schema established in Section 3.2.6. The resulting graph is queried using Cypher to evaluate the competency questions. The complete pipeline operates as a semi-automated workflow: NLP-based extraction stages (Steps 1–3) are executed algorithmically, while a manual validation layer is applied at the entity linking and relationship approval stages to ensure factual accuracy and domain consistency. This hybrid approach balances the scalability requirements of large-scale CH data processing with the interpretive rigour demanded by the complexity and contested nature of archaeological knowledge [31]. To support reproducibility, Table 2 reports the scale and validation outcomes of the pipeline to date. Coreference resolution, named entity linking, and relationship extraction (Steps 1–3) were executed using an LLM-based agent pipeline, in which each stage applied the rule-based heuristics and syntactic-dependency patterns described above through a dedicated, prompted agent step rather than a conventional NLP library. Each stage’s output was validated before being passed to the next. Across the 39 source documents processed, the pipeline yielded 1623 triples spanning 471 instances and 21 classes in the graph; entity resolution reduced 1423 candidate entity mentions to 1356 unique entities (4.71% pre-deduplication redundancy), and a relevance filter retained 590 of 1423 candidates as relevant to the Ephesus case, reflecting the intentionally broad scope of the expanded source corpus. This reflects the corpus processed at the time of writing and will be updated as source-document coverage expands. The output of the four-step pipeline is a populated Cultural Heritage Knowledge Graph in which nodes represent ontology-aligned entities (e.g., buildings, artefacts, historical periods, conservation activities, spatial locations) and edges represent semantically typed relationships between them (e.g., *constructed_during*, *has_material*, *found_in*, *carried_out_by*).

Table 2. Semantic enrichment pipeline: scale and validation summary.

Pipeline Stage	Input Scale	Manual/Automated Validation Outcome
Coreference Resolution	39 verified source documents	-----
Named Entity Linking	1723 candidate entity mentions	Entity resolution reduced 1423 candidate entities to 1356 unique (67 duplicates; 4.71% pre-dedup redundancy). Relevance filter retained 590 of 1423 candidates as relevant to the Ephesus case study (833 discarded as tangential to the target site or period); consolidated graph contains 1623 triples across 471 instances and 21 classes.
Relationship Extraction	11,392 candidate triples	

4. Results

In order to facilitate research in the CH field and promote the exchange of information among various stakeholders, the CH Ontology has been developed. The proposed ontology has been developed to provide a comprehensive semantic representation by helping to encode knowledge through the creation of a fundamental conceptual structure for domain knowledge (Figure 6). The use of the ontology enabled responding to queries and promoted the dissemination of reusable information. To ensure its effectiveness, the ontology is designed to support interoperability among a variety of heterogeneous systems and databases.

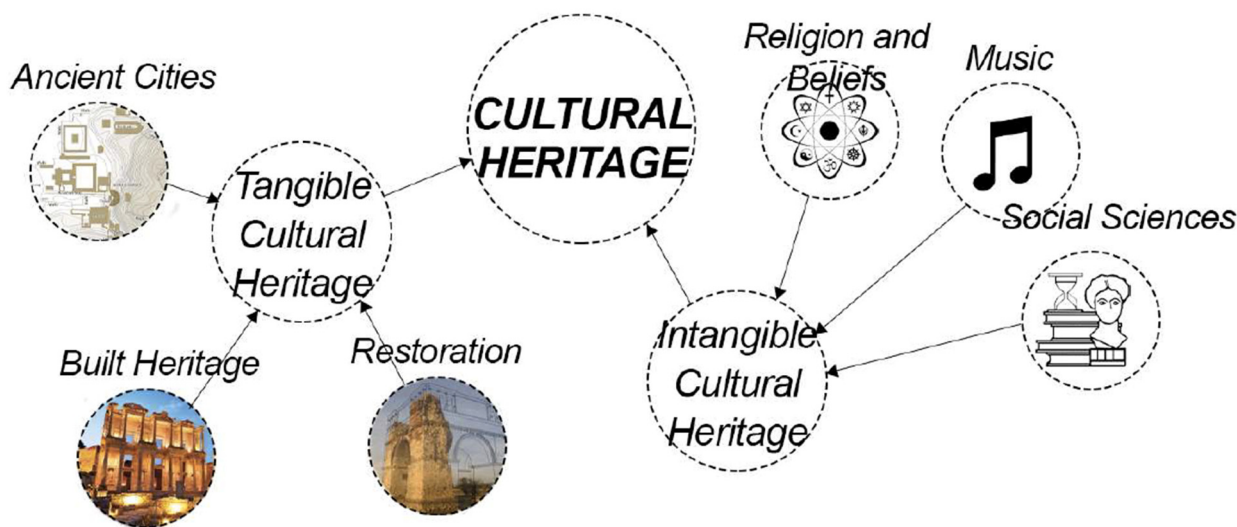


Figure 6. The Conceptual Cultural Heritage Ontology Approach.

4.1. Developed Cultural Heritage Ontology

The study produced a domain-specific ontology for ancient city buildings as CH assets. The model captures architectural elements, archaeological artifacts, historical events, and conservation activities in a unified structure. The final structure includes clearly defined classes and hierarchical relations. Core classes such as *Building*, *Artifact*, *Material*, and *Historical Event* form the backbone of the model. These classes support both descriptive and relational knowledge representation. Temporal and spatial aspects can explicitly be included. This allows the ontology to represent historical layering and support update-ready platforms for monitoring future site conditions [21,22]. The resulting ontology is modular. New classes can be added without changing the core structure. This design has the potential to improve long-term usability in digital heritage systems.

4.2. Semantic Enrichment Results

The semantic enrichment pipeline successfully integrated datasets into the ontology structure. Textual sources and spatial records were processed and linked through semantic mappings. NLP-based extraction identified key entities and relationships from verified archaeological sources, specifically excavation reports and scholarly publications about the CH asset (Figure 7). These entities were aligned with ontology classes. Only trusted web sources and peer-reviewed and institutionally validated documents were used to support factual accuracy. This process reduced ambiguity and contributed to data consistency. Similar results have been reported in recent NLP-driven heritage studies. Architectural elements were enriched with semantic attributes such as material type, condition state, and historical phase. This integration has the potential to improve the interpretability of spatial data. The enrichment process also revealed inconsistencies across data sources. These inconsistencies were resolved through manual validation and rule-based alignment. This hybrid approach is designed to support both accuracy and scalability. Overall, the enriched dataset became semantically consistent and machine-readable. It formed a reliable foundation for knowledge graph construction and digital twins integration.

Text example: Ephesus was one of the twelve cities of the Ionian League. It served as a major commercial and administrative centre of the Roman province of Asia. The city is situated on the Aegean coast of present-day Türkiye, near the mouth of the Caystros River. Among its most significant monuments are the Temple of Artemis, the Library of Celsus, the Great Theatre, and the State Agora. Ephesus was included in the UNESCO World Heritage List in 2015 as part of the Ephesus archaeological site (UNESCO, 2015). Excavations have been conducted continuously by the Austrian Archaeological Institute since 1895 except from brief interruptions due to world wars.

Figure 7. Sample text used for semantic enrichment.

The same expert panel, in its second round, reviewed the first version of the ontology seen in Figure 8. Panelists were asked to assess the ontology for completeness, correctness, and domain coverage during the second round. Rather than scoring independently and reconciling differences afterward, panelists assessed the ontology through a consensus-based approach [39]. The panel reviewed each class and relationship jointly in real time, and disagreements, where they arose, were resolved through discussion until a shared view was reached. The revisions agreed upon were incorporated directly into the ontology before finalization (Figure 9). As this procedure did not produce documented preliminary judgments per panelist separately, formal inter-rater agreement statistics could not be computed. This is noted as a limitation in Section 5.6.

4.3. Cultural Heritage Knowledge Graph

The enriched ontology data was transformed into a knowledge graph using a Neo4j-based implementation. Entities were represented as nodes, while relationships formed the edges of the graph structure. The graph integrated multiple data sources into a single interconnected network. This allowed complex relationships between structures, artifacts, and historical events to be queried efficiently. Query tests showed that the knowledge graph supports multi-hop reasoning. Relationships between architectural phases, materials, and spatial locations were retrieved without loss of semantic meaning. This confirms the suitability of graph-based models for cultural heritage applications [24,35].

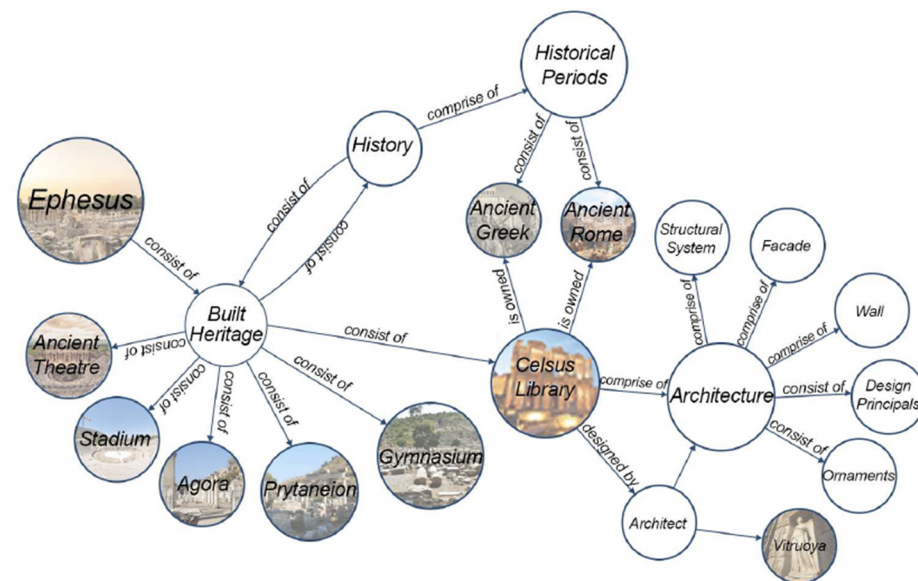


Figure 8. Ontology and Knowledge Graph Representation of the Cultural Heritage Asset.

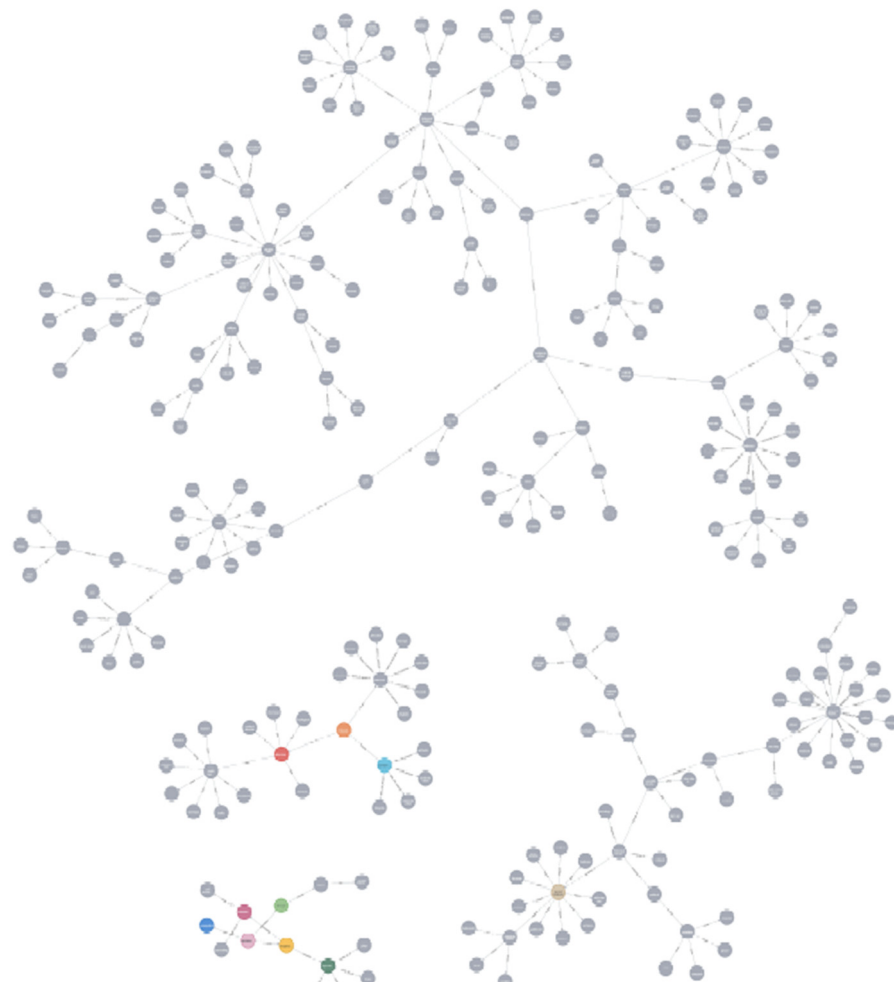


Figure 9. Developed Cultural Heritage Ontology—Class and Relationship Structure.

The final knowledge graph serves as the semantic backbone of the proposed semantic framework for future Heritage Digital Twins developments. It facilitates structured data access, contextual interpretation, and future scalability toward extensible updates. Despite these strengths, the system remains dependent on data completeness. Missing historical

records still affect certain query paths. This limitation is common in heritage digitalization efforts and has been widely reported in recent studies. The resulting knowledge graph serves as the semantic backbone of a potential digital twins environment. It enables structured data access, contextual interpretation, and scalability, forming a foundation upon which future dynamic and real-time digital twins functionalities can be developed.

4.4. Competency-Based Ontology Evaluation

The ontology was evaluated using competency questions and structural consistency checks (Table 3 and Figure 10). The competency questions tested whether the model can answer realistic domain queries related to CH assets. Five representative competency questions were formulated to reflect core use cases (Table 3). All five questions were successfully answered through Cypher queries executed against the Neo4j implementation of the ontology. This capability is not intended to demonstrate completeness, consistency, or scalability of the ontology, which remain open validation dimensions (Section 5.6).

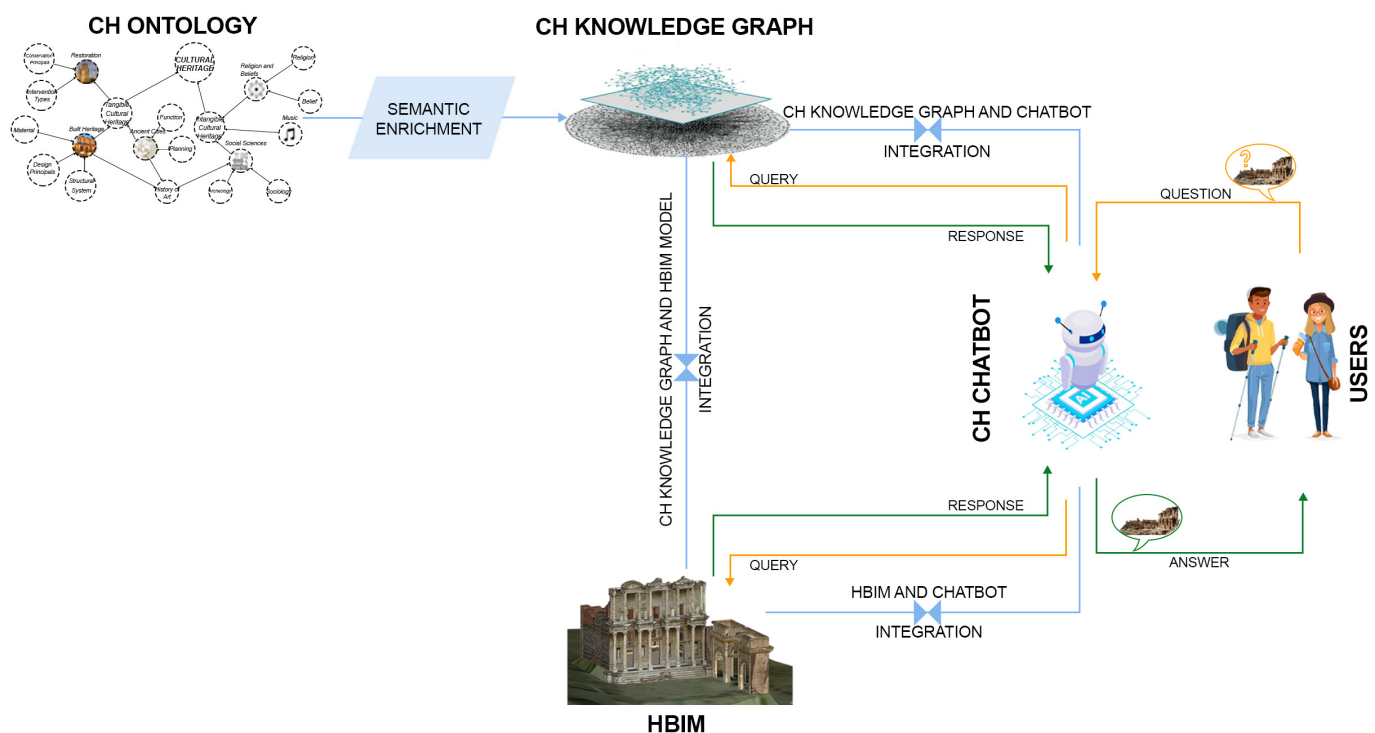


Figure 10. Conceptual Framework for Future Heritage Digital Twins Development.

Queries CQ1 through CQ3 tested entity-level retrieval and classification accuracy. CQ4 and CQ5 tested temporal and relational reasoning, which are the most demanding aspects of the ontology given the multi-period character of the site. The results confirmed that the class hierarchy and property definitions support meaningful inference across all five query types. Consistency checks showed no structural conflicts. Class inheritance and property constraints remained coherent across the model. These results indicate that the ontology maintains logical stability under different query conditions. A qualitative comparison with existing CH ontologies showed potential improvements in contextual representation. The model handles site-specific complexity better than generic heritage schemas. Recent studies also emphasize the importance of such targeted evaluation strategies in cultural heritage systems [23].

Table 3. Ontology-Driven Queries and Corresponding Results.

Competency Question	Query Description	Cypher Query	Ontology Elements Used	Expected Output	Result Type
CQ1. Which buildings were constructed during the Roman Period, and what materials were used?	Retrieves buildings associated with the Roman Period together with their construction materials.	cypher MATCH (structure:building)-[:CONSTRUCTED_DURING]->(:historical_period {title:"Roman Period"}) MATCH (structure)-[:HAS_MATERIAL]->(material:material) RETURN structure.title AS Building, material.title AS Material;	Classes: Building, Historical Period, Material Relationships: CONSTRUCTED_DURING, HAS_MATERIAL	List of Roman-period buildings and the materials used in their construction.	Building, Material
CQ2. Which conservation activities have been applied to the Library of Celsus, and who executed them?	Retrieves conservation activities associated with the Library of Celsus and the responsible organization or actor.	cypher MATCH (activity:conservation_activity)-[:APPLIED_TO]->(building:building {title:"Library of Celsus"}) MATCH (activity)-[:EXECUTED_BY]->(actor) RETURN activity.title AS ConservationActivity, actor.title AS Actor;	Classes: Conservation Activity, Building, Organization/Person Relationships: APPLIED_TO, EXECUTED_BY	Conservation activities performed on the Library of Celsus and their executors.	Activity, Actor
CQ3. Which artifacts were discovered in the State Agora, what materials are they made of, and from which historical period do they originate?	Retrieves artifacts excavated from the State Agora together with their material and associated historical period.	cypher MATCH (artifact:artifact)-[:FOUND_IN]->(:place {title:"State Agora"}) MATCH (artifact)-[:HAS_MATERIAL]->(material:material) MATCH (artifact)-[:CONSTRUCTED_DURING]->(period:historical_period) RETURN artifact.title AS Artifact, material.title AS Material, period.title AS HistoricalPeriod;	Classes: Artifact, Place, Material, Historical Period Relationships: FOUND_IN, HAS_MATERIAL, CONSTRUCTED_DURING	Artifacts found in the State Agora, their materials, and their historical periods.	Artifact, Material, Historical Period

Table 3. Cont.

Competency Question	Query Description	Cypher Query	Ontology Elements Used	Expected Output	Result Type
CQ4. Which places are located in a given region, and during which historical period were they occupied?	Retrieves places within a region together with their associated historical periods.	cypher MATCH (place:place)-[:LOCATED_IN]->(region:region) MATCH (place)-[:HAS_TIME_SPAN]->(timespan:time_span) MATCH (timespan)-[:FALLS_WITHIN]->(period:historical_period) RETURN place.title AS Place, region.title AS Region, period.title AS HistoricalPeriod;	Classes: Place, Region, Time Span, Historical Period Relationships: LOCATED_IN, HAS_TIME_SPAN, FALLS_WITHIN	Places grouped by region and their corresponding historical periods.	Place, Region, Historical Period
CQ5. Which historical events are associated with heritage entities, and when did they occur?	Retrieves historical events together with related heritage entities and their corresponding time spans.	cypher MATCH (event:historical_event)-[:HAS_TIME_SPAN]->(timespan:time_span) MATCH (event)-[:RELATED_TO]->(entity) RETURN event.title AS HistoricalEvent, entity.title AS RelatedEntity, timespan.title AS TimeSpan;	Classes: Historical Event, Heritage Entity, Time Span Relationships: HAS_TIME_SPAN, RELATED_TO	Historical events, related heritage entities, and their associated time spans.	Historical Event, Heritage Entity, Time Span

5. Discussion

Cultural heritage information is inherently heterogeneous, varying across disciplinary traditions, temporal periods, and geographical contexts. Semantic enrichment leverages the CH Ontology's expressiveness to enable the CHKG to capture relationships. The resulting CHKG facilitates broad access to CH information for diverse users, including researchers, visitors, and children. Leveraging Semantic Web technologies democratizes access, ensuring that even non-experts can engage with the knowledge base through structured query interfaces [40–42]. This accessibility fosters interdisciplinary collaboration and provides a shared platform for CH research. The integration of the CHKG with AI technologies holds significant potential for both CHKM and wider CH knowledge dissemination.

5.1. Key Findings

The study shows that ontology-driven modelling has the potential to improve how heritage data is structured and connected. In the Ephesus Ancient City case, semantic enrichment functions as a bridging layer between data sources and interpretable knowledge structures. The framework supports the preservation of contextual relationships, providing a structured basis for future digital twin applications rather than a complete operational system. Another important finding relates to knowledge continuity. The system supports long-term preservation of contextual relationships, not just isolated artifacts. This becomes critical in complex heritage environments, where spatial, temporal, and cultural dimensions overlap. Recent studies also emphasize that Heritage Digital Twins value increases when semantic layers are tightly integrated with knowledge graphs [29,30]. The results of this study align with that direction and extend it by providing an applied demonstration on primary heritage data.

The results demonstrate that the proposed semantic framework successfully addresses all four research questions. For RQ1, a site-specific ontology was created, giving structure to the entities, relationships, and contextual details that define the site. RQ2 enhanced this by transforming the ontology into a scalable knowledge graph that establishes meaningful connections between various heritage data sources and leaves room for growth as new data becomes available. RQ3 took things further through semantic enrichment. It created an interconnected dataset that is much easier to explore and work with across different systems by integrating data, linking entities, and populating the graph. RQ4 then focused on the bigger picture, outlining a conceptual framework for future Heritage Digital Twins architectures that integrates the ontology and knowledge graph at a system level, and illustrating how semantic technologies could support information exchange in future digital twins environments. These findings provide a strong rationale for the ontology-focused approach and point to its potential as a robust semantic backbone for Heritage Digital Twins systems.

5.2. Scientific Contribution

This work contributes to Heritage Digital Twins research by proposing a structured ontology, integrating semantic enrichment as a component of a future digital twins pipeline, and suggesting the applicability of the proposed semantic framework. The model moves beyond generic heritage data schemas and focuses on contextual richness, reduces the gap between data acquisition and knowledge generation, and validates the feasibility of constructing a knowledge-driven infrastructure that can support future Heritage Digital Twins development. Compared to the most closely related recent contributions [29], the study extends ontology applications from descriptive modelling toward cognitive interpretation layers.

5.3. Comparison with Literature

Existing studies in Heritage Digital Twins often focus on visualization or 3D reconstruction. These approaches are useful but remain limited at the semantic level. Recent research highlights a shift toward knowledge-driven systems [28,29]. Several works emphasize ontologies and knowledge graphs, yet many remain conceptual or partially implemented without site-specific validation [23,29]. The proposed approach differs in its full-cycle integration. Data is not only stored and visualized but also semantically processed and linked. This enables more structured querying and supports potential reasoning capabilities over heritage data when integrated with advanced digital twin systems. Compared with earlier frameworks, this model introduces stronger contextual binding between entities such as structures, events, and historical phases. This contributes to the interpretability and reuse of data across systems.

5.4. Implications for Heritage Digital Twins

The findings indicate that semantic layers should be treated as core components in the development of Heritage Digital Twins. In this study, the proposed framework represents a foundational layer rather than a complete digital twins system. It highlights the importance of establishing robust knowledge infrastructures prior to implementing extensible and update-ready functionalities. Semantic layers should be treated as core components, not optional extensions. Without them, digital twins remain descriptive systems. Interoperability improves when ontologies are aligned with domain-specific heritage concepts. This is especially relevant for large archaeological sites. In this context, decision support processes can be enhanced when such semantic layers are integrated into broader systems. Curators and researchers can trace relationships rather than isolated records. Recent research in digital heritage confirms that cognitive-level twins require structured semantic grounding to support advanced analytics and AI integration [30].

5.5. Conceptual Framework

The proposed conceptual framework positions ontology at the center of a conceptual architecture for future Heritage Digital Twins (Figure 10). Data flows from physical reality into acquisition layers. These layers feed the semantic enrichment module. The ontology then structures entities and relationships. A knowledge graph layer transforms these structures into queryable and inferable representations. On top of this, analytics and visualization modules support interpretation and interaction. The framework also defines a potential feedback loop through which future updates and new data can be incorporated into the ontology, enabling incremental system evolution in subsequent digital twins implementations. This aligns with perspectives on adaptive and learning digital twins emerging in the most recent literature [29].

5.6. Limitations

The ontology is developed using only a single heritage asset, which limits generalization to other sites. Data reliability is constrained by potential inconsistencies in long-term excavation records and gaps in the records. Recent work highlights the importance of combining textual and geometric data in heritage modelling [43,44]. However, the ontology remains primarily text-based and does not formally integrate geometric data (e.g., 3D point clouds, HBIM geometries). Because expert consensus was reached through joint discussion rather than independent scoring, inter-rater reliability statistics could not be computed for the ontology review process. The manual validation approach, while effective, limits scalability and replicability due to time intensity and expert bias. Computational scalability was not tested at city-scale, and reasoning remains rule-based without AI-driven inference.

The study does not implement an operational digital twins system, focusing instead on semantic infrastructure. Finally, the ontology is developed exclusively in English, limiting accessibility in non-English-speaking contexts.

5.7. Future Research

Future work should focus on expanding the ontology to multi-site heritage networks, including explicit temporal and spatial aspects to represent historical layering and support update-ready platforms [21,22]. Integration with large language models and multimodal AI systems represents another important direction, particularly Retrieval-Augmented Generation (RAG) for dynamic knowledge discovery [45]. Automatic validation and advanced AI techniques should also be explored to overcome the limitations of manual validation. Other research directions include real-time digital twins updates using IoT and sensor data, as well as adaptive ontologies that evolve automatically with new discoveries. Building upon the semantic backbone established in this study, future research should focus on integrating real-time data pipelines and adaptive components to realize fully functional Heritage Digital Twins environments.

6. Conclusions

This study presented an ontology-driven approach for modelling and semantically enriching heritage assets. The results show that structured semantic modelling has the potential to improve how cultural heritage data is organized, connected, and interpreted. The proposed ontology reduces fragmentation in heritage datasets and supports a more coherent representation of archaeological knowledge. The integration of semantic enrichment with a knowledge graph structure strengthens the link between raw data and contextual meaning. This connection is essential for Heritage Digital Twins that aim to move beyond visualization and toward knowledge-based understanding. Recent research increasingly supports this direction, highlighting the importance of semantic interoperability and cognitive-level digital twin architectures in cultural heritage applications. The case study indicates that ontology-based systems can capture temporal and cultural relationships in a unified structure. This has the potential to improve data traceability and support more consistent interpretation of complex archaeological environments. The framework also supports long-term knowledge continuity, which is critical for heritage preservation and reuse. Findings confirm that semantic layers should be considered a core component of Heritage Digital Twins. Without them, systems remain limited to descriptive or visualization-focused outputs. The study also shows that knowledge graph integration enhances the usability of heritage data for analysis, interpretation, and decision support. Despite its contributions, the approach has boundaries. The ontology is validated within a single site and requires extension for broader cultural and geographic contexts. Future systems should also integrate adaptive learning mechanisms and AI-driven semantic reasoning to increase autonomy and scalability. Recent advances in multimodal AI and knowledge graph learning provide strong directions for this evolution. The research results suggest a structured semantic foundation for Heritage Digital Twins. It positions ontology and knowledge graph integration as a prerequisite layer for future HDT systems, rather than as a complete implementation, thereby supporting the transition toward more advanced, dynamic, and intelligent digital twin environments.

Author Contributions: Conceptualization, G.B.O. and B.O.; methodology, G.B.O., B.O. and F.S.; validation, G.B.O., B.O. and F.S.; formal analysis, G.B.O., B.O. and F.S.; resources, G.B.O. and B.O.; writing—original draft preparation, G.B.O., B.O. and F.S.; writing—review and editing, G.B.O., B.O. and F.S.; visualization, G.B.O., B.O. and F.S.; supervision, G.B.O. and F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Aydin Adnan Menderes University Scientific Research Projects Unit, grant number MF23010 (project title: “Development of a Domain Ontology for Creating a Digital Twin Infrastructure for Cultural Heritage (CH)”). The APC was not covered.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Strlič, M. Heritage Science: A Future-Oriented Cross-Disciplinary Field. *Angew. Chem. Int. Ed.* **2018**, *57*, 7260–7261. [CrossRef] [PubMed]
2. Tibaut, A.; de Oliveira, S.G. A Framework for the Evaluation of the Cultural Heritage Information Ontology. *Appl. Sci.* **2022**, *12*, 795. [CrossRef]
3. European Commission. Commission Recommendation of 10/11/21 on a Common European Data Space for Cultural Heritage. 2021. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021H1970&from=EN> (accessed on 9 July 2026).
4. Ozturk, G.B.; Brilakis, I.; Ozen, B.; Soygazi, F. Ontology Research Fields in the Cultural Heritage Domain. In Proceedings of the 8th International Project and Construction Management Conference (IPCMC2024), Istanbul, Turkey, 2–4 October 2024. [CrossRef]
5. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [CrossRef]
6. Previtali, M.; Brumana, R.; Stanga, C.; Banfi, F. An ontology-based representation of vaulted system for HBIM. *Appl. Sci.* **2020**, *10*, 1377. [CrossRef]
7. Gruber, T.R. A translation approach to portable ontology specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [CrossRef]
8. Noy, N.F.; McGuinness, D.L. *Ontology Development 101: A Guide to Creating Your First Ontology*; Stanford Knowledge Systems Laboratory: Stanford, CA, USA, 2001; p. 25.
9. Doerr, M. Ontologies for Cultural Heritage. In *Handbook on Ontologies*; Staab, S., Studer, R., Eds.; International Handbooks on Information Systems; Springer: Berlin/Heidelberg, Germany, 2009. [CrossRef]
10. Messaoudi, T.; Véron, P.; Halin, G.; De Luca, L. An ontological model for the reality-based 3D annotation of heritage building conservation state. *J. Cult. Herit.* **2018**, *29*, 100–112. [CrossRef]
11. Elleuch, I.; Gargouri, B.; Hamadou, A.B. Lexical data mining-based approach for the self-enrichment of LMF standardised dictionaries: Case of the syntactico-semantic knowledge. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6312. [CrossRef]
12. Cursi, S.; Martinelli, L.; Paraciani, N.; Calcerano, F.; Gigliarelli, E. Linking external knowledge to heritage BIM. *Autom. Constr.* **2022**, *141*, 104444. [CrossRef]
13. Qi, Q.; Tao, F.; Zuo, Y.; Zhao, D. Digital twin service towards smart manufacturing. *Procedia CIRP* **2018**, *72*, 237–242. [CrossRef]
14. Lucchi, E. Digital twins for the automation of the heritage construction sector. *Autom. Constr.* **2023**, *156*, 105073. [CrossRef]
15. Liu, Z.; Wang, J. Protection and Utilization of Historical Sites Using Digital Twins. *Buildings* **2024**, *14*, 1019. [CrossRef]
16. Crognale, M.; De Iuliis, M.; Gattulli, V. A Framework for the Definition of Built Heritage Digital Twins. In *Handbook of Digital Twins*; CRC Press: Boca Raton, FL, USA, 2024; pp. 647–661. [CrossRef]
17. Simeone, D.; Cursi, S.; Acierno, M. BIM semantic-enrichment for built heritage representation. *Autom. Constr.* **2019**, *97*, 122–137. [CrossRef]
18. Bloch, T. Connecting research on semantic enrichment of BIM—Review of approaches, methods, and possible applications. *J. Inf. Technol. Constr.* **2022**, *27*, 416–440. [CrossRef]
19. Wang, J.; Mu, L.; Zhang, J.; Zhou, X.; Li, J. On intelligent fire drawings review based on building information modelling and knowledge graph. In *Construction Research Congress 2020: Computer Applications*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2020; pp. 812–820. [CrossRef]
20. Benabdellah, A.C.; Zekhnini, K.; Cherrafi, A.; Garza-Reyes, J.A.; Kumar, A. Design for the environment: An ontology-based knowledge management model for green product development. *Bus. Strategy Environ.* **2021**, *30*, 4037–4053. [CrossRef]
21. Hermon, S.; Niccolucci, F.; Bakirtzis, N.; Gasanova, S. A Heritage Digital Twin ontology-based description of Giovanni Baronzio’s “Crucifixion of Christ” analytical investigation. *J. Cult. Herit.* **2024**, *66*, 48–58. [CrossRef]
22. Hyvönen, E. Publishing and using cultural heritage linked data: Current trends and future directions. *Data Intell.* **2025**, *7*, 1–25. [CrossRef]

23. Desul, S.; Mahapatra, R.K.; Patra, R.K.; Sethy, M.; Pandey, N. Semantic technology for cultural heritage: A bibliometric-based review. *Glob. Knowl. Mem. Commun.* **2025**, *74*, 1356–1380. [CrossRef]
24. Syed, R.A.B.; Agliata, R.; Mecca, I.; Mollo, L. Semantic web technologies in construction facility management: A bibliometric analysis and future directions. *Buildings* **2025**, *15*, 3845. [CrossRef]
25. İnan, E.; Yönyül, B.; Tekbacak, F. A domain specific entity linking approach consuming multistore environment. *Akıllı Sist. Uygulamaları Derg.* **2018**, *1*, 46–52. [CrossRef]
26. Ozturk, G.B.; Ozen, B. Artificial Intelligence Enhanced Cognitive Digital Twins for Dynamic Building Knowledge Management. In *Handbook of Digital Twins*; CRC Press: Boca Raton, FL, USA, 2024; pp. 354–369. [CrossRef]
27. Ozturk, G.B.; Ozen, B.; Soygazi, F. Novel Technology Use for Digital Transformation of Cultural Heritage. *J. Constr. Eng. Manag. Innov.* **2024**, *7*, 172–188. [CrossRef]
28. Opoku, D.G.J.; Perera, S.; Osei-Kyei, R.; Rashidi, M. Digital twin application in the construction industry: A literature review. *J. Build. Eng.* **2021**, *40*, 102726. [CrossRef]
29. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Autom. Constr.* **2020**, *114*, 103179. [CrossRef]
30. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2025**, *10*, 1. [CrossRef]
31. Huggett, J. Archaeological Practice and Digital Automation. In *Digital Heritage and Archaeology in Practice*; Watrall, E., Goldstein, L., Eds.; University Press of Florida: Gainesville, FL, USA, 2022; pp. 275–303, ISBN 9780813069302.
32. Heath, S. Narrating Transitions and Transformation in Cultural Heritage Digital Workflows Using a JSON-Encoded Dataset of Roman Amphitheaters. In *Digital Heritage and Archaeology in Practice*; Watrall, E., Goldstein, L., Eds.; University Press of Florida: Gainesville, FL, USA, 2022; pp. 71–79. [CrossRef]
33. Alharbi, R.; Tamma, V.; Payne, T.R.; de Berardinis, J. A Comparative Study of Competency Question Elicitation Methods from Ontology Requirements. *arXiv* **2025**, arXiv:2507.02989. [CrossRef]
34. ISO 21127:2023; Information and Documentation—A reference Ontology for the Interchange of Cultural Heritage Information. International Organization for Standardization: Geneva, Switzerland, 2023. Available online: <https://www.iso.org/obp/ui/en/#iso:std:iso:21127:ed-3:v1:en> (accessed on 9 July 2026).
35. Rao, Z.; Wang, G. Smart Data-Enabled Conservation and Knowledge Generation for Architectural Heritage System. *Buildings* **2025**, *15*, 2122. [CrossRef]
36. Bratanic, T. Extract Knowledge from Text: End-to-End Information Extraction Pipeline with spaCy and Neo4j. Towards Data Science (Medium). 2021. Available online: <https://medium.com/data-science/extract-knowledge-from-text-end-to-end-information-extraction-pipeline-with-spacy-and-neo4j-502b2b1e0754> (accessed on 9 July 2026).
37. Çiftçi, O.; Soygazi, F.; Tekir, S. Enrichment of Turkish Question Answering Systems Using Knowledge Graphs. *Turk. J. Electr. Eng. Comput. Sci.* **2024**, *32*, 516–533. [CrossRef]
38. Kertkeidkachorn, N.; Ichise, R. T2kg: An end-to-end system for creating knowledge graph from unstructured text. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, San Francisco, CA, USA, 4–9 February 2017; AAAI Press: Palo Alto, CA, USA, 2017.
39. Bilgic, E.; Karacan, E. Clinical Safety and Reliability of Large Language Models in Answering Hemorrhoid-Related Patient Questions: A Comparative Study of ChatGPT, Gemini, and DeepSeek. *Healthcare* **2026**, *14*, 1909. [CrossRef] [PubMed]
40. Yeh, J.; Lin, S.; Lai, S.; Huang, Y.; Chen, Y.; Lee, Y.; Berkes, F. Taiwanese indigenous cultural heritage and revitalization: Community practices and local development. *Sustainability* **2021**, *13*, 1799. [CrossRef]
41. Collado, A.; Mora-Navarro, G.; Heras, V.; Lerma, J. A web-based geoinformation system for heritage management and geovisualisation in Cantón Nabón (Ecuador). *ISPRS Int. J. Geo-Inf.* **2021**, *11*, 4. [CrossRef]
42. Pellegrino, M.; Scarano, V.; Spagnuolo, C. Move cultural heritage knowledge graphs in everyone's pocket. *Semant. Web* **2022**, *14*, 323–359. [CrossRef]
43. Oprea, R.L.; Badea, A.C.; Badea, G. Multi-source 3D documentation for preserving cultural heritage. *Appl. Sci.* **2026**, *16*, 1834. [CrossRef]
44. Cheng, Y.; Mao, H.; Ho, P.; Wu, Y.; Zhang, C.; Xiao, T. A fusion framework to bridge expert and public perception gaps in cultural heritage conservation. *npj Herit. Sci.* **2026**, *14*, 46. [CrossRef]
45. Ozturk, G.B.; Soygazi, F. Generative AI Use in the Construction Industry. In *Applications of Generative AI*; Springer: Cham, Switzerland, 2024. [CrossRef]

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.