

Article

Semantic Integration and Automation of Cultural Heritage Risk Data: A CIDOC-CRM Workflow for Decision Support at the Territorial Scale

Sara Fiorentino , Matteo Lorenzini , Anna Casarotto , Alessandro Iannucci and Mariangela Vandini 

Department of Cultural Heritage, Alma Mater Studiorum University of Bologna, Via Degli Ariani 1, 48121 Ravenna, Italy; matteo.lorenzini7@unibo.it (M.L.); anna.casarotto3@unibo.it (A.C.); alessandro.iannucci@unibo.it (A.I.)

* Correspondence: sara.fiorentino2@unibo.it (S.F.); mariangela.vandini@unibo.it (M.V.)

Featured Application

The proposed semantic workflow supports interoperable and scalable cultural heritage risk management by integrating heterogeneous datasets into a CIDOC CRM-compliant knowledge graph environment. The approach can be applied to preventive conservation, territorial risk management, and decision-support systems for cultural heritage sites exposed to multi-risk scenarios.

Abstract

The increasing availability of digital documentation in cultural heritage has amplified the need for interoperable systems capable of integrating heterogeneous data and supporting risk-informed conservation strategies. In the field of Disaster Risk Management (DRM), the application of structured methodologies—such as the ICCROM-CCI ABC Method—is often hindered by fragmented data sources, inconsistent terminology, and limited interoperability across institutions. This study presents a semantic workflow for the harmonization, enrichment, and integration of cultural heritage risk assessment data within a CIDOC Conceptual Reference Model (CIDOC-CRM)-compliant environment. The proposed system is structured as an Extract–Transform–Load (ETL) pipeline that converts heterogeneous assessment records into interoperable semantic knowledge graphs. The workflow combines controlled vocabularies, project-specific thesauri for risk agents and heritage typologies, and formal ontology mapping implemented through the Mapping Memory Manager (3M) and executed with the X3ML engine. The resulting data are deployed within a ResearchSpace environment, enabling semantic querying, cross-dataset exploration, and integration with external knowledge infrastructures. The workflow was applied to a dataset comprising 295 cultural heritage sites in the municipality of Ravenna (Italy). The transformation process generated a CIDOC-CRM-compliant knowledge graph containing 134,611 RDF triples and 18,954 entities, integrating information on cultural assets, risk scenarios, actors, documentary resources, and quantitative risk assessments. Through the adoption of persistent identifiers and semantic mappings, the workflow also supports interoperability with external cultural heritage resources, including ArCo and GeoNames, facilitating the contextualization and enrichment of local risk assessment data. By transforming fragmented assessment records into structured and interoperable knowledge, the proposed workflow contributes to bridging semantic and information gaps in cultural heritage risk management. The study demonstrates the feasibility of integrating risk assessment data within an ontology-based semantic infrastructure and highlights its potential to support data integration, semantic interoperability, knowledge reuse, and future decision-support applications for preventive conservation and territorial risk management.



Academic Editors: Theodore Ganetsos and Antreas Kantaros

Received: 20 May 2026

Revised: 2 July 2026

Accepted: 6 July 2026

Published: 8 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.

Keywords: cultural heritage; risk management; CIDOC CRM; semantic interoperability; knowledge graph; preventive conservation; linked open data; decision-support systems

1. Introduction

Addressing data integration from both conceptual and methodological perspectives represents one of the central challenges in the development of information systems for the cultural heritage domain [1,2]. The rapid growth of digital documentation has significantly increased the availability of heterogeneous datasets describing historical sites, objects, conservation states, and associated documentation [3,4]. Ensuring that these datasets adhere to FAIR (Findable, Accessible, Interoperable, Reusable) data principles is essential to maximize their long-term usability and cross-domain integration [5].

While this expansion offers unprecedented opportunities for research and conservation planning, it also introduces substantial challenges related to interoperability, data integration, and semantic consistency. A notable example of efforts to address these challenges is the digital ecosystem developed for the multidisciplinary study of Notre-Dame de Paris [6]. In the context of the cathedral's restoration, De Luca and his team designed a digital infrastructure capable of managing and correlating data originating from multiple domains, including architectural surveys, historical documentation, material analyses, and conservation studies. This project illustrates how carefully designed digital frameworks can support the integration of diverse datasets while preserving the specificity of each disciplinary contribution. Within this line of research, De Luca identifies four major gaps that continue to affect digital cultural heritage practices: the semantic gap, referring to the fragmentation and insufficient semantic enrichment of digital resources; the memory gap, related to the difficulty of formalizing implicit knowledge and research protocols; the data correlation gap, which concerns the limited capability to automatically link heterogeneous datasets; and the technological gap, associated with the difficulty of connecting collaborative digital analysis with the physical heritage context.

From a complementary perspective, Bruseker et al. [7] emphasize the importance of formal ontologies—particularly CIDOC-CRM [8]—as a foundation for sustainable cultural heritage data management. Ontologies provide the conceptual clarity necessary to mediate between domain expertise and computational systems, enabling long-term interoperability, semantic consistency, and inferential capabilities. This approach has been operationalized in mapping models such as the Swiss Art Research Data Model (SRDM) [9], which translates CIDOC-CRM into structured mapping guidelines and practical data transformation workflows [10].

A further layer of complexity arises from the fact that cultural heritage documentation is typically produced across multiple institutional contexts and disciplinary perspectives. As a result, documentation practices vary widely in structure, terminology, and conceptual frameworks, making the integration of heterogeneous datasets particularly challenging for digital heritage infrastructures. These issues become even more pronounced in the field of cultural heritage risk management, where documentation processes often involve multiple actors and professional perspectives. Risk assessment typically requires the collaboration of emergency responders, conservation specialists, heritage authorities, and administrative bodies, each of whom documents risk scenarios according to their own operational priorities. For example, emergency responders may emphasize factors such as fire propagation, structural vulnerability, or accessibility in emergency situations, whereas heritage professionals may prioritize conservation conditions, material fragility, or historical significance.

Although this plurality of perspectives represents a valuable source of knowledge, it also introduces significant challenges for data integration.

Recent developments in digital cultural heritage research have increasingly explored the use of semantic infrastructures, knowledge graphs, and digital twins as mechanisms for integrating heterogeneous information and supporting advanced analytical processes. Digital twin approaches have been proposed as a means of linking physical heritage assets with continuously updated digital representations, enabling more dynamic forms of monitoring, interpretation, and decision support [11,12]. Similarly, recent work on semantic data modelling and knowledge graph construction has highlighted the importance of formal ontologies and reusable semantic frameworks for ensuring long-term interoperability, transparency, and sustainability of cultural heritage information systems [10]. These developments demonstrate the growing maturity of semantic technologies within the cultural heritage domain. However, their application to cultural heritage risk assessment and preventive conservation remains comparatively limited. In particular, the integration of risk evaluation records, environmental threats, and conservation-oriented decision-making processes within interoperable semantic infrastructures is still insufficiently explored, especially at the territorial scale. This gap motivates the present study.

While CIDOC-CRM, X3ML-based transformations, and Linked Open Data infrastructures have been extensively applied to cultural heritage documentation, their application to cultural heritage risk assessment remains comparatively underexplored. In particular, the semantic formalization of structured risk assessment methodologies and the representation of risk scenarios within an event-centric ontological framework and their integration with national heritage knowledge infrastructures have received limited attention. The originality of the present study lies in the development of a reproducible semantic workflow that bridges cultural heritage risk assessment practices and interoperable knowledge graph technologies, enabling the integration of ABC-based risk evaluation data within a CIDOC-CRM-compliant environment.

This study proposes the adoption of the CIDOC Conceptual Reference Model (CIDOC-CRM) for the semantic integration of cultural heritage risk assessment data [1]. By providing a formal ontology capable of modelling events, actors, objects, conditions, and their interrelations, CIDOC-CRM enables heterogeneous documentation practices to be aligned within a shared conceptual framework. Such an approach makes it possible to reconcile different institutional viewpoints while preserving their specificity, thereby supporting more coherent and interoperable cultural heritage risk management systems. Furthermore, the ontological nature of CIDOC-CRM not only enables data harmonization but also reasoning and knowledge inference. By formally representing actors, events, interventions, and conservation states within an explicit semantic structure, the model allows implicit relationships between heritage assets, risk factors, and evaluation processes to be analyzed through advanced queries and analytical reasoning, thus contributing to more informed decision-making processes and to the development of integrated, data-driven approaches to cultural heritage risk management. In this context, the present study proposes a semantic workflow for integrating cultural heritage risk assessment data within a CIDOC-CRM-compliant knowledge infrastructure.

2. Materials and Methods

2.1. Research Context and Dataset

The proposed workflow has been tested on a dataset of cultural heritage sites located in the municipality of Ravenna (Italy), collected in the framework of the SIRIUS—Strategie per la gestione del patrimonio culturale a rischio (Management strategies for cultural heritage at risk) [13] and RESTART—Resilienza e Sviluppo Territoriale: patrimonio a Rischio e Tutela

(Resilience and territorial development: Heritage at risk and its protection) [14] projects. Coordinated by the Department of Cultural Heritage at the University of Bologna, both initiatives aim to strengthen the capacity of local communities and institutions to understand, monitor, and manage the risks affecting cultural heritage at the territorial scale. The SIRIUS project, integrated into the PNRR CHANGES—Cultural Heritage Active Innovation for Sustainable Society programme [15], focuses on supporting local authorities and stakeholders in improving procedures for monitoring, prevention, and mitigation of risks affecting cultural heritage. In parallel, the RESTART project investigates cultural heritage as a resource for strengthening territorial resilience, particularly in the context of climate-related hazards. RESTART is funded within the Alma CaReS—Climate Change, Resilience, Sustainability initiative promoted by the University of Bologna following the severe floods that recently affected the Emilia-Romagna region. Both projects adopt Ravenna (Emilia-Romagna, Italy) as a case-study territory, implementing a bottom-up approach that combines local-scale risk analysis with the development of transferable heritage protection strategies.

Within this framework, a dataset of risk assessment records was collected through structured forms based on the ABC Method, developed by ICCROM and the Canadian Conservation Institute [16]. The ABC method is widely used in cultural heritage risk management as a structured approach for identifying, analyzing, and prioritizing risks affecting heritage assets. Although several methodologies have been proposed for evaluating and mitigating risks to cultural heritage [17–20], the ABC framework was selected for this study because of its heritage-oriented and stakeholder-inclusive approach, which facilitates the participation of both heritage professionals and non-specialist actors in risk assessment processes. The method is based on a systematic evaluation of risk magnitude through three quantitative components. Component A estimates the frequency or rate of occurrence of the damaging event, while components B and C jointly quantify the expected loss of value for the heritage asset. The combination of these three parameters defines the overall magnitude of the risk, enabling the comparison and prioritization of different threats. Beyond numerical evaluation, the ABC methodology emphasizes a cyclical risk management process that includes context definition, risk identification, evaluation, and the development of mitigation strategies. The ABC method has been applied in a variety of heritage contexts, including the assessment of water-related risks at San Clemente Church in Albenga [21], the management of natural and anthropogenic risks at Petra World Heritage Site [22], and the evaluation of climate-related threats at Nigerian cultural heritage sites [20]. Within the SIRIUS and RESTART projects, the methodology has been applied to several case studies in Italy, including the Arian Baptistry in Ravenna [23] and the Hypogeum Archaeological Site of Sigismund Street in Rimini [24].

Alternative frameworks and standards are available for cultural heritage documentation and metadata exchange, including Spectrum, LIDO, Getty vocabularies, and Wikidata-based approaches. However, CIDOC-CRM was selected because of its widespread adoption as the reference ontology for semantic interoperability in the cultural heritage domain and its event-centric structure, which is particularly suitable for representing risk as a dynamic process involving cultural assets, hazards, actors, observations, and evaluation activities. The ABC Method was selected because it combines a structured quantitative evaluation framework with a strong decision-support orientation and broad applicability across different categories of cultural heritage. Furthermore, its explicit articulation of risk through measurable components facilitates the semantic formalization of assessment records and their subsequent integration within ontology-based information systems. GeoNames and ArCo (Architettura della Conoscenza) were adopted as authoritative reference resources providing persistent identifiers and established Linked Open Data infrastructures, thereby facilitating semantic federation and interoperability beyond the local project context.

The records used for the present study describe a set of 295 cultural heritage sites, including architectural heritage, archaeological remains, and other cultural assets distributed across the Ravenna municipal territory. These 295 sites, which compose the dataset, were filtered and acquired through the API SPARQL endpoint of the Catalogo Generale dei Beni Culturali [25–27]. Each record includes descriptive information about the site, contextual documentation, and risk assessment data produced by different operators involved in risk evaluation processes. These operators include professionals from multiple institutional domains, such as heritage authorities, civil protection services, and emergency response units. As a result, the documentation reflects heterogeneous perspectives and disciplinary priorities, generating a dataset characterized by conceptual and terminological variability. The source data are structured as eXtensible Markup Language (XML) records derived from ABC assessment forms [28]. While these records provide rich documentation of risk scenarios and cultural heritage assets, their heterogeneity and limited semantic structure pose challenges for interoperability and large-scale analysis. The dataset therefore provides a suitable experimental context for testing semantic integration strategies aimed at harmonizing and operationalizing risk documentation within a structured knowledge framework.

2.2. From Data Lifting to Semantic Modelling and Integration Workflow

The first stage of the workflow represented in Figure 1 focuses on the harmonization of the source XML data. Because the assessment forms were compiled by different operators and institutions, the records exhibit variations in terminology, field usage, and data completeness. A preliminary harmonization process was therefore required to normalize the structure of the dataset and align the extracted fields with the core elements of the ABC risk assessment methodology.

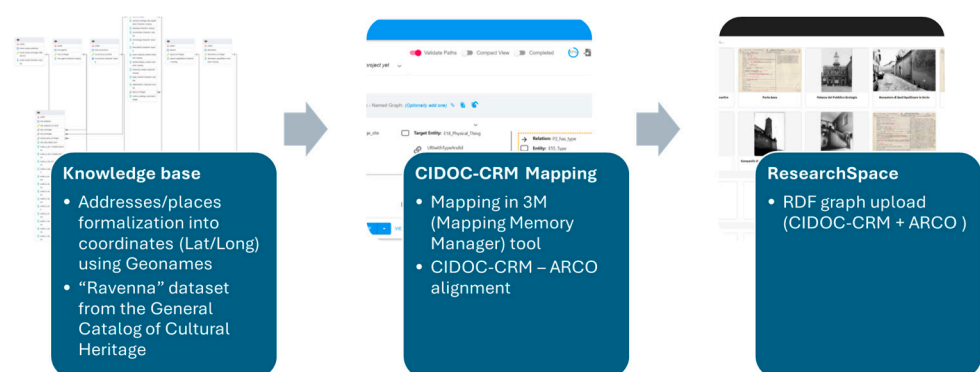


Figure 1. Harmonization and integration data workflow.

During this stage, the XML records were reorganized and standardized to ensure consistent representation of key entities such as cultural heritage assets, risk scenarios, and evaluation parameters. Attention was devoted to semantic alignment using controlled vocabularies and external reference resources. Geographic entities were linked to GeoNames [29], with the specific purpose of supporting the geolocation of cultural heritage sites through the assignment of standardized place identifiers. This alignment process enabled the enrichment of each site with precise latitude and longitude coordinates, ensuring consistent spatial referencing and improving the interoperability of the dataset with external geospatial datasets and services. Cultural heritage typologies were aligned with the controlled vocabulary defined by the Istituto Centrale per il Catalogo e la Documentazione (ICCD), which provides standardized terminology for the classification of cultural assets [30].

In addition, the dataset was enriched with the field corresponding to the Codice Catalogo Nazionale, which represents the identifier assigned to cultural heritage assets

within the Italian national catalogue of cultural heritage. This identifier enables semantic federation with the national catalogue dataset published through the ArCo ontology [31]. In our approach, ArCo was specifically used to ensure a shared and controlled denomination of cultural heritage assets and to support the harmonization and completion of key descriptive attributes, including typology, chronology, description, and property information. Moreover, ArCo was instrumental in guaranteeing reliable linkage to the associated visual documentation, particularly the images provided by the Catalogo Generale dei Beni Culturali. This integration step thus strengthens semantic consistency across datasets and provides a concrete mechanism for aligning local risk assessment data with authoritative national cultural heritage records within a Linked Open Data framework. The semantic modelling of the dataset is grounded in the CIDOC-CRM, an ISO standard ontology designed to facilitate the integration and interoperability of cultural heritage information. CIDOC-CRM provides a formal conceptual framework for representing entities such as physical objects, actors, events, places, and conceptual objects, as well as the relationships that link them. The event-centric structure of CIDOC-CRM makes it particularly suitable for modelling cultural heritage risk assessment processes. Within this framework, risk scenarios can be conceptualized as events affecting cultural assets, involving specific actors and producing evaluative documentation. This approach allows heterogeneous documentation practices to be aligned within a shared ontological structure, ensuring semantic coherence while preserving the complexity of the underlying information. To represent specific aspects of risk documentation and scientific observation processes, the modelling framework also incorporates two extensions of CIDOC-CRM. The CRMsci extension supports the formal representation of scientific observation and measurement processes, enabling the modelling of inspection activities and evaluative statements produced during risk assessment [32]. Within the proposed model, CRMsci is specifically employed to represent the quantification of risk through S21 Measurement, a class designed to document measurement activities and their outcomes. Risk values, scores, and severity assessments are therefore not modelled as intrinsic properties of the cultural asset or of the risk event itself, but as the result of an expert evaluation process. Together, these ontological components provide a flexible semantic framework capable of representing cultural assets, risk events, observations, and evaluative measurements within a unified knowledge model.

Within the proposed framework, cultural heritage risk is modelled using an event-centric representation derived from the CIDOC-CRM ontology. Rather than representing risk as a static attribute of a cultural asset, the model conceptualizes risk as a potential or actual transformation process affecting the asset. In this schema (Figure 2), the cultural heritage site is represented as an instance of E18 Physical Thing, which functions as the central reference entity within the knowledge graph. Risk scenarios are modelled as instances of E5 Event, representing either documented occurrences or potential future events identified through the assessment process, such as fire, flooding, seismic activity, or other hazardous phenomena. Although CIDOC-CRM was originally conceived to document historical events, its event-centric structure provides an effective mechanism for representing risk scenarios as possible transformation processes that are the subject of expert evaluation. The causes or generators of these risk scenarios are represented through instances of E39 Actor. This modelling choice was adopted to provide a uniform representation of entities participating in the risk event, including institutional actors, human activities, and natural or environmental agents identified through the risk assessment methodology. In this context, E39 Actor should be understood as a pragmatic modelling abstraction used to represent the source of the risk condition and to enable a consistent classification through the project-specific Risk Agent Thesaurus. This approach facilitates the semantic linking of risk events to their generating factors and supports comparative analysis across different

categories of hazards. Evaluative documentation, including structured assessment forms and analytical descriptions, is modelled as E73 Information Object, capturing the documentary layer of the risk analysis. The quantitative evaluation of risk is represented through S21 Measurement, enabling the explicit modelling of expert assessments, risk scores, and evaluation parameters associated with specific risk events. This approach makes explicit the methodological distinction between the risk scenario, represented as an E5 Event, and its magnitude or severity, represented through S21 Measurement. Consequently, the model preserves both the event-centric nature of CIDOC-CRM and the observational perspective provided by CRMsci, allowing risk assessments to be documented as measurable and reproducible analytical activities. This structure allows the modelling of relationships between cultural assets, risk scenarios, and evaluative measurements in a transparent and analytically robust manner. By representing risk assessment as a set of interconnected events, actors, and measurements, the ontology supports advanced querying and semantic reasoning across the dataset.

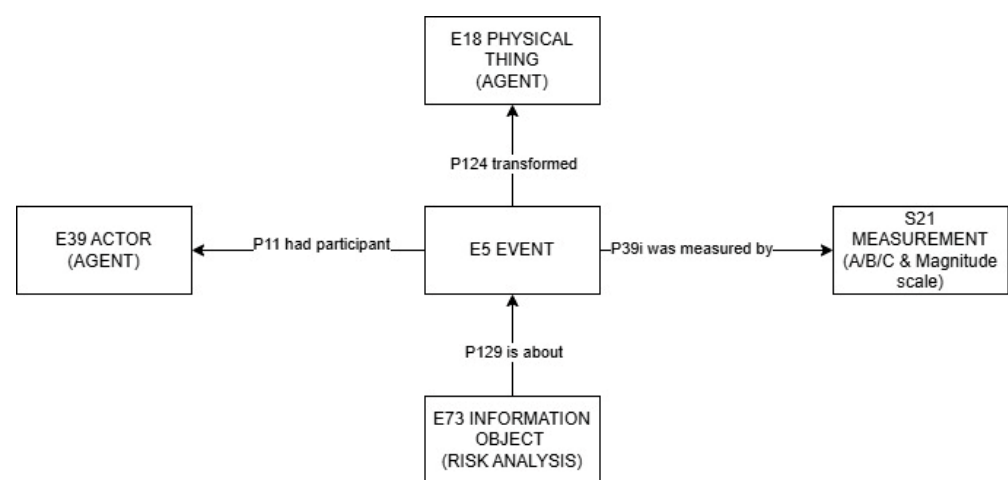


Figure 2. Event-based representation of cultural heritage risk within the CIDOC-CRM framework.

The semantic integration process is implemented through a complete ETL (Extract–Transform–Load) workflow that converts the harmonized XML records into Resource Description Framework (RDF) knowledge graphs compliant with the CIDOC-CRM ontology [33]. The semantic mapping between the XML source data and the CIDOC-CRM conceptual model was defined using the 3M editor, an environment designed to support the explicit modelling of correspondences between data structures and ontological entities. The mapping schema formalizes the relationships between XML elements and CIDOC-CRM classes and properties, capturing the semantic logic required to transform heterogeneous documentation into structured RDF triples. The mapping rules defined in the 3M editor are serialized as X3ML mapping files, which are subsequently executed using the X3ML Engine developed by the Institute of Computer Science of the Foundation for Research and Technology—Hellas (FORTH). The engine applies the mapping rules to the harmonized XML dataset and generates RDF triples representing the semantically enriched knowledge graph. The full ETL pipeline, including mapping definitions and execution artefacts, is publicly available in the dedicated repository [34]. The resulting RDF graphs are then imported into ResearchSpace, which serves as the semantic infrastructure for data storage, exploration, and querying. ResearchSpace provides a CIDOC-CRM-oriented environment that supports SPARQL querying, semantic visualization, and the integration of external Linked Open Data resources. Within this environment, the knowledge graph becomes accessible to both technical users and domain experts, enabling advanced exploration of cultural heritage risk data and facilitating cross-dataset interoperability. The full Re-

searchSpace application implementation, including configuration files, data integration components, and query templates, is publicly available in the dedicated repository [35].

The proposed workflow can be summarized in four main phases: (1) collection and harmonization of cultural heritage risk assessment records; (2) semantic modelling and ontology alignment through CIDOC-CRM and its extensions; (3) transformation of the source XML data into RDF knowledge graphs using the 3M mapping environment and the X3ML engine; and (4) deployment and exploration of the resulting semantic data within the ResearchSpace environment. This sequence provides a reproducible framework for the semantic integration of heterogeneous risk assessment datasets while maintaining traceability between source records, semantic mappings, and generated knowledge structures.

3. Results

3.1. Generation of the Cultural Heritage Risk Knowledge Graph

The semantic workflow described in the previous section was applied to the dataset of cultural heritage sites located in the municipality of Ravenna. Through the ETL pipeline, the harmonized XML records derived from the ABC risk assessment forms were transformed into RDF triples compliant with the CIDOC-CRM ontology.

The transformation process resulted in the generation of a semantic knowledge graph representing the relationships between cultural heritage sites, risk scenarios, actors involved in risk processes, and evaluation data. Each cultural heritage asset is represented as an entity within the graph and connected to documentary information, risk events, and measurement activities through explicitly defined semantic relationships.

The resulting dataset (Table 1) includes 134,611 RDF triples and 18,954 unique entities, structured across 295 cultural heritage sites, but currently contains only 36 actors and 86 events, indicating a relatively sparse CIDOC CRM event–actor representation within the knowledge graph.

Table 1. Overview of the RDF/CIDOC CRM knowledge graph statistics, including total triples, unique entities, and the distribution of actors and events across the dataset.

Metric	Value
Total Triples	134.611
Unique Entities	18.954
E5 Event	86
E39 Actor	36

These entities include cultural heritage assets, risk events affecting the assets, agents responsible for the risk conditions, and documentary resources describing the evaluation process. The RDF graphs produced through the X3ML transformation were successfully imported into the ResearchSpace environment, where they can be explored and queried through semantic interfaces and SPARQL queries.

The successful deployment of the resulting RDF graphs within the ResearchSpace environment demonstrates not only the feasibility of the proposed ETL workflow, but also its capacity to operationalize heterogeneous risk assessment records as semantically structured and queryable knowledge. By transforming fragmented documentation into an interconnected graph of cultural assets, risk events, actors, measurements, and documentary resources, the workflow establishes the conditions for advanced analytical exploration, cross-dataset integration, and decision-support activities. Furthermore, the use of persistent identifiers and ontology-based mappings enables the federation of local risk assessment data with external cultural heritage knowledge infrastructures, extending the analytical

potential of the resulting dataset beyond the boundaries of the original project. An additional contribution of the workflow lies in its reproducibility. By explicitly formalizing the semantic relationships between source records and ontology classes through the 3M mapping environment and the X3ML transformation engine, the methodology transforms domain knowledge and documentation logic into reusable and transparent semantic rules. This approach not only facilitates the consistent generation of RDF knowledge graphs from heterogeneous datasets but also contributes to reducing dependence on implicit expertise embedded in traditional documentation practices, thereby supporting more sustainable and interoperable cultural heritage risk management infrastructures.

3.2. Event-Based Representation of Cultural Heritage Risk

One of the central outcomes of the proposed modelling approach is the event-based representation of cultural heritage risk (Figure 2). Within the CIDOC-CRM framework, risk is not treated as a static property associated with a cultural asset, but rather as an event that may affect or transform the asset.

This modelling strategy enables a structured representation of the relationships between cultural assets, hazards, and evaluation processes. By explicitly modelling risk as a network of interconnected events, actors, and measurements, the knowledge graph supports more advanced analytical queries, such as the identification of recurring risk patterns, the comparison of risk scenarios across different sites, or the analysis of risk agents affecting specific categories of heritage assets.

3.3. Interoperability with External Cultural Heritage Datasets

Another significant result of the workflow concerns the semantic interoperability achieved through the integration of authoritative identifiers and external reference datasets. During the harmonization phase, each cultural heritage site record was enriched with the Codice Catalogo Nazionale, corresponding to the identifier assigned within the Italian national catalogue of cultural heritage.

Through this identifier, the knowledge graph establishes semantic links with the national dataset published via the ArCo ontology, which represents the official RDF implementation of the Italian cultural heritage catalogue. The mapping schema therefore enables the federation of the local risk dataset with the national cultural heritage knowledge graph, allowing cross-dataset queries and the integration of additional descriptive information about the cultural assets. In practical terms, this interoperability is implemented through CIDOC-CRM property alignments such as the mapping of *P2 has type* to *E55 Type*, where the typology values are directly aligned with the corresponding controlled IRI terms provided by ArCo.

This interoperability mechanism demonstrates how local risk documentation systems can be integrated within broader Linked Open Data ecosystems, enhancing the reusability and contextualization of risk assessment data. By linking the Ravenna dataset to external cultural heritage resources, the proposed workflow contributes to building a more interconnected digital infrastructure for cultural heritage documentation and risk management.

3.4. Example Semantic Query and Knowledge Retrieval

To demonstrate the analytical potential of the resulting knowledge graph, semantic queries can be used to retrieve and correlate information concerning cultural heritage assets, risk events, risk agents, and associated evaluation data. Unlike traditional record-based documentation systems, the semantic structure of the graph enables the exploration of explicit relationships between entities and supports cross-dataset analysis. The following example illustrates a typical risk-analysis query executed within the ResearchSpace environment.

Query 1—Risk event and agent retrieval (comparative analysis)

This component retrieves modification events associated with a given heritage asset and identifies the involved risk agents, enabling the comparative analysis of actor–event relationships across the dataset as illustrated in Figure 3.

EVENT	EVENTLABEL	ACTOR	ACTORLABEL
https://w3id.org/sirius/event/23/1/24/Regular_Monitoring_Not_Carried_Out 	Regular_Monitoring_Not_Carried_Out (xsd:string)	https://w3id.org/sirius/agent/Dissociation 	Dissociation (xsd:string)
https://w3id.org/sirius/event/22/1/23/UV_Causing_Disintegration 	UV_Causing_Disintegration (xsd:string)	https://w3id.org/sirius/agent/Light_UV_IR 	Light_UV_IR (xsd:string)
https://w3id.org/sirius/event/15/1/7/Use_of_Water_Activity 	Use_of_Water_Activity (xsd:string)	https://w3id.org/sirius/agent/Water 	Water (xsd:string)

Figure 3. Illustrates a snap of the actor–event relation related to a given target heritage site.

Query 2—Risk magnitude estimation and prioritization

This component computes an aggregated risk score per event based on multiple measurement dimensions. The resulting values can be ordered to support the prioritization of events or assets according to their estimated risk level.

Figure 4 illustrates the decision-support view. The graph ranks the risk events associated with a selected heritage asset according to their aggregated risk magnitude, with color coding highlighting relative severity (green to red). This enables conservation professionals to quickly identify the most critical risk factors and priorities interventions, inspections, or resource allocation based on quantitative evidence derived from the knowledge graph.

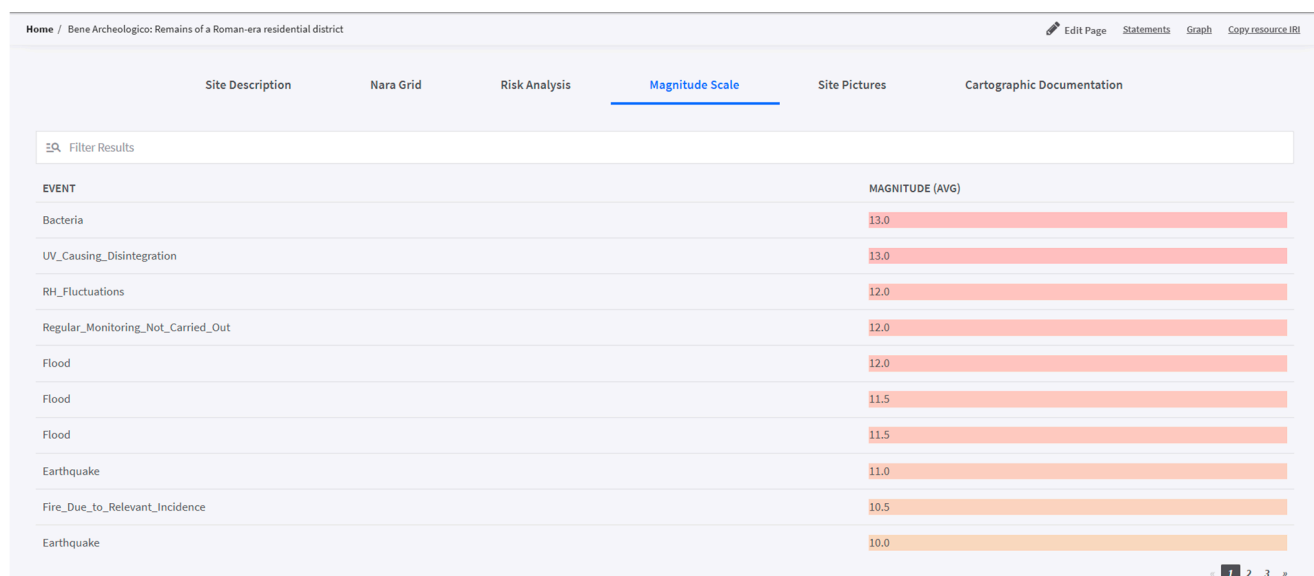


Figure 4. Decision-support graph ranking the risk events associated with a selected heritage asset by aggregated risk magnitude, enabling rapid identification of priority interventions.

Query 3—Uncertainty assessment (decision reliability)

This component extracts minimum and maximum uncertainty values associated with risk measurements, allowing the evaluation of confidence intervals and supporting uncertainty-aware decision-making as represented in Figure 5.

AGENT	EVENT	SCALE A	SCALE B	SCALE C	DESCRIPTION	MAGNITUDE	UNCERTAINTY
Dissociation	Regular_Monitoring_Not_Carried_Out	Val. 3 Val. 0 Val. 0 Description None	Val. 10 Val. 100 Val. 0 Description None	Val. 10 Val. 0 Val. 0 Description None	—	Avg. 13.666666666666667	Val. 1
Dissociation	No_Staff_Dedicated_to_Maintenance	Val. 0.01 Val. 0 Val. 0 Description None	Val. 10 Val. 100 Val. 0 Description None	Val. 10 Val. 100 Val. 0 Description None	—	Avg. 24.445555555555555	Val. 3.5
Dissociation	No_Staff_Dedicated_to_Maintenance	Val. 0.01 Val. 0 Val. 0 Description None	Val. 1 Val. 10 Val. 0 Description None	Val. 3 Val. 10 Val. 0 Description None	—	Avg. 2.6677777777777778	Val. 2.5
Dissociation	Poor_Organization_of_Collection	Val. 1 Val. 3 Val. 0 Description None	Val. 0.003 Val. 0.01 Val. 0 Description None	Val. 0.003 Val. 0.01 Val. 0 Description None	—	Avg. 0.4473333333333333	Val. 3

Figure 5. Extraction of minimum and maximum uncertainty bounds associated with risk measurements, enabling the estimation of confidence intervals and supporting uncertainty-aware decision-making.

The provided examples illustrate how semantically modelled relationships between heritage sites, risk events, agents, and measurements can be queried across the entire dataset. Such queries support comparative risk analysis, facilitate the identification of recurring patterns and critical risk factors, and contribute to evidence-based decision-making processes for preventive conservation and territorial risk management.

The examples presented above demonstrate not only the technical feasibility of the semantic workflow, but also its potential to support interoperable knowledge management, semantic querying, and decision-support activities in cultural heritage risk assessment. These aspects are further discussed in the following section.

The full SPARQL query implementation used in this study is publicly available [36].

4. Discussion

4.1. Bridging the Semantic Gap in Cultural Heritage Risk Data

The results obtained through the implementation of the proposed workflow demonstrate the potential of ontology-based infrastructures to address the semantic fragmentation that characterizes cultural heritage risk documentation. As highlighted in recent scholarship, the extensive production of digital resources in the heritage domain does not automatically translate into interoperable knowledge systems. In many cases, documentation remains fragmented across institutions, disciplinary traditions, and data formats, limiting the possibility of integrating and analyzing information at broader spatial or institutional scales.

The semantic workflow proposed in this study contributes to addressing this issue by transforming heterogeneous risk assessment records into a structured knowledge graph based on the CIDOC-CRM ontology. Through the formal mapping of entities such as cultural heritage assets, risk events, actors, and evaluative measurements, the approach enables previously disconnected documentation practices to be integrated within a shared conceptual framework. In this sense, the workflow does not simply harmonize data structures; it introduces a semantic layer that makes the relationships between heritage assets, hazards, and evaluation processes explicit and machine-readable.

The event-based modelling of risk plays a particularly important role in this process. By conceptualizing risk scenarios as events affecting cultural assets, the model captures the dynamic and processual nature of risk conditions. This representation allows heterogeneous risk factors and evaluation processes to be integrated within a coherent analytical structure, enabling the comparison of risk scenarios across multiple sites and facilitat-

ing the exploration of patterns that would remain difficult to identify within traditional documentation systems.

Compared with conventional relational database approaches, the proposed semantic workflow offers several advantages for the integration and analysis of cultural heritage risk data. Relational databases are highly effective for storing structured information, but they generally rely on predefined schemas and custom integration procedures when heterogeneous datasets need to be combined. In contrast, the ontology-based approach adopted in this study provides a shared conceptual framework capable of accommodating different documentation practices, vocabularies, and institutional perspectives. Furthermore, semantic querying through RDF and SPARQL enables the exploration of explicit relationships between cultural assets, risk events, actors, and evaluation processes, supporting analytical questions that are often difficult to formulate within conventional relational structures. The integration of authoritative identifiers and Linked Open Data resources further enhances interoperability by facilitating the federation of local datasets with external knowledge infrastructures.

4.2. Formalizing Documentation Practices and Addressing the Memory Gap

Beyond semantic interoperability, the proposed methodology also contributes to addressing what De Luca [6] describes as the memory gap in digital cultural heritage practices. In many heritage documentation workflows, key elements of interpretation, classification, and evaluation remain embedded in implicit professional knowledge rather than being formally represented within the data structure. This reliance on tacit expertise can hinder the reproducibility of documentation processes and complicate the long-term reuse of digital resources.

The semantic modelling strategy adopted in this study seeks to reduce this dependence on implicit knowledge by explicitly representing documentation logic within an ontological framework. Through the mapping schema defined in the 3M editor and executed via the X3ML engine, the conceptual relationships underlying the risk assessment process become formally encoded within the transformation rules that generate the knowledge graph. In this way, the methodological assumptions embedded in the ABC risk assessment workflow are translated into explicit semantic structures that can be shared, reused, and extended in other contexts.

This explicit modelling of documentation practices has important implications for the sustainability of digital heritage infrastructures. By transforming procedural knowledge into formal semantic representations, the workflow supports the development of reproducible and transparent data integration pipelines. Such an approach contributes to reducing the dependence on highly personalised expertise and facilitates collaborative data production across institutions and disciplinary communities.

4.3. Metadata Quality and Limitations of the Current Dataset

The experimentation conducted on the Ravenna dataset has also highlighted several limitations related to the quality and consistency of the underlying metadata. Inconsistencies in typological classifications within the national cultural heritage catalogue have occasionally generated redundancy or ambiguity in the semantic typing of cultural heritage sites. Similar issues emerged in the geocoding process, where non-standardized address expressions or incomplete location data prevented the automatic generation of geographic coordinates for some records.

These issues underline an important methodological consideration: while semantic infrastructures can significantly enhance interoperability and knowledge integration, their effectiveness ultimately depends on the quality and consistency of the source data. Semantic

modelling and ontology-based mapping can mitigate some forms of heterogeneity, but they cannot fully compensate for structural inconsistencies or incomplete metadata at the source level.

Consequently, the implementation of semantic integration workflows should be accompanied by systematic metadata governance strategies, including the adoption of controlled vocabularies, standardized documentation practices, and continuous data quality validation. In this sense, the experience described in this study suggests that the development of interoperable semantic infrastructures must be understood not only as a technological challenge but also as an organizational and institutional one, requiring coordinated efforts in data curation and standardization across cultural heritage institutions.

The present study should be regarded as a proof-of-concept implementation aimed at demonstrating the feasibility of integrating cultural heritage risk assessment data within a CIDOC-CRM-compliant semantic infrastructure. The primary objective was the design, implementation, and validation of a semantic workflow capable of transforming heterogeneous risk assessment records into interoperable knowledge graphs. Consequently, the research focused on data harmonization, ontology-based modelling, and semantic integration rather than on the evaluation of user interaction or system performance. Formal usability assessments, user-centered evaluations, and systematic benchmarking of query performance were therefore beyond the scope of the present work. Future developments will focus on the operational deployment of the infrastructure and its validation through real-world use cases involving cultural heritage professionals and institutional stakeholders, with particular attention to usability, scalability, query performance, and decision-support effectiveness.

Although the workflow was developed within the context of the SIRIUS and RESTART projects and relies on the ABC Method, ArCo, and ICCD documentation standards, its overall architecture is not intrinsically tied to these specific frameworks. The event-centric semantic model provided by CIDOC-CRM enables the representation of risk assessment processes at a conceptual level, allowing alternative methodologies to be accommodated through the remapping of evaluation entities, measurement procedures, and domain-specific vocabularies. Similarly, while ArCo and ICCD were adopted because they represent the official Italian cultural heritage information infrastructure, the proposed workflow is based on modular semantic components and ontology mappings that could be adapted to other national or institutional cataloguing systems, provided that appropriate semantic correspondences and persistent identifiers are available.

From an organizational perspective, the workflow is particularly suited to contexts where heterogeneous documentation sources need to be integrated and analyzed across multiple cultural heritage assets, such as regional heritage authorities, national heritage agencies, research institutions, and multi-site conservation programmes. Smaller institutions, including museums and local heritage organizations, could also benefit from the approach through shared infrastructures or consortium-based implementations. However, its adoption requires expertise in ontology modelling, semantic technologies, and data governance, highlighting the importance of interdisciplinary collaboration between domain experts, heritage professionals, and information specialists.

5. Conclusions

This study presented a semantic workflow for the integration and operationalization of cultural heritage risk assessment data within a CIDOC-CRM-compliant environment. By combining data harmonization and ontology-based semantic modelling, the proposed approach enables the conversion of heterogeneous ABC risk assessment records into interoperable RDF knowledge graphs.

The implementation carried out on the Ravenna dataset demonstrates the feasibility and scalability of the workflow. Through the integration of controlled vocabularies, project-specific thesauri, and formal ontology mapping, the system allows cultural heritage sites, risk events, agents, and evaluation processes to be represented within a coherent semantic structure. The resulting knowledge graph supports advanced querying, cross-dataset comparison, and the integration of risk documentation within broader cultural heritage data ecosystems.

The contribution of this study does not lie in the individual technologies employed, but in their integration within a coherent workflow specifically designed for cultural heritage risk assessment and semantic interoperability. Beyond the technical implementation, the study contributes to the broader discourse on semantic infrastructures for cultural heritage by addressing two key challenges identified in recent scholarship: the semantic gap and the memory gap in digital heritage documentation. By formalizing risk assessment practices within a transparent ontological framework, the proposed workflow transforms fragmented documentation records into structured, interoperable knowledge and reduces the reliance on implicit domain expertise embedded in traditional documentation processes.

At the same time, the experimentation highlights the importance of metadata quality and governance as critical factors for the success of semantic integration initiatives. While ontology-based infrastructures can significantly enhance interoperability, their effectiveness ultimately depends on the consistency and reliability of the underlying documentation practices.

Future developments may extend the proposed framework to larger territorial datasets and integrate additional analytical tools for risk evaluation and decision support. In this perspective, the adoption of ontology-based semantic infrastructures may play an increasingly important role in supporting coordinated, data-driven approaches to cultural heritage risk management at the territorial scale. In particular, such infrastructures should be further enhanced through the integration of a description logic layer, extending the existing RDF model into a well-defined OWL-DL graph, thereby enabling more robust semantic reasoning, consistency checking, and advanced inferencing capabilities. Additionally, the incorporation of Large Language Models (LLMs) may further support the reasoning process by enabling natural language interaction with the knowledge base, facilitating semantic querying, and assisting in the interpretation and contextualization of heterogeneous data sources. When combined with formal ontology-based reasoning, LLMs can contribute to hybrid approaches that enhance decision-making processes, support knowledge discovery, and improve the accessibility of complex cultural heritage information systems for both experts and non-expert users.

From a theoretical perspective, the study currently contributes to ongoing discussions on semantic interoperability in cultural heritage by demonstrating how risk assessment data—traditionally fragmented across disciplinary and institutional boundaries—can be formalized within a shared ontological framework. The proposed workflow extends existing applications of CIDOC-CRM beyond documentation and inventory management, showing its potential to support the representation of risk processes, evaluation activities, and decision-support knowledge structures. From a practical perspective, the workflow provides cultural heritage institutions with a reproducible framework for integrating heterogeneous risk assessment datasets, improving data consistency, facilitating information exchange, and supporting evidence-based conservation planning. By enabling semantic querying and interoperability with external knowledge infrastructures, the approach may contribute to more coordinated and informed risk management strategies at both site and territorial scales.

Although developed within the Italian cultural heritage context, the proposed workflow is based on modular semantic components and can be adapted to alternative documentation frameworks and cataloguing systems. Future work will focus on operational deployment, performance evaluation, and user-centered validation.

Author Contributions: Conceptualization, S.F.; methodology, M.L.; software, M.L.; validation, M.L., S.F., M.V. and A.C.; investigation, M.L., A.C. and S.F.; resources, M.V. and A.I.; data curation, M.L., A.C. and S.F.; writing—original draft preparation, S.F. and M.L.; writing—review and editing, M.V., A.I. and S.F.; visualization, M.V. and A.I.; supervision, M.V. and A.I.; project administration, M.V. and A.I.; funding acquisition, M.V. and A.I. All authors have read and agreed to the published version of the manuscript.

Funding: Project SIRIUS was funded by the European Union—NextGenerationEU under the National Recovery and Resilience Plan (PNRR)—Mission 4 Education and research—Component 2 From research to business—Investment 1.3, Notice D.D. 341 of 15 March 2022, en: PE0000020—CUPJ33C22002850006, entitled: Cultural Heritage Active Innovation for Sustainable Society, duration until 28 February 2026. Project RESTART is funded by “AlmaCaReS Cambiamenti Climatici, Resilienza, Sostenibilità”, an initiative promoted by the University of Bologna (5x1000) in response to the unfavourable weather conditions that struck the Emilia-Romagna area in May 2023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the paper.

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

References

- Doerr, M. The CIDOC conceptual reference model: An ontological approach to semantic interoperability of metadata. *AI Mag.* **2003**, *24*, 75–92.
- Hyvönen, E. Publishing and using cultural heritage linked data on the semantic web. In *Synthesis Lectures on the Semantic Web, Theory and Technology*; Morgan & Claypool Publishers: Morwen, UK, 2012; Volume 2, pp. 1–159.
- Kansa, E.C.; Kansa, S.W. We all know that a 14 is a sheep: Data publication and professionalism in archaeological communication. *J. East. Mediterr. Archaeol. Herit. Stud.* **2013**, *1*, 88–97. [\[CrossRef\]](#)
- Richards, J.D.; Hardman, C. Stepping Back from the Trench Edge: An Archaeological Perspective on the Development of Standards for Recording and Publication. Available online: <https://eprints.whiterose.ac.uk/id/eprint/7795/> (accessed on 6 May 2026).
- Scheffler, M.; Aeschlimann, M.; Albrecht, M.; Bereau, T.; Bungartz, H.-J.; Felser, C.; Greiner, M.; Groß, A.; Koch, C.T.; Kremer, K.; et al. FAIR data enabling new horizons for materials research. *Nature* **2022**, *604*, 635–642. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Luca, L. A digital ecosystem for the multidisciplinary study of Notre-Dame de Paris. *J. Cult. Herit.* **2024**, *65*, 206–209. [\[CrossRef\]](#)
- Bruseker, G.; Carboni, N.; Guillem, A. Cultural heritage data management: The role of formal ontology and CIDOC CRM. In *Heritage and Archaeology in the Digital Age*; Vincent, M., López-Menclero Bendicho, V., Ioannides, M., Levy, T., Eds.; Springer: Cham, Switzerland, 2017; pp. 93–131. [\[CrossRef\]](#)
- CIDOC CRM. Available online: <https://cidoc-crm.org/> (accessed on 6 May 2026).
- Swiss Art Research Infrastructure Documentation. Available online: <https://docs.swissartresearch.net/> (accessed on 6 May 2026).
- Bruseker, G.; Carboni, N. The semantic reference data modelling method: Creating understandable, reusable and sustainable semantic data models. *J. Open Humanit. Data* **2025**, *11*, 22. [\[CrossRef\]](#)
- Hermon, S.; Niccolucci, F.; Bakirtzis, N.; Gasanova, S. Digital twins in cultural heritage. *SSRN* **2022**. [\[CrossRef\]](#)
- Felicetti, A.; Niccolucci, F. Digital twin sensors in cultural heritage ontology applications. *Sensors* **2024**, *24*, 3978. [\[CrossRef\]](#) [\[PubMed\]](#)
- Patrimonio Culturale a Rischio. Available online: <https://site.unibo.it/patrimonioculturallearischio/it> (accessed on 22 April 2026).
- Resilienza Patrimonio Culturale. Available online: <https://site.unibo.it/resilienza-patrimonio-culturale/it> (accessed on 22 April 2026).
- Fondazione CHANGES. Available online: <https://www.fondazionechanges.org> (accessed on 22 April 2026).

16. Michalski, S.; Pedersoli, J.L., Jr. *The ABC Method: A Risk Management Approach to the Preservation of Cultural Heritage*; Canadian Conservation Institute: Ottawa, ON, Canada, 2016.
17. Ramalhinho, A.R.; Macedo, M.F. Cultural heritage risk analysis models: An overview. *Int. J. Conserv. Sci.* **2019**, *10*, 39–58.
18. Crowley, K.; Jackson, R.; O'Connell, S.; Karunarthna, D.; Anantasari, E.; Retnowati, A.; Niemand, D. Cultural heritage and risk assessments: Gaps, challenges, and future research directions for the inclusion of heritage within climate change adaptation and disaster management. *Clim. Resil.* **2022**, *1*, e45. [[CrossRef](#)]
19. Bonazza, A.; De Nuntiis, P.; Sardella, A.; Ciantelli, C.; Palazzi, E. ProteCHt2save: Valutazione del rischio e protezione sostenibile del patrimonio in ambienti in trasformazione. In *Monitoraggio e Manutenzione delle Aree Archeologiche: Cambiamenti Climatici, Dissesto Idrogeologico, Degrado Chimico-Ambientale; Atti del Convegno Internazionale di Studi, Roma, Italy, 20–21 March 2019*; L'Erma di Bretschneider: Rome, Italy, 2020; pp. 141–147.
20. Adetunji, O.; Daly, C. Climate risk management in cultural heritage for inclusive adaptation actions in Nigeria. *Heritage* **2024**, *7*, 1237–1264. [[CrossRef](#)]
21. Previtali, M.; Stanga, C.; Molnar, T.; Van Meerbeek, L.; Barazzetti, L. An integrated approach for threat assessment and damage identification on built heritage in climate-sensitive territories: The Albenga case study (San Clemente Church). *Appl. Geomat.* **2018**, *10*, 485–499. [[CrossRef](#)]
22. Paolini, A.; Vafadari, A.; Cesaro, G.; Santana Quintero, M.; van Balen, K.; Vileikis, O. *Risk Management at Heritage Sites: A Case Study of the Petra World Heritage Site*; UNESCO Amman Office: Amman, Jordan, 2012.
23. Fiorentino, S.; Casarotto, A.; Falbo, I.; Vandini, M. Building the knowledge base for cultural heritage risk assessment: The case of the Arian Baptistery, Ravenna (Italy). *Heritage* **2026**, *9*, 111. [[CrossRef](#)]
24. Casarotto, A.; Fiorentino, S.; Vandini, M. An approach to risk assessment and planned preventative maintenance of cultural heritage: The case of the hypogeum archaeological site of Sigismund Street (Rimini, Italy). *Heritage* **2025**, *8*, 344. [[CrossRef](#)]
25. SPARQL 1.1 Query Language. Available online: <https://www.w3.org/TR/sparql11-query/> (accessed on 6 May 2026).
26. SPARQL Endpoint—Ministero della Cultura. Available online: https://dati.cultura.gov.it/api-sparql_/ (accessed on 6 May 2026).
27. Catalogo Generale dei Beni Culturali. Available online: <https://catalogo.beniculturali.it/> (accessed on 6 May 2026).
28. Extensible Markup Language (XML) 1.0. Available online: <https://www.w3.org/TR/xml/> (accessed on 6 May 2026).
29. GeoNames. Available online: <https://www.geonames.org/> (accessed on 6 May 2026).
30. Istituto Centrale per il Catalogo e la Documentazione (ICCD). Available online: <http://www.iccd.beniculturali.it/> (accessed on 6 May 2026).
31. Progetto ArCo—Architettura Della Conoscenza. Available online: <https://dati.cultura.gov.it/progetto-ArCo-architettura-della-conoscenza/> (accessed on 6 May 2026).
32. CRMsci: Scientific Observation Model. Available online: <https://cidoc-crm.org/crmsci> (accessed on 6 May 2026).
33. Resource Description Framework (RDF). Available online: <https://www.w3.org/RDF/> (accessed on 6 May 2026).
34. Lorenzini, M. SIRIUS Mapping: CIDOC-CRM and X3ML Mapping Repository for Cultural Heritage Risk Assessment Data. Available online: https://github.com/matteoLorenzini/sirius_mapping/tree/main (accessed on 17 June 2026).
35. Lorenzini, M. Risk Analysis App: Semantic Risk Assessment and Knowledge Graph Exploration Repository. Available online: https://github.com/matteoLorenzini/risk_analysis_app/tree/main (accessed on 17 June 2026).
36. Lorenzini, M. Risk Analysis App Template: Semantic Risk Assessment Interface for Knowledge Graph Exploration. Available online: https://github.com/matteoLorenzini/risk_analysis_app/blob/main/data/templates/http%253A%252F%252Fwww.siriuplatform-risk-assessment%252Fresource%252Frisk-analysis.html (accessed on 17 June 2026).

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