

Article

Towards Heritage World Models

George Pavlidis ^{*}, Vasileios Sevetlidis  and Vasileios Arampatzakis 

Athena Research Center, 67100 Xanthi, Greece; vasiseve@athenarc.gr (V.S.);

vasilis.arampatzakis@athenarc.gr (V.A.)

^{*} Correspondence: gpavlid@athenarc.gr

Abstract

Digital twins have become a central paradigm for cultural heritage documentation, monitoring, and preventive preservation. Yet, when cultural heritage systems promise prediction, simulation, intervention planning, and decision support, a more explicit account is needed of the computational commitments behind such claims. This position paper proposes the notion of the *heritage world model* as a conceptual and architectural abstraction that uses the semantic digital twin as its representational layer and extends it toward prediction, memory, uncertainty-aware reasoning, and intervention evaluation. We define a heritage world model as a structured, temporally updated, semantically grounded, and action-aware model of a heritage asset and its preservation environment, capable of integrating observations, estimating latent risk states, predicting plausible future trajectories, and evaluating interventions under uncertainty. The paper does not present a validated deployed system. Rather, it clarifies the architectural conditions under which a decision-support digital twin infrastructure could support the kind of world-model-like preservation system proposed here. It further argues that such a model becomes operationally meaningful only when it includes a human-supervised controller layer that maps semantic state, predicted risk trajectories, uncertainty, memory, and institutional constraints into preservation-relevant actions, alerts, monitoring adaptations, or requests for expert review. Sensor data, remote sensing, computational models, risk assessments, policies, and conservation actions are interpreted as possible observational, dynamic, and intervention layers of a heritage world model. The paper reviews adjacent work in heritage digital twins, semantic and reactive ontologies, risk-aware preservation, agentic AI, and modern AI world models, and proposes a research agenda for moving toward predictive, memory-bearing, and intervention-aware preservation intelligence.

Keywords: world model; cultural heritage; digital twins; PANOPTES; ARGUS; semantic digital twins; agentic AI; ontology; risk memory; preventive preservation; decision support



Academic Editor: Marc Aurel Schnabel

Received: 1 May 2026

Revised: 21 May 2026

Accepted: 10 June 2026

Published: 13 June 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.

1. Introduction

Cultural heritage preservation increasingly operates under conditions of uncertainty, acceleration, and exposure. Heritage assets are affected by environmental, climatic, structural, anthropogenic, and institutional pressures that unfold across multiple temporal and spatial scales. Remote archaeological sites, exposed monuments, historic buildings, underground cultural landscapes, and fragile material remains are rarely threatened by a single isolated factor. More often, deterioration emerges through the interaction of humidity, temperature, rainfall, geological instability, visitor pressure, vegetation, pollution, structural weakness, maintenance strategies, and delayed decision-making.

Digital twins have been proposed as a response to this complexity [1–4]. In cultural heritage, they promise to move beyond static documentation by integrating three-dimensional models, geospatial context, sensor streams, remote sensing, semantic knowledge, predictive analytics, and visualization. In this sense, the digital twin is no longer simply a digital replica. It becomes a dynamic infrastructure, a living digital entity, for observing and managing the condition of a heritage asset over time.

However, a conceptual tension remains. When cultural heritage digital twins are described in terms of decision support, the computational meaning of this capacity is not always specified. The digital twin concept is broad enough to include monitoring, simulation, analytics, and decision-support ambitions, and heritage-oriented semantic/reactive models, such as HDTO/RHDTO, explicitly move the discussion toward sensors, reactivity, and decision processes [4–6]. However, the literature still lacks a clearly articulated architecture that combines semantic state estimation, latent risk modeling, uncertainty-aware memory, action-conditioned prediction, and policy-constrained control as an explicit closed preservation loop.

This paper introduces the term *heritage world model* for this stronger abstraction and studies it explicitly as a closed-loop preservation architecture. The term is intentionally adapted from contemporary artificial intelligence, where world models refer to internal representations that enable agents to understand, predict, simulate, and act within an environment [7–11]. In machine learning and autonomous systems, world models are commonly associated with learned dynamics, latent state spaces, action-conditioned prediction, imagined rollouts, planning, and memory. In cultural heritage, the same idea cannot be imported naively. Heritage preservation is not a video-game environment, a robotic control task, or a generic physical simulator. It is a socio-technical, material, historical, and value-sensitive domain in which observations, meanings, risks, and interventions must remain interpretable.

A related move from digital twins toward world models is also emerging in other technical domains, where digital twins are discussed as structured representations that can feed predictive world-model architectures for edge intelligence and action-oriented reasoning [12]. The heritage world model proposed here is therefore not a purely neural simulator. It is a hybrid semantic-predictive architecture that uses the digital twin as a structured representational layer and adds explicit state estimation, temporal memory, prediction, uncertainty representation, and intervention-aware reasoning.

The central claim of this paper is not that heritage world models replace digital twins, nor that all existing digital twins are passive. Rather, the claim is that a decision-support digital twin invokes an implicit world-model hypothesis whenever it promises to reason about future deterioration and intervention consequences. The proposed heritage world model abstraction makes that hypothesis explicit. A digital twin may represent, synchronize, and integrate evidence; a heritage world model organizes the broader closed-loop architecture required when the system is expected to estimate latent preservation states, predict possible futures, and compare the consequences of action. This also implies a *controller* layer within the heritage world model architecture. If the digital twin provides the structured representation of the heritage asset, the heritage world model uses this representation as part of a broader architecture for estimating latent preservation states, predicting how the preservation world may evolve, and supporting decisions about what should be done next. In cultural heritage, this controller layer should not be understood as a low-level motor controller or an autonomous actuator policy. It is better understood as a human-supervised preservation controller: a decision mechanism that transforms predicted risks, semantic evidence, uncertainty, previous interventions, and institutional

constraints into accountable preservation actions, alerts, recommendations, monitoring adaptations, or requests for further observation.

In this paper, *agentic AI* refers not to autonomous replacement of heritage experts, but to AI-supported reasoning systems capable of using memory, semantic context, policy constraints, and predicted futures to generate explainable recommendations under human supervision.

The PANOPTES ontology, developed within the ARGUS EU project¹, provides a concrete semantic basis for examining this shift [13–15]. PANOPTES was designed to integrate heritage assets, spatiotemporal asset states, measurements, instruments, diagnostic processes, computational models, predictions, threats, policies, decisions, and documentation within an interoperable semantic framework². Its later agentic extension embedded structured risk memory through the ICCROM/CCI ABC method, enabling the semantic encoding of risk episodes, policy responsiveness, and explainable preservation reasoning [16,17]. A further reflexive extension, based on evolving actor-network ontologies, introduced the possibility that semantic structures themselves may evolve in response to stakeholder constellations, contested meanings, and emergent preservation contexts [18–21]. *PANOPTES is used here as a concrete semantic basis* through which the proposed abstraction can be examined. It is not proposed as the definition of a heritage world model, and the concept is not intended to depend on PANOPTES specifically. From this perspective, PANOPTES can be interpreted as one possible semantic state space for a heritage world model, where ontology entities, temporal structures, provenance traces, risk representations, predictive components, and policy constraints jointly support preservation reasoning.

The contributions of this position paper are fourfold:

1. It introduces the concept of the *heritage world model* as an architectural abstraction for decision-support cultural heritage digital twins.
2. It synthesizes relevant work on heritage digital twins, semantic and reactive heritage ontologies, risk-aware decision support, agentic AI, and modern AI world models.
3. It maps PANOPTES and its agentic/reflexive extensions onto the structural components of the proposed heritage world model abstraction, while treating PANOPTES as a worked semantic basis rather than as the concept's definition.
4. It proposes illustrative scenarios, a synthetic operational sketch, and a research agenda for operationalizing the proposed heritage world model abstraction in preventive preservation.

This paper introduces the term *heritage world model* to describe a closed-loop preservation architecture that is not yet established as such in the cultural heritage literature. *The manuscript should therefore be read as a position paper and architectural proposal, not as a report of a validated deployed system.* Its aim is to define the conceptual scope of this proposed abstraction and to outline the computational commitments required for predictive, memory-bearing, uncertainty-aware, and intervention-aware preservation systems built over semantic digital twin infrastructures.

2. From Heritage Digital Twins to Predictive Preservation Systems

The digital twin concept originated in engineering and cyber-physical systems, where it denotes a digital counterpart of a physical asset or process that remains synchronized with its real-world referent [1,2]. In cultural heritage, the idea has been adopted to support documentation, monitoring, visualization, preventive conservation, and heritage management [3–5,15]. Heritage digital twins may combine 3D models, HBIM structures, GIS layers, sensor readings, environmental records, material information, inspection data, and conservation histories. This expansion has been important because conventional digital documentation remains largely static. A 3D model of a monument, an HBIM representa-

tion of a historic building, or a GIS inventory of heritage locations may support access, visualization, and documentation, but it does not necessarily provide dynamic understanding. Preventive preservation requires continuous observation, interpretation of change, prioritization of risks, and selection of interventions before irreversible damage occurs.

At the same time, it would be misleading to define the digital twin concept only as passive representation. In the broader digital twin literature, monitoring, simulation, analytics, and decision-support are often treated as possible ambitions of the paradigm, while heritage-oriented semantic and reactive approaches such as HDTO/RHDTO explicitly move the discussion toward sensors, reactivity, and decision processes [4–6]. The argument of this paper is therefore not that digital twins are passive or that they should be replaced by a different established paradigm. Rather, the paper introduces the term *heritage world model* as a proposed abstraction for making explicit the computational architecture required by decision-support digital twins when their claims depend on state estimation, latent risk modeling, action-conditioned prediction, uncertainty-aware memory, and policy-constrained control.

Semantic approaches to heritage digital twins address part of this landscape by introducing ontologies and knowledge graphs [4,5,22,23]. CIDOC CRM remains the foundational reference model for cultural heritage interoperability [24,25], while extensions and complementary models such as CRMdig, CRMsci, CRMba, LIDO, EDM, and domain-specific ontologies support documentation, provenance, scientific observation, and cultural data integration. These models are important because world-model-like reasoning should remain connected to semantic interoperability, provenance, and cultural documentation. However, descriptive semantic interoperability alone does not define a world model. The additional requirement is that semantic state, observation history, latent risk, prediction, uncertainty, policy, and action become part of an explicit closed-loop preservation architecture. The Heritage Digital Twin Ontology (HDTO) and related work by Niccolucci, Felicetti, Hermon, and collaborators mark an important step toward formalizing the heritage digital twin itself as an ontological construct [4,22]. HDTO connects cultural heritage documentation with the digital twin paradigm and introduces explicit semantic relations between heritage entities and their digital counterparts. The Reactive Heritage Digital Twin Ontology (RHDTO) further extends this direction by incorporating sensors, activators, and decision processes, thereby shifting the twin from a synchronic documentary structure toward a diachronic and reactive system [5,6]. This line of work provides a directly relevant semantic and reactive heritage-digital-twin foundation for the direction proposed here. The heritage world model abstraction extends this direction by making explicit the broader closed-loop architecture required for latent-state estimation, prediction, uncertainty-aware memory, policy-constrained reasoning, and accountable intervention.

Beyond ontology-centered digital twins, the proposed framing is also related to preventive preservation modeling and decision-analysis traditions. Risk-based preventive preservation has a long tradition in cultural heritage, including cultural property risk analysis and the ICCROM/CCI ABC method [16,26]. Hygrothermal monitoring and simulation have been used to study historic buildings and preventive conservation scenarios, including risk-based indoor climate analysis and validation of dynamic hygrothermal models for historical buildings [27–29]. Heritage Building Information Modeling has also been reviewed as a framework for documenting, analyzing, and managing heritage buildings [30]. Bayesian or dependency-based models have been proposed for risk and dependency modeling in preventive conservation [31], while multi-criteria decision-making approaches have been applied to heritage assessment, adaptive reuse, and preventive maintenance prioritization [32–34]. Agent-based modeling has similarly been used to understand tourism and visitor-flow dynamics in spatial settings, including sustainable tourism

and heritage-site management [35–37]. These approaches are relevant to the construction of the proposed heritage world model architecture over digital twin infrastructures because they provide possible implementations of transition dynamics T , latent risk estimation Z , uncertainty representation U , and controller objectives C . However, in the literature reviewed here, these methods are typically presented as analytical tools rather than as semantically integrated, provenance-aware, policy-constrained components of a live digital twin. The proposed heritage world model should therefore be understood not as a replacement for these methods, but as an architectural framework within which they can be connected, interpreted, and governed.

The work in the ARGUS EU project can be interpreted as occupying such an intermediate position [13–15]. Its digital twin vision is not restricted to 3D representation or documentation. It explicitly combines multimodal sensing, remote sensing, AI-based threat analysis, semantic modeling, risk assessment, decision support, and participatory heritage management. PANOPTES formalizes many of the entities required to make this operational. It represents heritage assets, asset states, measurements, instruments, protocols, physical quantities, diagnostic interpretations, computational models, predictions, threats, policies, rules, decisions, events, visualizations, and cultural documentation. It also aligns with CIDOC CRM, SOSA/SSN, PROV-O, GeoSPARQL, OWL-Time, and FAIR principles [13,17,24,38–42]. In this sense, PANOPTES already captures much of the semantic and procedural structure required for preservation-oriented reasoning, since observations are linked to diagnostic and computational processes, predictions are evaluated against policies, threats are identified, decisions are recommended, and provenance preserves the reasoning chain.

3. World Models in Contemporary Artificial Intelligence

In artificial intelligence, the term *world model* refers broadly to an internal model that allows an agent to represent, predict, and act within an environment [7–9]. The concept has a long history in cognitive science, reinforcement learning, control, and model-based planning, but it has gained renewed importance through modern deep learning, model-based reinforcement learning, self-supervised video representation learning, and joint-embedding predictive architectures [7,10,11,43–46].

Ha and Schmidhuber’s influential formulation treated world models as learned compressed representations of environments that allow agents to use imagined rollouts for policy learning [7,8]. In such architectures, high-dimensional observations are encoded into a lower-dimensional latent state; a recurrent or generative dynamics model predicts future latent states; and an agent can learn or evaluate behavior within this internally simulated environment. The key insight is that intelligent action does not require direct interaction with the external world at every step. An agent can use an internal model to anticipate possible futures.

LeCun’s more recent vision of autonomous machine intelligence similarly places predictive world models at the center of intelligent behavior [9]. In this view, intelligence requires systems that learn representations of the world, predict future states, reason over abstract representations, maintain memory, and plan across multiple time horizons. Recent work has made this direction more concrete through joint-embedding predictive architectures and video-based world models. V-JEPA 2 explores self-supervised learning from internet-scale video combined with a small amount of interaction data, aiming to support understanding, prediction, and planning in the physical world [44]. Its 2026 continuation, V-JEPA 2.1, further emphasizes dense, spatially structured, and temporally consistent representations for video understanding and embodied tasks [47]. Related JEPA-based work investigates physical planning in learned representation spaces, rather than directly in the

input space, emphasizing that useful world models may support action by abstracting away irrelevant visual detail while preserving task-relevant dynamics [45]. LeWorldModel further illustrates this trend by training a compact end-to-end JEPA from pixels and evaluating its latent space for planning speed, physical structure, and surprise detection [46]. World models are therefore not only simulators, but mechanisms for abstraction, prediction, planning, and control.

Across contemporary AI, several recurring elements define world-model thinking:

1. **State representation:** an internal representation of the current condition of the environment.
2. **Observation model:** a way of connecting raw sensory inputs to the inferred state.
3. **Latent variables:** hidden factors that are not directly observed but explain observations and future behavior.
4. **Transition dynamics:** a model of how states evolve over time.
5. **Action conditioning:** a model of how interventions or actions modify future states.
6. **Prediction or rollout:** the capacity to generate possible future trajectories.
7. **Memory:** the preservation of relevant past states, actions, and outcomes.
8. **Uncertainty:** the estimation of confidence, ambiguity, and unknowns.
9. **Planning or policy selection:** the use of predicted futures to select actions.

Closely related to the world model is the notion of the *controller*, which is central in control theory, model-based reinforcement learning, and planning architectures [10,11,43,45]. In control theory, reinforcement learning, and model-based agency, the controller is the component that selects actions by using the current state, predicted future states, goals, constraints, and feedback. The world model estimates or simulates the environment, while the controller uses that model to choose the next action. In some architectures, the controller is a learned policy. In others, it may be a rule-based mechanism, an optimization routine, a planner, or a human-supervised decision process. The distinction is important for cultural heritage because the preservation domain cannot be reduced to autonomous action selection. Instead, the controller must remain interpretable, policy-aware, and accountable to human experts and institutions.

A parallel recent concern is *evaluation*. As world models become more general and multimodal, their quality cannot be assessed only through pixel-level reconstruction or short-horizon prediction error. Recent work on evaluating world models with vision-language models points toward broader criteria, including semantic consistency, temporal plausibility, and action relevance [48]. This is especially important for cultural heritage, where the correctness of a predicted future is not only visual or statistical, but also semantic, causal, and policy-relevant. These elements are usually discussed in the context of robotics, reinforcement learning, embodied AI, autonomous agents, or video-based predictive learning. *Cultural heritage requires a different instantiation.*

Heritage systems should not reduce the world to opaque latent vectors when domain experts require interpretability, provenance, and accountability. Nor should they treat preservation as a simple reward-maximization problem, since heritage values are plural, negotiated, and institutionally constrained. A heritage world model should therefore be hybrid. It may include learned models, simulations, rule-based reasoning, probabilistic forecasts, symbolic ontologies, expert knowledge, and policy constraints. Its state space should be partly semantic and inspectable. Its predictions should be linked to evidence and provenance. Its actions should be conservation interventions, monitoring decisions, access-management policies, inspection recommendations, or communication strategies. Its objective should not be abstract reward alone, but risk reduction, value preservation, reversibility, feasibility, accountability, and alignment with cultural significance.

4. Defining the Heritage World Model

We define a *heritage world model* as follows:

A heritage world model is a structured, temporally updated, semantically grounded, and action-aware model of a heritage asset and its preservation environment, capable of integrating observations, estimating latent risk states, predicting plausible future trajectories, and evaluating interventions under uncertainty.

The proposed heritage world model uses the semantic digital twin as its representational layer while extending it toward explicit state estimation, prediction, uncertainty representation, memory, and action evaluation. The emphasis is therefore not on replacing digital twins, but on making explicit the computational architecture required for decision-support-oriented preservation systems.

A minimal formal abstraction can be introduced as:

$$HWM = \langle S, O, Z, A, T, R, M, U, P, C \rangle \quad (1)$$

where S is the semantic state space of the preservation world, O is the observation space, Z denotes latent risk or condition variables, A is the action or intervention space, T is the transition model, R is the risk, value, or objective model, M is memory and provenance, U is uncertainty, P denotes policies and institutional constraints, and C is the controller that maps state, prediction, uncertainty, memory, risk/value criteria, and policy constraints into preservation-relevant actions.

This tuple should be read as an architectural decomposition of computational responsibilities, not as a single closed-form algorithm. The components specify what must be organized around a semantic digital twin if it is to support the proposed heritage world model architecture. Different implementations may instantiate these components through physical simulation, statistical forecasting, Bayesian filtering, machine learning, expert rules, semantic reasoning, multi-criteria decision analysis, or hybrid combinations of these methods. Thus, the formalism is intentionally modular, as it identifies the required roles of state, observation, latent risk, transition dynamics, memory, uncertainty, policy, and control, while leaving their concrete implementation open to the data, site, and institutional context.

It should also be noted that the formulation is deliberately compatible with partially observable control models [49]. In typical heritage settings, the true preservation state is only partially observed, since sensors are sparse, inspections are intermittent, remote-sensing products are indirect, and expert interpretations may disagree. Therefore, the state S_t at time t should not be understood as a fully known physical state, but as a semantically structured belief state anchored in the active Asset State. The latent component Z_t captures preservation-relevant conditions that are inferred rather than directly observed, while U_t records uncertainty arising from sensor error, missing evidence, model limitations, and interpretive ambiguity. This makes the proposed heritage world model closer to a partially observable decision process than to a fully observed simulator.

Action-conditioned prediction means that future preservation risk is estimated not only under passive continuation of current conditions, but also under possible interventions and information-gathering actions. The resulting architectural distinction is summarized in Table 1.

Table 1. From heritage digital twins to heritage world models.

Dimension	Typical Heritage Digital Twin Emphasis	Heritage World-Model Emphasis
Primary role	Representation, synchronization, monitoring, and integration	Explicit state estimation, prediction, reasoning, and intervention evaluation
Object of modeling	Heritage asset, digital counterpart, observed condition, and documentation context	Asset, environment, latent risks, policies, actors, actions, and possible futures
Temporal scope	Past and present, with forecasting not architecturally required	Past, present, predicted, and counterfactual futures under alternative actions
Data role	Integration, visualization, interoperability, and monitoring	Evidence for belief-state update, transition modeling, uncertainty estimation, and decision support
Semantics	Documentation, interoperability, provenance, and domain meaning	Inspectable semantic state space for risk reasoning, prediction, memory, and policy-constrained action
Risk	Recorded, visualized, assessed, or forecast depending on the system	Estimated as a latent preservation state, remembered across episodes, predicted under actions, and used by the controller
Action	Often represented as an external management response or recommendation	Explicit intervention or information-gathering variable in action-conditioned prediction
Decision support	May be provided by attached analytics, rules, or expert workflows	Architecturally central: prediction, uncertainty, policy, memory, and controller logic form a closed loop

Within the heritage world model architecture, the controller layer can be expressed abstractly as:

$$a_t = C(S_t, Z_t, \hat{S}_{t+1:t+k}, U_t, M_t, R, \mathcal{P}), \quad a_t \in \mathcal{A} \quad (2)$$

where S_t is the current semantic state, Z_t is the estimated latent risk state, $\hat{S}_{t+1:t+k}$ denotes predicted future states over a horizon k , U_t represents uncertainty, M_t denotes memory and provenance, R represents the risk or value model, $\mathcal{P}$ denotes policies and institutional constraints, and a_t is the recommended or executable preservation action selected from the action space $\mathcal{A}$. The controller need not be implemented as a single learned policy. In early heritage deployments, C may be realized as a hybrid decision layer combining threshold rules, temporal logic, multi-criteria scoring, expert review, and policy constraints. The objective model R should therefore be understood as a constrained preservation utility rather than a free reward function. It may combine risk reduction, uncertainty reduction, intervention cost, reversibility, access impact, and conservation priority. The policy layer $\mathcal{P}$ restricts the admissible actions by encoding hard or soft constraints, such as avoiding irreversible intervention under high uncertainty, requiring expert approval for invasive action, or prioritizing information-gathering when evidence is conflicting.

Different forms of controller should also be distinguished. An *alerting controller* may only raise warnings when predicted risk or uncertainty exceeds a policy threshold. A *monitoring-scheduling controller* may recommend increased sampling, sensor calibration, inspection, or sensor redeployment. An *expert-recommendation controller* may propose conservation options but require human validation before any operational response. An *intervention-decision controller* concerns stronger preservation actions and should remain subject to explicit expert approval, institutional accountability, and documentation of reversibility, proportionality, and uncertainty. This typology is important because cultural heritage preservation is not an autonomous control problem, and the appropriate degree of automation depends on the reversibility, cost, uncertainty, and ethical significance of the action. Thus, in cultural heritage, a_t should be interpreted broadly. It may denote a physical intervention, an inspection request, a sensor redeployment, an increase in sam-

pling frequency, an access restriction, an alert escalation, a stakeholder communication, or a request for human review. Many preservation actions modify observation, governance, interpretation, or institutional response rather than the physical asset directly.

The abstraction used here is intentionally general. It can be instantiated by rule-based systems, probabilistic models, machine learning models, simulations, semantic reasoners, or hybrid architectures. What matters is not that every component is implemented through deep learning, but that the system can answer a specific class of preservation questions:

- What is the current state of the heritage asset and its environment?
- Which hidden risks best explain the observed evidence?
- How is the risk likely to evolve if no action is taken?
- Which intervention would most plausibly reduce future risk?
- How uncertain is the prediction, and what observations would reduce uncertainty?
- How does the current situation compare with previous episodes?
- Which decision can be justified under the applicable conservation policy?

These questions are representative of the questions that decision-support-oriented heritage digital twins are expected to address [3,5,15,17]. The heritage world model therefore makes explicit the computational commitments implied by such decision-support expectations.

5. PANOPTES as a Semantic Basis for Heritage World Modeling

PANOPTES is used in this paper as a worked semantic basis for examining how the proposed heritage world model abstraction could be operationalized [13–15,17]. It is not presented as the definition of a heritage world model, nor as evidence that a complete HWM has already been implemented. Rather, it provides an ontology-based case through which semantic state, observation provenance, latent risk, transition assumptions, uncertainty, policy constraints, memory, and decision traces can be inspected.

In the proposed interpretation, PANOPTES can provide one possible semantic state space S of a heritage world model. The state is not a single vector but a structured graph of entities and relations. A heritage asset is represented through one or more *Asset States*, each corresponding to a temporally and spatially situated snapshot of the asset and functioning as an instance of the digital twin. Measurements, documentation, installed devices, computational model executions, events, predictions, threats, policies, and decisions are then associated with the relevant active or historical asset state. The state therefore includes both material and informational entities, including physical assets, sensor observations, computational outputs, policies, rules, decisions, and provenance traces.

The observation space O is represented through measurements, instruments, protocols, physical quantities, remote sensing products, inspection records, and expert annotations. SOSA/SSN alignment allows sensor observations to be represented in interoperable form [38]. GeoSPARQL and OWL-Time provide spatial and temporal grounding [40,41]. PROV-O preserves provenance across observations, computations, and decisions [39]. The geospatial component of a heritage world model is not limited to storing site coordinates or asset locations. Heritage assets are spatially embedded entities whose preservation state may depend on terrain, hydrology, slope, vegetation, land cover, accessibility, exposure to neighboring hazards, and multi-scale remote-sensing evidence. In a fully operational version of the proposed architecture, GeoSPARQL-aligned geometries should support spatial relations, buffer-based exposure analysis, proximity to hazards, spatial uncertainty, and transitions between local asset states and broader environmental contexts. This is particularly important for relevant settings, where in-situ measurements and satellite-derived indicators must be fused across different spatial resolutions, temporal update rates, and uncertainty levels. The geospatial state is therefore part of the preservation world model itself, not merely a visualization layer.

The latent risk layer Z emerges through diagnoses, threats, risk assessments, and inferred conditions. A humidity measurement, for example, is not itself a risk. It becomes risk-relevant when interpreted through material vulnerability, temporal persistence, policy thresholds, previous episodes, and predicted consequences. The Agentic PANOPTES extension makes this more explicit by embedding the ICCROM/CCI ABC method into the semantic core [16,17]. Each risk assessment becomes a temporally situated memory unit with components corresponding to frequency, loss, fraction affected, and magnitude of risk [16,17]. In this context, diagnosis should be understood functionally, as it denotes the interpretive step through which observations and model outputs are transformed into preservation-relevant latent states, whether or not it is implemented as a distinct ontology class.

The action space $\mathcal{A}$ is represented through decisions, policies, rules, interventions, inspection actions, mitigation strategies, monitoring changes, and potentially actuator-based responses. In this layer, a decision is not merely a textual recommendation. It is a semantic entity linked to evidence, prediction, policy, and provenance.

The transition model T is only partly specified by the ontology itself. PANOPTES provides the scaffolding for transition modeling by representing computational models, predictions, events, and temporal updates. The actual dynamics may be implemented through physical simulation, statistical forecasting, machine learning, expert rules, or hybrid models. This distinction is important since the ontology does not replace predictive modeling, rather it structures the inputs, outputs, assumptions, and interpretability of predictive modeling.

The memory component M is distributed across cultural documentation, provenance, risk assessments, events, previous decisions, and historical measurements. This memory allows the system to compare present conditions with previous risk episodes, justify recommendations, and evaluate whether interventions changed future trajectories.

In an operational implementation, U should not be treated as a single scalar confidence value. It may be represented as a set of provenance-linked uncertainty attributes attached to observations, inferred risks, predictions, and decisions. At the observation level, uncertainty may encode missingness, sensor reliability, measurement error, spatial resolution, temporal latency, or disagreement between modalities. At the latent-risk level, it may encode posterior belief, expert disagreement, or ambiguity in the causal interpretation of observed change. At the prediction level, it may be produced by ensembles, Bayesian filtering, scenario spread, or model residuals. At the controller level, uncertainty constrains the admissible action space, such that under high uncertainty, the preferred output may be inspection, sensor calibration, or increased sampling rather than direct intervention. Existing PANOPTES structures can record provenance and context, but a full heritage world model should represent uncertainty more explicitly.

The controller C corresponds to the decision-support logic of the heritage world model, building on the policy, rule, decision, and risk-memory structures already introduced in PANOPTES and Agentic PANOPTES [13,17]. In a simple implementation, it may consist of SPARQL queries, threshold rules, and ABC-based prioritization. In a more advanced implementation, it may combine agentic reasoning, counterfactual scenario comparison, intervention planning, uncertainty-aware monitoring decisions, and explanation generation. In the PANOPTES setting, the controller is distributed across the Policy, Rule, Decision, RiskAssessment, Prediction, and agentic reasoning layers. Its role is to close the loop between observation and intervention, in such a way that observations update the semantic state, predictive components estimate possible futures, and the controller layer selects, recommends, or escalates the next accountable action within the heritage world model loop.

Operational deployment also requires attention to query performance, update latency, and scalability. A heritage world model should not assume that all reasoning must occur over a monolithic ontology graph in real time. Practical implementations may separate high-frequency ingestion pipelines, time-series databases, geospatial indexes, relational or graph stores, cached semantic views, and slower expert-facing reasoning layers. Semantic reasoning can therefore operate at selected decision points, while streaming anomaly detection, spatial preprocessing, and model execution may run in specialized computational services. This separation is important for remote monitoring deployments, where sensor frequency, satellite revisit times, network reliability, and expert-response workflows operate at different temporal scales.

The mapping between the proposed heritage world model components and the corresponding PANOPTES structures is summarized in Table 2. The Table identifies where the semantic, observational, predictive, policy, memory, uncertainty, and controller-related responsibilities of the proposed abstraction can be anchored in the existing ontology-based framework [13,14,17].

Table 2. Indicative mapping between the proposed heritage world model components and PANOPTES/ARGUS structures. The mapping identifies semantic anchoring points; it does not imply that PANOPTES or ARGUS already implement a complete heritage world model.

World-Model Component	PANOPTES/ARGUS Counterpart
S: Semantic state space	Asset, AssetState, Location, temporal context, cultural documentation, condition, event context
O: Observations	Measurement, Instrument, MeasurementProtocol, PhysicalQuantity, remote sensing, expert annotation, inspection records
Z: Latent risk state	Diagnosis/diagnostic interpretation, Threat, RiskAssessment, vulnerability, deterioration, ABC components
A: Action space	Decision, Policy, Rule, intervention, inspection, mitigation, monitoring adaptation
T: Transition model	ComputationalModel, ComputationalModelExecution, Prediction, event progression, risk trajectory, deterioration forecast
R: Risk/objective model	MagnitudeOfRisk, preservation priority, policy threshold, conservation objective
M: Memory/provenance	CulturalDocumentation, PROV-O traces, historical measurements, risk episodes, previous decisions
U: Uncertainty	sensor reliability, missing data, confidence estimates, model uncertainty, contested interpretation, spatial uncertainty
P: Policy constraints	Policy, Rule, conservation policies, institutional rules, access constraints, reversibility principles, stakeholder requirements
C: Controller	Policy, Rule, Decision, RiskAssessment, Prediction, agentic reasoning, SPARQL rules, alerting, intervention recommendation, monitoring adaptation

The next step is to operationalize these components as a cyclic architecture:

observe → update state → infer risk → predict trajectories → evaluate actions →
 → controller decision → act/escalate → record outcome →
 → memory/state update → observe (3)

6. Illustrative Toy Scenarios

The concept of a heritage world model can be clarified through small, controlled scenarios. These are not presented as validated models, but as minimal examples of the type of reasoning that the proposed heritage world model makes explicit, including latent risk interpretation, action-conditioned prediction, uncertainty-aware recommendation, and policy-constrained decision support.

6.1. Moisture Accumulation in a Historic Building

Consider a wall painting or masonry surface in a historic building affected by chronic moisture. A digital twin representation may include the location of the wall, its geometry, material type, and recent humidity readings. The heritage world model perspective asks what additional reasoning is needed when these observations are used for risk interpretation and preservation decision support.

- The state S includes the asset, material vulnerability, spatial zone, historical measurements, previous conservation events, and applicable microclimate policy.
- The observation space O includes relative humidity, temperature, wall moisture proxy, rainfall data, and inspection notes.
- The latent risk state Z represents chronic moisture accumulation and possible biological growth.
- The transition model T estimates how risk evolves under continued humidity exposure.
- The action space A includes ventilation adjustment, inspection, temporary access limitation, and additional sensor deployment.

The heritage world model view adds two elements beyond monitoring. It estimates whether the observed moisture pattern corresponds to a latent deterioration risk, and it compares possible responses such as continued monitoring, increased sampling, expert inspection, or ventilation adjustment. This scenario is expanded into a minimal formal instantiation in Section 7, while Section 8 presents a broader relevant operational simulation based on multimodal late fusion and adaptive control.

6.2. Tourism Pressure at an Archaeological Site

At an exposed archaeological site, visitor pressure may contribute to path erosion, surface abrasion, uncontrolled access, and localized structural stress. A digital twin representation may include visitor routes and sensitive zones. Within a heritage world model, visitor movement becomes an action-sensitive component of the preservation world.

- The state includes visitor density, path geometry, sensitive asset zones, surface condition, weather, seasonality, and previous management actions.
- Observations include visitor counts, UAV imagery, inspection records, erosion indicators, and possibly participatory reports.
- Latent variables include abrasion risk and access-induced degradation.
- Actions include rerouting visitors, temporary closure, signage, path reinforcement, and monitoring intensification.

The world model can generate counterfactual futures. What happens if the current visitor route remains unchanged, if visitors are redirected, or if access is seasonally limited? The decision is not simply triggered by a threshold. It is justified by comparing predicted preservation outcomes under alternative management policies. Here the controller operates over management actions rather than material interventions. It may recommend visitor rerouting, temporary access limitation, signage, additional monitoring, or communication with site authorities. The decision depends not only on the predicted deterioration trajectory, but also on reversibility, visitor impact, conservation priority, and institutional feasibility.

6.3. Underground Cellar Microclimate

In an underground cellar town, microclimatic stability is essential but fragile. Temperature, humidity, airflow, and condensation interact with geology, construction material, and human access. A monitoring dashboard may show current readings. A heritage world model estimates whether an observed drift is normal seasonal variation, sensor malfunction, or emerging deterioration risk.

- The state includes cellar geometry, material properties, airflow structure, historical seasonal cycles, access patterns, and known vulnerabilities.
- Observations include temperature, humidity, air movement, visual inspection, and local weather.
- Latent risks include condensation, salt crystallization, and structural weakening.
- Actions include ventilation adjustment, access control, inspection, and sensor redeployment.

Here the key contribution is uncertainty-aware interpretation. The system should not only issue an alert but explain whether the alert arises from a known seasonal pattern, a deviation from previous memory (anomaly in the data), a sensor anomaly, or an emerging threat. This is where semantic memory and provenance become essential. In this case, the controller may choose an information-gathering action before a physical intervention. If the model cannot distinguish seasonal variation from sensor drift or emerging deterioration, the appropriate action may be to request inspection, deploy an additional sensor, or compare the episode with historical microclimatic memory. This highlights a distinctive property of the proposed heritage world model framing, by which, sometimes, the most responsible action is not to intervene physically, but to reduce uncertainty.

7. Minimal Operational Instantiation

To clarify how the proposed abstraction can be operationalized without requiring a full deployment, this section sketches a minimal instantiation for the moisture-accumulation scenario presented in the previous section. The example is intentionally simple, not presented as a validated conservation model, but as a transparent template showing how a heritage world model can connect semantic state, observations, latent risk, prediction, uncertainty, policy, and controller logic.

Let the active semantic state S_t be anchored by an Asset State corresponding to a wall painting or masonry surface at time t . The observation space O_t contains relative humidity RH_t , temperature $Temp_t$, rainfall proxy $Rain_t$, wall-moisture proxy WM_t , and inspection notes. The latent risk variable Z_t denotes the probability or graded belief that the asset is undergoing moisture-induced biological or material deterioration. This latent state is not directly observed. It is inferred from sustained humidity exposure, material vulnerability, prior episodes, and the reliability of available measurements. A minimal transition model may be expressed as:

$$Z_{t+1} = f(Z_t, RH_t, Temp_t, WM_t, Rain_t, V, a_t) + \epsilon_t \quad (4)$$

where V denotes material vulnerability, a_t denotes the selected action, and ϵ_t captures model uncertainty. In the simplest rule-based implementation, f may increase Z_{t+1} when relative humidity remains outside an acceptable range for more than a policy-defined duration. In a probabilistic implementation, Z_t may be represented as a Bayesian belief updated from observations [31,49]. In a physics-informed implementation, f may be coupled to a hygrothermal model or to empirical moisture-response curves [27–29]. The policy layer $\mathcal{P}$

defines acceptable conditions and action thresholds, consistent with risk-based preventive conservation and the ABC method [16,26]. For example:

$$\mathcal{P}_1 : RH \notin [45\%, 60\%] \text{ for more than 48 h} \Rightarrow \text{humidity stress event} \quad (5)$$

$$\begin{aligned} \mathcal{P}_2 : \Pr(Z_{t+7} = \text{high risk} \mid O_{1:t}, M_t) > 0.7 \\ \Rightarrow \text{recommend preventive inspection or ventilation adjustment} \end{aligned} \quad (6)$$

The controller C maps the predicted latent risk, uncertainty, memory, and policy constraints into a preservation action:

$$C(S_t, Z_t, \hat{Z}_{t+1:t+k}, U_t, M_t, \mathcal{P}) \rightarrow a_t. \quad (7)$$

For example, the controller may use the following transparent decision template:

$$a_t = \begin{cases} \text{increase sampling frequency,} & \text{if } U_t \text{ is high and } Z_t \text{ is moderate} \\ \text{request sensor calibration,} & \text{if observations conflict with historical memory} \\ \text{recommend expert inspection,} & \text{if } Z_t \text{ is high and confidence is sufficient} \\ \text{recommend ventilation adjustment,} & \text{if } \hat{Z}_{t+7} \text{ remains high under no action} \\ \text{continue monitoring,} & \text{otherwise} \end{cases}$$

The explanatory trace produced by the system should include provenance-preserving links across observations, model executions, predictions, policies, and decisions [13,17,39]. In PANOPTES terms, this trace can be represented through *AssetState*, *Measurement*, *ComputationalModel*, *Event*, *Prediction*, *Policy*, *ThreatContext*, *DecisionContext*, and *Decision*. Thus, even this minimal example illustrates the additional reasoning commitments emphasized by the heritage world model framing, by which the system does not merely report that humidity is high, rather it estimates latent deterioration risk, predicts future risk under possible actions, and recommends either an information-gathering action or a preventive intervention under explicit policy constraints.

8. Illustrative Operational Simulation

This section presents a synthetic operational sketch of the proposed heritage world model abstraction in a PANOPTES-like setting. The simulation is not intended to validate an operational deployment, evaluate conservation effectiveness, or demonstrate superiority over existing digital twin or expert-rule systems. Its purpose is narrower, focusing on making the proposed closed-loop architecture concrete by showing how heterogeneous evidence streams can be transformed into fused preservation-risk signals, latent ABC-style risk estimates, action-conditioned short-horizon predictions, and policy-constrained controller recommendations. In this sense, the simulation provides an illustrative operational instantiation of the proposed heritage world model, rather than a calibrated conservation model for any specific site.

8.1. Simulation Setup

Five ARGUS-like pilot sites were simulated over a period of 180 days: Delos island in Greece, Sant'Antonio di Ranverso Abbey in Italy, Cellar town of Baltanás in Spain, Schenkenberg Castle in Switzerland, and Monti Lucretili open sites in Italy. Each site was assigned a fixed vulnerability value and three synthetic evidence streams, including a local sensor anomaly score, a satellite-derived environmental stress score, and an image-derived anomaly score. Missing observations were also simulated in order to reflect the incompleteness expected in real remote-monitoring settings. An uncertainty score was

computed from missingness and disagreement between modalities. At each time step, the observation vector was represented as:

$$O_t = [s_t, r_t, i_t, v, u_t] \quad (8)$$

where s_t denotes the sensor anomaly score, r_t the satellite stress score, i_t the image anomaly score, v the site vulnerability, and u_t the estimated uncertainty. These variables correspond to the observation and uncertainty components of the heritage world model, while v acts as a simple static proxy for material or contextual susceptibility.

A transparent late-fusion layer was then used to compute a fused preservation-risk signal in a heuristic manner:

$$F_t = 0.45s_t + 0.35r_t + 0.10i_t + 0.10v \quad (9)$$

Although heuristic, the weighting scheme intentionally gives priority to local sensor evidence and satellite-derived stress, while allowing image anomaly and vulnerability to provide additional contextual information. These weights are illustrative and were not calibrated from empirical conservation data.

The fused signal was converted into a simplified ABC-style latent risk estimate. The three ABC-inspired components were represented as operational proxies, where A_t approximates persistence or frequency through a moving average of the fused signal, B_t approximates expected loss using vulnerability and satellite stress, and C_t approximates the fraction affected using image and sensor anomaly evidence. The latent risk score was computed as:

$$Z_t = \sigma(0.40A_t + 0.35B_t + 0.25C_t - \theta) \quad (10)$$

where $\sigma(\cdot)$ is the logistic function and θ is a risk-offset parameter. This should not be interpreted as a formal implementation of the ICCROM/CCI ABC method, where the magnitude of risk is classically expressed as the direct additive combination $MR = A + B + C$. Instead, the present formulation uses an ABC-inspired computational proxy in which the three components are combined through a weighted latent-risk function in order to illustrate how risk reasoning can be embedded into a heritage world model loop.

The transition component was implemented as a seven-day action-conditioned rollout. For each day, the model predicted the future risk under passive continuation and under the action selected by the controller. The available actions were:

$$A = \{\text{continue monitoring, increase sampling, request inspection,} \\ \text{recommend preventive intervention}\} \quad (11)$$

The controller was implemented as a conservative adaptive controller. It begins with predefined action preferences, but updates these preferences from simulated outcomes, using a simple bandit-like action-value update rule inspired by reinforcement-learning formulations of action selection [50,51]. The reward favors predicted risk reduction and uncertainty reduction, while penalizing unnecessary intervention cost and missed high-risk situations. However, the controller remains bounded by policy constraints, in which strong intervention is restricted under high uncertainty unless risk is sufficiently high, costly actions are avoided under low risk, and conflicting evidence favors information-gathering actions. The controller can therefore be summarized as:

$$a_t = C(Z_t, \hat{Z}_{t+1:t+k}, u_t, m_t, \mathcal{P}) \quad (12)$$

where Z_t is the current latent risk, $\hat{Z}_{t+1:t+k}$ is the predicted future risk trajectory, u_t is uncertainty, m_t is the accumulated memory of previous actions and outcomes, $\mathcal{P}$ denotes policy constraints, and a_t is the selected preservation action.

8.2. Illustrative Outputs

The generated outputs are shown in Figures 1–5. Figure 1 shows the synthetic evidence streams and the resulting late-fused risk signal for a Baltanás-like underground heritage structure. The fused signal follows the broad structure of the sensor and satellite signals while smoothing modality-specific fluctuations. Figure 2 shows the estimated ABC-style latent risk, the seven-day predicted risk under no action, the predicted risk under the selected action, and uncertainty. The difference between the no-action and selected-action trajectories illustrates the action-conditioned character of the model. Figure 3 shows the controller decisions over time for this site, while Figure 4 shows the evolution of the controller’s learned action preferences. Finally, Figure 5 summarizes the distribution of controller actions across all five simulated pilot sites.

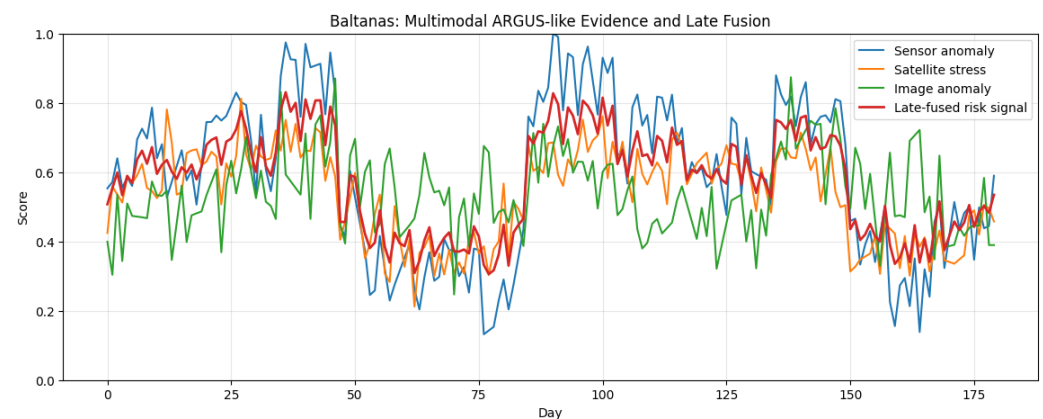


Figure 1. Synthetic ARGUS-like multimodal evidence and late-fused risk signal for a Baltanás-like site. Sensor anomaly, satellite-derived stress, image anomaly, and site vulnerability are combined into a single fused preservation-risk signal.

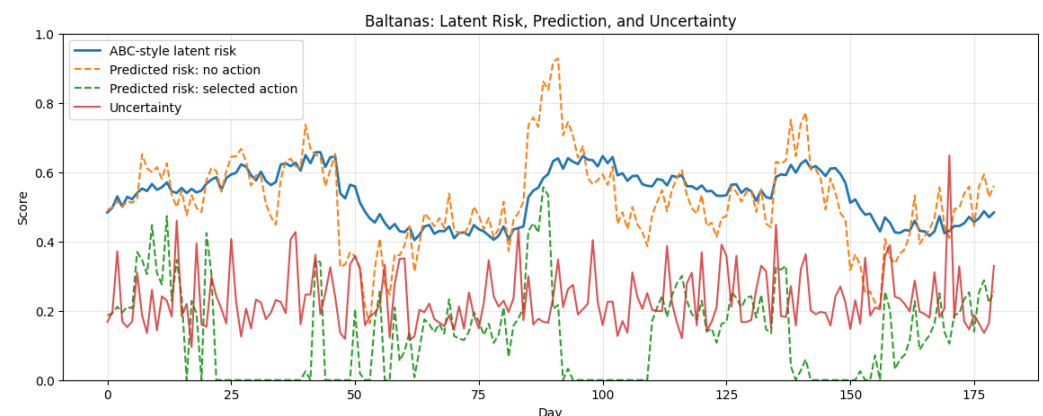


Figure 2. ABC-style latent risk, seven-day predicted risk under no action, predicted risk under the controller-selected action, and uncertainty for a Baltanás-like site. The separation between no-action and selected-action trajectories illustrates action-conditioned prediction.

The same controller-action distribution is also reported numerically in Table 3. In this particular illustrative run, the controller selected *request inspection* most frequently, followed by *recommend preventive intervention*, while *increase sampling* was selected less often. No *continue monitoring* actions were selected in this parameterization, because the simulated

risks remained mostly in a moderate-to-high range and the controller was initialized with conservative preservation preferences.

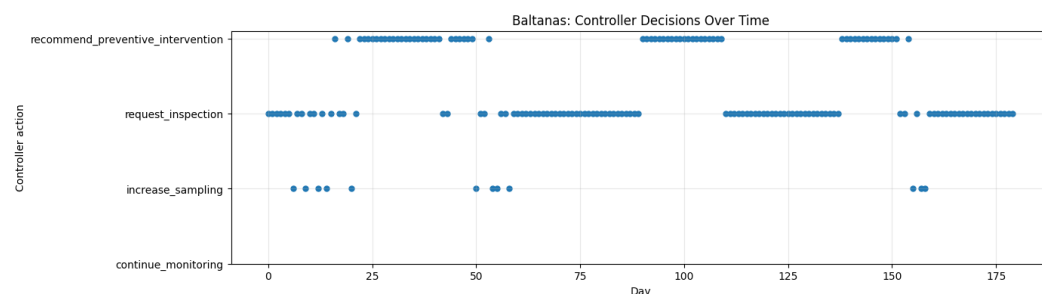


Figure 3. Controller decisions over time for a Baltanás-like site. The controller behaves conservatively in this synthetic run, relying mostly on inspection and preventive intervention, with increased sampling used in selected uncertainty-sensitive cases.

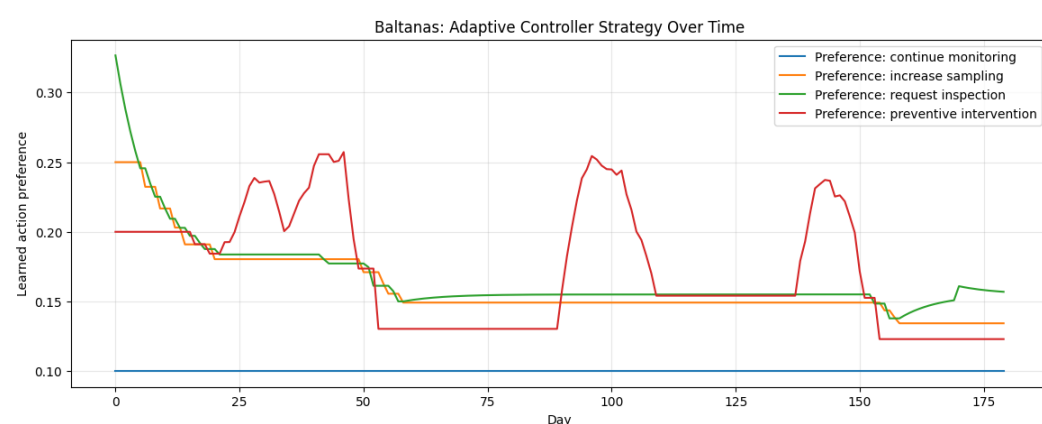


Figure 4. Adaptive controller strategy over time for a Baltanás-like site. The learned action preferences vary as simulated outcomes accumulate, illustrating how the controller can be modeled as more than a fixed threshold rule, but an adaptive policy-constrained decision layer.

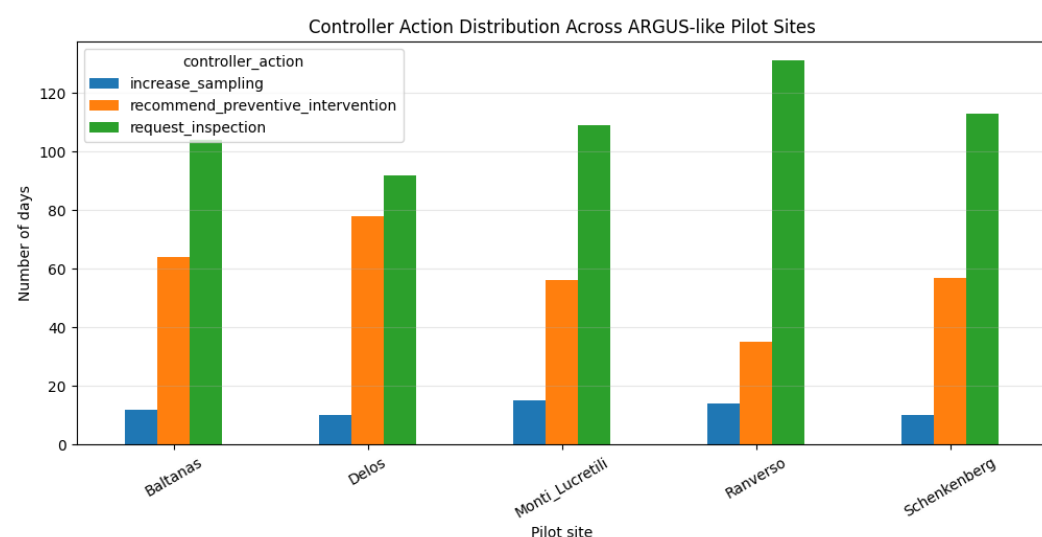


Figure 5. Controller-action distribution across the five synthetic ARGUS-like pilot sites over 180 simulated days. The distribution reflects the conservative parameterization used in this illustrative run, with inspection being the dominant recommendation.

The decision trace further illustrates the internal logic of the controller. For example, in the initial Baltanás-like site sequence, the ABC-style risk lies between approximately 0.48 and 0.57, uncertainty remains generally moderate, and sensor-satellite agreement is high.

Under these conditions, the controller mostly recommends inspection, with occasional increased sampling or preventive intervention. A compact excerpt is shown in Table 4.

Table 3. Controller-action distribution across the five synthetic ARGUS-like pilot sites over 180 simulated days.

Site	Sampling	Intervention	Inspection
Baltanás-like	12	64	104
Delos-like	10	78	92
Monti Lucretili-like	15	56	109
Ranverso-like	14	35	131
Schenkenberg-like	10	57	113

Table 4. Excerpt from the simulated decision trace for a Baltanás-like site. Values are generated from synthetic data and are reported only to illustrate the controller logic.

Day	Risk	Uncertainty	Agreement	Action	Predicted Risk
0	0.484	0.168	0.871	inspection	0.186
1	0.497	0.195	0.987	inspection	0.194
2	0.530	0.372	0.895	inspection	0.212
6	0.540	0.310	0.897	sampling	0.370
16	0.540	0.220	0.922	intervention	0.000
19	0.547	0.160	1.000	intervention	0.000

8.3. Scope and Limitations of the Simulation

The simulation should be interpreted qualitatively and illustratively. It uses synthetic time series, manually assigned site vulnerability values, heuristic fusion weights, simplified ABC-style risk proxies, assumed action effects, and a conservative controller parameterization. No real sensor streams, inspection records, satellite products, conservation logs, or site-management decisions are used. Therefore, the numerical values and action distributions reported in Tables 3 and 4 should not be interpreted as indicative of real ARGUS-like site behavior.

The purpose of the simulation is methodological rather than evidential. It shows how the architectural components proposed in this position paper can be connected in a transparent computational loop, where multimodal evidence is fused, latent preservation risk is estimated, short-horizon futures are compared under different actions, policy constraints are applied, and an auditable decision trace is produced. It does not demonstrate that the proposed controller outperforms threshold alerting, expert rules, fixed inspection schedules, no-action monitoring, or other decision-support baselines.

The present parameterization is intentionally conservative and should not be read as an empirically optimized preservation policy. In particular, the absence of passive monitoring decisions and the near-complete risk reduction predicted for some preventive-intervention cases indicate that future versions should calibrate action effects, costs, lower risk bounds, fusion weights, ABC-style proxies, and policy thresholds using real pilot-site evidence and expert feedback.

Despite these limitations, the simulation illustrates the operational distinction between monitoring-oriented digital twins and the proposed heritage world model abstraction. Rather than only reporting observations or anomaly scores, the simulated architecture estimates latent risk, predicts short-horizon futures under alternative actions, applies policy constraints, and records auditable recommendations. The simulation is therefore intended only as an illustrative operationalization of the proposed closed-loop architecture, not as a validated conservation or controller-performance study.

9. Research Agenda

Moving from semantic digital twins toward the heritage world model abstraction proposed here requires progress along several directions, building on semantic digital twin research, agentic AI, risk-aware preservation, and reflexive ontology engineering [5,6,17,21]. However, these directions should not be treated as equally immediate. A practical roadmap should distinguish near-term components that can be implemented within existing infrastructures, medium-term components that require calibration and accumulated evidence, and longer-term components that require broader community validation.

9.1. Near-Term Stage: Semantic Ingestion, Uncertainty Annotation, and Auditable Traces

The first step is to make existing semantic digital twins more explicit as *state-updating systems*. The proposed abstraction requires methods for updating semantic state from heterogeneous observations. Sensor measurements, satellite data, inspection reports, photographs, historical documentation, and expert annotations must be integrated into temporally grounded state representations. This requires not only data ingestion but interpretation, deciding what observations mean for the current state of the heritage asset and its preservation environment. At this stage, *uncertainty* should also be represented explicitly. Heritage data are often incomplete, irregular, noisy, spatially heterogeneous, and institutionally fragmented. A heritage world model should represent not only what is known, but also what is unknown, uncertain, contested, or poorly observed. Uncertainty should inform both recommendations and future monitoring priorities. In practical terms, this means recording missing observations, sensor reliability, spatial resolution, temporal latency, model confidence, disagreement between modalities, and expert uncertainty as part of the semantic and provenance layer. The same near-term stage should establish *auditable decision traces*. Every recommendation should be traceable to evidence, previous events, risk assessments, policies, and assumptions. Memory is not only a historical archive, it is a reasoning resource. It enables comparison between episodes, evaluation of interventions, and justification of decisions. For ARGUS-like deployments, this stage is feasible through semantic ingestion pipelines, explicit uncertainty annotations, provenance-preserving model executions, and decision logs connected to the active Asset State.

9.2. Medium-Term Stage: Latent Risk Modeling, Action-Conditioned Prediction, and Controller Evaluation

The second step is to move from enriched semantic state representations toward *predictive and intervention-aware reasoning*. Many preservation-relevant variables are not directly observed. Vulnerability, deterioration, material fatigue, biological activity, visitor-induced stress, and policy urgency are inferred constructs. Heritage world models need explicit mechanisms for representing such latent states, linking them to observations, and updating them over time. Decision support also requires more than forecasting what will happen. It requires forecasting what may happen under different actions. This is the key difference between trend analysis and intervention-aware modeling. Heritage world models should compare plausible futures under alternative conservation strategies, including passive continuation, increased monitoring, inspection, access-management changes, preventive intervention, or expert escalation. A heritage world model becomes operational only when coupled to a *controller*. Future research should therefore investigate how preservation controllers can be designed, evaluated, and governed. Unlike controllers in robotics or industrial automation, heritage controllers must operate under cultural, ethical, institutional, and epistemic constraints. Their outputs may include physical interventions, monitoring actions, expert-review requests, alerts, reports, or stakeholder communications. This requires hybrid architectures combining symbolic rules, risk thresholds, learned predictions,

optimization, argumentation, expert validation, and human approval. Evaluation should not focus only on whether the controller selects an apparently optimal action, but also on whether the recommendation is explainable, reversible, proportionate, policy-compliant, and acceptable to heritage professionals. *Evaluation* should proceed at several levels. Recent world-model research increasingly recognizes that evaluation cannot be reduced to visual fidelity or short-horizon prediction accuracy [48]. For heritage world models, evaluation must be broader. A predicted trajectory should be assessed not only for numerical accuracy, but also for semantic validity, causal plausibility, policy relevance, uncertainty calibration, and usefulness to heritage professionals. A practical evaluation protocol should include at least four levels. First, predictive evaluation should compare forecasted risk trajectories against historical measurements, inspection reports, or simulated ground truth. Suitable metrics include prediction error, calibration error, time-to-detection, false alert rate, and missed-risk rate. Second, decision evaluation should compare the controller against reactive baselines, such as threshold-only alerting, no-action monitoring, traditional expert rules, or fixed inspection schedules. Metrics may include expected risk reduction over a time horizon, intervention timeliness, number of unnecessary interventions, and robustness under missing data. Third, explanation evaluation should assess whether experts judge the system's recommendation trace as complete, understandable, and policy-compliant. Fourth, governance evaluation should verify that each recommendation preserves provenance, records the applied policy, supports human override, and remains auditable. Because real heritage failures are rare, heterogeneous, and ethically difficult to reproduce, evaluation may combine retrospective site logs, synthetic stress scenarios, expert-in-the-loop assessment, and controlled simulation. For ARGUS-like deployments, a realistic medium-term step would be offline policy evaluation on historical or accumulated pilot-site episodes, where alternative controllers can be replayed against recorded sensor, image, satellite, and inspection evidence without affecting the real heritage asset. This mixed strategy is appropriate for an agenda-setting paper: it does not claim full validation, but defines how future heritage world models can be compared against reactive digital twins and conventional decision-support workflows.

9.3. Longer-Term Stage: Reflexive Semantic Evolution, Cross-Site Transfer, and Community Benchmarks

The third step concerns capabilities that go *beyond a single project or pilot deployment*. Heritage meanings, categories, and priorities may change over time. A mature heritage world model should be able to accommodate not only new data but also shifts in interpretive frameworks. The EANO direction suggests that *reflexive digital twins* may eventually update not only instance values but also semantic roles, relations, and classifications under controlled and accountable conditions [21]. This requires explicit governance, because reflexivity should not become uncontrolled semantic drift. *Cross-site transfer* is another longer-term challenge. A model calibrated for an underground cellar town, an island archaeological landscape, a historic abbey, or an exposed castle ruin cannot be transferred naively to another setting. Future heritage world models will need mechanisms for distinguishing reusable semantic structures from site-specific risk dynamics, sensor configurations, policy thresholds, and conservation priorities. This also implies the need for community benchmarks, shared competency questions, reusable preservation scenarios, and reference datasets that allow different HWM implementations to be compared.

9.4. Governance, Auditability, and Human Oversight Across All Stages

Because heritage world models may influence preservation decisions, their recommendations must be governed through explicit human oversight mechanisms. Each controller output should be accompanied by an auditable trace linking the active Asset State,

observations, model execution, prediction, uncertainty estimate, policy constraint, and recommended action. Human experts should be able to approve, reject, override, or defer a recommendation, and such responses should themselves be recorded as part of the system memory. Under high uncertainty or conflicting evidence, conservative policies should be preferred, including requests for further observation rather than irreversible intervention. The reflexive ontology direction introduces an additional governance challenge. If semantic roles, categories, or relations evolve over time, versioning becomes essential. Ontology evolution should therefore preserve longitudinal comparability through explicit version identifiers, mappings between schema versions, change justifications, and validation by domain experts. Reflexivity should not mean uncontrolled semantic drift. It should mean accountable semantic adaptation under documented evidence and institutional review. More broadly, *heritage world models must remain accountable* to conservators, heritage managers, communities, and institutions. The goal is not autonomous replacement of expert judgment, but structured augmentation, consistent with the need for accountable, interpretable, and value-sensitive cultural AI [17,20,21]. A system may predict, compare, and justify, but final preservation decisions often involve ethical, cultural, legal, and political dimensions that must remain under human governance.

10. Discussion

The proposed heritage world model clarifies the broader computational architecture required when cultural heritage digital twins are expected to support prediction, intervention evaluation, uncertainty-aware reasoning, and accountable preservation decisions [3–5,15]. The contribution of the paper is therefore not to reject digital twins, but to make explicit the architectural commitments implied by decision-support-oriented preservation systems.

Within this progression, the controller turns prediction into accountable decision support. It structures the space of possible responses by asking which action is justified by the evidence, which action would reduce uncertainty, which intervention is proportionate to the predicted risk, and which recommendation complies with conservation policy. This is where a digital twin infrastructure becomes part of an operationally decision-supportive heritage world model. Not only by displaying, integrating, or forecasting aspects of the preservation world, but by helping human actors decide how to act within it.

There are risks in adopting the term, especially because “world model” is increasingly overloaded across AI, robotics, generative simulation, and agentic systems [52,53]. It could become fashionable vocabulary detached from implementation. It could be misunderstood as implying a large neural simulator of cultural heritage. It could also overstate the maturity of current systems. For this reason, the term must be defined carefully. A heritage world model is not simply a digital twin renamed. It is a broader closed-loop architecture built over a semantic digital twin representation and made explicit around state estimation, latent risk modeling, action-conditioned prediction, memory, uncertainty, and policy-aware reasoning.

The concept also creates an important bridge between cultural heritage and contemporary AI [7,9,17]. Instead of treating AI merely as a tool for classification, segmentation, reconstruction, or chatbot interaction, heritage world models place AI inside a broader semantic and decision-oriented architecture. This is especially important for cultural heritage, where black-box predictions are insufficient and where provenance, semantic interoperability, and cultural accountability remain central requirements [17,20,24,39]. The model must be interpretable, accountable, provenance-aware, and aligned with cultural values.

Several limitations should be acknowledged. First, the proposed formulation remains architectural and does not validate a deployed implementation of the proposed heritage world model architecture in an operational site. Second, the moisture-driven instantia-

tion and the ARGUS-like operational simulations are illustrative rather than empirically calibrated. Third, the mapping to PANOPTES identifies semantic compatibility and architectural potential, but additional schema-level and implementation work is required to represent uncertainty, confidence, disagreement, controller outputs, spatial uncertainty, ontology versioning, and performance constraints in a fully operational implementation. Fourth, the paper does not claim comparative superiority over advanced digital twins, threshold-based alerting, expert-rule systems, or other decision-support approaches. These limitations are consistent with the paper's position-paper role, but they define the immediate next steps for implementation, benchmarking, and evaluation.

11. Conclusions

This paper introduced the concept of the *heritage world model* as an architectural abstraction for building closed-loop, decision-support-oriented preservation systems over cultural heritage digital twins. The central argument is not that digital twins are inherently passive, nor that prediction or recommendation functions are incompatible with the digital twin paradigm. Rather, the argument is that when a digital twin is expected to anticipate deterioration, interpret observations, reason over risk, compare interventions, and justify recommendations, it implicitly requires a model of the preservation world and its possible futures.

The paper defined a heritage world model as a structured, temporally updated, semantically grounded, and action-aware model of a heritage asset and its preservation environment, capable of integrating observations, estimating latent risk states, predicting plausible future trajectories, and evaluating interventions under uncertainty. This definition adapts the world-model concept from contemporary AI while respecting the interpretability, provenance, and value-sensitive requirements of cultural heritage.

PANOPTES was used as a worked semantic basis for this direction [13–15,17,21]. In this framing, ontology-based semantic structures can provide the state space, predictive models can estimate possible futures, and a human-supervised controller can link those futures to accountable preservation recommendations. PANOPTES defines many of the entities required for such a semantic state space, while its agentic and reflexive extensions introduce risk memory, policy responsiveness, and controlled semantic evolution [17,21]. Reinterpreted through the heritage world model lens, PANOPTES is therefore not presented as a complete HWM implementation, but as one possible semantic basis for predictive, memory-bearing, and intervention-aware preservation intelligence.

The proposed direction should be understood as a research agenda rather than a completed implementation. Future work should formalize uncertainty, strengthen the geospatial and multi-scale remote-sensing dimensions, implement action-conditioned predictive models, calibrate risk and action-effect assumptions from real pilot evidence, compare controllers against baseline decision-support approaches, and evaluate how heritage professionals interact with such systems in practice. The long-term goal is not to replace expert judgment, but to make preservation reasoning more anticipatory, explainable, auditable, and operationally grounded.

The resulting vision is therefore not merely a smarter digital twin, but a closed-loop heritage intelligence system, in which semantic representation, world modeling, and human-supervised control jointly support preventive preservation.

Author Contributions: Conceptualization, G.P.; methodology, G.P., V.A. and V.S.; formal analysis, V.A., V.S. and G.P.; investigation, V.A. and V.S.; resources, G.P.; writing—original draft preparation, G.P.; writing—review and editing, G.P., V.S. and V.A.; visualization, V.A.; supervision, G.P.; project administration, G.P.; funding acquisition, G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union’s Horizon Europe programme under grant agreement No 101132308 (ARGUS—Non-destructive, scalable, smart monitoring of remote cultural treasures).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new empirical datasets were generated or analyzed in this study. The numerical results reported in the illustrative operational simulation are based on synthetic data generated solely for explanatory purposes.

Acknowledgments: AI-assisted language refinement tools (OpenAI; model: GPT-5.5 Thinking) were used during manuscript preparation. All scientific content and conclusions remain the responsibility of the authors.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Notes

- ¹ ARGUS—Non-destructive, scalable, smart monitoring of remote cultural treasures, Grant agreement ID 101132308, is an EU project on preventive preservation of remote built heritage. Find out more at <https://cordis.europa.eu/project/id/101132308> and <https://www.argus-project.eu> (accessed on 29 April 2026).
- ² A graphical representation of the original PANOPTES can be found at <https://webvowl.athenarc.gr> (accessed on 29 April 2026).

References

1. Grieves, M.; Vickers, J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems*; Springer: Cham, Switzerland, 2017; pp. 85–113. [\[CrossRef\]](#)
2. Grieves, M. Intelligent Digital Twins and the Development and Management of Complex Systems. *Digit. Twin* **2023**, *1*, 8. [\[CrossRef\]](#)
3. Jouan, P.; Hallot, P. Digital Twin: Research Framework to Support Preventive Conservation Policies. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 228. [\[CrossRef\]](#)
4. Niccolucci, F.; Markhoff, B.; Theodoridou, M.; Felicetti, A.; Hermon, S. The Heritage Digital Twin: A Bicycle Made for Two. The Integration of Digital Methodologies into Cultural Heritage Research. *Open Res. Eur.* **2023**, *3*, 64. [\[CrossRef\]](#)
5. Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. [\[CrossRef\]](#)
6. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2025**, *10*, 1. [\[CrossRef\]](#)
7. Ha, D.; Schmidhuber, J. World Models. *arXiv* **2018**, arXiv:1803.10122. [\[CrossRef\]](#)
8. Ha, D.; Schmidhuber, J. Recurrent World Models Facilitate Policy Evolution. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31, pp. 2451–2463.
9. LeCun, Y. A Path Towards Autonomous Machine Intelligence. *Open Rev.* **2022**, *62*, 1–62.
10. Hafner, D.; Lillicrap, T.; Ba, J.; Norouzi, M. Dream to Control: Learning Behaviors by Latent Imagination. In *Proceedings of the International Conference on Learning Representations*, Addis Ababa, Ethiopia, 26–30 April 2020.
11. Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; et al. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *Nature* **2020**, *588*, 604–609. [\[CrossRef\]](#)
12. Zheng, J.; Niyato, D.; Zhao, C.; Kang, J.; Wang, J. From Digital Twins to World Models: Opportunities, Challenges, and Applications for Mobile Edge General Intelligence. *arXiv* **2026**, arXiv:2603.17420. [\[CrossRef\]](#)
13. Pavlidis, G.; Sevetlidis, V.; Arampatzakis, V. PANOPTES: A Digital Twin Ontology for Cultural Asset Management. In *Proceedings of the 2025 IEEE International Conference on Cyber Humanities, Workshop on Protecting the Future of Cultural Heritage: Secure Digital Twins for Sustainable Preservation*; IEEE: New York, NY, USA, 2025. [\[CrossRef\]](#)
14. Pavlidis, G.; Sevetlidis, V.; Arampatzakis, V. Defining a New Digital Twin Ontology for Cultural Heritage Preservation—The Case of ARGUS. In *Proceedings of the Digital Heritage 2025*; Eurographics Association: Eindhoven, The Netherlands, 2025. [\[CrossRef\]](#)
15. Pavlidis, G.; Koutsoudis, A.; Tsiafaki, D.; Karta, M.; Sevetlidis, V.; Arampatzakis, V.; Sarris, A.; Polidorou, M.; Klinkenberg, V.; Boukhers, Z.; et al. Future-Proofing Heritage with ARGUS: A Multimodal Digital Twin Approach for Sustainable Preservation. In *Proceedings of the 52nd Computer Applications and Quantitative Methods in Archaeology Conference, CAA2025 Proceedings*, Athens, Greece, 5–9 May 2025. [\[CrossRef\]](#)

16. Michalski, S.; Pedersoli, J.L. *The ABC Method: A Risk Management Approach to the Preservation of Cultural Heritage*; Canadian Conservation Institute and ICCROM: Ottawa, ON, Canada, 2016.
17. Pavlidis, G. Agentic AI for Cultural Heritage: Embedding Risk Memory in Semantic Digital Twins. *Computers* **2025**, *14*, 266. [CrossRef]
18. Latour, B. *Reassembling the Social: An Introduction to Actor-Network-Theory*; Oxford University Press: Oxford, UK, 2005.
19. Callon, M. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *Sociol. Rev.* **1984**, *32*, 196–233. [CrossRef]
20. Smith, L. *Uses of Heritage*; Routledge: Abingdon, UK, 2006.
21. Pavlidis, G.; Arampatzakis, V.; Sevetlidis, V.; Koutsoudis, A.; Arnaoutoglou, F.; Ioannakis, G.A.; Kiourt, C. Towards Evolving Actor–Network Ontologies: Enabling Reflexive Digital Twins for Cultural Heritage. *Information* **2025**, *16*, 892. [CrossRef]
22. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [CrossRef]
23. Karabulut, E.; Su, J.; Zia, H.; McMeekin, D.A.; Gonzalez, L.F.; Khoshelham, K. Ontologies in Digital Twins: A Systematic Literature Review. *Future Gener. Comput. Syst.* **2024**, *153*, 442–456. [CrossRef]
24. Doerr, M. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Mag.* **2003**, *24*, 75–92. [CrossRef]
25. ISO 21127:2023; Information and Documentation—A Reference Ontology for the Interchange of Cultural Heritage Information, Edition 3. ISO: Geneva, Switzerland, 2023.
26. Waller, R. Conservation Risk Assessment: A Strategy for Managing Resources for Preventive Conservation. *Stud. Conserv.* **1994**, *39*, 12–16. [CrossRef]
27. Silva, H.E.; Henriques, F.M.A. Preventive Conservation of Historic Buildings in Temperate Climates. The Importance of a Risk-Based Analysis on the Decision-Making Process. *Energy Build.* **2015**, *107*, 26–36. [CrossRef]
28. Huerto-Cardenas, H.E.; Leonforte, F.; Aste, N.; Del Pero, C.; Evola, G.; Costanzo, V.; Lucchi, E. Validation of Dynamic Hygrothermal Simulation Models for Historical Buildings: State of the Art, Research Challenges and Recommendations. *Build. Environ.* **2020**, *180*, 107081. [CrossRef]
29. Dias Pereira, L.; Saraiva, N.B.; Soares, N. Hygrothermal Behavior of Cultural Heritage Buildings and Climate Change: Status and Main Challenges. *Appl. Sci.* **2023**, *13*, 3445. [CrossRef]
30. Penjor, T.; Banihashemi, S.; Hajirasouli, A.; Golzad, H. Heritage building information modeling (HBIM) for heritage conservation: Framework of challenges, gaps, and existing limitations of HBIM. *Digit. Appl. Archaeol. Cult. Herit.* **2024**, *35*, e00366. [CrossRef]
31. Taylor, J.; Blades, N.; Cassar, M. Dependency Modelling for Cultural Heritage. In *Proceedings of Safeguarded Cultural Heritage—Understanding & Viability for the Enlarged Europe*; Dradcky, M., Chapuis, M., Eds.; Institute of Theoretical and Applied Mechanics, Academy of Sciences of Czech Republic: Prague, Czech Republic, 2006; Volume 6, pp. 51–59.
32. Fu, L.; Pan, J.; Tang, Y.; Zhang, Q. A multi-criteria decision-making approach to assessing the health risk of rockeries in Chinese classical gardens. *Herit. Sci.* **2023**, *11*, 92. [CrossRef]
33. Olimpio, L.C.M.; Campos, V.R.; Mesquita, E.F.T. Multicriteria Decision-Making to Support Preventive Maintenance of Built Heritage in the Historic Center of Sobral, Brazil. *J. Cult. Herit. Manag. Sustain. Dev.* **2023**, *13*, 814–836. [CrossRef]
34. Maselli, G.; Cucco, P.; Nesticò, A.; Ribera, F. Historical heritage—MultiCriteria Decision Method (H-MCDM) to prioritize intervention strategies for the adaptive reuse of valuable architectural assets. *MethodsX* **2024**, *12*, 102487. [CrossRef]
35. Amelung, B.; Student, J.; Nicholls, S.; Lamers, M.; Baggio, R.; Boavida-Portugal, I.; Johnson, P.; de Jong, E.; Hofstede, G.J.; Pons, M.; et al. The value of agent-based modelling for assessing tourism–environment interactions in the Anthropocene. *Curr. Opin. Environ. Sustain.* **2016**, *23*, 46–53. [CrossRef]
36. Wallinger, S.; Arnberger, A.; Brandenburg, C. Agent-Based Modelling for Sustainable Tourism. In *Handbook of e-Tourism*; Springer: Cham, Switzerland, 2023. [CrossRef]
37. Bai, N.; Nourian, P.; Xie, A.; Pereira Roders, A. Towards a Finer Heritage Management: Evaluating the Tourism Carrying Capacity Using an Agent-Based Model. In *Proceedings of the RE: Anthropocene, Design in the Age of Humans—Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA 2020)*; The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA): Hsinchu City, Taiwan, 2020; Volume 1, pp. 305–314. [CrossRef]
38. Haller, A.; Janowicz, K.; Cox, S.; Phuoc, D.; Taylor, K.; Lefrancois, M. (Eds.) *Semantic Sensor Network Ontology*; W3C: Wakefield, MA, USA, 2017. Available online: <https://www.w3.org/TR/vocab-ssn/> (accessed on 29 April 2026).
39. Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; Zhao, J. *PROV-O: The PROV Ontology*; W3C: Wakefield, MA, USA, 2013. Available online: <https://www.w3.org/TR/prov-o/> (accessed on 29 April 2026).
40. Open Geospatial Consortium. OGC GeoSPARQL: A Geographic Query Language for RDF Data. 2024. Available online: <https://www.ogc.org/standards/geosparql/> (accessed on 29 April 2026).

41. Cox, S.; Little, C. *Time Ontology in OWL*; W3C: Wakefield, MA, USA, 2017. Available online: <https://www.w3.org/TR/owl-time/> (accessed on 29 April 2026). .
42. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
43. Hafner, D.; Pasukonis, J.; Ba, J.; Lillicrap, T. Mastering diverse control tasks through world models. *Nature* **2024**, *640*, 647–653. [[CrossRef](#)]
44. Assran, M.; Bardes, A.; Fan, D.; Garrido, Q.; Howes, R.; Komeili, M.; Muckley, M.; Rizvi, A.; Roberts, C.; Sinha, K.; et al. V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. *arXiv* **2025**, arXiv:2506.09985. [[CrossRef](#)]
45. Terver, B.; Yang, T.Y.; Ponce, J.; Bardes, A.; LeCun, Y. What Drives Success in Physical Planning with Joint-Embedding Predictive World Models? *arXiv* **2025**, arXiv:2512.24497. [[CrossRef](#)]
46. Maes, L.; Le Lidec, Q.; Scieur, D.; LeCun, Y.; Balestrierio, R. LeWorldModel: Stable End-to-End Joint-Embedding Predictive Architecture from Pixels. *arXiv* **2026**, arXiv:2603.19312. [[CrossRef](#)]
47. Mur-Labadia, L.; Muckley, M.; Bar, A.; Assran, M.; Sinha, K.; Rabbat, M.; LeCun, Y.; Ballas, N.; Bardes, A. V-JEPA 2.1: Unlocking Dense Features in Video Self-Supervised Learning. *arXiv* **2026**, arXiv:2603.14482. [[CrossRef](#)]
48. Hendriksen, M.; Rashid, T.; Bignell, D.; Georgescu, R.; Lemkhenter, A.; Hofmann, K.; Devlin, S.; Parisot, S. Adapting Vision-Language Models for Evaluating World Models. *arXiv* **2025**, arXiv:2506.17967. [[CrossRef](#)]
49. Kaelbling, L.P.; Littman, M.L.; Cassandra, A.R. Planning and Acting in Partially Observable Stochastic Domains. *Artif. Intell.* **1998**, *101*, 99–134. [[CrossRef](#)]
50. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.
51. Lattimore, T.; Szepesvári, C. *Bandit Algorithms*; Cambridge University Press: Cambridge, UK, 2020.
52. Li, K.; Viégas, F.; Wattenberg, M. What Does it Mean for a Neural Network to Learn a “World Model”? *arXiv* **2025**, arXiv:2507.21513. [[CrossRef](#)]
53. Parmar, M. Safety, Security, and Cognitive Risks in World Models. *arXiv* **2026**, arXiv:2604.01346. [[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.