



# The Heritage Digital Twin

Beyond cultural heritage digitalization  
and why the HDT is perfect for it



## Table of contents

1. Niccolucci, Franco, Achille Felicetti, and Sorin Hermon. 2023. 'Populating the Data Space for Cultural Heritage with Heritage Digital Twins'. *Data* 7 (8): 105. <https://doi.org/10.3390/data7080105>.
2. Niccolucci, Franco, Beatrice Markhoff, Maria Theodoridou, Achille Felicetti, and Sorin Hermon. 2023. 'The Heritage Digital Twin: a bicycle made for two. The integration of digital methodologies into cultural heritage research'. arXiv, <https://doi.org/10.48550/arXiv.2302.07138>.
3. Felicetti, Achille, and Franco Niccolucci. 2024. 'Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data'. *Data* 10 (1): 1. <https://doi.org/10.3390/data10010001>.
4. Niccolucci, Franco, and Achille Felicetti. 2024. 'Il Futuro digitale di 4CH: il gemello digitale dei Beni Culturali'. *Digitalia* 19 (2): 117–25. <https://doi.org/10.36181/digitalia-00105>.
5. Niccolucci, Franco, and Achille Felicetti. 2024. 'Digital Twin Sensors in Cultural Heritage Ontology Applications'. *Sensors* 24 (12): 3978. <https://doi.org/10.3390/s24123978>.
6. Niccolucci, Franco, and Achille Felicetti. 2025. 'Digital Twins and the Stendhal Syndrome'. *Computers* 14 (4): 136. <https://doi.org/10.3390/computers14040136>.
7. Felicetti, Achille, Aida Himmiche, and Miriana Somenzi. 2025. 'Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins'. *Appl. Sci.* 2025, 15(18), 10061; <https://doi.org/10.3390/app151810061>.





## Article

---

# Populating the Data Space for Cultural Heritage with Heritage Digital Twins

---

Franco Niccolucci, Achille Felicetti and Sorin Hermon

## Special Issue

A European Approach to the Establishment of Data Spaces

Edited by

Dr. Alexander Kotsev, Dr. Marco Minghini, Dr. Massimo Craglia, Dr. Stefano Nativi and  
Dr. Carlos Granell



## Article

# Populating the Data Space for Cultural Heritage with Heritage Digital Twins

Franco Niccolucci <sup>1,\*</sup> , Achille Felicetti <sup>1</sup>  and Sorin Hermon <sup>2</sup><sup>1</sup> PIN—University Center City of Prato, Piazza Ciardi 25, 59100 Prato, Italy; achille.felicetti@gmail.com<sup>2</sup> STARC—The Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, 20 Konstantinou Kavafi Street, Aglantzia 2121, Cyprus; s.hermon@cyi.ac.cy

\* Correspondence: franco.niccolucci@gmail.com

**Abstract:** The present paper concerns the design of the semantic infrastructure of the data space for cultural heritage as envisaged by the European Commission in its recent documents. Due to the complexity of the cultural heritage data and of their intrinsic inter-relationships, it is necessary to introduce a novel ontology, yet compliant with existing standards and interoperable with previous platforms used in this context as Europeana. The data space organization must be tailored to the methods and the theory of cultural heritage, briefly summarized in the introduction. The new ontology is based on the Digital Twin concept, i.e., the digital counterpart of cultural heritage assets incorporating all the digital information pertaining to them. This creates a Knowledge Base on the cultural heritage data space. The paper outlines the main features of the proposed Heritage Digital Twin ontology and provides some examples of its application. Future work will include completing the ontology in all its details and testing it in other real cases and with the various sectors of the cultural heritage community.



**Citation:** Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. <https://doi.org/10.3390/data7080105>

Academic Editors: Alexander Kotsev, Marco Minghini, Massimo Craglia, Stefano Nativi and Carlos Granell

Received: 30 January 2022

Accepted: 26 July 2022

Published: 29 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cultural heritage data space; digital twin; cultural heritage semantics

## 1. Introduction

With its recent *Recommendation of 10/11/21 on a common European data space for cultural heritage* [1] the EU Commission has set the foundations and defined the features of a framework supporting and hosting the digital transformation of cultural heritage, i.e., the process and the results of leveraging digital technologies to transform how heritage organisations operate and deliver value. This section will introduce the general concept of cultural heritage to explain what the objective of the digital transformation should be, and which is the perspective and the methodology of the heritage community. Other relevant EU policy documents include [2–4].

The concept of cultural heritage and of its safeguard, conservation, and valorisation, has evolved in time since the 18th century. We may date back to that period the creation of the present most important museums (Uffizi, 1737; Hermitage, 1764; British Museum, 1753; Louvre, 1793); the first large-scale archaeological excavations (first systematic excavation of Pompeii, 1749); and, in general, the interest in antiquities by antiquarians (English Society of Antiquarians chartered in 1780). The “Family Pact” by the last member of the Medici family and Electress of the Palatinate Anna Maria Luisa de’ Medici, i.e., her testament signed in 1737 and eventually leading to the creation of the Uffizi Gallery, stated that all heritage assets in Florence should remain there as “ornament of the state, for the utility of the public and to attract foreigners’ curiosity”. This is perhaps the first definition of the mission of cultural heritage institutions.

Since then, an innumerable set of statements, discussion, papers, and books has been dedicated to this topic. It has been the subject of international charters and has informed heritage practice and methodology. Hence, it would be impossible to address here such

theoretical background in detail. However, it has been clearly outlined in the EU report *Innovation in Cultural Heritage Research* [4], which describes the so-called cultural *regimes* and the progress of the notion of cultural heritage. The following list is derived from this report.

- The *First Regime* (circa 1860–1960s) is determined by national and local heritage conservation regulations and it lasts until the codification of international cultural heritage protection. It is mainly based on national initiatives, with international charters starting to appear in its final stage. Actions address the physical conservation and enhancement of monuments, sometimes allowing substantial modification as in Viollet-Le-Duc ‘reconstructions’, later deprecated in the Venice Charter (1964) for philological fidelity to the original; and the exhibition of the atemporal beauty of art masterpieces. The public is mainly formed by ‘visitors’, guided by the experts, archaeologists, museum curators and other heritage professionals.
- The *Second Regime* (1960s–1990s) corresponds to the institutionalisation of cultural heritage as an international value. UNESCO and its related institutions (ICOM, ICOMOS) are the chief actors. Diverse communities with their cultures can express and see their own testimonies recognized, including intangible ones. Deeper understanding of heritage by visitors is fostered with various means. The role of heritage in understanding past and present societies is enhanced.
- The *Third Regime* (late 1990s to present) corresponds to the renewed institutionalisation of cultural heritage characterised by its expansion in terms of concepts, significance, number and diversity of heritage assets and elements. A significant feature is the involvement of communities and their contribution to the definition and construction of cultural heritage, sometimes described as co-creation.

The development of digital technologies in the present century is producing deep effects on our lifestyle, on our work and on our communication attitude. Nevertheless, in most cultural heritage-related activities the digital component has not been interiorized by institutions and professionals, using it as an appliance to perform the same tasks and to achieve the same goals as before. It is now high time to fully incorporate the digital component into heritage methods and practice, moving the concept of heritage towards Heritage 4.0, i.e., the *Fourth Regime*, characterized by the digital transformation effects on cultural heritage activities, institutions and people, both professionals and citizens.

The digital transformation is not simply the deployment of advanced digital technologies to aid traditional practices and business models. Careful consideration needs to be given to how the traditional core of the organizations can benefit of digitization and explore and capture new ways of creating value. This new approach will progressively add to and partially replace the traditional one, without superseding the values and goals of heritage institutions, summarized in the current ICOM definition of a museum [5], (Art. 3 Section 1) as “a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment”. This definition dates back to 1974 and was updated several times, the last one in 2017. At the 2019 Tokyo ICOM Extraordinary General Assembly a new one has been presented, but it raised a still unresolved controversy and is currently suspended. Our approach seems to fit also with the discussed new definition, which stresses the social values of cultural heritage.

Since 2017, when this definition was updated, social networks have become commonplace, and the digital component is an integral part of our daily life. Digital identities—legal and functional/social ones—are replacing real ones under many regards. Communication is social, quick, often superficial, and broadcast to many. This impacts both on the traditional definition of museums, here used as representatives of the whole heritage sector, and on the one newly proposed including multivocality and social aspects in the museum mission. People increasingly speak digital. Talking and listening to them needs to establish novel communication channels, in which Creative Industries play a substantial role together with

‘transformed’ heritage professionals and institutions. Moreover, the nowadays available advanced digital technologies for cultural heritage offer incredible opportunities for fulfilling other aspects of heritage institutions mission as, among others, conservation. They need to become part of the heritage methodology, on a par with more traditional ones.

To be up to this new role, digital technologies must support access to accumulated knowledge and be capable of improving their contribution to cultural heritage. Technologies are of no value by themselves, even if advanced: it is when they are used to help making processes related to heritage knowledge, conservation, and communication more effective, and to deliver a better experience, that real value is created.

The digital evolution requires tools to operate on digital artefacts, which form and populate the data space of cultural heritage, conceived as a meta-space where digital artefacts duplicate real ones; and through services, commanded by humans or activated by some kind of artificial intelligence (AI), which reproduces heritage processes as curation, conservation and preservation.

The present paper intends to contribute to the definition of such digital artefacts, to outline how they may be (digitally) manipulated and propose some exemplary applications among the many that the move to digital will enable. To define these artefacts, it is necessary to be aware that heritage comes in very different shapes. They range from immovable property, as historic buildings, sites, landscapes, environments; to movable ones, as works of art or technology and science. They also include immaterial resources as music, performances, traditions, practices, representations, expressions, knowledge, skills—as well as the instruments, objects, and cultural spaces associated with them and recognized as such by a community or group of people. These aspects are intertwined with each other, linking for example a building to its content and its history. To denote each heritage item, we will use the term *heritage asset*, which may refer to any kind of heritage, either tangible, both movable and immovable, or intangible.

Digital models of heritage assets must therefore be able to encompass all such diversity and incorporate the relationships among the various components of the assets they represent. They must also be able to react to external impulses, reproducing the effects these would have on their real counterpart. This digital framework is already in use in several manufacturing sectors [6]. News in the press report that every Tesla car has a digital duplicate—with which it continuously dialogues—checking the proper functionality of the vehicle and its parts. Bioinformatics allows us to describe, study and make in silico experiments on complex objects as proteins and viruses. The design and engineering of spacecrafts are digitally tested in simulations on digital models: actually, this was the first domain in which this approach was extensively introduced by NASA in 2010, being often impossible to reproduce the conditions of space travels. The digital model used in all such cases is called a *digital twin*, a virtual representation that serves as the real-time digital counterpart of a physical object. The digital twin concept has made its way also in the official documents of the European Commission, for example in the Commission’s Communication *Shaping Europe’s digital future* [7], which envisages a digital twin of the Earth. The digital twin is also quoted as a substantial component of the Data Space for Smart Communities initiative within the Digital Europe programme.

The same approach is being applied in a domain closer to cultural heritage, the building industry, incorporating in the design of buildings not only the geometry of the object being modelled, but also the information about its physical and functional characteristics, e.g., the material; services as electricity, plumbing and so on; and the construction process, management and control. This methodology is called Building information Modelling (BIM). In BIM applications, information is attached on a 3D model. The digital representation of architectural, structural and all other elements is based on an international metadata schema, an ISO standard: ISO (2018) 16739-1:2018 [8], called Industry Foundation Classes (IFC). The IFC schema is a standardised data model that codifies the identity (name, machine-readable unique identifier, object type or function), the characteristics or attributes (as material, colour, and physical properties) and relationships

(including locations and connections) of objects (as columns or slabs), abstract concepts (performance, costing), processes (installation, operations), and people (owners, designers, contractors, suppliers, etc.). With this approach, a 3D CAD model becomes the pivot for searching and the preferential entry point for accessing the knowledge about an asset. Among others, the Centre for Digital Built Britain (CDBB), a partnership between the Department for Business, Energy & Industrial Strategy (BEIS) of the UK Government and the University of Cambridge, has started the application of digital twins for the built environment in UK on behalf of the UK government, publishing in a report [9] a summary of the principles a digital twin must comply with, which they call the *Gemini Principles*.

In the literature there has been a very active thread proposing to use BIM as the entry point to access heritage documentation and as the main constituent of the digital twin, renaming this technology as Heritage BIM (HBIM). Google Scholar lists about 1400 papers on the subject in 2020 and 2021 only, ranging from the illustration of individual case studies to more general ones. Many different proposals have been advanced to enrich the foundation classes of an HBIM model to incorporate the necessary information concerning shape and materials. All such proposals share a common approach derived from the architectural/civil engineering origin of the model: they are all based on ‘enriched’ 3D models. With this approach, however, the field of application of digital twins is restricted to immovable physical assets and even for them it does not consider all the features that are relevant in a cultural heritage context, for example the intangible value of a monument. The next sections propose an improved approach which addresses these limitations. To the best of our knowledge, the approach we propose is completely new in the domain of cultural heritage documentation. It is completely innovative as it does not consider the graphical representation of heritage assets as the root of related information, to which any other data must be appended. Instead, it introduces a holistic approach to digital documentation where all the components have the same relevance and may be used as access points and as links to other digital objects, including for example intangible components that may not have a visual representation.

## 2. Results

### 2.1. The Definition of the Heritage Digital Twin

The new model proposed in this paper builds on the existing digital twin concept to extend it to the whole domain of cultural heritage, in its diverse facets of immovable, movable and immaterial/intangible aspects. It reformulates what seems to be a partially distorted perspective totally relying on the object shape and architectural characteristics, derived from the building applications developed so far. The current model is usually regarded as a very rich BIM model, making it unsuitable for movable heritage, which has a material component without the features considered in BIM models, and especially for intangible or born-digital heritage, which lack a physical component at all. Instead, in our approach, we will develop a semantic model of the digital twin where the knowledge about each asset is organised in a knowledge graph, altogether being regarded as the digital twin of the heritage asset. This leads to the following definition:

The *Heritage Digital Twin (HDT)* of a *Heritage Asset (HA)* is the digital representation of the complex of knowledge about that heritage asset, organised according to a specific ontology, called the *HDT ontology*.

The HDT includes, but is not limited to, 3D models of the asset. Thus, visual models as point-cloud ones or HBIM models may be important components of a Heritage Digital Twin but have no privileged status compared to other attributes. Their role ranges from primary (but not exclusive) for architectural assets, to modest or non-existing for intangible assets. From the user’s perspective, the information contained in the HDT ontology can be mapped to the IFC schema and produce a HBIM view on HDT via software, in real time or generating permanent HBIM models. The other way round is also feasible, by transforming the HBIM data to HDT content via the inverse mapping of the above-mentioned one. Since all the IFC classes and properties will have corresponding HDT classes and properties,



both transformations are lossless. Thus, the HDT of a heritage asset may be considered an extension of the HBIM model, which maps one-to-one to a part of it: but it is indeed much more.

On the other hand, the HDT approach is not only a clever way of organizing information for retrieval, allowing one to access data for consultation by a human agent from any point of the knowledge graph that documents the heritage asset. Information may also be used for direct machine processing by means of simulation models integrated with the knowledge base. Thus, messages from sensors or from external big data systems may trigger automated actions. For example, as explained in greater detail in Section 3, a weather threat alert coming from a (Big Data) meteo forecast system connected with the heritage knowledge base will launch a software service that retrieves and analyses the heritage assets located in the threatened area and identifies the ones at flood risk, for example comparing the height of their location with the expected level of the flood water. This system may activate automated actions, for example raising barriers, and alert human actors to intervene to mitigate the adverse effects of this natural disaster, according to the different nature of the assets and of their content as resulting from the collection of all the relevant data provided in HDT, for example, moving paintings upstairs.

## 2.2. Introducing the Heritage Digital Twin Ontology

The following description of classes and properties of the HDT ontology aims at being readable also by non-technical but digitally aware heritage professionals and researchers; it is also intended to give a preliminary insight into the ontology. A more technical description is provided in Section 3, including a summary table. The ontology is described through the explanation of (some of) its classes, i.e., the concepts it organizes by grouping similar instances, i.e., individual occurrences; and by illustrating properties, i.e., the relationships between these concepts. To visually identify them, HDT ontology class and property names will be indicated in *italic*, with class names capitalized. The inverse property of a property will be parenthesized in the definition: thus, the property name *is part of (has part)* means that statements such as “X *is part of* Y” and “Y *has part* X” have the same meaning and are the same property, only swapping the subject with the predicate.

Notice that the same words used for classes and properties may be used in the discourse: in such cases they will not be italicized. As customary in semantics, class names are used in the singular only; when used as plain words, names may obviously be in the singular or plural as required by the context.

All HDT classes are considered as subclasses of an overarching one, *Heritage Entity*, which has no instances. The pivot concepts are the class *Heritage Asset*, corresponding to actual heritage assets of any nature (physical, both movable or immovable, immaterial or born digital), and the class *Heritage Digital Twin*, to indicate the whole of the digital information pertaining to *Heritage Asset*. The *Heritage Digital Twin* is related to the *Heritage Asset* via the property *is digital twin of (has digital twin)*. We will start describing the hierarchical structure of *Heritage Asset* and then introduce the one for *Heritage Digital Twin*.

Any instance of the *Heritage Asset* class may be related to other instances of the same class, for example being a part of them, as the tower of a castle, or forming with them a collection or a more general asset, as the paintings of an art gallery. Another recent example is the new UNESCO World Heritage site “The Great Spa towns of Europe” [10], which is made of eleven towns, each one developed around a spa, far from each other but belonging to a common cultural framework. Each one will have its own *Heritage Asset* instance, and there will also be a collective *Heritage Asset* instance representing the whole UNESCO World Heritage site.

Information about the parts forming an asset is relevant also to the whole, as it describes important details as, for example, the material of which the part is made. Thus, parts of a heritage asset may be considered as assets as well, and correspondingly the HDT of the whole will incorporate the HDTs of all the parts of the entire heritage asset which are identified as heritage asset on their own. For example, a church, i.e., the ‘main’ heritage

asset, has many parts as chapels, paintings, architectural components, furniture, and so on, each one a heritage asset on its own and with an own heritage digital twin. The heritage digital twin of the church is then the assemblage of all these heritage digital twins, plus features concerning the whole, for example the style, the cult, the architect and so on. HAs being parts of an HA are related to the whole by the property *is part of (has part)*, which produces a similar property *is digital part of (has digital part)* linking the corresponding HDTs to the HDT of the whole. A more generic connection than part-whole is described instead by the property *is related to*, relating heritage assets to other assets. Being symmetric, this property coincides with its inverse.

The instances of all classes *have* an *identifier*, i.e., a code attached to them, for example an inventory number; and may *have* one (or more) *Title*, often according to language, e.g., “Mona Lisa” (English), “Monna Lisa” or “La Gioconda” (Italian), “La Joconde” (French) etc., all referring to the same famous Leonardo’s painting. Thus, the properties *is identified by (identifies)* and *is titled (titles)* have *Heritage Entity* as domain, since identifiers may be defined also for digital twins, e.g., their URL, as well as names if one likes to do so for digital artefacts.

The *Heritage Document* class includes all the documentation items pertaining to a *Heritage Asset*. The documentation may consist in digital or analogic objects as printed or handwritten documents, old analogic photos, drawings, and so on; or digital, either born digital like digital photos or digitised from analogic one.

The property linking *Heritage Asset* to *Heritage Document* is the property *is documented in (documents)*, which may apply to the whole asset or to specific parts of it.

Among others, we may distinguish among the following ones:

- *3D model*, resulting from any of the various techniques available as 3D scanning, wire-frame modelling and so on. Practice will assess if it is more convenient to distinguish among them introducing different types, e.g., 3D point-cloud model, 3D CAD model and so on.
- *Imagery*, as photos and videos, but also special imagery as X-ray images, spectra of chemical and physical analyses, and so on. Among them, particularly relevant are the Virtual Reality (VR) and Augmented Reality (AR) models, other types of visual digital artefacts pertaining to *Heritage Asset*. Both VR and AR models rely on 3D models of the related heritage asset, but may add or remove parts of it, or require further digital input as in AR, so they should be catalogued separately from 3D models. 3D models may correspond to actual objects–artefacts or built structures–or to conceptual ones, often representing the reconstruction of what is presumed to be the original configuration (and often, the use) of the reconstructed object. Similar models are named Virtual Reconstructions and are a commonplace in archaeology to communicate to researchers or, more frequently, to the public, the interpretation of past appearance. In general, Virtual Reality (VR) models enable virtual visits, while models incorporating the present appearance of heritage assets. i.e., Augmented Reality (AR) models can be viewed only on site, as they need the real time acquisition of the current asset appearance.

More types/subclasses may be introduced according to needs. For example, it may be worth considering a *Conservation Document*, data about the asset conservation both in terms of the documentation of past interventions, the materials, and the analyses carried out on it. It consists in text files, numeric files (e.g., the results of analyses), images, videos, and special 3D objects, for example the results of tomography. In general, different types of models may generate a subclass if they have special properties that apply only to them. Otherwise, it is simpler to consider their *Type* only, associated via the property *has type (is type of)*.

A *Heritage Asset* pertaining to a tangible asset *is located in (is location of)* a *Place*. The location may have different levels of precision, as a determinate position or area, a town, a region, and so on, for example “Room 1 of the Uffizi Museum”, “Athens”, “Cyprus”. It may also vary in time, if the related asset is moved elsewhere. A concept similar to

location may be considered also for some intangible assets, as is done for some members of the UNESCO Intangible Heritage Representative List [11]. For others, it is intrinsic to the intangible asset to be on the move or to have no location, for example music. Thus, the (important) location property domain does not coincide with all the *Heritage Asset* class and is somehow different between tangible and intangible heritage: for the former it defines where the asset is located, while for the latter it indicates where the asset manifests itself. Therefore, the *Heritage Asset* class must split into the two subclasses *Tangible Asset* and *Intangible Asset*. For the former, *is located in (is location of)* specifies where the asset is placed. For the latter, the sister property is *happened at location (was location for)*. Both have *Heritage Location* as the range. Defining the location of a tangible asset may have various degrees of difficulty: if it is easy for most of them, in some cases it relies on research as there is no physical evidence confirming the location. This is the case of battlefields, for instance, when no finds or traces exists. An example is the Cannae site, where the battle (216 BC) between Romans and Carthaginians took place during the Second Punic War. Historical reports by Polybius and Titus Livius are available and enable a trustworthy identification of the location.

A *Heritage Asset* has many a *Heritage Story* that are associated to it. A *Story* includes any kind of witness related to the asset: it can be a narrative, a historical source, a popular attribution, co-created content and so on. A very special case concerns a literary itinerary as Leopold Bloom's route through Dublin in James Joyce's *Ulysses*, or—in a light-hearted perspective—real or almost imaginary places featured in popular novels or TV series, for example crime series as George Simenon's *Commissaire Maigret*, in Paris; or Andrea Camilleri's *Inspector Montalbano*, based in southern Sicily; both generating a substantial flow of 'cultural' tourism. In such cases, it is the *Heritage Story* that creates the *Heritage Asset*. Similar 'fake' cultural heritage is discussed in detail in [12].

In general, a *Heritage Story* relates tangible heritage assets to their intangible components and to their reference communities. They are therefore of paramount importance also for the asset physical conservation and the safeguard of its intangible value. Before starting a conservation intervention or the evaluation of an activity from the heritage conservation perspective, it is necessary to consider the impact on the intangible component especially when the impact on the tangible one is irrelevant. For example, fast food shops are often banned from historic centres, although their visual impact may be negligible; there are, instead, provisions to preserve the permanence of historic shops. Locating a McDonald's in the basement of a historic palace would be unthinkable even if the building statics is unaffected. Ignoring these intangible aspects as it happens in BIM and HBIM models is a serious shortcoming for any heritage application of these approaches and one of the main reasons to propose a generalisation as the HDT ontology. A very nice example concerns Orthodox sacred icons. The devotion to a particular icon is manifested by lighting candles in front of it, which in time causes the blackening of the painting. Thus, the icon blackening level is the evidence of the believers' devotion and a confirmation of its religious value. Therefore, cleaning it would damage this intangible value: a common and much valued restoration practice for paintings would have an unexpected adverse effect if this intangible component is not considered. There are of course stories about intangible heritage as well, often in a much greater amount than for the tangible one.

To clarify the difference between a *Heritage Document* and a *Heritage Story*, let us give a more precise explanation of both.

A *Heritage Document* consists of data produced by a dedicated activity, finalised to produce knowledge about one or more heritage assets. Such production activities include research, management, conservation and restoration, and display of the assets. In some cases, these activities have also an interest independent from the heritage they address, for example as regards the technology or the materials used.

Documents span over many types of data. They include, among others: research reports and data; all kinds of imagery, from black and white photos to 3D models, drawings and maps; videos and other recordings, for example interviews with artists and architects as



well as with people and communities. They also include historical reports and descriptions, both ancient and modern, e.g., from Vasari's "*Vite*" (*The Lives of the Most Excellent Painters, Sculptors, and Architects*, 3rd edition 1568) to a recent paper on *Art History*.

The activity of producing a *Heritage Document* is called *Heritage Documenting*. It is characterised by the intentionality of creating new knowledge. We deliberately avoid using the term "documentation" in the class names as it has an ambiguous meaning, the activity and its result: "Documentation (i.e., the activity) produces documentation (the outcomes)". The term can be disambiguated only by the context, obviously unavailable in an isolated name.

On the other hand, a *Heritage Story* is the account of facts about a *Heritage Asset*, including but not limited to descriptions based on documents and on the interpretation of these documents. They are usually formulated in an attractive and accessible way to facilitate visitors' understanding and stimulate their curiosity. Frequently they avail of communication techniques as drawings, physical or digital reconstructions, and increasingly use VR and AR technology.

The level of factualness of a story may vary from interpretation based on research and told as a story, to a somehow arbitrary reconstruction, and to a mostly imaginary legend. For example, the historical existence of King Arthur is debated among historians, probably with no conclusive fact on his real existence due to the lack of historical documents and the doubts existing on the value of mediaeval texts. Conclusions on this regard are disputed as it is widely (but not unanimously) considered as legendary by scholars. Locating Arthur's birthplace and possibly his royal court at Tintagel Castle in Cornwall is nowadays refused by historians. Nevertheless, stories continue to be told about these early mediaeval ruins excavated in the 20th century, probably as they attract tourists more fascinated by Arthur's legend and the link to the Arthurian cycle of literary legends concerning him, his wife Guinevere, his sword Excalibur and the Knights of the Round Table, rather than by factual accounts about a sixth century small castle and trading point, of which very little remains. The activity of creating and disseminating a story is known as storytelling and thus we will name the class including these activities with the same name, *Storytelling*.

In conclusion, stories are an integral component of the heritage framework as they may make it more understandable, interesting, and attractive for the public. They are distinct from documents, as they do not create—nor aim to create—new knowledge on the assets they concern but rather organise it into a coherent account mixing research results with interpretation based on research at a variable degree. Instead, a *Heritage Document* is aimed at creating new knowledge and is made using a scholarly methodology based on factual evidence, deduction and inference, regardless of the correctness of the premises and the conclusions. In the case of intangible heritage, the border between a *Story* and a *Document* is blurred because the story may be part of the asset value. In this case, we would distinguish between them according to the intentionality of the production activity—research for *Documenting* and explanation, valorisation, communication for *Storytelling*. Regarding to the cultural heritage data space as a source of content for creative industries, stories with their visual apparatus are perhaps the most valuable information concerning heritage. Creative Industries also produce new stories, which might be linked to the heritage assets concerned.

The association between a heritage asset and its heritage digital twin induces properties relating the latter to the various attributes of the former: digital objects, e.g., a 3D model, related to a heritage asset, are also components of the digital twin of that asset. For example, the property *has visual documentation* (*is visual documentation of*) for *Heritage Asset* establishes a property *includes visual documentation* (*visual documentation is included in*) for the related *Heritage Digital Twin*; *is documented in* (*documents*) for *Heritage Asset* establishes *includes document* (*document is included in*) for its twin; and so on. Therefore, induced properties are shortcuts, i.e., they are equivalent to the combination of *is digital twin of* and the corresponding heritage asset property: for instance *Heritage Digital Twin includes document* is a shortcut for *Heritage Digital Twin is digital twin of Heritage Asset (which) is documented*.

in *Heritage Document*. Since the heritage digital twin is the complex of all information, all the twin properties are subproperties of a superproperty *has as component (is component of)* having *Heritage Digital Twin* as a domain.

To conclude this discursive introduction to the Heritage Digital Twin ontology, we notice that the level of ontology definition is still initial and needs further work. Nevertheless, the approach seems promising as it takes into account aspects relevant for cultural heritage data which are instead ignored by other models. It supports the construction of a data space more functional than relying on other simpler knowledge organization schemes that overlook several important features. A preliminary assessment of the HDT ontology, summarized below, shows that it is suitable for heritage assets as resulting from the above summary description of its classes and properties.

### 2.3. The Semantics of the HDT Ontology

This section will introduce the HDT ontology. Classes will be indicated with a label formed by HC (for heritage class) followed by a progressive number accompanying the class name. e.g., *HC1 Heritage Entity*. Likewise, properties will be denoted by HP and a number, preceding the property name. As before, class and property names will be italicized, and class names also capitalized.

We note that the HDT ontology is a compatible extension—more precisely, an application profile—of the CRM (CRM stands for Conceptual Reference Model. It designates the semantic model, an ontology, developed and maintained by a team from CIDOC, the documentation committee of ICOM, the International Council of Museums; equivalently, it may be indicated as CIDOC CRM. The CRM allows for extensions and specializations, usually indicated appending as a suffix to the CRM acronym a few letters related to the extension purpose, for example CRMdig that stands for the CRM extension for digital objects.) conceptual reference model [13] for heritage documentation, an ISO 21127:2014 standard. As such, it makes the cultural heritage data space interoperable with thousands of repositories of heritage documentation which adopt CRM-compliant data models, from which information can be directly retrieved. Moreover, with this approach, the HDT ontology is developed to be compatible with EDM, the Europeana Data Model [14]—actually to be an extension of it to accommodate with a broader use—thus enabling the direct transfer of information to and from the current Europeana system. Several classes and properties of the HDT ontology are indeed equivalent to CRM ones. Referred CRM classes and properties will be denoted by the prefix *crm:*, which applies also to other CRM extensions, as CRMdig, the CRM extension for digital objects; CRMarchaeo, the extension to archaeological investigations; CRMba, the one for building archaeology; CRMsci, the one for heritage science, and so on. These extensions are characterized by different encoding for the labels, which for instance include E and P for the base CRM, D and L for CRMdig, and so on: thus, class and property labels directly identify the relevant extension they belong to. Reference to these extensions is available on the already mentioned CIDOC CRM site [13]. We also notice that the definition of the HDT ontology builds on pre-existing work that has already developed, or is currently developing, parts re-usable for it, mainly in the ARIADNE [15] and in the 4CH [16] EU projects.

The ontology structure is synthesized in the following two diagrams (Figures 1 and 2), illustrating the relationships among classes and properties. A more formal—but still preliminary—list of the HDT essential classes and properties, with synthetic descriptions is provided in the Appendix A. They are intended to give a more technically-oriented idea of the HDT ontology, to just illustrate the data space structure. The Appendix A may be used for reference if need be.

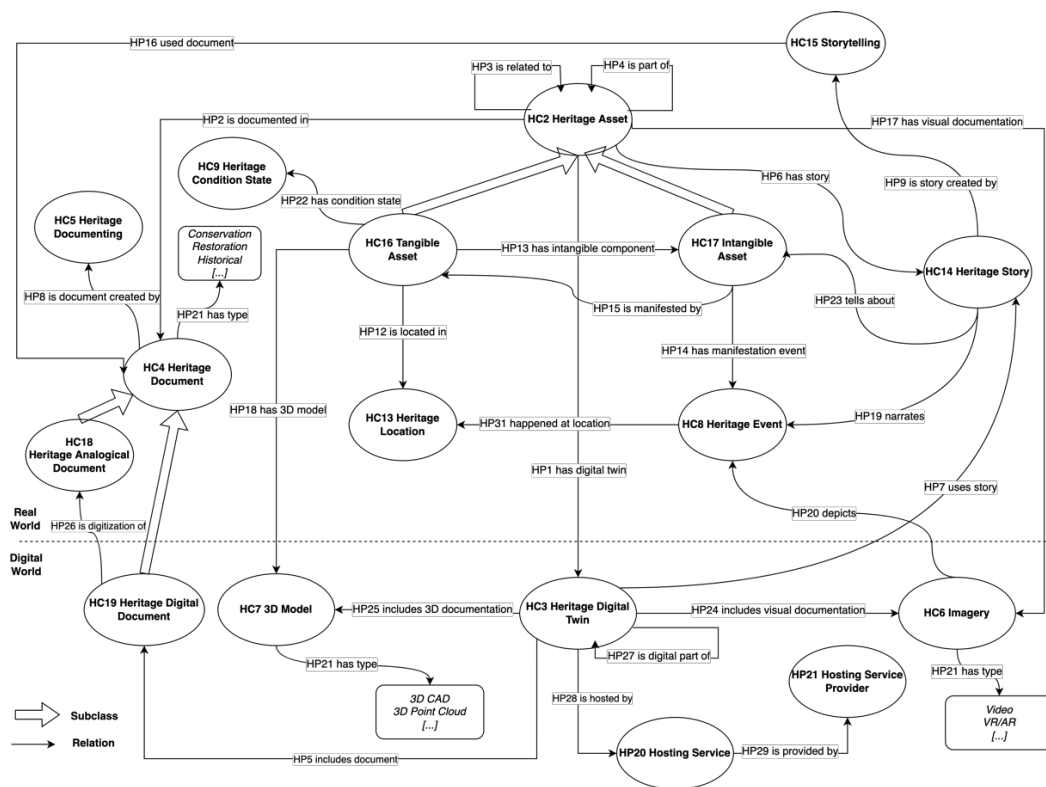


Figure 1. The semantic graph of the HDT ontology.

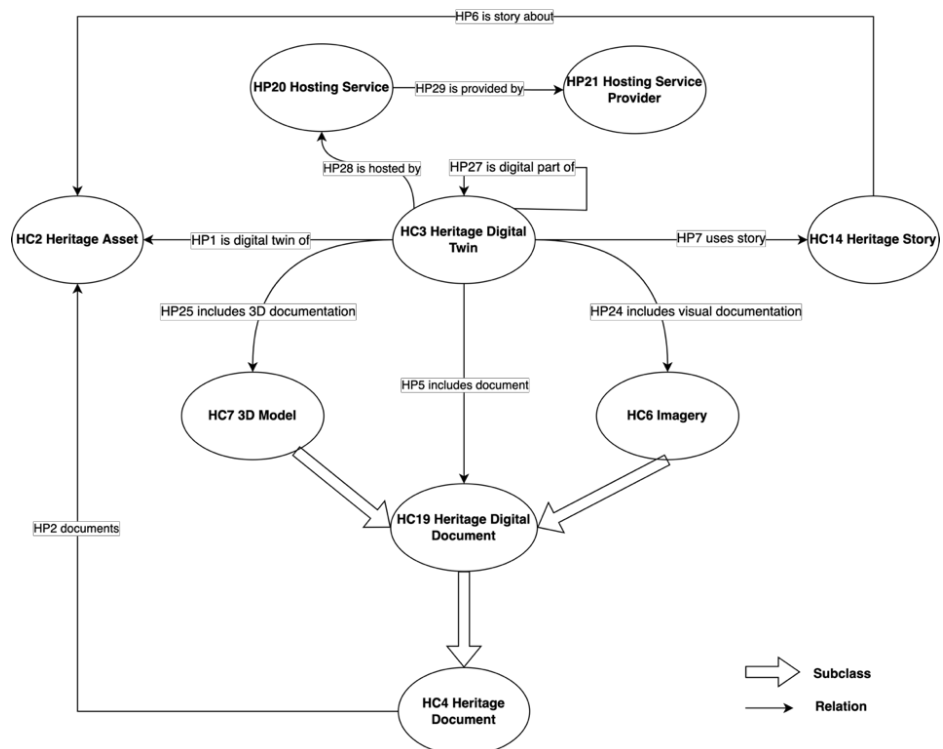


Figure 2. The semantic graph of HC19 Heritage Digital Twin.

The following diagram (Figure 1) shows the structure of the HDT ontology with its most important classes and properties.

The structure of HC3 Heritage Digital Twin, a pivot concept for the HDT ontology, and its related classes are outlined in Figure 2.

### 3. Discussion

#### 3.1. Suitability of the HDT Ontology

The distinction used below for different aspects of cultural heritage by no means implies a strict classification of assets as belonging to any of the categories listed below. It rather analyses different facets of heritage, which may be present at the same time in the asset.

- Monuments appear to be well described by the HDT ontology.
- Sites are more difficult to deal with because of the vague definition of a “site”. It may be an archaeological site, where remnants of the past were discovered and, in many cases, there are still standing structures. For archaeological sites, the corresponding digital twin must incorporate reference to all the discoveries made there, as they make the place a “site”. It must also be capable of dealing with standing structures and their individual digital twins. Notable examples are Pompeii, where both assets, i.e., finds and standing structures, are present, and Stonehenge, where only the standing structures remain. Complex sites will be better represented by a global digital twin being the collection of the digital twins of the relevant parts of the site. In other cases, no structure survives, e.g., in very ancient sites, and only finds may contribute to the heritage digital twin. Thus, the site heritage digital twin consists of all relevant information as resulting from site investigations, and of the digital twins of surviving structures. In this case VR and AR models, for example virtual reconstructions, may store in a concise way plenty of information, besides being a help for visitors and an enhancement of the tourist attraction of the site. In this case, the models must comply with the requirements of the London Charter [17], establishing principles for the computer visualisation of cultural heritage. Other examples of sites, potentially even more complicated, are historic centres. They include individual monuments but the characteristic of being a site stays with the complex, with its own digital twin, and often incorporates typical activities. Natural and cultural landscapes mix heritage and natural features which must be separately documented. An example from the UNESCO WH list [10], the landscape of Val d’Orcia in Tuscany, Italy, combines anthropogenic features with a particular natural and cultivated environment. Finally, other heritage sites are such only because of their history, as in the already mentioned case of battlefields that have neither remains nor finds.
- For movable artefacts, 3D models have a different significance depending on the nature of the object and from the way the multidimensional visual data are produced. 3D point-cloud models may be used to represent statues and other human-made objects, but for intrinsically 2-dimensional objects such as paintings or photos 2D imaging is in general better suited. 3D in the usual meaning of the term is in this case significant only in some special applications, for example to appreciate the paint technique in Pollock’s *Alchemy* [18], or the deformation of the surface of Leonardo da Vinci’s *The Adoration of the Magi* [19]. Other multidimensional imagery comes, among others, from scientific analyses as tomography, layered non-invasive analyses, and so on. The resulting images supports interpretation and conservation activities. In some cases, AR/VR can help as well, when for example reconstructing the shape of an ancient musical instrument, accompanied with the sound it is supposed to have produced.
- Intangible heritage (and the intangible component of every heritage asset) is the most demanding component in terms of the HDT ontology development. As outlined above, it mainly relies on stories, and on the documentation of practices and individual or collective beliefs and traditions. To set up all the required classes and properties, we will start from what is already available here, and improve the fit by analysing and assessing the results of previous applications to intangible heritage.
- Born-digital heritage will have its own specifications, also part of the HDT ontology. The heritage digital twin will coincide in some parts with the asset, adding additional information to it.

As recommended in the 2021 Commission's Recommendation [1], special attention must be paid to so-called "High Value Datasets" (HVDS), which include the results of scientific analyses on artefacts for conservation or to improve the knowledge about the artefact under study, for example to date it. The semantic organization of such datasets has been studied in [20–23], and is compliant with the HDT ontology, into which it may be included.

In conclusion, the HDT ontology appears as a suitable foundation to build a knowledge base of cultural heritage, the main constituent of the heritage data space.

To incorporate legacy data in the knowledge base, mappings to the HDT ontology need to be produced as necessary. A possibility to be explored concerns the automatic production of parts of the HDT by means of artificial intelligence. This enrichment includes, for example, text automatic annotation using Natural Language Processing (NLP), to recognize named entities in an unstructured text. An example of an application of Named Entity Recognition (NER) is described in [23], where NER is applied to archaeological reports and to the documentation of scientific analyses. Another example discusses automatic segmentation of a 3D model to recognize its parts [24].

The next section outlines how to build a data space on the HDT ontology and exemplifies some services operating on it.

### 3.2. Using the Heritage Digital Twin: The Data Space for Cultural Heritage as a Knowledge Graph

The HDT ontology enables the creation of a heritage Knowledge Base (KB), where information about heritage is organized in a knowledge graph, which can be managed and searched using a NoSQL DBMS, for example Ontotext GraphDB [25], which has already been applied in the cultural heritage domain. The Research Space project at the British Museum [26] is based on OntoText.

For instance, the Museo del Prado has developed a CIDOC CRM-based semantic graph for its collection [27]; the Rijksmuseum collections have been organized as Linked Open Data [28]; and a semantic graph has been used for the creation of virtual exhibitions [29]. The main operation enabled by these DBs on the graph is searching the knowledge base and surfing it by means of Linked Data, where the links are represented by the common content for a class, i.e., by the same value for instances in that class and by properties relating classes with each other. With this approach, the heritage data space is not just a plain repository where cultural heritage data are stored and managed by a traditional database management system. Instead, it consists of the abstractions—the Heritage Digital Twins—of actual heritage assets, which model the real assets by selecting relevant features at the desired (or available) level of information granularity. In this regard, a compromise is reached between the detail, consisting of the raw data, and the abstraction, represented by the synthesis at ontology class level. For example, the material of a bronze object may be used as such, i.e., "bronze", to relate the object to other bronze objects; it may take into account the alloy composition, i.e., the percentage of copper, tin and other metals; in some cases it may refer to further details e.g., to classify the object provenance; finally, in very rare cases, the raw data of the analyses are used for linking and comparisons. All the required data are however stored in the heritage data space.

In conclusion, our vision of the data space for cultural heritage consists in a federation of (raw) data repositories distributed in distinct digital locations, possibly located in distinct geographical locations; and in a centralised system, the Knowledge Base (KB), which aggregates and manages synthetic information organised according to the HDT ontology approach. The KB content incorporates the metadata of the aggregated data, but also includes the relationships among them. This distributes the raw data storage load among repositories, operating mainly on the data synthesis stored in the KB by accessing raw data only when required.

The KB may be centralised in the same geographical place or split in distributed parts, all interconnected to make up the same knowledge base. The balance between centralization and decentralisation depends on the expected access and use: since cultural

heritage is a location-based commodity, it may be worthwhile to evaluate the expected processing at the edges of the system, i.e., a use restricted to local data in a prevalent way, to optimise the efficiency of the global system through a balanced distribution of the data and the setup of edge computing processes. In any case, the federal approach for repositories also satisfies an aspect much appreciated by heritage institutions, i.e., the feeling that their data are kept under their control.

### 3.3. Using the Heritage Digital Twin: Automatic Processing in the Data Space

The heritage knowledge base is intended to be used by humans searching for information, but it is also the base for machine-actioned activities, e.g., triggered by sensors, which may send to a specific digital twin a continuous flow of signals or react to events or conditions requiring an intervention.

Under this regard, a digital twin is not a static concept, updated from time to time; it is instead a living replica of the real asset and is able to react autonomously, if necessary, according to specific conditions. Thus, sensors of humidity, for example, may trigger the system to limit the number of accesses to a monument, and an inexplicable strong increase in temperature may trigger a fire alarm and alert the fire brigade. Sensor data may also be stored for time-series analysis. Such processes require the use of AI systems, monitoring the HDT status to trigger actions on the real asset, warn custodians to act, or just record the event for future use. Thus, advanced digital technologies provide data to the cultural heritage data space, send it messages, and receive messages triggering for actions. The extent and effect of this reaction depend on how the simulation model analyses and reacts to external inputs.

As an example of automatic reaction, we can consider the following one.

The HDT system is connected to a system evaluating the risk of flood in some regions, based on meteo forecasts and a hydrogeological digital model. When a risk occurs, an alert activates a service, an automatic analyser searching for heritage asset locations placed in the risk region and comparing the expected flood height with asset basement height, calculated from the location and from the asset's 3D models. For those at risk of flooding, an analysis of the nature of the involved assets can indicate the most appropriate emergency measures, for example moving movable assets (e.g., paintings) at a higher position, putting barriers to water, and so on. As shown in Figure 3, instructions can then be sent to human officers in charge, and automatic systems may be activated when available, e.g., to raise barriers in the flood example.

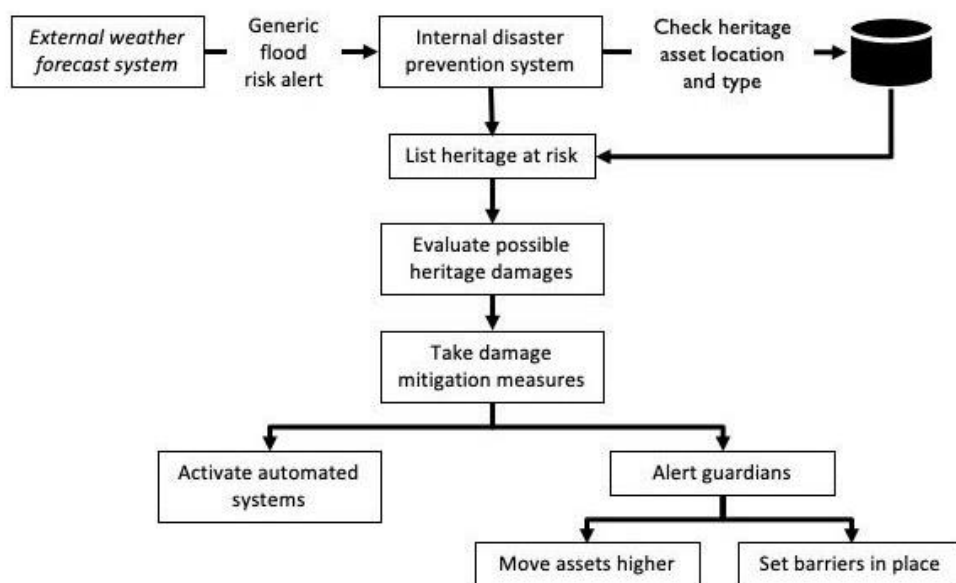


Figure 3. Schema of direct action activated by an automatic alert system based on HDT data.



Another important opportunity provided by the heritage data space organization proposed here concerns the creation of Virtual Research Environments (VRE). A VRE is the digital simulation of a laboratory, where experiments are carried out on the digital representation of real items, i.e., on the digital twins, using software that simulates on the digital twins the effects of actions and events on their real counterpart. Thus, a VRE is the digital evolution of laboratories where real events are simulated in small-scale analogic models, such as a wind gallery or a ship model basin. Another feature of VREs is the collaboration they enable among different researchers who may access shared data spaces.

Examples concern the simulation of destruction caused by earthquake on monuments, or the impact of sea level raising due to global warming on monuments located at low height on the sea level, for example those in Venice.

### 3.4. An Example: Vasari's "Last Supper"

In this section we develop a simple example to show how the HDT ontology may be used to create the digital twin of a heritage asset and to document some of its features, including a story regarding its location in different times and how this can be recorded availing of digital documentation.

Unfortunately, semantic information tends to be verbose, as it is conceived to be machine-processed and not read by humans. Therefore, we will not write the full semantic description as it would appear when codified in RDF or equivalent formats. Also the Turtle format [30] includes information going beyond the purpose of this example, so we will use a "simplified Turtle" enabling to write human-understandable statements omitting the information required for machine-processing it.

The simple rules of the description are the following ones:

- Classes and properties are written in italic.
- Instances of classes are written in bold and are followed by the class name they belong to. They are designated with short sentences to let the reader understand what they are, e.g., **The Pisa Leaning Tower** (*HC2 Heritage Asset*) means the instance of *HC2* which is known as the (famous) Pisa leaning tower.
- Properties relating classes are shown as relationships among instances. The subject, i.e., the instance of the property domain, is not repeated as long as it is the same. A change of subject is indicated by further indenting the statement. For example **The Pisa Leaning Tower** (*HC2 Heritage Asset*) *HP12 is located in* **Pisa** (*HC13 Heritage Location*) *HP2 is documented in* **Pisa Tourist Guide** (*HC4 Heritage Document*) *HP5 includes document* **Map of Pisa** (*HC4 Heritage Document*)

means that the **Pisa Leaning Tower** is a heritage asset located in **Pisa**. It is described in the **Pisa Tourist Guide**. The latter includes a **Map of Pisa**.

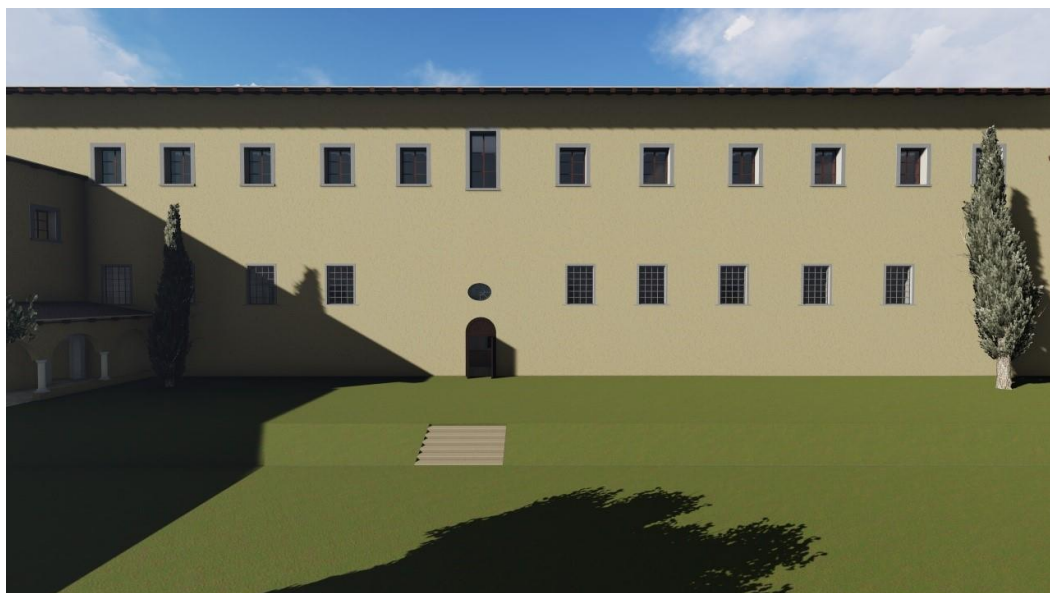
- Any reference to the actual world, e.g., inventory numbers, addresses, DOI, links etc. is used for the sake of the example and may, or may not, be a real one. Therefore no access date is indicated for web references as they may be (and actually in most cases are) fictitious ones.
- Use of classes and properties directly taken from CIDOC CRM are indicated with *crm:* the namespace to which they belong. Note that *crm:D* and *crm:L* are used respectively for classes and properties defined in the CRMdig CRM extension [13].

The example concerns a 16th century painting by Giorgio Vasari titled "The Last Supper" and exhibited in the Santa Croce Museum in Florence. The painting was badly damaged by the 1966 flooding of the Arno River (emergency interventions are shown in Figure 4) and after a long restoration it was placed back in this location. At the time of placing it back in the museum, a study was carried out to determine its original location in a nearby convent, the Murate Convent, from which it was removed in the 19th century following the Napoleonic convent suppression, confirmed by all the governments that ruled Tuscany afterwards and eventually turned the convent into a jail. The study was based on documents kept in the Florence State Archive, including architectural drawings

and the 16th century handwritten diary of a nun named Suor Giustina reporting relevant details about the convent spaces. The study enabled to reconstruct the whole original convent structure (shown in Figure 5) and the position of the refectory where the painting was originally placed, eventually leading to a reconstruction of the room as it was in the 16th century including the furniture and the painting in the original location, as shown in Figure 6. The details of this reconstruction process are reported in [31]. The reconstruction also produced a video with the story, currently visible on YouTube at the link included in the text below.



**Figure 4.** A screenshot from the YouTube video with 1966 black and white TV footage documenting emergency interventions on the painting. PRISMA 2016, CC-BY.



**Figure 5.** A view of the 3D reconstruction of the Murate Convent inner garden, mentioned in the semantic description. N. Amico & F.Niccolucci 2016, CC-BY.





**Figure 6.** A view of the 3D reconstruction of the Murate Convent Refectory as it appeared in the 16th century, with the *Last Supper* painting at the end of the room. This figure and Figure 5 are shown here to give an idea of the visual content included in the Heritage Digital Twin. N. Amico & F.Niccolucci 2016, CC-BY.

The text below is the semantic description of this story using the HDT ontology.

**S. Croce Church and Convent** (HC16 Tangible Asset)

HP12 is located in **Piazza S. Croce, Florence, Italy** (HC13 Heritage Location)

HP4 is part of **Florence Historic Centre** (HC1 Heritage Entity)

*crm:P2 has type "UNESCO WHS" (crm:E55 Type)*

**Refectory in S. Croce Church and Convent** (HC16 Tangible Asset)

HP12 is located in Cloister of S. Croce Church, Piazza Santa Croce, Florence, Italy (HC13 Heritage Location)

HP4 is part of **S. Croce Church and Convent** (HC16 Tangible Asset)

**Murate Convent** (HC16 Tangible Asset)

HP12 is located in **Via Ghibellina 2, Florence, Italy** (HC13 Heritage Location)

HP4 is part of **Florence Historic Centre** (HC1 Heritage Entity)

*crm:P2 has type "UNESCO WHS" (E55 Type)*

**Painting Ultima Cena** (HC16 Tangible Asset)

HP30 is titled "**Ultima Cena**" (HC12 Heritage Title)

HP30 is titled "**Last Supper**" (HC12 Heritage Title)

HP10 is identified by "**1234–5678**" (HC 11 Heritage Identifier)

HP12 is located in **Refectory in S. Croce Church and Convent, Florence, Italy** (HC13 Heritage Location)

HP32 was affected by **Florence Flooding of the Arno River in 1966** (crm:E5 Event)

*crm:P31 was modified by **Restoration of Vasari's Last Supper** (crm:E11 Modification)*

*crm:P14 carried out by **Opificio delle Pietre Dure** (crm:E39 Actor)*

HP2 is documented in **Vasari's Last Supper Restoration** (HC4 Heritage Document)

HP22 has condition state **Restored** (HC9 Heritage Condition State)

*crm:P27 moved from **location of Murate Convent in Florence, Italy** (HC13 Heritage Location)*

*crm:P28 moved to **location of S.Croce Church and Convent in Florence, Italy** (HC13 Heritage Location)*

*crm:P28 moved to refectory location within the S. Croce Church and Convent (HC13 Heritage Location)*

#### **Digital Painting Ultima Cena (HC3 Heritage Digital Twin)**

*HP1 is digital twin of **Painting Ultima Cena** (HC16 Tangible Asset)*

*HP7 uses story **The Original Location of Vasari's Ultima Cena** (HC14 Heritage Story)*

*HP25 includes 3D documentation **3D Model of Murate Convent** (HC7 3D Model)*

#### **Digital Murate Convent (HC3 Heritage Digital Twin)**

*HP1 is digital twin of **Murate Convent** (HC16 Tangible Asset)*

*HP5 includes document **Scan of 1826 Elevation of the Murate convent** (HC19 Heritage Digital Document)*

*HP26 is digitization of **1826 Elevation of the Murate convent** (HC18 Heritage Analogical Document)*

*HP10 is identified by "Florence State Archive (ASF), SFF, FL, 2109, II, ins.132bis, N35" (HC11 Heritage Identifier)*

*HP21 has type "Architectural Elevation" (crm:E55 Type)*

*HP5 includes document "Scan of 1837 Plan of the Murate convent (HC19 Heritage Digital Document)*

*HP26 is digitization of **1837 Plan of the Murate convent** (HC18 Heritage Analogical Document)*

*HP10 is identified by "Florence State Archive (ASF), FL, 12399" (HC11 Heritage Identifier)*

*HP21 has type "Architectural Plan" (crm:E55 Type)*

*HP5 includes document **Digital Diary of Suor Giustina** (HC19 Heritage Digital Document)*

*HP26 is digitization of **Diary of Suor Giustina** (HC18 Heritage Analogical Document)*

*HP10 is identified by "Florence State Archive folder 32399" (HC11 Heritage Identifier)*

*HP21 has type "Handwritten diary" (crm:E55 Type)*

*HP25 includes 3D documentation **3D Model of Murate Convent** (HC7 3D Model)*

*HP7 uses story **Ultima Cena Original Location** (HC14 Heritage Story)*

*HP9 is story created by **Ultima Cena Storytelling** (HC15 Storytelling)*

*HP16 used document Youtube video <https://www.youtube.com/watch?v=P1Uv4Zf5xKk> (HC19 Heritage Digital Document)*

*HP16 used document **3D Model of Murate Convent** (HC7 3D Model)*

*HP6 is story about **Painting Ultima Cena** (HC16 Tangible Asset)*

#### **3D Model of Murate Convent (HC7 3D Model)**

*crm:P2 has type **Reconstruction 3D Model** (E55 Type)*

*crm:L11 was output of **Creation of Murate 3D model** (crm:D7 Digitization Machine Event)*

*crm:L23 used software **3ds Max** (crm:D14 Software)*

*crm:L10 had input **Scan of 1826 Elevation of the Murate convent** (HC19 Heritage Digital Document)*

*crm:L10 had input **Scan of 1837 Plan of the Murate convent** (HC19 Heritage Digital Document)*

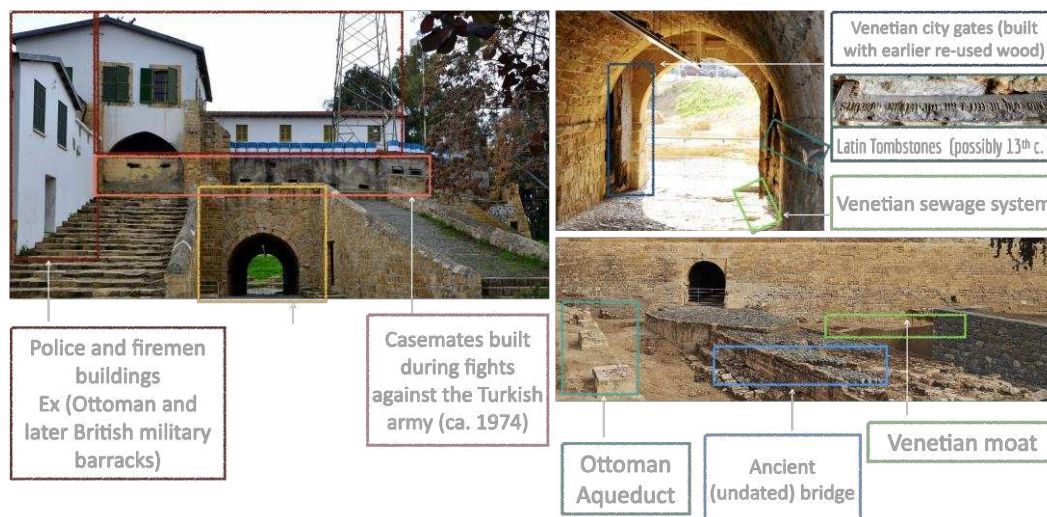
*crm:L10 had input **Digital Diary of Suor Giustina** (HC19 Heritage Digital Document)*

### **3.5. A Case Study: The Pafos Gate in Nicosia, Cyprus and Its Heritage Digital Twin**

The Pafos gate (or "The Upper Gate", being at a slightly higher elevation than the others, at ca. 150 m above the sea level) is one of the three accesses to the walled old city of Nicosia. It connects the access routes to the city from the west, through the main commercial area of the city, with the routes towards east, accessible via the Famagusta gate. Its history can be traced back to the Venetian administration of Francesco Barbaro, Proveditore (military administrator of the kingdom), when the Italian military engineer

Giulio Savorgnan was assigned the role of building the defense system of the capital, in light of the imminent ottoman threats. He designed and supervised the construction of the defense system for eight months in 1567, which continued afterwards under the guidance of his assistant Leonardo Roncone until 1568.

The gate complex has two broad staircases leading from its internal access to the walls. A covered sewage system runs along the edges of the access tunnel. The gate complex is illustrated in Figure 7.



**Figure 7.** The Pafos gate with its main components. (The Cyprus Institute, CC-BY).

The riverbed of the Pedeios river, in earlier periods crossing the city, was diverted so as its waters would fill the stellar-shaped moat, ca. 80 m wide, surrounding the city. Earlier medieval fortifications, along with churches and buildings from the same period located across the perimeter of the new defense wall (for example the San Domenico monastery), were demolished, stones being re-used in the construction of the walls. These consist of a hendecagonal wall primarily made of earth filling and stone cover, ca. 7 km in diameter and eleven bastions of similar size and shape located at ca. 260 m intervals one from the other. The walls were built with a slope at an angle chosen to reduce as much as possible the forces of the bullets. The rampart consists of earth, with a facing of stone half-way up its height forming a retaining wall for the lower portion. Above this retaining wall the scarped face of the earthwork was intended to form a grassy slope [32]. Whereas early modern photographs show that the gate tunnel had wooden doors on both its ends [33], nowadays only one is in place (at the external end of the tunnel) while the fate of the other is unknown. The surviving door is made of vertical beams of wood (predating the Venetian period) reinforced by horizontal metal long plates nailed into the wood with iron spikes (Figure 7, top right).

During the Ottoman rule on the island (1571–1878), the walls of Nicosia were reinforced—a barely discernible inscription on the gate’s internal arch, possibly from the reign of Sultan Mahmut the 2nd (ca. 1821) was added on the wall just above the entrance to the gate’s access tunnel; military barracks were built on the walls above and adjacent to the gate. Recent excavations just outside the gate revealed, among others, remains of the Venetian mote, a bridge and parts of the Ottoman period Arab Ahmet Pasha aqueduct. The quarter inside the walls and in from the gate were known as Tabakhane (i.e., Tannery) Quarter. Further changes occurred during the British Protectorate of the island (1878–1914), when the ground inside the gate was raised by circa half a meter to avoid flooding, the gate was closed and a breach in the wall next to the gate was opened so as to allow increased traffic into the city. A final addition on top of the gate is a series of concrete casemates built in haste during the fights against the invading Turkish troops in 1974. In recent years, as part of an initiative to re-open the gate as part of a cultural tourism circuit of the city,

excavations were held in front of the gate and an access bridge was constructed over this area and connect the sidewalk with the gate, above the excavated features.

This example shows how the HDT ontology may connect very different aspects of cultural heritage documentation. It includes, among others, the aspects listed above, i.e.,

- The historical documentation of the asset by Giulio Savorgnan and Fra' Stefano Lusignano, both as actual documents and their digital counterpart;
- The visual documentation consisting in drawings and 2D and 3D images;
- The list of its different components;
- The overall monument of which it is part, i.e., the Nicosia fortifications;
- The names under which it is known and its official identifier, e.g., in the heritage inventory;
- The location where it is placed;
- Stories about it, useful for communication and storytelling;
- Intangible components belonging to it.

The above information is encoded below, using the HDT ontology. The first list includes the value of the instances used in the encoding, which is presented below using again the simplified Turtle. As the descriptions of the instances, i.e., the actual content, are relevant here—it is a real case—they are provided separately to avoid messing up the Turtle encoding. The list of these instances—i.e., pieces of information—is possibly long, and readers may skip it if they wish, but it shows that the proposed ontology can host actual data—as many as they can be—and is not just a theoretical semantic exercise. As already noted, machine-readable data appear rather verbose to a human, but they are presented here to demonstrate the potential of the HDT ontology in organizing disparate information in a well-structured system. Thus, the ontology may be used according to different interests and perspectives., selecting the information suitable for the specific interests of a very diverse category of users.

#### **Pafos Gate (HC16 Tangible Asset)**

*HP1 has digital twin **Pafos Gate Digital Twin** (HC3 Heritage Digital Twin)*  
*HP2 is documented in **Pafos Gate Documentation Folder** (HC4 Heritage Document)*  
*HP2 is documented in **Giulio Savorgnan's Letters** (HC18 Heritage Analogical Document)*  
*HP2 is documented in **Giulio Savorgnan Drawings** (HC18 Heritage Analogical Document)*  
*HP2 is documented in **Fra' Stefano Lusignano's Chronography** (HC18 Heritage Analogical Document)*  
*HP3 is related to **Nicosia Venetian Fortification System** (HC16 Tangible Asset)*  
*HP4 is formed by part **Pafos Gate Doors** (HC16 Tangible Asset)*  
*HP4 is part of **Nicosia Venetian Fortification System** (HC16 Tangible Asset)*  
*HP4 is part of **Police/Firemen Headquarters** (HC16 Tangible Asset)*  
*HP6 has story **Pafos Gate Story** (HC14 Heritage Story)*  
*HP10 is identified by "DoAIN2022" (HC11 Heritage Identifier)*  
*HP30 is titled "The Pafos Gate" (HC12 Heritage Title)*  
*HP30 is titled "High Gate" (HC12 Heritage Title)*  
*HP30 is titled "Porta San Domenico" (HC12 Heritage Title)*  
*HP12 is located in **Nicosia, Cyprus** (HC13 Heritage Location)*  
*HP13 has intangible component **Tanners and Dyers Commerce** (HC17 Intangible Asset)*  
*HP14 has manifestation event **Activity of Tanners and Dyers Workshops** (HC8 Heritage Event)*  
*HP31 happened at location **Nicosia, Cyprus** (HC13 Heritage Location)*  
*HP17 has visual documentation **Pafos Gate Tunnel Image** (HC6 Imagery)*  
*HP18 has 3D model **Pafos Gate Laser Scanning** (HC7 3D Model)*

#### **Pafos Gate Digital Twin (HC3 Heritage Digital Twin)**

*HP5 includes document **Pafos Gate Documentation** (HP19 Heritage Digital Document)*  
*HP8 is document created by **Pafos Gate Documenting Activity** (HC5 Heritage Documenting)*



HP5 includes document **Scan of Giulio Savorgnan's Letters** (HP19 Heritage Digital Document)

HP26 is digitization of **Giulio Savorgnan's Letters** (HC18 Heritage Analogical Document)

HP7 uses story **Pafos Gate Story** (HC14 Heritage Story)

HP24 includes visual documentation **Pafos Gate Tunnel Image** (HC6 Imagery)

HP25 includes 3D documentation **Pafos Gate Laser Scanning** (HC7 3D Model)

HP21 has type **3D Point Cloud** (crm:E55 Type)

HP28 is hosted by **STARC Repository** (HC20 Hosting Service)

HP29 is provided by **The Cyprus Institute** (HC21 Hosting Service Provider)

**Pafos Gate Story** (HC14 Heritage Story)

HP9 is story created by Correspondence between Giulio Savorgnan and Venetian Officials (HC15 Storytelling)

HP16 used document **Giulio Savorgnan's Letters** (HC18 Heritage Analogical Document)

HP9 is story created by **Fra Stefano Lusignano's Narration** (HC15 Storytelling)

HP16 used document **Fra Stefano Lusignano's Chronography** (HC18 Heritage Analogical Document)

HP9 is story created by **Giulio Savorgnan's fortifications building plan** (HC15 Storytelling)

HP16 used document **Giulio Savorgnan Drawings** (HP18 Heritage Analogical Document)

HP9 is story created by **Reconstruction of Pafos Gate Uses** (HC15 Storytelling)

HP16 used document **Pafos Gate Documentation** (HP19 Heritage Digital Document)

HP16 used document **Pafos Gate Architectonic Documentation** (HP19 Heritage Digital Document)

HP8 is document created by **Pafos Gate 3D Modelling Activity** (HC5 Heritage Documenting)

### Instances, Classes and Descriptions

(boldface denotes Class instances; text within "" is a character string; all web references are real ones and were accessed on 29 January 2022)

#### Instance: **Pafos Gate Digital Twin**

Instance of Class: *HC3 Heritage Digital Twin*

Instance Description: <http://public.cyi.ac.cy/starcRepo/explore/objects> Pafos Gate collection

#### Instance: **Pafos Gate Documentation Folder**

Instance of Class: *HC4 Heritage Document*

Instance Description: Folder containing all material on the study of the gate in order to develop a 3D virtual environment, as part of the rehabilitation plan of the Nicosia municipality to transform the area into a visitable archaeological park.

#### Instance: **Pafos Gate Documenting Activity**

Instance of Class: *HC5 Heritage Documenting*

Instance Description: The activity of collecting or digitizing documents concerning the Pafos Gate

#### Instance: **Pafos Gate 3D Modelling Activity**

Instance of Class: *HC5 Heritage Documenting*

Instance Description: Activity of creating 3D models of the various architectonic components of the Gate

#### Instance: **Pafos Gate Tunnel Image**

Instance of Class: *HC6 Imagery*

Instance Description: <http://public.cyi.ac.cy/starcRepo/details/show/2d131880fdb75dbcb3c35cf8df74b1ca>

#### Instance: **Pafos Gate Laser Scanning**

Instance of Class: *HC7 3D Model*

Instance Description: <http://public.cyi.ac.cy/starcRepo/details/show/5a628b796fc6100f5a3a2e2dce1dc175>

Instance: **Activity of Tanners and Dyers Workshops**

Instance of Class: *HC8 Heritage Event*

Instance Description: Specific events manifesting the activity of the Tanners and Dyers Workshop near the Pafos Gate

Instance: **REVEEL 3D**

Instance of Class: *HC10 Heritage Event Model*

Instance Description: <https://avl.ncsa.illinois.edu/realworld-software/riveel-3d>

Instance: **DoAIN2022**

Instance of Class: *HC11 Heritage Identifier*

Instance Description: inventory number of the gate in the Department of Antiquities catalogue of monuments.

Instance: **"The Pafos Gate"**

Instance of Class: *HC12 Heritage Title*

Instance Description: The Pafos Gate (ex. High Gate, earlier Porta San Domenico)

Instance: **Nicosia Cyprus**

Instance of Class: *HC13 Heritage Location*

Instance Description: Nicosia, Cyprus 35.1737° N, 33.3568° E

Instance: **Pafos Gate Road**

Instance of Class: *HC13 Heritage Location*

Instance Description: Road which leads to the Pafos Gate from the centre of the city, part of the commercial route that crossed the city connecting the Pafos Gate with the Famagusta gate.

Instance: **Pafos Gate Story**

Instance of Class: *HC14 Heritage Story*

Instance Description: Story of the Pafos Gate throughout the centuries

Instance: **Correspondence between Giulio Savorgnan and Venetian Officials**

Instance of Class: *HC15 Storytelling*

Instance Description: Story of the correspondence between Giulio Savorgnan and Venetian officials

Instance: **Giulio Savorgnan's fortifications building plan**

Instance of Class: *HC15 Storytelling*

Instance Description: Giulio Savorgnan's plans to build the fortifications

Instance: **Fra Stefano Lusignano's Narration**

Instance of Class: *HC15 Storytelling*

Instance Description: Storytelling of the Chronography of the Cyprus island by Fra Stefano Lusignano

Instance: **Reconstruction of Pafos Gate Uses**

Instance of Class: *HC15 Storytelling*

Instance Description: Reconstruction of the history of the various uses of the Pafos Gate throughout the centuries

Instance: **Pafos Gate**

Instance of Class: *HC16 Tangible Asset*

Instance Description: The Pafos Gate

Instance: **Nicosia Venetian Fortification System**

Instance of Class: *HC16 Tangible Asset*

Instance Description: Nicosia Venetian fortification system; commercial routes from the west to the east

Instance: **Police/Firemen Headquarters**

Instance of Class: *HC16 Tangible Asset*

Instance Description: Police/Firemen Headquarters; part of a new cultural tourism path in Nicosia

Instance: **Pafos Gate Doors**

Instance of Class: *HC16 Tangible Asset*

Instance Description: The wooden doors reinforced by iron beams.

Instance: **Tanners and Dyers Commerce**

Instance of Class: *HC17 Intangible Asset*

Instance Description: Tanners and Dyers workshops were located along the road which leads to the Pafos Gate from the centre of the city.

Instance: **Giulio Savorgnan Drawings**

Instance of Class: *HP18 Heritage Analogical Document*

Instance Description: Drawings of the fortifications made by Giulio Savorgnan.

Instance: **Giulio Savorgnan's Letters**

Instance of Class: *HC18 Heritage Analogical Document*

Instance Description: Collection of letters written by Giulio Savorgnan to Venetian officials, describing the design and building plans of the Venetian fortifications (Pafos gate part of them)

Instance: **Fra Stefano Lusignano's Chronography**

Instance of Class: *HC18 Heritage Analogical Document*

Instance Description: The *Chorographia et brevis historia universale dell'isola de Cipro* of Fra Stefano Lusignano, a Dominican friar from the royal family. Printed first in Bologna, 1573.

Instance: **Scan of Giulio Savorgnan's Letters**

Instance of Class: *HP19 Heritage Digital Document*

Instance Description: Digital collection of letters written by Giulio Savorgnan to Venetian officials, describing the planning and building of the Venetian fortifications (Pafos gate being part of them) and including measurements, plan drawings and textual descriptions.

Instance: **Pafos Gate Documentation**

Instance of Class: *HP19 Heritage Digital Document*

Instance Description: Documentation of the gate, RTI on specific details (the inscription, the lapidary stones), dendrochronology analysis of the wooden beams of the doors.

Instance: **Pafos Gate Architectonic Documentation**

Instance of Class: *HP19 Heritage Digital Document*

Instance Description: Models of the various architectonic components of the Gate

Instance: **STARC Repository**

Instance of Class: *HC20 Hosting Service*

Instance Description: STARC Repo—<http://public.cyi.ac.cy/starcRepo/>

Instance: **The Cyprus Institute**

Instance of Class: *HC21 Hosting Service Provider*

Instance Description: The Cyprus Institute (CYI)

#### 4. Conclusions

The ontology introduced in the present paper needs further refinement, which may only come from testing it in practice and getting feedback from the heritage community and stakeholders. It is also particularly important to refine the documentation of relevant features concerning, for instance, the quality and detail of 3D models, the procedure and settings used in scientific analyses, and so on. On this regard, ongoing work within the ARIADNEplus project concerning application profiles and the 4CH project, especially as regards 3D modelling and conservation activities, provides components that are already usable within the HDT ontology.

The HDT ontology is primarily aimed at organizing and managing digital information about heritage, in order to produce what the EU Commission calls “high quality records” in the *Recommendation on a common European data space for cultural heritage* [1], for which it envisages an increase of about 60% by 2025—a very ambitious target indeed. It is clear that high quality records require rich, high-quality metadata, as those implemented by the HDT ontology proposed here. It relies on and is fully compliant with the CIDOC CRM international standard and its compatible models, as the one adopted by Europeana in the EDM ontology [14] and those adopted by a large and increasing number of galleries, museums, libraries and other cultural institutions. It is also compatible with the ARIADNE catalogue of about two million archaeological datasets and the forthcoming 4CH knowledge base. Actually, the EDM is equivalent to a subset of HDT so the improvement of the quality

of existing records, also recommended by the Commission, is progressively feasible without disrupting the compatibility with the current wealth of records provided by Europeana.

With such rich metadata, the system will enable users to perform the activities typical of working with cultural heritage, including the access and retrieval of information and its use to know, understand, preserve, communicate and share the assets. Adopting this scheme and its semantic graph guarantees retrievable and accessible data. Using the HDT ontology enables interoperability and reuse, making heritage data FAIR. It supports the creation of a distributed system, a federation of storage and cloud facilities concerning heritage across Europe, as indicated by the EU Commission for the Data Space for Cultural Heritage in its *Communication* [1]. It enables linking with other economic sectors as tourism and creative industries. It supports the development of advanced services, as sketched in Section 3.2. It is ready to link or to incorporate related initiatives as the forthcoming Competence Centre for Cultural Heritage to be designed by the 4CH project [16], focusing on a digital approach to conservation and preservation. It enables collaboration among institutions and professionals, as well as with citizens, via the establishment of VREs. In sum, it has all the characteristics required for a starting point when building a cultural heritage data space as envisaged by European strategies.

**Author Contributions:** All authors have equally contributed to all the activities leading to the preparation of the present paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## Appendix A. Annex–Ontology Definitions

### Appendix A.1. Ontology Description

As mentioned above, this section may be skipped at first reading and used just as a reference in the examples

### Appendix A.2. Class Definitions

#### HC1 Heritage Entity

Subclass of: *crm:E1 CRM Entity*

Description: This class comprises all things in the universe of discourse of cultural heritage data. It has no instances. Its use is mainly to define general properties that are inherited by all its subclasses, i.e., all the classes of the HDT ontology, for example *crm:P1 is identified by* or *crm:P2 has type*.

#### HC2 Heritage Asset

Subclass of: *HC1 Heritage Entity*, *crm:E77 Persistent Item*

Description: Actual heritage asset of any nature: physical, both movable and immovable, immaterial, or born digital.

#### HC3 Heritage Digital Twin

Subclass of: *HC1 Heritage Entity*, *crm:D1 Digital Object*

Description: Collection of all the digital objects pertaining to an *HC2 Heritage Asset*.

#### HC4 Heritage Document

Subclass of: *HC1 Heritage Entity*, *crm:E31 Document*

Description: A piece of the collection of all documents concerning an *HC2 Heritage Asset*.

#### HC5 Heritage Documenting

Subclass of: *HC1 Heritage Entity*, *crm:E7 Activity*



Description: The activity of collecting or digitizing documents concerning an *HC2 Heritage Asset*.

*HC6 Imagery*

Subclass of: *HC5 Heritage Document, crm:D1 Digital Object*

Description: a *HC5 Heritage Document* consisting in digital imagery of any kind about an *HC2 Heritage Asset*.

*HC7 3D Model*

Subclass of: *H5 Heritage Document, crm:D1 Digital Object*

Description: an *HC5 Heritage Document* consisting of a 3D (digital) model of the heritage asset. AR and VR models are related by an appropriate function to the place where they are to be used.

*HC8 Heritage Event*

Subclass of: *HC1 Heritage Entity, crm:E5 Event*

Description: Any external event related or susceptible to have an impact on an *HC2 Heritage Asset*.

*HC9 Heritage Condition State*

Subclass of: *HC1 Heritage Entity, crm:E3 Condition State*

Description: Condition of an *HC2 Heritage Asset* caused by an *HC8 Heritage Event* as documented in an *HC4 Heritage Document*: for example, the condition of “destroyed” following an earthquake event.

*HC10 Heritage Event Model*

Subclass of: *HC1 Heritage Entity, crm:D1 Digital Object*

Description: Digital model of any *HC4 Heritage Event* susceptible to have an effect on heritage assets. It may range from plain description to simulation software, and is not part of the *HC3 Heritage Digital Twin*. It is used, for example, in simulations.

*HC11 Heritage Identifier*

Subclass of: *HC1 Heritage Entity, crm:E42 Identifier*

Description: Identifier for an *HC1 Heritage Entity*, e.g., the DOI for digital items.

*HC12 Heritage Title*

Subclass of: *HC1 Heritage Entity, crm:E35 Title*

Description: A title under which an *HC2 Heritage Asset* is known.

*HC13 Heritage Location*

Subclass of: *HC1 Heritage Entity, crm:E53 Place*

Description: The location an *HC2 Heritage Asset*.

*HC14 Heritage Story*

Subclass of: *HC1 Heritage Entity, crm:E89 Propositional Object*

Description: Account about an *HC2 Heritage Asset*.

*HC15 Storytelling*

Subclass of: *HC1 Heritage Entity, crm:E7 Activity*

Description: Activity of creating an *HC14 Heritage Story*.

*HC16 Tangible Asset*

Subclass of: *HC2 Heritage Asset, crm:E18 Physical Thing*

Description: Tangible part of an *HC2 Heritage Asset*.

*HC17 Intangible Asset*

Subclass of: *HC2 Heritage Asset, crm:E28 Conceptual Object*

Description: Intangible part of an *HC2 Heritage Asset*.

*HP18 Heritage Analogical Document*

Subclass of: *HC4 Heritage Document*

Description: A piece of the collection of all analogical documents of an *HC2 Heritage Asset*.

*HP19 Heritage Digital Document*

Subclass of: *HC4 Heritage Document, crm:D9 Data Object*

Description: A piece of the collection of all digital documents of an *HC2 Heritage Asset*, which form its digital twin.

*HP20 Hosting Service*

Subclass of: *HC1 Heritage Entity, crm:PE5 Digital Hosting Service*

Description: The digital platform or infrastructure on which an *H3 Heritage Digital Twin* is hosted and operated.

*HP21 Hosting Service Provider*

Subclass of: *HC1 Heritage Entity, crm:E39 Actor*

Description: The owner or administrator of the digital infrastructure providing the *HP20 Hosting Service* for operating an *HC3 Heritage Digital Twin*.

### *Appendix A.3. Property Definitions*

*HP1 has digital twin (is digital twin of)*

Domain: *HC2 Heritage Asset*

Range: *HC3 Heritage Digital Twin*

*HP2 is documented in (documents)*

Domain: *HC2 Heritage Asset*

Range: *HC4 Heritage Document*

*HP3 is related to (relates)*

Domain: *HC2 Heritage Asset*

Range: *HC2 Heritage Asset*

*HP4 is part of (is formed by part)*

Domain: *HC2 Heritage Asset*

Range: *HC2 Heritage Asset*

*HP5 includes document (digital document is included in)*

Domain: *HC3 Heritage Digital Twin*

Range: *HC4 Heritage Document*

*HP6 has story (is story about)*

Domain: *HC2 Heritage Asset*

Range: *HC14 Heritage Story*

*HP7 uses story (story is used by)*

Domain: *HC3 Heritage Digital Twin*

Range: *HC14 Heritage Story*

*HP8 created document (is document created by)*

Domain: *HC5 Heritage Documenting*

Range: *HC4 Heritage Document*

*HP9 creates story (is story created by)*

Domain: *HC15 Storytelling*

Range: *HC14 Heritage Story*

*HP10 identifies (is identified by)*

Domain: *HC1 Heritage Entity*

Range: *HC11 Heritage Identifier*

*HP11 has component (is component of)*

Domain: *HC3 Heritage Digital Twin*

Range: *HC4 Heritage Document*

*HP12 is located in (is location of)*

Domain: *HC16 Tangible Asset*

Range: *HC13 Heritage Location*

*HP13 has intangible component (is intangible component of)*

Domain: *HC16 Tangible Asset*

Range: *HC17 Intangible Asset*

*HP14 has manifestation event (event is manifestation of)*

Domain: *HC17 Intangible Asset*

Range: *HC8 Heritage Event*

*HP15 is manifested by (is manifestation of)*

Domain: *HC17 Intangible Asset*

Range: *HC16 Tangible Asset*

HP16 used document (document used for)  
 Domain: HC15 Storytelling  
 Range: HC4 Heritage Document  
 HP17 has visual documentation (is visual documentation of)  
 Domain: HC2 Heritage Asset  
 Range: HC6 Imagery  
 HP18 has 3D model (is 3D model of)  
 Domain: HC16 Tangible Asset  
 Range: HC7 3D Model  
 HP19 narrates (is narrated by)  
 Domain: HC14 Heritage Story  
 Range: HC8 Heritage Event  
 HP20 depicts (is depicted by)  
 Domain: HC6 Imagery  
 Range: HC8 Heritage Event  
 HP21 has type (is type of)  
 Domain: HC1 Heritage Entity  
 Range: crm:E55 Type  
 [controlled vocabulary, e.g., Getty AAT]  
 HP22 has condition state (is condition state of)  
 Domain: HC16 Tangible Asset  
 Range: HC9 Heritage Condition State  
 HP23 tells about (is told by)  
 Domain: HC14 Heritage Story  
 Range: HC17 Intangible Asset  
 HP24 includes visual documentation (visual documentation is included in)  
 Domain: HC3 Heritage Digital Twin  
 Range: HC6 Imagery  
 HP25 includes 3D documentation (3D documentation is included in)  
 Domain: HC3 Heritage Digital Twin  
 Range: HC7 3D Model  
 HP26 is digitization of (was digitised to create)  
 Domain: HC19 Heritage Digital Document  
 Range: HC18 Heritage Analogical Document  
 HP27 is digital part of (has digital part)  
 Domain: HC3 Heritage Digital Twin  
 Range: HC3 Heritage Digital Twin  
 HP28 is hosted by (hosts)  
 Domain: HC3 Heritage Digital Twin  
 Range: HP20 Hosting Service  
 HP29 is provided by (provides)  
 Domain: HP20 Hosting Service  
 Range: HP21 Hosting Service Provider  
 HP30 is titled (titles)  
 Domain: HC1 Heritage Entity  
 Range: HC12 Heritage Title  
 HP31 happened at location (was location for)  
 Domain: HC8 Heritage Event  
 Range: HC13 Heritage Location  
 HP32 was affected by (affected)  
 Domain: HC2 Heritage Asset  
 Range: crm:E5 Event

## References

1. European Commission. Commission Recommendation of 10/11/21 on a Common European Data Space for Cultural Heritage. 2021. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021H1970&from=EN> (accessed on 29 January 2022).
2. European Commission. The New Renaissance. In *Report of the Comité des Sages on Bringing Europe's Cultural Heritage Online*; 2011. Available online: <https://op.europa.eu/s/pGpJ> (accessed on 29 January 2022).
3. CHCFE EU Project. 2015. Cultural Heritage Counts for Europe. Available online: <https://www.europanostra.org/our-work/policy/cultural-heritage-counts-europe/> (accessed on 29 January 2022).
4. European Commission. Innovation in Cultural Heritage Research. 2018. Available online: <https://op.europa.eu/en/publication-detail/-/publication/1dd62bd1-2216-11e8-ac73-01aa75ed71a1/language-en/format-PDF/source-128623197> (accessed on 29 January 2022).
5. ICOM. Statute. 2017. Available online: [https://icom.museum/wp-content/uploads/2018/07/2017\\_ICOM\\_Statutes\\_EN.pdf](https://icom.museum/wp-content/uploads/2018/07/2017_ICOM_Statutes_EN.pdf) (accessed on 29 January 2022).
6. Singh, M.; Fuenmayor, E.; Hinchy, E.P.; Qiao, Y.; Murray, N.; Devine, D. Digital Twin: Origin to Future. *Appl. Syst. Innov.* **2021**, *4*, 36. [CrossRef]
7. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—Shaping Europe's Digital Future. 2020. Available online: [https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020\\_en\\_3.pdf](https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_3.pdf) (accessed on 29 January 2022).
8. ISO 16739-1:2018. Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries—Part 1: Data Schema. ISO: London, UK, 2018. Available online: <https://www.iso.org/standard/70303.html> (accessed on 29 January 2022).
9. Bolton, A.; Lorraine, B.; Dabson, I.; Enzer, M.; Evans, M.; Fenemore, T.; Harradence, F.; Keaney, E.; Kemp, A.; Luck, A.; et al. The Gemini Principles: Guiding Values for the National Digital Twin and Information Management Framework. Available online: <https://www.cdbb.cam.ac.uk/system/files/documents/TheGeminiPrinciples.pdf> (accessed on 29 January 2022).
10. UNESCO. World Heritage List. Available online: <https://whc.unesco.org/en/list> (accessed on 29 January 2022).
11. UNESCO. Intangible Cultural Heritage List. Available online: <https://ich.unesco.org/en/lists> (accessed on 29 January 2022).
12. Niccolucci, F. Virtual museums and archaeology: An international perspective. *Archeol. Calc.* **2007**, *17*, 15–30.
13. CIDOC-CRM. 2021. Available online: <http://www.cidoc-crm.org/> (accessed on 29 January 2022).
14. EuropeanaPro. The Europeana Data Model. 2017. Available online: <https://pro.europeana.eu/page/edm-documentation> (accessed on 29 January 2022).
15. ARIADNE. Available online: <https://ariadne-infrastructure.eu/> (accessed on 29 January 2022).
16. 4CH. Available online: <https://4ch-project.eu/> (accessed on 29 January 2022).
17. The London Charter. Available online: <https://The-London-Charter.org/> (accessed on 29 January 2022).
18. Callieri, M.; Dellepiane, M.; Pavoni, G.; Pingi, P.; Potenziani, M.; Scopigno, R. Alchemy in 3D: A digitization for a journey through matter. In *Proceedings of the International Congress on Digital Heritage, Granada, Spain, 28 August–2 October 2015*; pp. 223–231. [CrossRef]
19. Palma, G.; Pingi, P.; Siotto, E.; Bellucci, R.; Guidi, G.; Scopigno, R. Deformation analysis of Leonardo da Vinci's "Adorazione dei Magi" through temporal unrelated 3D digitization. *J. Cult. Herit.* **2019**, *38*, 173–185. [CrossRef]
20. Felicetti, A.; Niccolucci, F. A CIDOC CRM-based Model for the Documentation of Heritage Sciences. In *Proceedings of the 3rd International Congress on Digital Heritage (DH2018), San Francisco, CA, USA, 26–30 October 2018*; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2018. [CrossRef]
21. Castelli, L.A.; Felicetti, A. Modelli e Strumenti Semantici per la Documentazione dell'Heritage Science, Atti della Conferenza GARR 2018, Cagliari, 3–5 Ottobre. 2018. Available online: <https://www.garr.it/it/news-e-eventi/pubblicazioni/atti-delle-conferenze/atti-della-conferenza-garr-2018> (accessed on 29 January 2022).
22. Castelli, L.; Felicetti, A.; Proietti, F. Heritage Science and Cultural Heritage: Standards and tools for establishing cross-domain data interoperability. *Int. J. Digit. Libr.* **2021**, *22*, 279–287. [CrossRef]
23. Bombini, A.; Castelli, L.; dell'Agnello, L.; Felicetti, A.; Giacomini, F.; Niccolucci, F.; Taccetti, F. CHNet cloud: An EOSC-based cloud for physical technologies applied to cultural heritages. In *Proceedings of the GARR Conference 2021, Online, 7–16 June 2021*. [CrossRef]
24. Gonizzi Barsanti, S.; Guidi, G.; De Luca, L. Segmentation of 3D models for cultural heritage structural analysis—some critical issues. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017 Volume IV-2/W2*. In *Proceedings of the 2017-26th International CIPA Symposium 2017, Ottawa, Canada, 28 August–1 September 2017*; pp. 115–122.
25. Ontotext. Available online: <https://graphdb.ontotext.com/> (accessed on 29 January 2022).
26. Research Space. Available online: <https://researchspace.org/> (accessed on 29 January 2022).
27. Museo del Prado. The Museo del Prado's Ontological Model. Available online: <https://www.museodelprado.es/en/grafico-de-conocimiento/modelo-ontologico> (accessed on 29 January 2022).
28. Dijkshoorn, C.; Jongma, L.; Aroyo, L.; van Ossenbruggen, J.; Schreiber, G.; ter Weele, W.; Jan Wielemaker, J. The Rijksmuseum Collection as Linked Data. *Semant. Web* **2018**, *9*, 221–230. [CrossRef]

29. Danthine, B.; Hiebel, G.; Posch, C.; Stadler, H. The Integration of Semantic Networks in Virtual Exhibitions The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021 Volume XLVI-M-1-2021. In Proceedings of the 28th CIPA Symposium “Great Learning & Digital Emotion”, Beijing, China, 28 August–1 September 2021; pp. 165–172.
30. W3C. RDF 1.1 Turtle. 2014. Available online: <https://www.w3.org/TR/turtle/> (accessed on 29 January 2022).
31. Niccolucci, F.; Amico, N. Implementing the London Charter in Virtual Reconstructions. Virtual Archaeology (from Air, on Earth, under Water and at Museum). In *Proceedings of the International Forum held at the State Hermitage Museum, 28-30 May 2018*; The State Hermitage Publishers: St. Petersburg, Russia, 2018; pp. 132–142.
32. Jeffery, G. *Historic Monuments of Cyprus: Studies in Archaeology and Architecture of the Island*; Zeno: London, UK, 1983.
33. CVAR: Costas and Rita Severis Foundation. Paphos Gate, Gelatin Silver Print. Available online: <https://cvar.severis.org/en/collections/item/paphos-gate-nicosia-cyprus/3069/> (accessed on 29 January 2022).





Thank you to arXiv for use of its open access interoperability. This reprint of a paper from the arXiv archive was not reviewed or approved by, nor does it necessarily express or reflect the policies or opinions of, arXiv.





# The Heritage Digital Twin: a bicycle made for two. The integration of digital methodologies into cultural heritage research

Franco Niccolucci<sup>\*1</sup>, Béatrice Markhoff<sup>2</sup>, Maria Theodoridou<sup>3</sup>, Achille Felicetti<sup>1</sup>, Sorin Hermon<sup>4</sup>

<sup>1</sup> PIN, Servizi Didattici e Scientifici per l'Università di Firenze, Piazza Ciardi 25, I-59100 Prato, Italy

<sup>2</sup> Université de Tours, 64 Avenue Jean Portalis, F-37200 Tours, France

<sup>3</sup> FORTH, N. Plastira 100, Vassilika Vouton, GR-70013 Heraklion, Crete, Greece

<sup>4</sup> STARC-The Cyprus Institute, 20 Konstantinou Kavafi Street, CY-2121 Aglantzia, Nicosia, Cyprus

**\*Corresponding author:** franco.niccolucci@gmail.com

## Abstract

The paper concerns the definition of a novel ontology for cultural heritage based on the concept of digital twin. The ontology, called Heritage Digital Twin ontology, is a compatible extension of the well-known CIDOC CRM ISO standard for cultural heritage documentation and incorporates all the different documentation systems presently in use for cultural heritage documentation. In the authors' view, it supports documentation interoperability at a higher level than the ones currently in use and enables effective cooperation among different users.

**Keywords:** digital twin, heritage documentation systems, CIDOC CRM extension, heritage semantics, art history semantics, archaeology semantics, scientific data for cultural heritage, conservation

## Plain Language Summary

The paper addresses the issue of managing the digital documentation of cultural heritage in an effective way. For this purpose, it introduces the concept of heritage digital twin, which is inspired by the similar one used in the industry and in other domains. A digital twin is the digital replica of a real-world object. It includes all the necessary information and is able to simulate - in a digital environment - the characteristics and the behaviour of its real counterpart. Two examples are fully developed to demonstrate the value of this novel approach. This research was carried out in the framework of the EU-funded projects ARIADNEplus and 4CH, both concerning the use of digital technology to study and preserve cultural heritage.

## 1. Introduction

The title of this paper is inspired by the similar one of a paper by Pollard and Bray [1] concerning how the contribution of archaeological sciences should be integrated into the process of archaeological interpretation. According to the authors, such integration is like riding a bicycle made for two, also known as a tandem. This kind of vehicle requires a strong collaboration between the two riders to pedal synchronically and the one in front must be able and willing to drive the tandem towards a common destination, on which both riders agree. The structure of the bicycle should suit a diversity of users: tall and short; married couples and perfect strangers; sportspeople and lazy ones. The way it can be used must adapt to any kind of road, dirt trails and urban well-paved streets alike.

Cycling metaphors aside, the convergence and integration of two different disciplines puts requirements to the method and the attitude of both and of all participants. Each party involved must consider the other one's perspective and adapt its own accordingly. A similar situation takes place about the interdisciplinary approach known as digital cultural heritage. Adopting such a collaborative attitude by heritage professionals and institutions is the nucleus of what is usually called digital transformation of cultural heritage. The upskilling of professionals and the allocation of resources to information technology by heritage managers is of course required, but without such change of perspective they might be ineffective.

In the present paper we report the results of research triggered by the activities carried out within two European highly interdisciplinary projects, which probably extend well beyond the project objectives.

The first project is the ARIADNE group of projects comprising ARIADNE (Advanced Research Infrastructure for Archaeological Data Networking in Europe), an FP7 project operating from 2013 to the end of 2016 [2], and its continuation ARIADNEplus<sup>1</sup>, an H2020 project operating from 2018 until the end of 2022. The two projects are collectively indicated as ARIADNE.

The original ARIADNE objective was stated as “To turn the sparse existing archaeological repositories, into a pan-European Integrated Research Infrastructure, with easily available and harmonised access, responding to the demand of the archaeological research community of the European Research Area and beyond” while the ARIADNEplus one is simply “Extending and focusing ARIADNE”. The break between the two projects, determined by the schedule of EU calls, was bridged maintaining the main project result, the ARIADNE catalogue, and preparing for the continuation.

The most visible part of the project outcomes is the ARIADNE catalogue, also called the ARIADNE Knowledge Base, accessible from the ARIADNE portal<sup>2</sup>, with various searching options including time, location and content of the catalogued datasets. ARIADNE aggregated more than 2,000,000 archaeological datasets from a large part of Europe. ARIADNEplus has expanded ARIADNE in geographic coverage, now extending to a larger part of Europe and including datasets from Japan, Israel, Africa, USA and Latin America; and in content type, now including many more datasets concerning archaeological sciences, images, 3D models, archaeological GIS, and so on. The total number of items in the ARIADNEplus version of the catalogue is more than 3,300,000 aggregated datasets, with another hundred thousand datasets still in the aggregation queue. Such results have raised the enthusiasm of the archaeological research community as witnessed in two publications, *The ARIADNE Impact* [3] concerning ARIADNE and a forthcoming one about the added value of ARIADNEplus. This has created a vibrant community that has been nurtured in parallel by training and dissemination activities.

Such achievements have been attained thanks to research results that in ARIADNE operate behind the scenes. There are two main strands: information technology, supporting all the steps of aggregation, and providing the technical instruments to make data FAIR (Findable, Accessible, Interoperable and Reusable); and advanced work in semantics, leading among others to the development of CRMarchaeo<sup>3</sup>, an archaeological extension of the well-known CIDOC CRM standard ontology for cultural heritage<sup>4</sup>. As it is well known, CIDOC CRM – also mentioned in brief as the CRM – stands for Conceptual Reference Model supported by CIDOC, the International Committee for Documentation of ICOM, the International Council of Museums and since December 2006 is an official ISO standard (ISO 21127:2014).

The ARIADNE research achievements form one of the pillars supporting the work presented in this paper.

The ARIADNE Knowledge Base is based on a subset of the CRM ontology, to which the metadata schemas of the catalogued items are mapped. The project also developed application profiles for specific subdomains such as epigraphy, scientific analyses, excavations and so on. Thus, it suggested the idea of an overarching ontology as the one presented in the present paper.

The second project supporting the present paper is the H2020 project 4CH – Competence Centre for the Conservation of Cultural Heritage<sup>5</sup>. This ongoing project addresses the conservation of monuments and sites with digital methods, in view of the creation of a European Competence Centre on the subject, which the project is in charge of designing. Among others, the project aims at setting up a cloud-based Knowledge Base with conservation data. This aspect has a key but not exclusive role in the project work plan. With regard to the present paper, 4CH has brought the need of extending the scope of semantic research beyond archaeology to include tangible cultural heritage as well as its intangible component. Within 4CH, implementing a suitable documentation system is still ongoing work. We anticipate that the ontology presented here will be adopted in 4CH to build its Knowledge Base.

Thus, the semantic team of both projects has been induced to start conceiving an overarching paradigm that could include the digital representation of all heritage assets, tangible or intangible alike, for any purpose.

---

<sup>1</sup> <https://ariadne-infrastructure.eu/>

<sup>2</sup> <https://portal.ariadne-infrastructure.eu/>

<sup>3</sup> <https://www.cidoc-crm.org/crmarchaeo/home-3>

<sup>4</sup> <https://cidoc-crm.org/>

<sup>5</sup> <https://www.4ch-project.eu/>

Illustrating this approach and its main concept, the Heritage Digital Twin, forms the body of the present paper.

In the rest of this paper, Section 2 is a summary of the cultural heritage digital documentation's history with its current state of the art. Section 3 motivates the need of a heritage digital twin ontology. Section 4 brings the basics of the heritage digital twin concept and shows how it goes further than 3D-based data models. Section 5 introduces the overall methodology and main concepts of the Heritage Digital Twin (HDT) Ontology, and more technical descriptions are provided in Section 6 and Section 7. Examples of digital documentations pertaining to the same heritage digital twin are developed in Section 8, one for scientific analyses and another one for art history analyses, performed on the same heritage entity. The section also includes an example of use for built heritage.

## 2. Background: the state of the art in the digital documentation of cultural heritage

An even concise description of how the digital revolution of the last part of the XX century affected also documenting cultural heritage would probably require an encyclopaedia. For example, Google scholar lists about 1,900,000 papers under the search key "digital documentation of cultural heritage" and it is likely that titles wrongly included in the search results are less than titles wrongly excluded. For instance, does Google AI consider "digital archaeology" as also pertaining to "digital documentation of cultural heritage"?

Archaeology started to be digitally documented in the second half of 1900 using relational databases, while in general similar IT applications to cultural heritage were less frequent at that time. We will follow the development of digital archaeological documentation as an indicator of the progress of such techniques.

At the beginning, digital methods were also called computer applications or quantitative methods, hence the name of the most important conference on the subject: Computer Applications and quantitative methods in Archaeology, CAA<sup>6</sup>. This annual event has been for a long time the main place to communicate such applications, together with other initiatives such as the dedicated journal "*Archeologia e Calcolatori*" (*Archaeology and Computers*, a vintage name indeed), which started publications in 1990 and is still published in Italy. This journal published in 2019 [4, 5] a historical account of the development of computing applications, their perspectives and their impact on archaeological theory and methods. Many papers published in the CAA proceedings in the last decade of the XX century describe examples of database use to digitally store archaeological data. The number of such applications to individual case studies continued to increase until such applications became standard practice and developing a database to collect the data of a specific archaeological investigation was no more a contribution to advancing research in this field, but just another example. At this point, the annual number of related papers collapsed. In parallel, the interest on GIS (Geographic Information Systems) increased, as it appeared that this technique might capture the locational characteristic and spatial relationships of archaeology. Non-canonical threads as the system proposed in 1979 by Jean-Claude Gardin denominated *Logicisme* [6] received less favour possibly also because they were out of the current mainstream in digital archaeology. The interest in GIS was also an obstacle to acceptance, as they need only very simple databases to document archaeological assets and related spatial information.

Two recent papers, one by Dallas [7] and the other one by Moscati [4] have brought *Logicisme* on the forefront again, and an excavation documented according to this system has recently been published [8, 9]. In the meanwhile, the concept of Linked Open Data has made its way into archaeology, starting from the first decade of the XXI century. In the same period, semantic concepts started to be applied to archaeological documentation and domain ontologies showed up.

This very summary sketch of the archaeological computing history explains why it is now necessary to propose new semantic tools for the archaeological documentation, based on accepted domain standards to support interoperability. This principle has been considered in ARIADNE, where a set of common aggregation metadata are used for all datasets.

Another important thread in archaeological documentation is represented by 3D models. Initially used for communication purposes, to facilitate visitors' understanding of complex archaeological sites – an application still alive today – it was realised that such 3D models might also serve for a more complete

---

<sup>6</sup> <https://caa-international.org/>

information about archaeological assets, both objects and remains. A recent book [10] combines such a 3D approach with GIS. This thread may be considered as part of the general interest for using 3D as a documentation tool in Cultural Heritage.

On the general cultural heritage side, digital documentation was limited to inventories for a long period. The turning point coincided with the definition of the CIDOC CRM<sup>7</sup>.

The CRM is the culmination of more than two decades of standards development work by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Work on the CRM itself began in 1996 under the auspices of the ICOM-CIDOC Documentation Standards Working Group. Since 2000, development of the CRM has been officially delegated by ICOM-CIDOC to the CIDOC CRM Special Interest Group (SIG). The SIG, in turn, collaborates with the ISO working group ISO/TC46/SC4/WG9 to bring the CRM to the form and status of an International Standard. This set of collaborations has resulted in the production of ISO21127:2004 and ISO21127:2014, the ISO standard editions of the CIDOC CRM.

The CRM was initially conceived for museum applications but then it gradually evolved into a general-purpose cultural heritage documentation system. The CRM is an extendable system, enabling to develop extensions compliant with the CRM to consider the diverse aspects of cultural heritage, such as monuments, sites, movable or immovable assets, tangible assets in general and finally intangible ones. There are several ongoing developments of CRM compatible models<sup>8</sup> covering different specialised fields

Additionally, 3D models have gradually assumed an increasing importance in cultural heritage. 3D models are yet another digital representation for museum objects, but their importance is especially increasing for architectural heritage, mainly due to the architects' practice. Some actually consider the 3D model as the root and build around it the documentation system. This explains why BIM (Building Information Modelling) is being increasingly proposed as the cornerstone of such systems. BIM is a methodology in use to design new buildings which incorporates in the building project all the necessary information about services such as water, electricity, heating networks, and more. It develops a very simple semantic of objects and of their parts, to be able to avail of graphical libraries in the 3D model construction where the necessary information such as material, characteristics, industrial producer (in new buildings) and assembling/construction phases is attached to the building parts, as walls, pillars, floors, ceilings, and so on. BIM models are based on an industry standard, IFC<sup>9</sup> (Industry Foundation Classes), a standardized, digital description of the built asset industry, and an open, international standard ISO 16739-1:2018. A pillar of a BIM model is the 3D representation of the building as a wireframe/solid model, i.e. a skeletal representation of it as produced by well-known tools, among others by Autocad. It consists of various points, arcs, lines, circles, and curves to clearly denote object edges and depth. Information about the various components – structured according to the IFC standard – is attached to this model. The application of BIM to cultural heritage has led to HBIM, standing for Heritage BIM. An extensive survey of HBIM applications is provided by [11], while [12] focuses on applications to archaeology.

In parallel, there has been a number of proposals to use a point-cloud 3D model as the graphical support and add annotations on it concerning various aspects of the object. This kind of application is usefully applied to conservation and restoration. Among others, it is worth quoting the model used for the restoration of the Nettuno fountain in Bologna [13] and the one currently in use for the Notre Dame restoration [14]. A recent paper proposes to reconcile HBIM and point-cloud models via AI [15].

In conclusion, heritage documentation based on 3D models, either point-cloud ones or BIM-based ones, considers the volume and the shape of the heritage asset to be documented as the root of all the documentation systems. This approach does not allow – at least so far – to consider a large part of documentation that cannot be appended to any physical component and severely limits the interoperability of such documentation systems. Documentation based on the shape/volume of assets generates the fragmentation of the documentation system, as each item is documented separately. It does not allow searching across different assets, for example comparing materials, restoration techniques and so on. Also the intangible component of heritage is generally overlooked.

---

<sup>7</sup> <https://cidoc-crm.org/>

<sup>8</sup> <https://www.cidoc-crm.org/collaborations>

<sup>9</sup> <https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/>

Finally, no complete semantic model is available so far for intangible heritage or for the intangible component of the tangible one.

### **3. The way forward**

In conclusion, there appears to be a need to improve the semantic apparatus of cultural heritage documentation. We suggest starting from and extending the CRM that is nowadays an internationally accepted standard. This will preserve the data interoperability across different documentation systems as they will all be compliant and based on the CRM. Even if two documentation systems introduce different specialization of CRM concepts, it is always possible to link elements across the two systems, availing, if necessary, of the more general CRM concept from which the two specializations are derived. This is what ARIADNE and 4CH are doing, the former by creating mappings between the original metadata schema of each dataset collection to be aggregated and a common ARIADNE one. 4CH is still in the design phase of its knowledge base and needs to incorporate provisions for heritage science and conservation data.

The need for an overarching documentation schema is motivated also by more general considerations. If different ontologies were used, interoperability might be put at risk unless mappings are provided. But the most important disadvantage would concern the methodological integration as advocated by the tandem metaphor: an overarching knowledge organization approach, common to the entire community of use, suitable for the different needs and adapted to the many existing research questions, is the first step (the bicycle, in the metaphor) for a real digital transformation of cultural heritage towards a common methodology. Such an approach must be suitable for all heritage documentation applications, from archaeological research to conservation. It needs to include the semantics required by a diverse research community. It must reconcile the different starting points and perspectives, from the documental one to the one organizing contributions of sciences like chemistry and physics, and moreover incorporate the architect's graphical approaches putting at the centre of documentation the shape, almost always interpreted as a 3D replica. We propose a unified methodology using the digital twin concept.

### **4. What is a digital twin**

The concept of digital twin is not new. Digital twins have been applied in many industrial fields where the idea was born to test components, devices and, later on, to simulate the real behaviour of complex appliances in a digital way [16]. Then, digital twins made their way in machinery control applications, by using sensors surveying the behaviour of a device and sending an alert when an anomalous value is measured, or directly activating specific components to return to a normal condition. This kind of application required the use of simulation processes within the model, which were eventually incorporated in the concept of digital twin. Thus, for industry digital twins, the data component has a relatively simple schema, while the process part is more complex. The stress is on how the system behaves rather than on how the information about the system is structured. The European Commission has recently proposed to create a digital twin of the Earth [17] to evaluate complex environmental processes and their impact on the whole system, and to forecast the effects of mitigating measures against, for instance, global warming.

There are many digital twin definitions which put in evidence different characterizations according to the intended use. Two papers, [18] and [19], analyse many of them, relate them to different purposes and try to summarize them into a single overarching approach. None of the uses considered concerns cultural heritage. The 2020 paper by David Jones et al. titled "Characterising the Digital Twin: A systematic literature review" [18] mainly considers industrial applications, analysing in great details the features of such applications, with a short section dedicated to BIM. This thorough study may be useful in the future also for cultural heritage applications, as it analyses many aspects potentially to be considered when the interconnections between the dynamics of the real world of cultural heritage and those in the virtual world of digital twins will be modelled. Although the paper does not state a formal definition, its systematic approach collects and compares different naming conventions and may be instrumental to setting up a precise naming system for heritage applications.

Also the 2021 paper by E. VanDerHorn and M. Sankaran titled "Digital Twin: Generalization, characterization and implementation" [19] is based on a systematic literature analysis. The paper rightfully argues that "many

of the definitions in the literature combine a definition with specific characterizations about Digital Twins that are unique to the use case(s) that they are describing”, which creates confusion about the digital twin general definition. The overarching definition they propose is the following one:

*[A digital twin is] a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual system.*

This shows that an information update is a substantial component of a digital twin system. However, we believe that the two phases of “virtual representation” and “update” are better analysed separately, in our case, where there is no straightforward way even for the first one.

The digital twin (DT) concept may also be described using the five-dimension description introduced by Qinglin Qi *et al.* [20]. Although still in a manufacturing perspective, their definition of a DT model may be useful also in the cultural heritage domain. They define a Digital Twin as made across five dimensions: Physical Entities (PE), Virtual Models (VM), Services (Ss), Data (DD), and Connections (CN). We would rename PE as *Real-world Entities* (RE), since this also encompasses the immaterial components of a real-world object. *Virtual Models* and *Services* include the behaviour of the Digital Twin according to the impulse of services: for example, how the Virtual Model of a building behaves when it receives the digital effects of an earthquake simulated by a Service. For cultural heritage, this kind of interaction is so far considered and processed outside the digital environment, making the virtual model still a descriptive one rather than a dynamic one. It is likely that in the future also simulation services will make their way into the heritage digital twin applications. *Connections* are the mutual interaction between any of the other four dimensions.

Digital twins are nowadays extensively used among others in mechanical engineering, architecture and especially in the building industry, where they belong to the BIM approach. A recent important project combining BIM and digital twins is the UK Gemini project, proposing to use digital twins nationwide for town planning [21].

Since the architecture domain has a close similarity with the cultural heritage (CH) one, such proximity has promoted the development of HBIM, i.e. Heritage BIM, which incorporates the BIM approach enriching it with additional classes pertaining to the heritage domain, but still within a flat data system. It seems that at present it is almost impossible to incorporate a much wider set of concepts – including some that are of a non-physical nature, the so-called intangible heritage – and relations among them. Recently, some authors [22, 23] suggested that an HBIM model is actually a Digital Twin and used it to develop applications in conservation of built heritage assets. For references to HBIM, see e.g. the bibliography published in [12] for archaeological applications and in general the survey [11].

In our opinion, the heritage domain requires much greater attention on how information is organised, an essential step before continuing to computational modelling that simulates real-world processes, the primary focus of most industrial digital twins. That is why we focus the present research on the data organisation, i.e. the semantics of digital twins, and for the time being set apart the digital simulation component so popular in the manufacturing applications. Thus, in this paper a *digital twin* is considered as *the complex of information about digital counterparts of real-world heritage objects, both material and immaterial ones*. We consider simulation as a distinct aspect, which includes processing heritage data in a computer environment, by creating virtual experiments on digital heritage objects in a virtual research environment to simulate the behaviour of the corresponding assets in the real world, or by availing of processes that get data from the real world via sensors, process such input in the virtual world using the features and data of the virtual models and, according to the results, update the virtual models and trigger real world actions via actuators. The data organisation within such a digital twin system has a relatively stable – but not immutable – design. Processes based on consuming the data to perform a simulation or other required operations may instead be created ad-hoc to address specific problems. Thus, the Heritage Digital Twin (HDT) discussed here does not (yet) consider dynamic interactions between Virtual Models activated by Services. It instead creates static Virtual Models based only on the knowledge derived by physical entities.

In sum, at present the Heritage Digital Twin is formed only by the knowledge about Real-world Entities, stored in digital format in the Virtual Model and the inclusion or extraction of such knowledge. Nevertheless, the information stored in the HDT is rather complex and needs a correct semantic structure, which is the purpose of the present paper. Services and Connections are not addressed so far. Indeed, introducing such dimensions would be useful for cultural heritage documentation and management, above all for its

conservation. We plan to address these facets when the ontology is stable and has been reviewed by the community.

Our approach improves the current HBIM approach, strongly influenced by its CAD derivation and its architecture-civil engineering provenance. In HBIM, the Virtual Model consists of a (virtual) shape, almost always a 3D model, optionally divided into parts – e.g. the roof, the columns, the walls, etc. Each of them has some attributes such as colour, material, technical features, and so on. The corresponding semantic graph is a tree, having as root the graphical entity, typically a 3D model. With this approach, a set of information about a building would be worthless if no drawing is included, regardless of the richness of such information, while a 3D model would be worth consideration even if no information is included besides shape. This approach is equally present in the literature about CAD-based HBIM and in the other branch based on point clouds, where virtual models are called “augmented objects” and the respective non-graphical data often described as “annotations”, a name that perhaps underlines their consideration as ancillary information compared to the graphical one. An interesting approach is proposed by CHER-Ob [24], which adds the need for additional information to the usual 3D annotation-based system. This approach is based on the concept of Cultural Heritage Entity (CHE) which corresponds to our Heritage Digital Twin. On the other hand, the main goal of CHER-Ob is more limited than ours, as it focuses on producing visual content for storytelling from available data, in what they call a *project*, i.e. a study on one or more CHEs addressing specific research questions. Information within a CHE is labelled according to the Getty Categories for the Description of Work of Art (CDWA) [25].

Differently from the 3D-based data models described in the current literature on heritage applications, we present here a semantic model, the Heritage Digital Twin ontology, in which there is no privileged class. The shape of an object is an important feature when it exists, but with no higher rank compared to others. For the above-mentioned reasons, we define the whole digital representation of the real-world CH ‘object’ as its digital twin, which consists of the aggregation of different components, among others its shape represented by a visual model. This approach as an extension of the CRM model also allows the creation of digital twins of immaterial ‘objects’ – having no shape by definition – i.e., in our case, intangible cultural heritage; of stories about heritage; and of people’s relationships with tangible assets. Our proposed ontology incorporates the HBIM approach, which covers only a part of the information, and the 3D annotation model. A final consideration concerns the impact of time on digital twins. In industrial applications what matters is the present state, which can change but only the latest state is relevant. On the contrary, in cultural heritage applications information about past states may be relevant as the information of current ones. For this reason, we will timestamp all the information, i.e. consider a time span of validity. By default, this time span is “always”, unless stated otherwise. At implementation, some simple service may automatically change this default assignment, for example always putting as beginning the construction date or putting as end the destruction date. Other cases concern documented changes, which may also be automatically dated. Time is defined according to ISO 8601-1 and ISO 8601-2. Note also that PeriodO<sup>10</sup> offers solutions to convert named periods (e.g. *The Renaissance*) into time intervals.

The approach presented here was initiated in a seminal paper [26] by some of the authors of the present one.

## 5. The Heritage Digital Twin: a discursive introduction

A necessary premise concerns the overall approach to the definition of the Heritage Digital Twin (HDT) ontology. The HDT ontology is based on the CIDOC CRM and its ecosystem, using whenever possible, the classes and properties of the CIDOC CRM and its extensions and defining new ones only to describe more specific concepts such as cultural entities, digital twins, 3D models, which do not find an exact match in the CIDOC CRM. In any case, as far as possible, we have always tried to derive the new entities from those of the CIDOC CRM in order to keep our model completely compatible, consistent and aligned with it. Introducing new classes maintains compatibility and interoperability when the new class is a (proper) subclass of an already existing one: if not possible at the subclass detail level newly introduced, interoperability will still be maintained at the superclass existing one.

---

<sup>10</sup> <https://perio.do/en/>



In the real world there are **Heritage Entities**, regarded as valued by a community – from the whole of humanity represented, for instance, by worldwide organizations as UNESCO, to smaller ones such as a group of believers. We do not enter into the debate of what is cultural heritage: whatever it is, the system is able to take it into account. A distinction is made between Tangible Entities and Intangible ones. A Heritage Entity may be composed from both tangible and intangible entities. A purely intangible heritage entity may be recognized as it has no tangible component.

The corresponding complex of (digital) information about a Heritage Entity is its **Heritage Digital Twin (HDT)**. The term “complex” used above does not imply that the information must be stored in the same digital device: it can be distributed in different storage as long as they are accessible from the “main” system managing the HDT and possibly integrating on demand such a distributed knowledge. Thus, in such a main system information may consist in the actual data or in the data URI, resolved on demand when necessary. This is the result of a holistic approach, incorporating, putting into order, and relating to it all the digital information pertaining to that Heritage Entity, and possibly to others as well. Thus, the Heritage Digital Twin is actually a network of relationships among data putting into evidence the connections between those data and a real-world Heritage Entity. There must obviously exist some item belonging to a Heritage Digital Twin, although it might conceptually be considered also when empty, a sort of placeholder for forthcoming digital information concerning some Heritage Entity. For example, a folder in a computer server containing photos, descriptions, e-books, and so on, of a particular monument is an embryonic heritage digital twin: not yet a full-fledged one because the relationships between each folder item and the heritage entity is not explained, it is just “the file (or a symlink to it) that stays in the folder named after the heritage entity”.

The Heritage Digital Twin will then consist of pieces of its own information (e.g. the Twin’s identifier) and of other digital information pertaining to it. Such information is organized as follows.

The main class is **Heritage Entity**, comprising tangible and intangible entities of the real-world regarded as valuable because of their contribution to society, knowledge and/or culture. The tangible and intangible aspects of the *same* Heritage Entity are recorded as **Tangible Aspect** and **Intangible Aspect**. While all Heritage Entities have an intangible aspect, some may not have the tangible one.

As already mentioned, the complex of the information concerning a Heritage Entity forms its Heritage Digital Twin. The latter includes **Digital Representations**, i.e. digital representations of a Heritage Entity such as text documents or visual ones, i.e. a photo, a video, a 3D model, each one with its own class. It also includes **Stories** about the Heritage Entity, i.e. narratives, modelled according to the NOnt ontology [27, 28], which is based on narratology, a formal way to describe narratives. It distinguishes the *fabula*, i.e. the thematic content of a narrative, and the *narration* (also called *syuzhet*), the chronological structure of the events within the narrative.

The resulting HDT ontology is compatible with the CRM and is also extensible. For example, further subdivisions of text documents might be introduced if necessary, distinguishing e.g. among historical documents, scientific documents and so on. 3D models may be further characterized according to type and have their own metadata and paradata as required. Here we introduce only the most general classes, and are planning to present such extensions in a forthcoming paper.

## 6. Technical description of the HDT ontology: introduction

### 6.1 General note on classes and properties

As already mentioned, the HDT ontology is based on the CIDOC CRM and its ecosystem, directly using the classes and properties of the CIDOC CRM and its extensions for entities and properties having identical meaning and conceptual scope also in our domain, such as places, agents, physical objects. In this case the class or property name is preceded by the namespace prefix, which identifies the relevant ontology.

### 6.2 Ontological models used in HDT ontology



The following Table 1 shows the ontological models of the CIDOC CRM ecosystem used to build the HDT ontology and the namespaces (prefixes) used to indicate their classes and properties throughout this document<sup>11</sup>.

Table 1. Ontological models used in HDT ontology definitions

Model	Version	Name space prefix	Description	Classes prefix	Properties prefix
HDT	1.0	hdt	The ontology described in the present paper (i.e. new classes and properties)	HC	HP
CIDOC CRM	6.2.1	crm	A formal ontology for modelling Cultural Heritage information	E	P
CRMsci	1.2.6	crmsci	The scientific observation model	S	O
CRMdig	3.2	crmdig	Model for provenance metadata	D	L
CRMpe	3.1.2	crmpe	The PARTHENOS Entities model	PE	PP
CRMinf	0.10.1	crminf	An Extension of CIDOC-CRM to support argumentation	I	J
CRMba	1.4	crmba	An extension of CIDOC CRM to support buildings archaeology documentation	B	BP
FRBRoo	2.4	frbr	Functional requirements for bibliographic records	F	R
NOnt	1.0	nont	The MINGEI Narrative Ontology		

### 6.3 Events

To model cultural events, traditions and practices, typical of the intangible heritage, we have defined the HC4 Intangible Aspect class, which is declared as a subclass of crm:E89 Propositional Object and is referred to by an HC3 Tangible Aspect class by the property HP5 has intangible aspect. While all tangible heritage has always an intangible aspect, intangible heritage does not necessarily have a tangible aspect.

Instances of HC4 Intangible Aspect describe generic (template) events such as the Palio di Siena and not the individual occurrences of the Palio. The actual individual occurrences are instances of crm:E5 Event, and property crm:P129 is about (is subject of) may be used to link an individual Event (E5) to the generic description of the cultural heritage event (HC4), such as the Palio di Siena race of the present year (E5) which is about (P129) the Palio tradition (HC4). Nevertheless, it is important to define a new property to specify this special link between the HC4 Intangible Aspect and its punctual manifestations. This property is HP6 has manifestation event (event is manifestation of). Notice that, with this minimal requirement model, recurrent event series as specified in [29] may be computed, instantiated and associated to an HC4 instance if needed. The following is an example:

The Palio di Siena (HC1 Heritage Entity), is a horse race (HC4 Intangible Aspect) that is held twice each year, on 2 July and 16 August, in Siena, Italy (crm:E53 Place). Ten horses and riders (crm:E39 Actor), represent ten of the seventeen *contrade* (crm:E74 Group), or city wards.

<sup>11</sup> All the referred models are described here: <http://cidoc-crm.org/collaborations>. The NOnt ontology is described here: <https://dinarratives.eu/ontology/>.

The historical horse race Palio di Siena (HC4 Intangible Aspect) was held again (crm:E5 Event) on 17/8/2022 (crm:E52 Time-Span) after a two-year pause because of the COVID-19 pandemic. The winner was jockey Giovanni Atzeni (crm:E39 Actor).

For events and activities that are not strictly “cultural” and therefore fall outside the immediate scope of HC4 Intangible Entity, but which in any case concern, affect and remain somehow connected to the various cultural entities, the crm:E5 Event class can be used instead. These are activities like conservation, restoration, reconstruction, and natural events like earthquakes and floods. The type of events for which instances of crm:E5 are used can be specified by means of the crm:P2 which has type property of the CIDOC CRM. Open vocabularies, containing the most common descriptions of these cultural heritage activities, can be defined and released within the ontology to be used in combination with the above-mentioned crm:P2 property for a more standardized and complete description.

#### 6.4 Conditions and states

An important information about heritage assets is their condition state. In the CRM this is documented using E3 Condition State followed by P2 has type E55 Type, choosing the latter in a vocabulary of possible states. If this way of assessing the state of an asset seems too generic, a more precise solution is offered by E14 Condition Assessment. This is the activity dedicated to the evaluation of the condition, and its outcome is a report about the condition. For example, a paper titled “Three-Dimensional Creep Analyses of The Leaning Tower of Pisa” on the condition of the Pisa leaning tower was published in 1997 by Dryden and Wilson. This fact can be expressed as follows:

The “Analysis on Pisa Tower” (E14 Condition Assessment) in 1997 was carried out by (P14 carried out by) “Dryden & Wilson” (E39 Actor) in (P4 has time span) 1997 (E52 Time-Span). It is documented (P70 documents) in the document “Three-Dimensional Creep Analyses of The Leaning Tower of Pisa” (E31 Document).

#### 6.5 Stories and Storytellings

In our model, stories are considered as accounts of facts about a certain Heritage Entity, including (but not limited to) descriptions based on documents and on their interpretation. Stories are an integral aspect of the framework of a Heritage Entity (and thus, of its Digital Twin) since they contribute to the construction of its intangible part as they may make it more understandable, interesting, and attractive for the public. A story can be seen as the core of a series of facts and how they happened, as for example the story of the Knossos Palace and its discovery by Arthur Evans or the story of Falconry over the centuries. From a conceptual point of view, a story is equivalent to the concept of *narrative*, i.e., “a story as it happened in reality or in fiction”, as defined in the Narrative Ontology (NOnt) of the MINGEI project<sup>12</sup>. Since this definition is perfectly suited to our concept of story, and furthermore, given that Narrative in the Narrative Ontology is a subclass of E73 Information Object of CIDOC CRM, which provides it an additional level of formal compatibility, this class will be used in our model as it is.

A storytelling, instead, is considered in our model as the way in which the facts composing a story are actually narrated, presented and disseminated. Storytellings, in fact, comprise social and cultural activities of telling, writing and disseminating stories, for the purpose of education, cultural preservation or entertainment, both in oral form and by means of simple or sophisticated techniques aimed at making the narration of a story effective. An example of storytelling is how the history of Falconry is narrated in “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II. From a conceptual point of view, a result of storytelling can be seen as equivalent to the concept of narration, which in the Narrative Ontology “represents the narration of a narrative, i.e. an individual work that tells the events of the narrative through some form of media (text, video, audio, etc.)”. As stated in [27, 28] there can be many narrations of the same

---

<sup>12</sup> <https://www.mingei-project.eu>. Narrative Ontology documentation is available at: <https://dlnarratives.eu/ontology/>

story, focusing on different aspects of the *fabula*, or presenting events in a different order. In the Mingei Narrative Ontology, Narration is a subclass of F14 Individual Work of the FRBRoo Ontology, another extension of CIDOC CRM. Since also in this case, as for the story, we have verified the perfect conceptual and formal overlap of this class with our idea of storytelling, the Narration class will also be used in our model as it is.

## 7. Scope notes of Classes and Properties

### 7.1 Classes

#### HC1 Heritage Entity

Subclass of: crm:E77 Persistent Item

Superclass of:

Scope Note: This class comprises tangible and intangible entities of the real-world regarded as valuable because of their contribution to society, knowledge and/or culture. Instances of HC1 Heritage Entity may refer to real assets of any nature: physical, both movable and immovable, immaterial, or born digital. They can also refer to cultural events, traditions and practices, typical of the intangible heritage, and can be used to describe their features and their extent in space and time. In the case of events, we can create instances of event types. An instance of HC1 can be considered as the entry point for inferring the content of its corresponding HC2 instance, even if by using crm:P148 has component property it can also be used for denoting HC2's components.

Examples: the Knossos Palace, part of the Knossos WH archaeological site  
the Pafos Gate in Nicosia  
the "Palio di Siena"  
the Florence Historical Centre, a WH Site  
the Stonehenge Complex, a WH site  
the Bauhaus style.

Properties: HP1 has digital twin (is digital twin of): HC2 Heritage Digital Twin  
HP2 has story (is story about): nont:narrative  
crm:P70 is documented in (documents): HC6 Digital Heritage Document  
HP9 has visual representation (is visual representation of): HC7 Digital Visual Object  
crmdig:L1 was digitized by (digitized) D2 Digitization Process  
crmdig:L11 had output (was output of): HC8 3D Model<sup>13</sup>

#### HC2 Heritage Digital Twin

Subclass of: crm:E89 Propositional Object

Scope Note: The class consists of the information available in a given system and pertaining to an HC1 Heritage Entity. Every instance of HC1 Heritage Entity is linked to one instance of HC2 Heritage Digital Twin, which provides an archive of the documented history of the corresponding HC1 Heritage Entity. It includes digital representations of that Heritage Entity (e.g. 3D models, images, videos), textual descriptions (e.g. digital documents, narrations or stories), information of the effects on the related HC1 Heritage Entity of events that influenced or/and are related in any way to its state of (e.g. earthquakes, floods etc.) and of activities (e.g. restorations, conservations etc.) carried out on it.

---

<sup>13</sup> This means that an HC1 instance is related to an HC8 instance via the property path crmdig:L1/crmdig:L11.

Examples: the HDT of Pisa Leaning Tower  
the HDT of the Neptune Fountain in Bologna  
the HDT of Knossos Palace  
the HDT of the Pafos Gate in Nicosia  
the HDT of the “Palio di Siena”  
the HDT of the Florence Historical Centre

Properties: HP3 is digital twin component of (has digital twin component): HC2 Heritage Digital Twin

### **HC3 Tangible Aspect**

Subclass of: HC1 Heritage Entity  
crm:E18 Physical Thing

Scope Note: This class comprises tangible, material entities of the real-world, both movable (e.g. archaeological, artistic and cultural objects) and immovable (e.g., built heritage like monuments, buildings, cities and other complexes), regarded as valuable because of their contribution to society, knowledge and/or culture. The “tangible” term in the name of this class does not exclude that its instances also possess an intangible aspect, which is specified through the HP5 has intangible aspect property.

Examples: the Neptune Fountain in Bologna (Italy)  
the Pisa Leaning Tower, a UNESCO World Heritage (WH) Site  
the Nike of Samothrace of the Louvre Museum in Paris (France)

Properties: HP5 has intangible aspect (is intangible aspect of): HC4 Intangible Aspect  
HP7 is manifestation of (is manifested by): HC4 Intangible Aspect

### **HC4 Intangible Aspect**

Subclass of: HC1 Heritage Entity  
crm:E89 Propositional Object.

Scope Note: This class comprises cultural events, traditions and practices having particular social, historical and cultural significance, including practices and expressions, memories and oral traditions about events, things, people.

Examples: the Mediterranean diet  
Falconry  
the Rebetiko music tradition  
the “Palio di Siena”

Properties: HP6 has manifestation event (event is manifestation of): crm:E5 Event

### **HC5 Digital Representation**

Subclass of: crmdig:D1 Digital Object

Scope Note: This class comprises the digital virtual representations of an HC1 Heritage Entity such as e-texts, images, audio or video items, 3D models, etc., that are documented as single units.

Examples: the digital version of Vasari’s “Vite”

the video <https://www.youtube.com/watch?v=P1Uv4Zf5xKk>  
the Pafos Gate laser scanning 3D model

### **HC6 Digital Heritage Document**

Subclass of: HC5 Digital Representation

Scope Note: This class comprises pieces or collections of digital, non-visual documents, either born-digital or digitised from physical, real-world ones, typically containing textual or numerical information regarding an HC1 Heritage Entity and intended to become part of the related HC2 Heritage Digital Twin. Documentation of this kind may include scientific data, research results and interpretation, as well as historical and cultural information, including textual descriptions related to the nature, conditions, positioning and to the whole set of events in which the cultural entity has been involved and the actors who have participated in them.

### **HC7 Digital Visual Object**

Subclass of: HC5 Digital Representation

Scope Note: This class comprises digital visual objects, such as photos and videos, but also special imagery such as X-ray images, spectra of chemical and physical analyses, and so on, intended to become part of the HC2 Heritage Digital Twin of an HC1 Heritage Entity. Digital documentation of this kind can be born digital or digitised from physical objects (such as paper photographs, drawings and so on). Particularly relevant digital visual objects are also Virtual Reality (VR) and Augmented Reality (AR) models, other types of visual digital artefacts pertaining to a HC1 Heritage Entity. Both VR and AR models rely on 3D models of the related heritage entity, but may add or remove parts of it, or require further digital input as in AR, so they should be catalogued separately from 3D models.

Examples: The Europeana digital version of the paper picture of the Pisa Leaning Tower taken by Paolo Monti in 1960  
([https://www.europeana.eu/it/item/9200369/webclient\\_DeliveryManager\\_pid\\_6363979\\_custom\\_att\\_2\\_simple\\_viewer](https://www.europeana.eu/it/item/9200369/webclient_DeliveryManager_pid_6363979_custom_att_2_simple_viewer)).

### **HC8 3D Model**

Subclass of: HC5 Digital Representation

Scope Note: This class is used for rendering in detail the 3D model of HC1 Heritage Entity and intended as a particular crmdig:D1 Digital Object having its definite identity and resulting from operations such as digitization, acquisition, processing and other actions typical of the three-dimensional modelling world (e.g., 3D scanning, wireframe modelling and so on). The particular features of a 3D model (e.g., its type, format, resolution, size, etc.) and its relationships with the series of activities carried out for its creation and manipulation are modelled through the properties inherited from its superclass HC5, which in turn inherits from crmdig:D1 Digital Object, and through the other classes and properties of CRMdig.

Examples: The 3D model of the Neptune Fountain produced by ISTI-CNR (Pisa, Italy) as part of the documentation used for the restoration of the Neptune Fountain in Bologna (Italy).  
<https://www.cnr.it/en/focus/074-43/3d-supported-restoration-the-neptune-fountain-in-bologna>.

## 7.2 Properties

### HP1 has digital twin (is digital twin of)

Domain: HC1 Heritage Entity  
Range: HC2 Heritage Digital Twin

Scope Note: This property links an instance of HC1 Heritage Entity with an instance of its related HC2 Heritage Digital Twin in a given system.

Examples: The Pafos Gate in Nicosia, Cyprus (HC1) has digital twin (HP1) the Pafos Gate digital twin (HC2) created by Cyprus Institute.

### HP2 has story (is story about)

Domain: HC1 Heritage Entity  
Range: nont:Narrative

Scope Note: This property links an instance of HC1 Heritage Entity with an instance of a nont:Narrative that refers to it.

Examples: Falconry has story (HP2) the history of Falconry over the centuries.

### HP3 is digital twin component of (has digital twin component)

Domain: HC2 Heritage Digital Twin  
Range: HC2 Heritage Digital Twin

Scope Note: This property associates an instance of HC2 Heritage Digital Twin with another HC2 of which is component. The term 'component' here is not limited to physical or geographical relationships (see examples), but encompasses any kind of main-associated relationship.

Examples: The HC2 Digital Twin of Pafos Gate in Nicosia (Cyprus) HP3 is a digital twin component of the HC2 Digital Twin of Nicosias' City Walls.  
The HC2 Digital Twin of the "Cento Camini" Medici Villa in Artimino (Florence) HP3 is a digital twin component of the HC2 Digital Twin of the UNESCO WHS Medici Villas in Tuscany.  
The HC2 Digital Twin of Vichy is a HP3 digital twin component of the HC2 Digital Twin of the UNESCO WHS The Great Spa Towns of Europe  
The HC2 Digital Twin of the "Basilica of San Salvatore in Spoleto, Italy" is a HP3 digital twin component of the HC2 Digital Twin of "Spoleto", which is a HP3 digital twin component of the UNESCO WHS "Longobards in Italy. Places of Power"

### HP4 narrates (is narrated through)

Domain: nont:Narration  
Range: nont:Narrative

Scope Note: This property links an instance of nont:Narration with an instance of a nont:Narrative which has this narration. It is similar to the nont:hasNarration property, but is not a subproperty of crm:P148 has component.

Examples: The "*De Arte Venandi Cum Avibus*" treatise by the Holy Roman Emperor Frederick II narrates (HP4) the history of Falconry.



**HP5 has intangible aspect (is intangible aspect of)**

Domain: HC3 Tangible Aspect  
Range: HC4 Intangible Aspect

Scope Note: This property associates an instance of HC3 Tangible Aspect with its intangible aspects (HC4), i.e. the cultural, social and historical value it incorporates.

Examples: The “Theotokos of Vladimir” (HC3) icon HP5 has intangible aspect the secular veneration that is addressed to it (HC4).  
The UNESCO WHS site “Routes of Santiago de Compostela” (HC3) has intangible aspect (HP5) pilgrimage to Santiago (HC4).

**HP6 has manifestation event (event is manifestation of)**

Domain: HC4 Intangible Aspect  
Range: crm:E5 Event  
SubPropertyOf: crm:P129 is about (is subject of)

Scope Note: This property associates an instance of HC4 Intangible Aspect with the instances of the crm:E5 Event (or of the unique and specific crm:E5 Event) through which the intangible entity manifests itself in the physical world.

Examples: The Palio di Siena (HC4) has manifestation event (HP6) the historical horse race that was held in Siena on 17/8/2022 (E5)

**HP7 is manifestation of (is manifested by)**

Domain: HC3 Tangible Aspect  
Range: HC4 Intangible Aspect

Scope Note: This property associates instances of HC3 Tangible Aspect with the HC4 Intangible Aspect of which they are the manifestation in the physical world.

Examples: The set of devotional graffiti engraved on the walls of the Church of the Holy Sepulchre in Jerusalem (HC3) is the manifestation of (HP7) the pilgrimage of which the church is the final destination (HC4).

**HP8 is narrated in document (document used for narration)**

Domain: nont:Narration  
Range: crm:E31 Document

Scope Note: This property associates an instance of nont:Narration with instances of E31 Document used to implement it.

Examples: The “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II (nont:Narration) is narrated in document (HP8) the “MS. Lat. 419” manuscript , now in the library of the University of Bologna E31.

**HP9 has visual representation (is visual representation of)**

Domain: HC1 Heritage Entity  
Range: HC7 Digital Visual Object

Scope Note: This property associates an instance of HC1 Heritage Entity with instances of HC7 Digital Visual Object in which it is represented.

Examples: The Pisa Leaning Tower (HC1) has visually representation (HP9) the Europeana digital version of the paper picture of the Pisa Leaning Tower taken by Paolo Monti in 1960 ([https://www.europeana.eu/it/item/9200369/webclient\\_DeliveryManager\\_pid\\_6363979\\_custom\\_att\\_2\\_simple\\_viewer](https://www.europeana.eu/it/item/9200369/webclient_DeliveryManager_pid_6363979_custom_att_2_simple_viewer)) HC7

#### **HP10 tells about (is told by)**

Domain: nont:Narrative

Range: crm:E5 Event

Scope Note: This property is intended to identify the specific events (E5) to which a nont:Narrative relates.

Examples: The history of Falconry (nont:Narrative) tells about (HP10) the writing of “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II (E5).

### **8. Use cases**

In this section we develop two complete examples of HDT. In the first one, the focus is on art history and on scientific analyses carried out on a painting. The second one describes the monastery of St. John Lampadistis in Kalopanayotis, Cyprus, including the description of intangible aspects.

#### **8.1 The analyses on the portrait of Caterina Cornaro from the Leventis museum of Nicosia**

##### **8.1.1 History of the painting and results of the analyses carried out on it**

The painting was acquired by a private collector (Michael Zeippeki) who donated it to the Leventis museum. It is a 19<sup>th</sup> century portrait of Caterina Cornaro, the last queen of Cyprus [30]. Visual investigations hinted at the existence of an underlayer of paint. Further heritage science investigations (x-ray imaging, digital microscopy, multi-spectral imaging, XRF) confirmed the existence of an earlier, 16<sup>th</sup> century painting [31]. Tobias Lange, a restorer from Dresden, Germany, removed the 19<sup>th</sup> century painting and exposed the 16<sup>th</sup> century one (Figure 1). Many details changed: the crown, turban, breast cover and right shoulder size and orientation.



Figure 1. (a; b; c; d: clockwise from top left): The painting's development: a – as appeared before latest restorations; b – infra-red imaging reveals an underlayer paint with a modified pictorial composition; c – X-Ray imaging displaying in more details these differences: shoulder in a different position, a cloth covering the chest, the veil position shifted, a pointed, un-decorated crown, the turban broader and richer, the ear more to the left; d – the painting as of today, after restorations that removed the upper pictorial layer (added in the 19<sup>th</sup> century) and highlighting the 16<sup>th</sup> century original.

Art historical investigations indicate a lost Titian portrait, which was the source of inspiration for several copies, made by different artists during the 16<sup>th</sup> century and which circulated in Europe. Common to all is the shape of the pointed crown, the queen depicted as a mourning widow with characteristic clothes and a veil.

A close examination of the painting revealed some details implying relevant information about the painting. **Boldface** indicates the parts developed in the semantic example.

Table 2. Results deduced from observation of the painting and from literary references, in boldface those modelled with HDT in the example

Observation	Inference
<b>Painting canvas nailed on stretches</b>	<b>Painting cut from a larger one</b>
No signatures	A 19 <sup>th</sup> century painting overpaint on a 16 <sup>th</sup> century one, unknown artists
Old adhesive on the retro - GK9654	GK = Royal Prussian Cataloguing system
Old adhesive on the retro in German – mentioning the collectors	Edward Solly was an English collector focusing on Renaissance paintings – he donated his collection in the 19 <sup>th</sup> century to the Royal Prussian house.
Observations on a cross-section sample from the stretcher with a transmitted light microscope (technical details mentioned)	Wood structure indicates poplar ( <i>Populus</i> ), widely used in S. Europe, but not N. Europe, to make stretchers, indicating that it was probably made in Italy.
References in various books and collection inventories (Vasari's Lives, Colbertaldo's Storia di Caterina Cornaro) on paintings	Portraits of the queen by several Venetian artists, such as Titian, Giorgione, Tintoretto, Paolo Veronese, etc., done from imagination.
<b>Carlo Ridolfi writes in his Le Meraviglie dell' Arte on Titian's work "... with the same way the Queen Caterina Cornaro is depicted in widow's dress that lets the whiteness of the skin rise through the black..."</b>	<b>Indication of a now-lost painting of Titian with an iconography of "Caterina Cornaro as a widow".</b>
Marin Sanuto – chronicler and member of the Major Council of Venice mentions in his diary that upon her arrival in Venice, the queen wore "a black velvet dress, a veil on her head and Cypriote jewellery".	Hanging pearls as "Cypriote jewellery", the black dress, the almost transparent veil and the chest cover became a fashion of depicting Cypriot nobility during the 16 <sup>th</sup> century paintings in Cyprus.
Use of gesso on a tabby waved canvas, brown ochre ( <i>imprimatura</i> ) on the preparation layer, paint layer applied with fine brush strokes.	Techniques of a Venetian Renaissance painter.
Canvas of thin linen fabric in simple tabby weaves.	Produced in the Low Countries and widely used in Venice during the 16 <sup>th</sup> century.
Thin gesso layer on the canvas.	Fills the depressions in the weave.
Brown underpaint layer on top of the gesso.	Typical of 16 <sup>th</sup> century paintings.
No underdrawing beneath the painted surface following IRR and X-ray analysis.	Outlining of forms with a brush and fluid black was common to Titian's practice as evidenced on his autographed paintings.
No outlining of the canvas for a duplicating process, which would serve to duplicate the portrait to another surface.	This suggests a creation after a prototype, perhaps a repetition, worked with a free hand.
Later overpaints in various areas of the painting, such as re-positioning of the ear, change in the shape of the crown and turban, shoulders and the covering of the chest with a white fabric.	Efforts to bring the composition to the fashion of the 19 <sup>th</sup> century, as well as to centralize better the figure, since it was initially cut from a larger painting.
Overall composition of the 16 <sup>th</sup> century painting is of high quality.	As described by C. Ridolfi, this painting probably coincides with the times of Titian, but unclear by whom and is among the earliest among the queen's portraits.

Table 3. List of Heritage Science experiments.

Instruments and environmental conditions		Reason
Imaging	Ricoh WG-30 digital camera	Delineation of composition details on pictographic layers.
	Hirox KH-8700 digital microscope with a magnification range of 5×–2500× with dual illumination revolver zoom lenses.	
Spectroscopy	XRF - ARTAX-200 $\mu$ -XRF Bruker, Mo X-ray tube operated at 50 kV and 700 $\mu$ A, CCD camera with sample illumination and laser spot, a silicon drift detector with a resolution of <150 eV and a 0.65 mm collimator. Acquisition time = 90 s. Measurements taken in air atmosphere and with a Mo filter. The energy to channel calibration was done with a bronze standard, using the Cu- and Sn-K $\alpha$ lines; the Mn-K $\alpha$ line of a manganese standard for the full width at half maximum calibration. Spectra acquisition and evaluation were done with Spectra 5.3 software.	Identification of pigments used for all pictorial and preparation layers.
SEM	SEM-EDS - TM3000 Hitachi, with a Backscattered Electron (BSE) detector coupled with EDS for elemental analysis. The system was operated at 15 kV at a working distance of 8.5 mm; images were taken at magnifications of 15× to 30,000×.	Painting stratigraphy

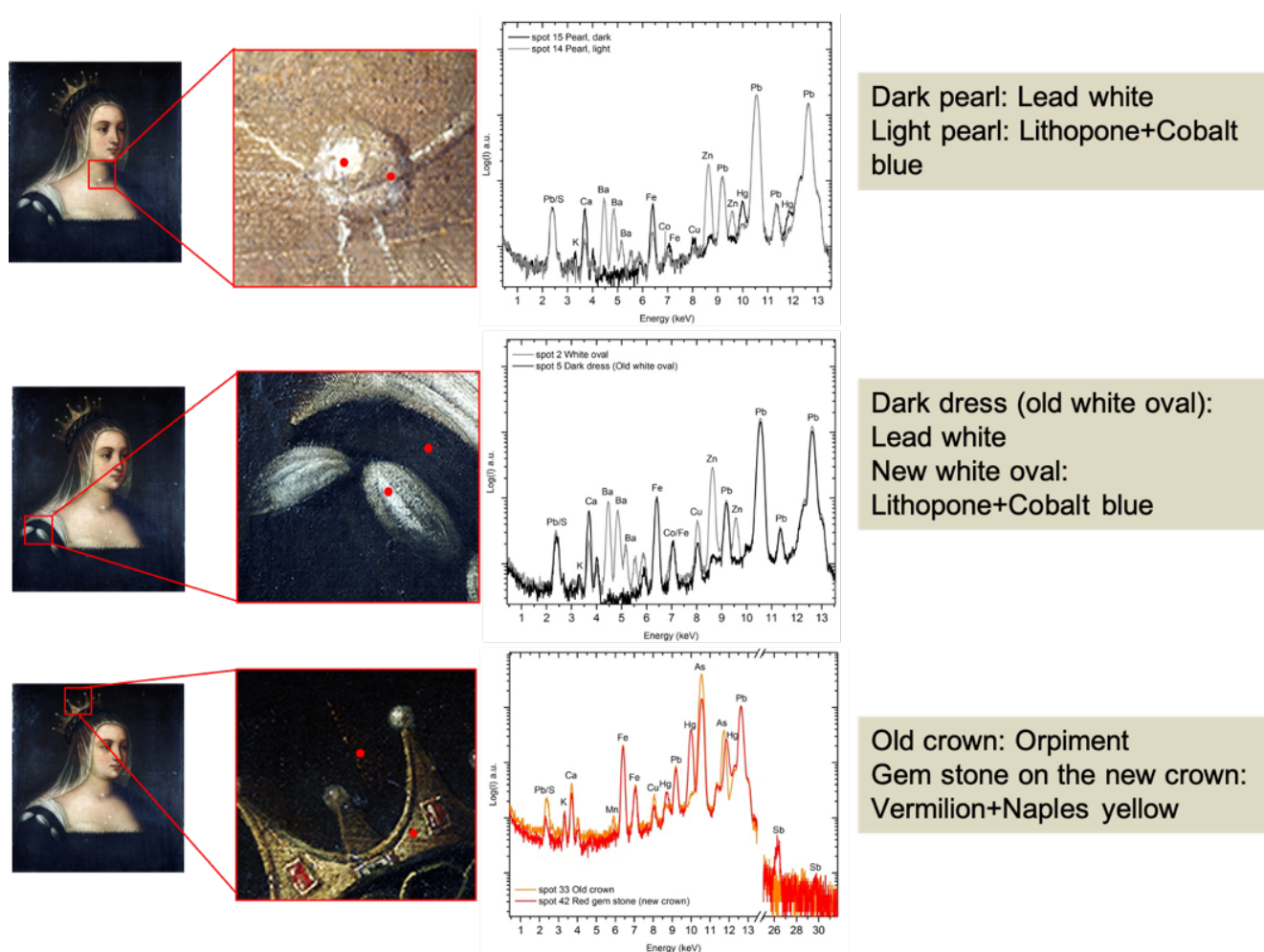


Figure 2. X-Ray fluorescence analyses of the various spots on the painting and suggested pigments used.

Table 4. Results of the analyses and conclusions based on them.

Observation		Inference
EDS	Intense Ca and S lines	Preparation layer of canvas– a thick white layer of calcium sulphate, either anhydrite ( $\text{CaSO}_4$ ) or gypsum ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ).
	Pb, Fe, Ca, Mg, Na, Al and P.	Brown preparation layer called <i>imprimatura</i> , characteristic of Renaissance paintings consisting of a mixture of lead white, earth pigments (ochre), ivory black and dolomite (?).
A thin organic layer.		Residues of varnish from the original painting.
Two uppermost transparent layers, recognizable under UV light microscopy.		Layers of varnish of the overpainting.
Staining tests on the painting cross- sections detected a proteinaceous material in a paint layer of the overpainting as well as in the Ca-containing white ground layer.		Presence of a glue binder.
XRR	Bright areas.	X-ray opaque pigments composed of heavy elements.
	Dark areas.	X-ray pigments composed of light elements.
XRF	Basic lead carbonate, intense Pb lines	<b>Lead white (<math>2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2</math>), characteristic of Renaissance paintings.</b>
	Intense Ba and Zn lines.	Lithopone (barium sulfate-zinc sulfide, $\text{BaSO}_4 \cdot \text{ZnS}$ ). Pigment used in the second half of the 19 <sup>th</sup> century (thus dating the overpaint).

Table 5. Colour Analysis with XRF

	Original paint	Overpaint
White	Lead white ( $2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$ )	A mixture of Lead white with Lithopone ( $\text{BaSO}_4 \cdot \text{ZnS}$ )
Carnation	A mixture of Lead white, Vermilion ( $\text{HgS}$ ), Ochre (iron yellow, brown) and Copper green.	
Yellow	Mainly Orpiment ( $\text{As}_2\text{S}_3$ ). Small intensity of Sn lines might be related to small additions of a yellow tin pigment, such as Lead-tin yellow ( $\text{Pb}_2\text{SnO}_4$ ). Small intensity of Ag lines might be related to admixtures of silver in natural orpiment.	Mainly Lead yellow (Litharge $\text{PbO}$ )
Red	none	Vermilion for the lighter and paler red, Vermilion with an antimony pigment such as Antimony red ( $2\text{Sb}_2\text{S}_3 \cdot \text{Sb}_2\text{O}_3$ , superposition or mixture) used to highlight the borders of the gemstones which appear brighter and darker.
Green		Copper green
Black	Iron-Manganese black (Umber) and Copper (green) pigments	

Each statement from the tables above is supported by available datasets.



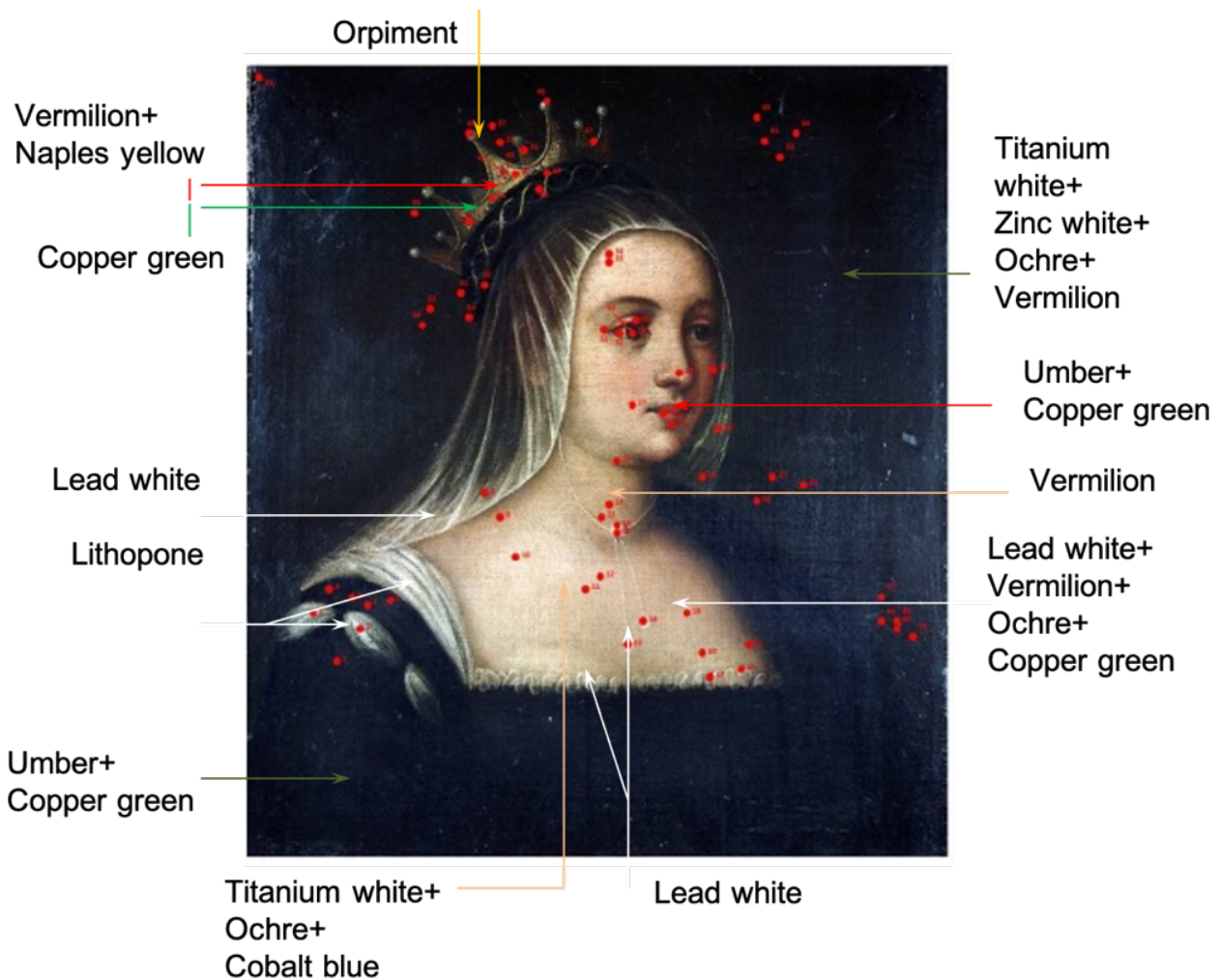


Figure 3. Distribution of pigments on the painting, demonstrating a mixture of original 16<sup>th</sup> century ones and later, 19<sup>th</sup> and 20<sup>th</sup> century pigments

### 8.1.2 Modelling Art History information

As a preliminary test, we modelled two of the results proposed above (highlighted in boldface). The first concerns a direct observation of the condition of the painting, the second concerns information derived from a book in which the painting is probably mentioned.

In particular, in the first case, some physical peculiarities are observed ( $HC1 \rightarrow O8 \rightarrow S20$ ) on the painting canvas ( $O9 \rightarrow S9$ ) which allow researchers to make hypotheses ( $J2 \rightarrow I2$ ) about past conditions and provenance of the object.

In the second case, from Carlo Ridolfi's book (E31), in which the object is mentioned ( $HC3 \rightarrow P70$ ), we read that Titian painted a portrait of Caterina Cornaro, now lost ( $J7 \rightarrow I7$ ). This leads to the hypothesis ( $J2 \rightarrow I2$ ) that this portrait could be the lost one painted by Titian.

Finally, the two  $I2$  Belief(s) resulting from the above observations (together with the others from Heritage Science described below, and others deriving from further investigations and analyses not modelled here) can then be used as a premise ( $J1$ ) to support further inferences ( $I5$ ), such as the attribution of the painting to Titian ( $I2$ ).

The full model is shown in Figure 4.

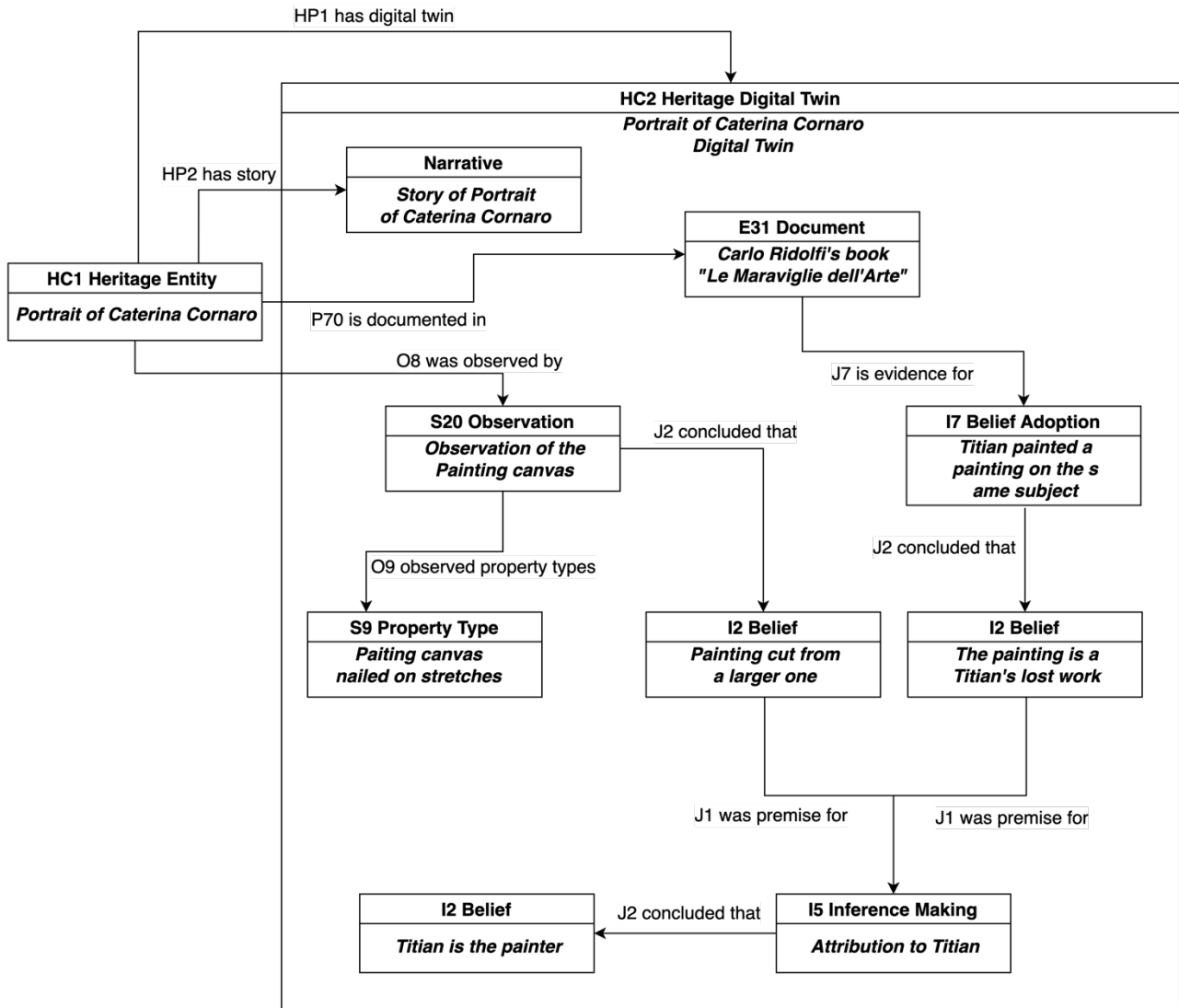


Figure 4. Modelling Art History research activities

### 8.1.3 Modelling Heritage Science results

For the Heritage Science test, we chose spectroscopy and x-ray imagery as representative examples of scientific investigations (also highlighted in boldface in this document, see above). In this case, the information derives from a spectroscopic examination of the painting (HC1 → P39 → D11), intended to identify specific pigments (P17 → E7), that was performed using a specific device (L12 → D8) equipped with additional physical accessories (P16 → E19), operated by specific software (L23 → D14) and using a series of settings and environmental parameters (L10 → D9).

The spectroscopic analysis produced a series of results (L20 → D9) which, once evaluated (O16 → S6 → O11 → S9), led the experts to some conclusions (J2) regarding the chemical elements used by the artist as characteristic of a certain historical period (I2), in this case the “Renaissance”.

The painting was also analysed by means of x-ray imaging techniques (L1 → D2) which created an x-ray image (HC7) of it, through which bright areas present on the painting are observed (O8 → S20 → S9) and are associated (J2) by the experts with the use of pigments composed of heavy elements (I2).

Also in this case, as for Art History, the final scientific results can constitute the bases (J1) for the attribution of the painting to a certain author (I5 → J2 → I2), in this case to Titian. The full model is shown in Figure 5.

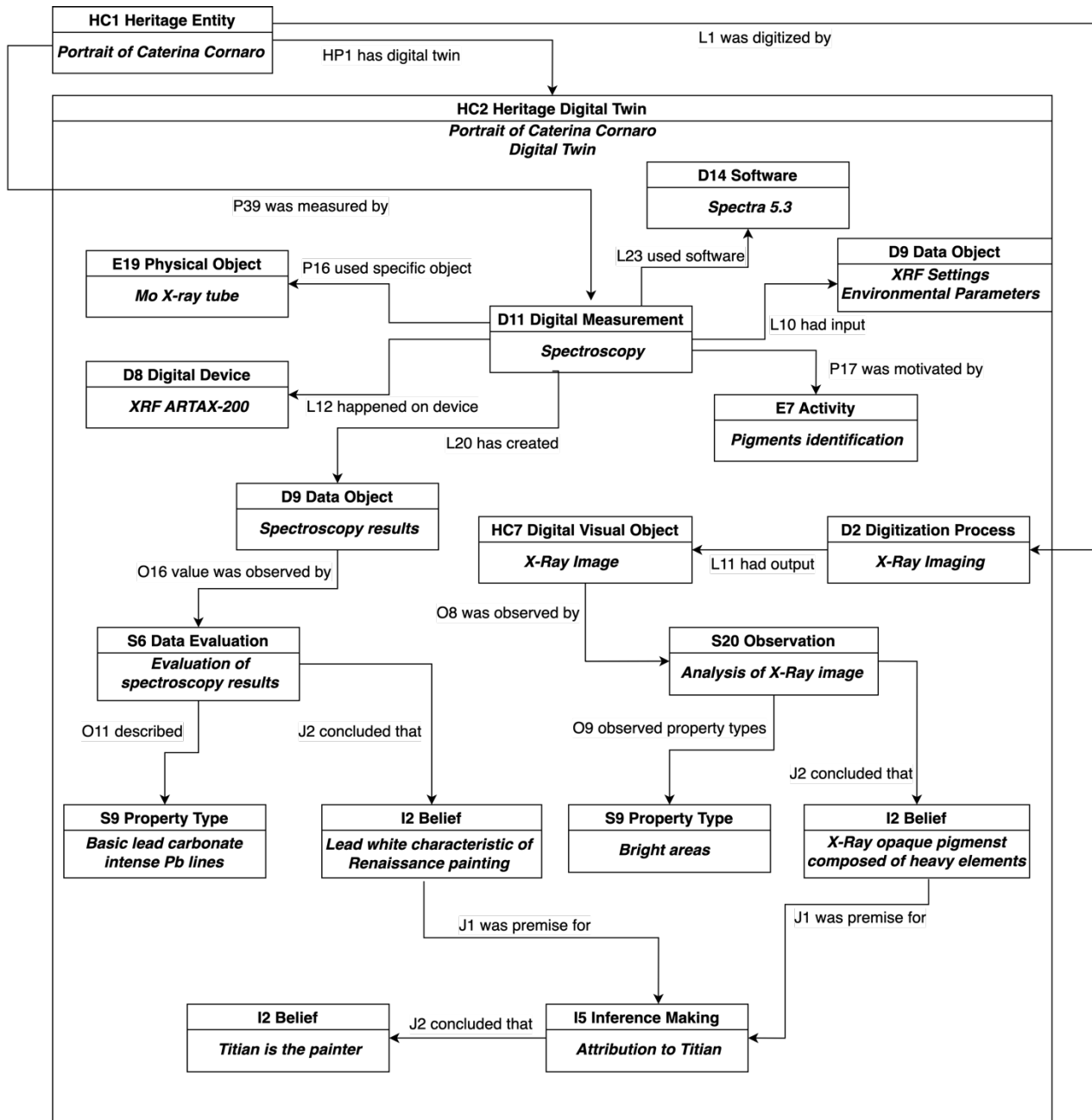


Figure 5. Modelling Heritage Science research activities and their results

## 8.2 The monastery of St. John Lampadistis in Kalopanayotis, Cyprus.

### 8.2.1 Tangible aspects

The monastery of St. John Lampadistis consists today of a group of buildings built and renovated in different periods. While its founding date is unclear, the *katholicon* (the monastery church), dedicated to Saint Herakleidios, born in the village of Lampadistis and the first bishop of Cyprus, is dated to the 11th century. Among the wall-paintings of the narthex an inscription, dated to the 15th century, describes this church as a “*katholiki*”. While written sources indicate the monastery was in use until the 19th century, afterwards it has been used as the main church of the village. Apart from the complex of the three churches under one roof, a phenomenon unique for Cyprus, there are other monastic buildings including cells, auxiliary rooms and an oil press. One of the rooms is used today to house icons from the monastery as well as other churches of the village of Kalopanagiotis [31].

The main monastery church is a domed cross-in-square structure, dated to the 11th century. In the 12th century the vaulted chapel of St. John Lampadistis was added to the north of the first church, above the tomb of the Saint. Relics of the saint, kept and displayed in a rich reliquary, and his tomb are now incorporated in the church as part of this chapel. This second chapel collapsed and was almost entirely rebuilt in the 18th century. In the middle of the 15th century a common narthex was built to the west of the two churches.

During the second half of the 15th century a vaulted chapel was added to the north of that of St. John. It became known as the 'Latin chapel' because of the assumption that it was built for the Latins (Catholics) and indicating the possible co-existence of the two rites in the same church, in line with the atmosphere of tolerance which prevailed in Cyprus after the Council of Florence (1439). Sometime between the 15th and the beginnings of the 18th century), a timber roof covered with flat hooked tiles sheltered the entire roof complex. As a result of its tripartite character, the building acquired an external image of a large building covered with a timber roof.

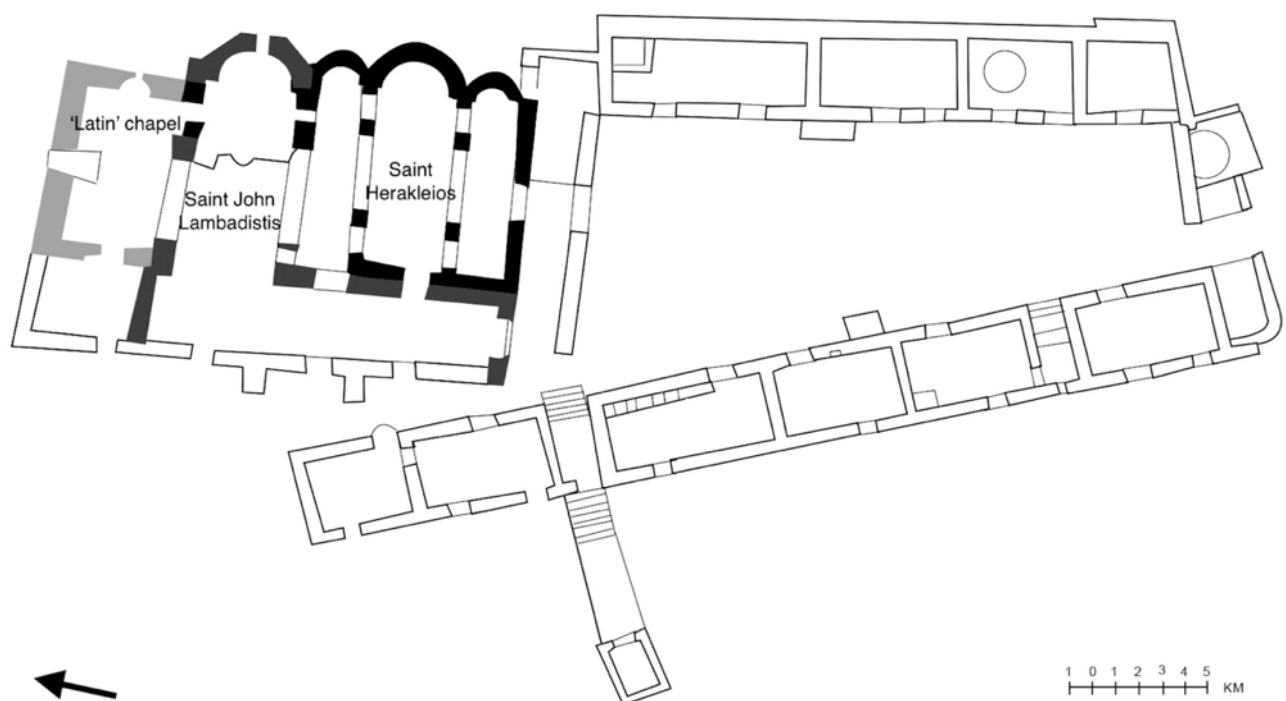


Figure 6. The plan of St. John Lampadistis complex (after [32], Figure 4)

The wall-paintings of the church are primarily from the 11th to the 15th centuries. The apse of the church, as well as some other parts, preserve fragmentary scenes dated to the 11th and 12th century. The rest of the church was painted in the 13th and 14th century. The narthex was decorated in a later period and is the work of an artist from Constantinople, who fled to Cyprus after the fall in 1453. These wall-paintings follow the trends of the Byzantine capital. The 15th century frescoes of the 'Latin' chapel display Byzantine and Italian Renaissance elements. A wooden templon screen, the oldest in Cyprus and dated to the 13th-14th century, is painted with decoration imitating coats-of-arms.

### 8.2.2 Intangible aspects

#### *The life of St. John Lampadistis*

A now lost 1640 manuscript written by a priest named Savvas from the village of Agios Theodoros Agrou and later copied by monk named Kirililos, of the Stavrovouni Monastery in 1903 narrates the life of St. John Lampadistis and the prayer said in church on his feast day on 4th of October. St. John was born in the village of Lampadistis, a now extinct village presumably located somewhere between the modern villages of Galata and Kakopetria, being the last offspring of the priest's village, Papa Kyriakos and his wife, Anna. His birthdate is unknown, probably sometime in the early 17th century. Early in his life, he was sent to learn to read and

write through the study of the Holy Scriptures, the child showing great aptitudes. At the time of his adulthood, a marriage was arranged for him, but the parents of his future wife served him poisoned fish, causing him to lose his eyesight. Now being unsuitable to marriage, St. John turned to spiritual life and spent his days in prayer. After 12 years of living as a blind person, he died. Soon people reported light beaming from his grave. Word spread and people known to be possessed came seeking John's body to pray to it and be cured. Following pressure from the people, his father agreed to open his son's tomb and found his relic's, while his heart was preserved "like a dry fig". Hence, John's remains were deposited in the chapel of St. Herakleidios at Kalopanayiotis. The many miracles attributed to John earned him the status of a saint. His father, hearing a voice telling him to build a church dedicated to his son, which he did and the church of St. John Lampadistis was built adjacent to the one of St. Herakleidios. St. John's reputation as a miracle maker expanded and the inhabitants of the island decided to view him as their guardian saint.

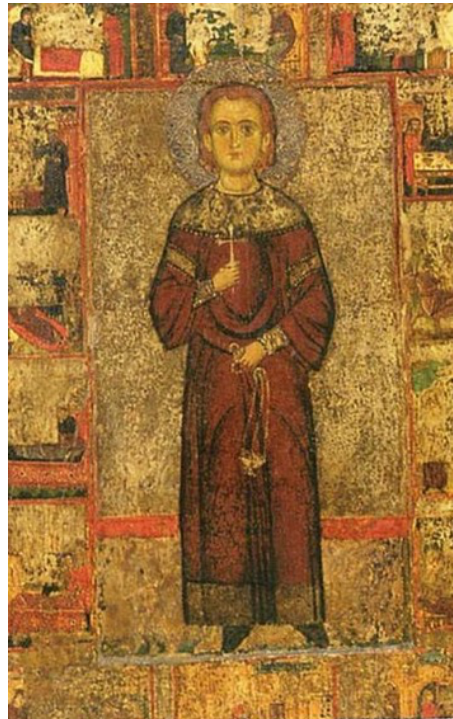


Figure 7. Vita icon depicting St. John and central episodes from his life

#### *Pilgrims and their testimonies - the Karamanlides*

The Karamanlika are a Turkish-speaking population of Greek Orthodox rite originally from Asia minor. During the 18th century a group of them, while on pilgrimage to the Holy Land, passed by the monastery of St. John Lampadistis and left several written testimonies of their passage on the church's wall close to its reliquary, the earliest being dated to 1749 and the latest to 1880, with a highest frequency between 1770 and 1780.





Figure 8. The Saint John Lampadistis reliquary, showing graffiti carved on the wall by Karamanlika pilgrims

### 8.2.3 Digital models

The 3D model of the monastery and related historical imaging documentation are available here:

<https://modelier.us.aldryn.io/models/d10a14ee-4192-45b6-b54b-8d11ecd6f70e/v2/embed/>

and a 360 panoramic walk-through is available here:

<https://dioptra.cyi.ac.cy/sites/360/lampadistis/>

part of a broader online-accessible database of the Troodos complex monasteries registered as World Heritage Sites <http://ihat.cyi.ac.cy/?q=Collection>

### 8.2.4 Ontological modelling examples

From the ontological point of view, the information about the structure of the monastery of St. John Lampadistis can be modelled by combining the classes and properties of the HDT Ontology with those of CIDOC CRM and CRMba. The monastery (HC1) is in fact made up of various buildings (P46 → B1) including the central church (*katholicon*) where the frescoes are found. Each building can in turn contain several building sections (BP1 → B2) clearly identifiable by their morphology, and of each of them it is possible to specify and describe the construction phases (P108 → E12) and the relative dating (P4 → E52). As far as the frescoes on the walls are concerned, it is possible to identify those pertaining to each section of the building (P56 → E26) so that even for them the phases of realization (P108 → E12) and the dating (P4 → E52) can be clearly distinguished. Figure 9 illustrates the case of the modelling of the Latin Chapel and its frescoes, highlighting, through the distinction between the two different production activities (E12 → P4 → E52), how the painting of this section of the building is coeval with the construction of the chapel itself (both belonging to the 15th century, in this case). The monastery is also visually represented through various digital technologies (3D models, 360 panoramic views, etc.) which can be perfectly represented through our modelling tools either with (P129 → HC8) as in Figure 9, or with (L1/L11 → HC8) when describing the digitisation process is needed.

The HDT Ontology and the CIDOC CRM ecosystem can be also employed to model the intangible aspects of the monastery, represented by its history (HP5 → HC4 and HP2 → Narrative) and linked in particular to the life of St. John Lampadistis, narrated by the priest Savvas (HP4 → Narration and HP8 → E31) in a manuscript (P128 → E18) that is now lost (P44 → E34). It is also possible to model the information related to the graffiti engraved on the walls of the monastery church by the Karamanlika pilgrims (P56 → E26), to specify their approximate dating (the 18th century in this case)(P4 → E52), and to link the images that depict them (P138 → E36).



Finally, the complex of information described above and their mutual relationships form the elements that make up the Heritage Digital Twin of the monastery of St. John Lampadistis. The HDT Ontology establishes this fundamental relationship via the property HP1 has digital twin, used to bind each Heritage Entity (HC1) to its digital representation (HC2 Heritage Digital Twin).

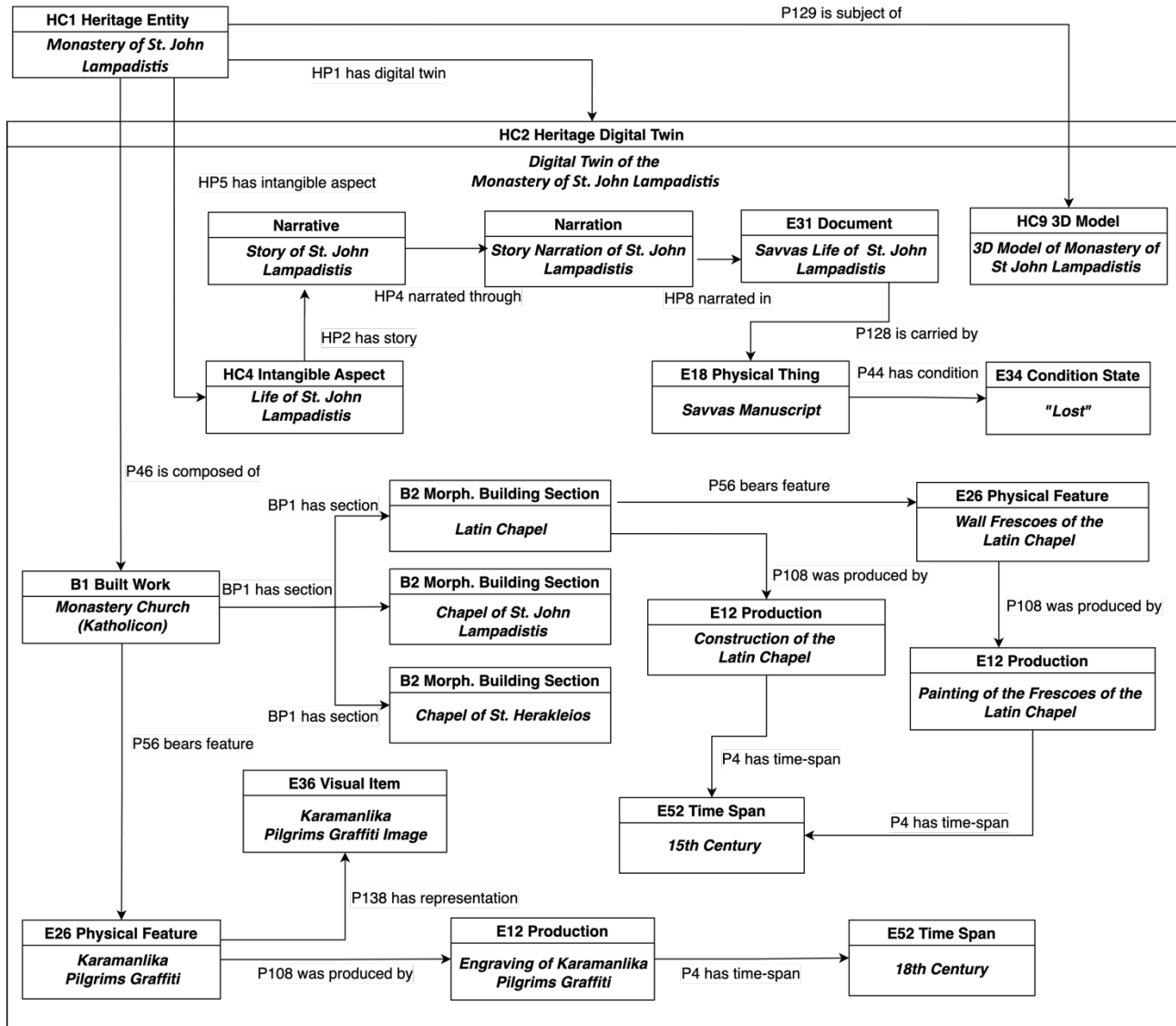


Figure 9: The ontological model of (part of) the Heritage Digital Twin of St. John Lampadistis Monastery

### 8.3 Final considerations on the examples

The patterns built by combining the different extensions of the CIDOC CRM and our new classes and properties are used to model different situations and scenarios. They can be replicated to describe all the other observations, analyses, hypotheses, and any inferences and conclusions that can be drawn from them. The conclusions in turn can serve to support new hypotheses and theories, the graph can be extended without any limitation.

Finally, as seen in the diagrams, all the semantic information acquired about the Heritage Entity and modelled through the ontology eventually becomes part of the Digital Twin (HC2), in which all the knowledge about the object is organized.

## 9. Conclusions and further work

This paper is the outcome of a research activity which is still ongoing. We decided to publish it as it is now, in our opinion a stable and coherent version, to open a discussion and improve it with the contributions of the users' community. We believe that the proposed solution reconciles the different and sometimes incompatible approaches summarily sketched in section 3 and enables an approach a la carte according to different user needs while preserving overall interoperability. Existing documentation may be easily mapped to the HDT ontology and different systems may be integrated into it, combining the existing information under the HDT overarching umbrella. Doing so will enhance the interoperability of different systems and enable the creation of a well-organized data space to support the work by researchers, heritage professionals and the public. Moreover, the HDT is ready for simulation – the second step of digital twin systems – such as answering to “what if” questions as well as causing automatic reactions to actual external events. The CRM community (CRM SIG) is currently working on a new extension for representing not only past activities but also future ones, CRMact<sup>14</sup>, that could be the way towards modelling such dynamics.

Our plans for future work include polishing the HDT ontology with the suggestions coming from the community. In parallel, we will start mapping existing schemas, such as the Europeana EDM ontology<sup>15</sup> or those based on HBIM or 3D annotations, on the HDT ontology adding specific classes and properties if needed, to ultimately create a framework where everybody can recognize their own documentation scheme but can also interoperate with those created by others.

We believe that this objective is a necessary condition to support collaborative scenarios on digital cultural heritage as those envisaged by the EU strategies on digitization.

#### Grant information

The research leading to the present paper was partially funded by the following grants:

ARIADNEplus: European Commission programme H2020 (INFRAIA-01-2018-2019) Grant no. 823914

4CH: European Commission programme H2020 (DT-TRANSFORMATIONS-20-2020) Grant no. 101004468

#### References

1. Pollard, A. M. and Bray, P. (2007) “A Bicycle Made for Two? The Integration of Scientific Techniques into Archaeological Interpretation”, *Annu. Rev. Anthropol.* 2007. 36:245–59
2. Meghini, C., Scopigno, R., Richards, J.D., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavrili, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C. and Vlachidis, A. (2017) “ARIADNE: A Research Infrastructure for Archaeology”. *JOCCH Special Issue on Digital Infrastructure for Cultural Heritage*, Part 2 Volume 10 Issue 3, Article No. 18
3. Richards, J and Niccolucci, F. (eds.) (2019) *The ARIADNE impact*. Budapest, Archaeolingua.
4. Moscati, P. (2019) “30 anni di Archeologia e Calcolatori. tra memoria e progettualità”, *Archeologia e Calcolatori* 30, 2019, 1-12
5. Djindjian, F. (2019) “Archaeology and computers: a long story in the making of modern archaeology”, *Archeologia e Calcolatori* 30, 2019, 13-20
6. Gardin, J-C. (1979) *Pour une archéologie théorique* Paris, Hachette, 1979. English translation: *Archaeological Constructs*, Cambridge, Cambridge University Press, 1980
7. Dallas, C. (2016) “Jean-Claude Gardin on archaeological data, representation and knowledge: Implications for digital archaeology”, *Journal of Archaeological Method and Theory* 23 (1), 305-330
8. Marlet, O., Zadora-Rio, E., Buard, P.Y., Markhoff, B. and Rodier, X. (2019): “The Archeological Excavation Report of Rigny, an Example of an Interoperable Logician Publication.” *Heritage* 2019, 2(1), 761-773; <https://doi.org/10.3390/heritage2010049>
9. Zadora-Rio, E. and Galinié, H. (2020) *L'église de Rigny et ses abords. De la colonia de Saint-Martin de Tours au transfert du centre paroissial (600-1865)*. Presses universitaires de Caen
10. Dell'Unto N. and Landeschi G. (2022) *Archeological 3D GIS*. London & New York, Routledge

<sup>14</sup> <https://cidoc-crm.org/crmact/node/8732>

<sup>15</sup> <https://pro.europeana.eu/page/edm-documentation>

11. Bastem, S. S. and Cekmis, A. (2021) Development of historic building information modelling: a systematic literature review, *Building Research & Information*, <https://doi.org/10.1080/09613218.2021.1983754>.
12. Diara, F. and Rinaudo, F. (2021). ARK-BIM: Open-Source Cloud-Based HBIM Platform for Archaeology *Applied Sciences* 11, no. 18: 8770. <https://doi.org/10.3390/app11188770>.
13. Apollonio, F. I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A. R., Scopigno, R. and Sobrà, G. (2017) "A 3D-centered information system for the documentation of a complex restoration intervention", *Journal of Cultural Heritage*, 2017, Vol 29, 89-99
14. De Luca, L., Abergel, V., Guillem, A., Malavergne, O., Manuel, A., Néroulidis, A., Roussel, R., Rousset, M. and Tournon S. (2022) "L'écosystème numérique n-dame pour l'analyse et la mémorisation multi-dimensionnelle du chantier scientifique Notre-Dame-de-Paris", <https://hal.archives-ouvertes.fr/hal-03826931>
15. Croce, V, Caroti G., De Luca L., Jacquot K., Piemonte A. and Véron, P. (2021) "From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning" *Remote Sens.* 2021, 13(3), 461
16. Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021) "Digital twin paradigm: A systematic literature review." *Computers in Industry*, 130, 103469
17. European Commission (2022). *Destination Earth*. <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>
18. Jones, D., Snider, C., Nassehi, A., Jason Yon, J. and Hicks, B. (2020) "Characterising the Digital Twin: A systematic literature review" *CIRP Journal of Manufacturing Science and Technology* 229 (2020) 36-52
19. VanDerHorn, E. and Sankaran M. (2021) "Digital Twin: Generalization, characterization and implementation", *Decision Support Systems* 145 (2021) 113524
20. Qinling Qi, Tianliang Hu, Fei Tao and Nabil Anwer (2021) "Enabling technologies and tools for digital twin". *Journal of Manufacturing Systems* March 2021. <https://doi.org/10.1016/j.jmsy.2019.10.001>
21. Bolton, A., Lorraine, B., Dabson, I., Enzer, M., Evans, M, Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A. and Makri, C. (2017) *The Gemini Principles: Guiding values for the national digital twin and information management framework*. <https://doi.org/10.17863/CAM32260>
22. Shahzad, M., Shafiq, M.T., Douglas, D. and Kassem, M. (2022) "Digital Twins in Built Environments: An Investigation of the Characteristics, Applications, and Challenges." *Buildings* 2022,12,120. <https://doi.org/10.3390/buildings12020120>
23. Jouan, P., and Hallot, P. (2020) "Digital twin: Research framework to support preventive conservation policies" *ISPRS International Journal of Geo-Information*, 9(4), 228
24. Zeyu Wang, Weiqi Shi, Kiraz Akoglu, Kotoula, E., Ying Yang, and Rushmeier, H. (2018) "CHER-Ob: A Tool for Shared Analysis and Video Dissemination" *ACM J. Comput. Cult. Herit.* 11, 4, Article 18 (November 2018)
25. Baca, M. and Harpring, P (eds.) (2022) *Categories for the Description of Works of Art*, Getty Trust and College Art Association. [https://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](https://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)
26. Niccolucci, F., Felicetti, A. and Hermon, S. (2022) "Populating the Digital Space for Cultural Heritage with Heritage Digital Twins" *Data* 2022, 7(8), 105; <https://doi.org/10.3390/data7080105>
27. Bartalesi, V., Meghini, C. and Metilli, D. (2017) "A conceptualisation of narratives and its expression in the CRM" *Int. J. Metadata Semant. Ontol.* 2017, 12, 35–46, <https://doi.org/10.1504/IJMSO.2017.087692>
28. Meghini, C., Bartalesi, V., and Metilli, D. (2021) "Representing Narratives in Digital Libraries: The Narrative Ontology". *Semantic Web*, vol.12, no. 2, pp. 241 – 264, 2021
29. Carriero, V. A., Gangemi, A., Nuzzolese, A. G. and Presutti, V. (2019) "An Ontology Design Pattern for Representing Recurrent Events" *10th Workshop on Ontology Design and Patterns (WOP 2019), co-located with 18th International Semantic Web Conference (ISWC 2019)*, Auckland, New Zealand, October 27, 2019
30. E. Poyiadji-Richter (ed.) (2019) *The Secrets of a Royal Portrait*, Nicosia, Leventis Foundation

31. Hermon, S. and Bakirtzis, N. (2019) "More than Meets the Eye: Integrating Art History, Digital Technologies and Natural Sciences", in E. Poyiadji-Richter (ed.) *The Secrets of a Royal Portrait*, Nicosia, Leventis Foundation, pp. 37-46
32. Irakleous, S. and Bakirtzis, N., 2020. "Karamanlides at the monastery of Saint John Lampadistis in Kalopanagiotis: pilgrimage votives, architecture and sacred landscape" *Cahiers du Centre d'Études Chypriotes*, (50), pp. 383-405.

# The Heritage Digital Twin: a bicycle made for two. The integration of digital methodologies into cultural heritage research

Franco Niccolucci<sup>\*1</sup>, Béatrice Markhoff<sup>2</sup>, Maria Theodoridou<sup>3</sup>, Achille Felicetti<sup>1</sup>, Sorin Hermon<sup>4</sup>

<sup>1</sup> PIN, Servizi Didattici e Scientifici per l'Università di Firenze, Piazza Ciardi 25, I-59100 Prato, Italy

<sup>2</sup> Université de Tours, 64 Avenue Jean Portalis, F-37200 Tours, France

<sup>3</sup> FORTH, N. Plastira 100, Vassilika Vouton, GR-70013 Heraklion, Crete, Greece

<sup>4</sup> STARC-The Cyprus Institute, 20 Konstantinou Kavafi Street, CY-2121 Aglantzia, Nicosia, Cyprus

**\*Corresponding author:** franco.niccolucci@gmail.com

## Abstract

The paper concerns the definition of a novel ontology for cultural heritage based on the concept of digital twin. The ontology, called Heritage Digital Twin ontology, is a compatible extension of the well-known CIDOC CRM ISO standard for cultural heritage documentation and incorporates all the different documentation systems presently in use for cultural heritage documentation. In the authors' view, it supports documentation interoperability at a higher level than the ones currently in use and enables effective cooperation among different users.

**Keywords:** digital twin, heritage documentation systems, CIDOC CRM extension, heritage semantics, art history semantics, archaeology semantics, scientific data for cultural heritage, conservation

## Plain Language Summary

The paper addresses the issue of managing the digital documentation of cultural heritage in an effective way. For this purpose, it introduces the concept of heritage digital twin, which is inspired by the similar one used in the industry and in other domains. A digital twin is the digital replica of a real-world object. It includes all the necessary information and is able to simulate - in a digital environment - the characteristics and the behaviour of its real counterpart. Two examples are fully developed to demonstrate the value of this novel approach. This research was carried out in the framework of the EU-funded projects ARIADNEplus and 4CH, both concerning the use of digital technology to study and preserve cultural heritage.

## 1. Introduction

The title of this paper is inspired by the similar one of a paper by Pollard and Bray [1] concerning how the contribution of archaeological sciences should be integrated into the process of archaeological interpretation. According to the authors, such integration is like riding a bicycle made for two, also known as a tandem. This kind of vehicle requires a strong collaboration between the two riders to pedal synchronically and the one in front must be able and willing to drive the tandem towards a common destination, on which both riders agree. The structure of the bicycle should suit a diversity of users: tall and short; married couples and perfect strangers; sportspeople and lazy ones. The way it can be used must adapt to any kind of road, dirt trails and urban well-paved streets alike.

Cycling metaphors aside, the convergence and integration of two different disciplines puts requirements to the method and the attitude of both and of all participants. Each party involved must consider the other one's perspective and adapt its own accordingly. A similar situation takes place about the interdisciplinary approach known as digital cultural heritage. Adopting such a collaborative attitude by heritage professionals and institutions is the nucleus of what is usually called digital transformation of cultural heritage. The upskilling of professionals and the allocation of resources to information technology by heritage managers is of course required, but without such change of perspective they might be ineffective.

In the present paper we report the results of research triggered by the activities carried out within two European highly interdisciplinary projects, which probably extend well beyond the project objectives.

The first project is the ARIADNE group of projects comprising ARIADNE (Advanced Research Infrastructure for Archaeological Data Networking in Europe), an FP7 project operating from 2013 to the end of 2016 [2], and its continuation ARIADNEplus<sup>1</sup>, an H2020 project operating from 2018 until the end of 2022. The two projects are collectively indicated as ARIADNE.

The original ARIADNE objective was stated as “To turn the sparse existing archaeological repositories, into a pan-European Integrated Research Infrastructure, with easily available and harmonised access, responding to the demand of the archaeological research community of the European Research Area and beyond” while the ARIADNEplus one is simply “Extending and focusing ARIADNE”. The break between the two projects, determined by the schedule of EU calls, was bridged maintaining the main project result, the ARIADNE catalogue, and preparing for the continuation.

The most visible part of the project outcomes is the ARIADNE catalogue, also called the ARIADNE Knowledge Base, accessible from the ARIADNE portal<sup>2</sup>, with various searching options including time, location and content of the catalogued datasets. ARIADNE aggregated more than 2,000,000 archaeological datasets from a large part of Europe. ARIADNEplus has expanded ARIADNE in geographic coverage, now extending to a larger part of Europe and including datasets from Japan, Israel, Africa, USA and Latin America; and in content type, now including many more datasets concerning archaeological sciences, images, 3D models, archaeological GIS, and so on. The total number of items in the ARIADNEplus version of the catalogue is more than 3,300,000 aggregated datasets, with another hundred thousand datasets still in the aggregation queue. Such results have raised the enthusiasm of the archaeological research community as witnessed in two publications, *The ARIADNE Impact* [3] concerning ARIADNE and a forthcoming one about the added value of ARIADNEplus. This has created a vibrant community that has been nurtured in parallel by training and dissemination activities.

Such achievements have been attained thanks to research results that in ARIADNE operate behind the scenes. There are two main strands: information technology, supporting all the steps of aggregation, and providing the technical instruments to make data FAIR (Findable, Accessible, Interoperable and Reusable); and advanced work in semantics, leading among others to the development of CRMarchaeo<sup>3</sup>, an archaeological extension of the well-known CIDOC CRM standard ontology for cultural heritage<sup>4</sup>. As it is well known, CIDOC CRM – also mentioned in brief as the CRM – stands for Conceptual Reference Model supported by CIDOC, the International Committee for Documentation of ICOM, the International Council of Museums and since December 2006 is an official ISO standard (ISO 21127:2014).

The ARIADNE research achievements form one of the pillars supporting the work presented in this paper.

The ARIADNE Knowledge Base is based on a subset of the CRM ontology, to which the metadata schemas of the catalogued items are mapped. The project also developed application profiles for specific subdomains such as epigraphy, scientific analyses, excavations and so on. Thus, it suggested the idea of an overarching ontology as the one presented in the present paper.

The second project supporting the present paper is the H2020 project 4CH – Competence Centre for the Conservation of Cultural Heritage<sup>5</sup>. This ongoing project addresses the conservation of monuments and sites with digital methods, in view of the creation of a European Competence Centre on the subject, which the project is in charge of designing. Among others, the project aims at setting up a cloud-based Knowledge Base with conservation data. This aspect has a key but not exclusive role in the project work plan. With regard to the present paper, 4CH has brought the need of extending the scope of semantic research beyond archaeology to include tangible cultural heritage as well as its intangible component. Within 4CH, implementing a suitable documentation system is still ongoing work. We anticipate that the ontology presented here will be adopted in 4CH to build its Knowledge Base.

Thus, the semantic team of both projects has been induced to start conceiving an overarching paradigm that could include the digital representation of all heritage assets, tangible or intangible alike, for any purpose.

---

<sup>1</sup> <https://ariadne-infrastructure.eu/>

<sup>2</sup> <https://portal.ariadne-infrastructure.eu/>

<sup>3</sup> <https://www.cidoc-crm.org/crmarchaeo/home-3>

<sup>4</sup> <https://cidoc-crm.org/>

<sup>5</sup> <https://www.4ch-project.eu/>

Illustrating this approach and its main concept, the Heritage Digital Twin, forms the body of the present paper.

In the rest of this paper, Section 2 is a summary of the cultural heritage digital documentation's history with its current state of the art. Section 3 motivates the need of a heritage digital twin ontology. Section 4 brings the basics of the heritage digital twin concept and shows how it goes further than 3D-based data models. Section 5 introduces the overall methodology and main concepts of the Heritage Digital Twin (HDT) Ontology, and more technical descriptions are provided in Section 6 and Section 7. Examples of digital documentations pertaining to the same heritage digital twin are developed in Section 8, one for scientific analyses and another one for art history analyses, performed on the same heritage entity. The section also includes an example of use for built heritage.

## **2. Background: the state of the art in the digital documentation of cultural heritage**

An even concise description of how the digital revolution of the last part of the XX century affected also documenting cultural heritage would probably require an encyclopaedia. For example, Google scholar lists about 1,900,000 papers under the search key "digital documentation of cultural heritage" and it is likely that titles wrongly included in the search results are less than titles wrongly excluded. For instance, does Google AI consider "digital archaeology" as also pertaining to "digital documentation of cultural heritage"?

Archaeology started to be digitally documented in the second half of 1900 using relational databases, while in general similar IT applications to cultural heritage were less frequent at that time. We will follow the development of digital archaeological documentation as an indicator of the progress of such techniques.

At the beginning, digital methods were also called computer applications or quantitative methods, hence the name of the most important conference on the subject: Computer Applications and quantitative methods in Archaeology, CAA<sup>6</sup>. This annual event has been for a long time the main place to communicate such applications, together with other initiatives such as the dedicated journal "*Archeologia e Calcolatori*" (*Archaeology and Computers*, a vintage name indeed), which started publications in 1990 and is still published in Italy. This journal published in 2019 [4, 5] a historical account of the development of computing applications, their perspectives and their impact on archaeological theory and methods. Many papers published in the CAA proceedings in the last decade of the XX century describe examples of database use to digitally store archaeological data. The number of such applications to individual case studies continued to increase until such applications became standard practice and developing a database to collect the data of a specific archaeological investigation was no more a contribution to advancing research in this field, but just another example. At this point, the annual number of related papers collapsed. In parallel, the interest on GIS (Geographic Information Systems) increased, as it appeared that this technique might capture the locational characteristic and spatial relationships of archaeology. Non-canonical threads as the system proposed in 1979 by Jean-Claude Gardin denominated *Logicisme* [6] received less favour possibly also because they were out of the current mainstream in digital archaeology. The interest in GIS was also an obstacle to acceptance, as they need only very simple databases to document archaeological assets and related spatial information.

Two recent papers, one by Dallas [7] and the other one by Moscati [4] have brought *Logicisme* on the forefront again, and an excavation documented according to this system has recently been published [8, 9]. In the meanwhile, the concept of Linked Open Data has made its way into archaeology, starting from the first decade of the XXI century. In the same period, semantic concepts started to be applied to archaeological documentation and domain ontologies showed up.

This very summary sketch of the archaeological computing history explains why it is now necessary to propose new semantic tools for the archaeological documentation, based on accepted domain standards to support interoperability. This principle has been considered in ARIADNE, where a set of common aggregation metadata are used for all datasets.

Another important thread in archaeological documentation is represented by 3D models. Initially used for communication purposes, to facilitate visitors' understanding of complex archaeological sites – an application still alive today – it was realised that such 3D models might also serve for a more complete

---

<sup>6</sup> <https://caa-international.org/>



information about archaeological assets, both objects and remains. A recent book [10] combines such a 3D approach with GIS. This thread may be considered as part of the general interest for using 3D as a documentation tool in Cultural Heritage.

On the general cultural heritage side, digital documentation was limited to inventories for a long period. The turning point coincided with the definition of the CIDOC CRM<sup>7</sup>.

The CRM is the culmination of more than two decades of standards development work by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Work on the CRM itself began in 1996 under the auspices of the ICOM-CIDOC Documentation Standards Working Group. Since 2000, development of the CRM has been officially delegated by ICOM-CIDOC to the CIDOC CRM Special Interest Group (SIG). The SIG, in turn, collaborates with the ISO working group ISO/TC46/SC4/WG9 to bring the CRM to the form and status of an International Standard. This set of collaborations has resulted in the production of ISO21127:2004 and ISO21127:2014, the ISO standard editions of the CIDOC CRM.

The CRM was initially conceived for museum applications but then it gradually evolved into a general-purpose cultural heritage documentation system. The CRM is an extendable system, enabling to develop extensions compliant with the CRM to consider the diverse aspects of cultural heritage, such as monuments, sites, movable or immovable assets, tangible assets in general and finally intangible ones. There are several ongoing developments of CRM compatible models<sup>8</sup> covering different specialised fields

Additionally, 3D models have gradually assumed an increasing importance in cultural heritage. 3D models are yet another digital representation for museum objects, but their importance is especially increasing for architectural heritage, mainly due to the architects' practice. Some actually consider the 3D model as the root and build around it the documentation system. This explains why BIM (Building Information Modelling) is being increasingly proposed as the cornerstone of such systems. BIM is a methodology in use to design new buildings which incorporates in the building project all the necessary information about services such as water, electricity, heating networks, and more. It develops a very simple semantic of objects and of their parts, to be able to avail of graphical libraries in the 3D model construction where the necessary information such as material, characteristics, industrial producer (in new buildings) and assembling/construction phases is attached to the building parts, as walls, pillars, floors, ceilings, and so on. BIM models are based on an industry standard, IFC<sup>9</sup> (Industry Foundation Classes), a standardized, digital description of the built asset industry, and an open, international standard ISO 16739-1:2018. A pillar of a BIM model is the 3D representation of the building as a wireframe/solid model, i.e. a skeletal representation of it as produced by well-known tools, among others by Autocad. It consists of various points, arcs, lines, circles, and curves to clearly denote object edges and depth. Information about the various components – structured according to the IFC standard – is attached to this model. The application of BIM to cultural heritage has led to HBIM, standing for Heritage BIM. An extensive survey of HBIM applications is provided by [11], while [12] focuses on applications to archaeology.

In parallel, there has been a number of proposals to use a point-cloud 3D model as the graphical support and add annotations on it concerning various aspects of the object. This kind of application is usefully applied to conservation and restoration. Among others, it is worth quoting the model used for the restoration of the Nettuno fountain in Bologna [13] and the one currently in use for the Notre Dame restoration [14]. A recent paper proposes to reconcile HBIM and point-cloud models via AI [15].

In conclusion, heritage documentation based on 3D models, either point-cloud ones or BIM-based ones, considers the volume and the shape of the heritage asset to be documented as the root of all the documentation systems. This approach does not allow – at least so far – to consider a large part of documentation that cannot be appended to any physical component and severely limits the interoperability of such documentation systems. Documentation based on the shape/volume of assets generates the fragmentation of the documentation system, as each item is documented separately. It does not allow searching across different assets, for example comparing materials, restoration techniques and so on. Also the intangible component of heritage is generally overlooked.

---

<sup>7</sup> <https://cidoc-crm.org/>

<sup>8</sup> <https://www.cidoc-crm.org/collaborations>

<sup>9</sup> <https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/>

Finally, no complete semantic model is available so far for intangible heritage or for the intangible component of the tangible one.

### **3. The way forward**

In conclusion, there appears to be a need to improve the semantic apparatus of cultural heritage documentation. We suggest starting from and extending the CRM that is nowadays an internationally accepted standard. This will preserve the data interoperability across different documentation systems as they will all be compliant and based on the CRM. Even if two documentation systems introduce different specialization of CRM concepts, it is always possible to link elements across the two systems, availing, if necessary, of the more general CRM concept from which the two specializations are derived. This is what ARIADNE and 4CH are doing, the former by creating mappings between the original metadata schema of each dataset collection to be aggregated and a common ARIADNE one. 4CH is still in the design phase of its knowledge base and needs to incorporate provisions for heritage science and conservation data.

The need for an overarching documentation schema is motivated also by more general considerations. If different ontologies were used, interoperability might be put at risk unless mappings are provided. But the most important disadvantage would concern the methodological integration as advocated by the tandem metaphor: an overarching knowledge organization approach, common to the entire community of use, suitable for the different needs and adapted to the many existing research questions, is the first step (the bicycle, in the metaphor) for a real digital transformation of cultural heritage towards a common methodology. Such an approach must be suitable for all heritage documentation applications, from archaeological research to conservation. It needs to include the semantics required by a diverse research community. It must reconcile the different starting points and perspectives, from the documental one to the one organizing contributions of sciences like chemistry and physics, and moreover incorporate the architect's graphical approaches putting at the centre of documentation the shape, almost always interpreted as a 3D replica. We propose a unified methodology using the digital twin concept.

### **4. What is a digital twin**

The concept of digital twin is not new. Digital twins have been applied in many industrial fields where the idea was born to test components, devices and, later on, to simulate the real behaviour of complex appliances in a digital way [16]. Then, digital twins made their way in machinery control applications, by using sensors surveying the behaviour of a device and sending an alert when an anomalous value is measured, or directly activating specific components to return to a normal condition. This kind of application required the use of simulation processes within the model, which were eventually incorporated in the concept of digital twin. Thus, for industry digital twins, the data component has a relatively simple schema, while the process part is more complex. The stress is on how the system behaves rather than on how the information about the system is structured. The European Commission has recently proposed to create a digital twin of the Earth [17] to evaluate complex environmental processes and their impact on the whole system, and to forecast the effects of mitigating measures against, for instance, global warming.

There are many digital twin definitions which put in evidence different characterizations according to the intended use. Two papers, [18] and [19], analyse many of them, relate them to different purposes and try to summarize them into a single overarching approach. None of the uses considered concerns cultural heritage. The 2020 paper by David Jones et al. titled "Characterising the Digital Twin: A systematic literature review" [18] mainly considers industrial applications, analysing in great details the features of such applications, with a short section dedicated to BIM. This thorough study may be useful in the future also for cultural heritage applications, as it analyses many aspects potentially to be considered when the interconnections between the dynamics of the real world of cultural heritage and those in the virtual world of digital twins will be modelled. Although the paper does not state a formal definition, its systematic approach collects and compares different naming conventions and may be instrumental to setting up a precise naming system for heritage applications.

Also the 2021 paper by E. VanDerHorn and M. Sankaran titled "Digital Twin: Generalization, characterization and implementation" [19] is based on a systematic literature analysis. The paper rightfully argues that "many

of the definitions in the literature combine a definition with specific characterizations about Digital Twins that are unique to the use case(s) that they are describing”, which creates confusion about the digital twin general definition. The overarching definition they propose is the following one:

*[A digital twin is] a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual system.*

This shows that an information update is a substantial component of a digital twin system. However, we believe that the two phases of “virtual representation” and “update” are better analysed separately, in our case, where there is no straightforward way even for the first one.

The digital twin (DT) concept may also be described using the five-dimension description introduced by Qinglin Qi *et al.* [20]. Although still in a manufacturing perspective, their definition of a DT model may be useful also in the cultural heritage domain. They define a Digital Twin as made across five dimensions: Physical Entities (PE), Virtual Models (VM), Services (Ss), Data (DD), and Connections (CN). We would rename PE as *Real-world Entities* (RE), since this also encompasses the immaterial components of a real-world object. *Virtual Models* and *Services* include the behaviour of the Digital Twin according to the impulse of services: for example, how the Virtual Model of a building behaves when it receives the digital effects of an earthquake simulated by a Service. For cultural heritage, this kind of interaction is so far considered and processed outside the digital environment, making the virtual model still a descriptive one rather than a dynamic one. It is likely that in the future also simulation services will make their way into the heritage digital twin applications. *Connections* are the mutual interaction between any of the other four dimensions.

Digital twins are nowadays extensively used among others in mechanical engineering, architecture and especially in the building industry, where they belong to the BIM approach. A recent important project combining BIM and digital twins is the UK Gemini project, proposing to use digital twins nationwide for town planning [21].

Since the architecture domain has a close similarity with the cultural heritage (CH) one, such proximity has promoted the development of HBIM, i.e. Heritage BIM, which incorporates the BIM approach enriching it with additional classes pertaining to the heritage domain, but still within a flat data system. It seems that at present it is almost impossible to incorporate a much wider set of concepts – including some that are of a non-physical nature, the so-called intangible heritage – and relations among them. Recently, some authors [22, 23] suggested that an HBIM model is actually a Digital Twin and used it to develop applications in conservation of built heritage assets. For references to HBIM, see e.g. the bibliography published in [12] for archaeological applications and in general the survey [11].

In our opinion, the heritage domain requires much greater attention on how information is organised, an essential step before continuing to computational modelling that simulates real-world processes, the primary focus of most industrial digital twins. That is why we focus the present research on the data organisation, i.e. the semantics of digital twins, and for the time being set apart the digital simulation component so popular in the manufacturing applications. Thus, in this paper a *digital twin* is considered as *the complex of information about digital counterparts of real-world heritage objects, both material and immaterial ones*. We consider simulation as a distinct aspect, which includes processing heritage data in a computer environment, by creating virtual experiments on digital heritage objects in a virtual research environment to simulate the behaviour of the corresponding assets in the real world, or by availing of processes that get data from the real world via sensors, process such input in the virtual world using the features and data of the virtual models and, according to the results, update the virtual models and trigger real world actions via actuators. The data organisation within such a digital twin system has a relatively stable – but not immutable – design. Processes based on consuming the data to perform a simulation or other required operations may instead be created ad-hoc to address specific problems. Thus, the Heritage Digital Twin (HDT) discussed here does not (yet) consider dynamic interactions between Virtual Models activated by Services. It instead creates static Virtual Models based only on the knowledge derived by physical entities.

In sum, at present the Heritage Digital Twin is formed only by the knowledge about Real-world Entities, stored in digital format in the Virtual Model and the inclusion or extraction of such knowledge. Nevertheless, the information stored in the HDT is rather complex and needs a correct semantic structure, which is the purpose of the present paper. Services and Connections are not addressed so far. Indeed, introducing such dimensions would be useful for cultural heritage documentation and management, above all for its

conservation. We plan to address these facets when the ontology is stable and has been reviewed by the community.

Our approach improves the current HBIM approach, strongly influenced by its CAD derivation and its architecture-civil engineering provenance. In HBIM, the Virtual Model consists of a (virtual) shape, almost always a 3D model, optionally divided into parts – e.g. the roof, the columns, the walls, etc. Each of them has some attributes such as colour, material, technical features, and so on. The corresponding semantic graph is a tree, having as root the graphical entity, typically a 3D model. With this approach, a set of information about a building would be worthless if no drawing is included, regardless of the richness of such information, while a 3D model would be worth consideration even if no information is included besides shape. This approach is equally present in the literature about CAD-based HBIM and in the other branch based on point clouds, where virtual models are called “augmented objects” and the respective non-graphical data often described as “annotations”, a name that perhaps underlines their consideration as ancillary information compared to the graphical one. An interesting approach is proposed by CHER-Ob [24], which adds the need for additional information to the usual 3D annotation-based system. This approach is based on the concept of Cultural Heritage Entity (CHE) which corresponds to our Heritage Digital Twin. On the other hand, the main goal of CHER-Ob is more limited than ours, as it focuses on producing visual content for storytelling from available data, in what they call a *project*, i.e. a study on one or more CHEs addressing specific research questions. Information within a CHE is labelled according to the Getty Categories for the Description of Work of Art (CDWA) [25].

Differently from the 3D-based data models described in the current literature on heritage applications, we present here a semantic model, the Heritage Digital Twin ontology, in which there is no privileged class. The shape of an object is an important feature when it exists, but with no higher rank compared to others. For the above-mentioned reasons, we define the whole digital representation of the real-world CH ‘object’ as its digital twin, which consists of the aggregation of different components, among others its shape represented by a visual model. This approach as an extension of the CRM model also allows the creation of digital twins of immaterial ‘objects’ – having no shape by definition – i.e., in our case, intangible cultural heritage; of stories about heritage; and of people’s relationships with tangible assets. Our proposed ontology incorporates the HBIM approach, which covers only a part of the information, and the 3D annotation model. A final consideration concerns the impact of time on digital twins. In industrial applications what matters is the present state, which can change but only the latest state is relevant. On the contrary, in cultural heritage applications information about past states may be relevant as the information of current ones. For this reason, we will timestamp all the information, i.e. consider a time span of validity. By default, this time span is “always”, unless stated otherwise. At implementation, some simple service may automatically change this default assignment, for example always putting as beginning the construction date or putting as end the destruction date. Other cases concern documented changes, which may also be automatically dated. Time is defined according to ISO 8601-1 and ISO 8601-2. Note also that PeriodO<sup>10</sup> offers solutions to convert named periods (e.g. *The Renaissance*) into time intervals.

The approach presented here was initiated in a seminal paper [26] by some of the authors of the present one.

## 5. The Heritage Digital Twin: a discursive introduction

A necessary premise concerns the overall approach to the definition of the Heritage Digital Twin (HDT) ontology. The HDT ontology is based on the CIDOC CRM and its ecosystem, using whenever possible, the classes and properties of the CIDOC CRM and its extensions and defining new ones only to describe more specific concepts such as cultural entities, digital twins, 3D models, which do not find an exact match in the CIDOC CRM. In any case, as far as possible, we have always tried to derive the new entities from those of the CIDOC CRM in order to keep our model completely compatible, consistent and aligned with it. Introducing new classes maintains compatibility and interoperability when the new class is a (proper) subclass of an already existing one: if not possible at the subclass detail level newly introduced, interoperability will still be maintained at the superclass existing one.

---

<sup>10</sup> <https://perio.do/en/>

In the real world there are **Heritage Entities**, regarded as valued by a community – from the whole of humanity represented, for instance, by worldwide organizations as UNESCO, to smaller ones such as a group of believers. We do not enter into the debate of what is cultural heritage: whatever it is, the system is able to take it into account. A distinction is made between Tangible Entities and Intangible ones. A Heritage Entity may be composed from both tangible and intangible entities. A purely intangible heritage entity may be recognized as it has no tangible component.

The corresponding complex of (digital) information about a Heritage Entity is its **Heritage Digital Twin (HDT)**. The term “complex” used above does not imply that the information must be stored in the same digital device: it can be distributed in different storage as long as they are accessible from the “main” system managing the HDT and possibly integrating on demand such a distributed knowledge. Thus, in such a main system information may consist in the actual data or in the data URI, resolved on demand when necessary. This is the result of a holistic approach, incorporating, putting into order, and relating to it all the digital information pertaining to that Heritage Entity, and possibly to others as well. Thus, the Heritage Digital Twin is actually a network of relationships among data putting into evidence the connections between those data and a real-world Heritage Entity. There must obviously exist some item belonging to a Heritage Digital Twin, although it might conceptually be considered also when empty, a sort of placeholder for forthcoming digital information concerning some Heritage Entity. For example, a folder in a computer server containing photos, descriptions, e-books, and so on, of a particular monument is an embryonic heritage digital twin: not yet a full-fledged one because the relationships between each folder item and the heritage entity is not explained, it is just “the file (or a symlink to it) that stays in the folder named after the heritage entity”.

The Heritage Digital Twin will then consist of pieces of its own information (e.g. the Twin’s identifier) and of other digital information pertaining to it. Such information is organized as follows.

The main class is **Heritage Entity**, comprising tangible and intangible entities of the real-world regarded as valuable because of their contribution to society, knowledge and/or culture. The tangible and intangible aspects of the *same* Heritage Entity are recorded as **Tangible Aspect** and **Intangible Aspect**. While all Heritage Entities have an intangible aspect, some may not have the tangible one.

As already mentioned, the complex of the information concerning a Heritage Entity forms its Heritage Digital Twin. The latter includes **Digital Representations**, i.e. digital representations of a Heritage Entity such as text documents or visual ones, i.e. a photo, a video, a 3D model, each one with its own class. It also includes **Stories** about the Heritage Entity, i.e. narratives, modelled according to the NOnt ontology [27, 28], which is based on narratology, a formal way to describe narratives. It distinguishes the *fabula*, i.e. the thematic content of a narrative, and the *narration* (also called *syuzhet*), the chronological structure of the events within the narrative.

The resulting HDT ontology is compatible with the CRM and is also extensible. For example, further subdivisions of text documents might be introduced if necessary, distinguishing e.g. among historical documents, scientific documents and so on. 3D models may be further characterized according to type and have their own metadata and paradata as required. Here we introduce only the most general classes, and are planning to present such extensions in a forthcoming paper.

## 6. Technical description of the HDT ontology: introduction

### 6.1 General note on classes and properties

As already mentioned, the HDT ontology is based on the CIDOC CRM and its ecosystem, directly using the classes and properties of the CIDOC CRM and its extensions for entities and properties having identical meaning and conceptual scope also in our domain, such as places, agents, physical objects. In this case the class or property name is preceded by the namespace prefix, which identifies the relevant ontology.

### 6.2 Ontological models used in HDT ontology

The following Table 1 shows the ontological models of the CIDOC CRM ecosystem used to build the HDT ontology and the namespaces (prefixes) used to indicate their classes and properties throughout this document<sup>11</sup>.

Table 1. Ontological models used in HDT ontology definitions

Model	Version	Name space prefix	Description	Classes prefix	Properties prefix
HDT	1.0	hdt	The ontology described in the present paper (i.e. new classes and properties)	HC	HP
CIDOC CRM	6.2.1	crm	A formal ontology for modelling Cultural Heritage information	E	P
CRMsci	1.2.6	crmsci	The scientific observation model	S	O
CRMdig	3.2	crmdig	Model for provenance metadata	D	L
CRMpe	3.1.2	crmpe	The PARTHENOS Entities model	PE	PP
CRMinf	0.10.1	crminf	An Extension of CIDOC-CRM to support argumentation	I	J
CRMba	1.4	crmba	An extension of CIDOC CRM to support buildings archaeology documentation	B	BP
FRBRoo	2.4	frbr	Functional requirements for bibliographic records	F	R
NOnt	1.0	nont	The MINGEI Narrative Ontology		

### 6.3 Events

To model cultural events, traditions and practices, typical of the intangible heritage, we have defined the HC4 Intangible Aspect class, which is declared as a subclass of crm:E89 Propositional Object and is referred to by an HC3 Tangible Aspect class by the property HP5 has intangible aspect. While all tangible heritage has always an intangible aspect, intangible heritage does not necessarily have a tangible aspect.

Instances of HC4 Intangible Aspect describe generic (template) events such as the Palio di Siena and not the individual occurrences of the Palio. The actual individual occurrences are instances of crm:E5 Event, and property crm:P129 is about (is subject of) may be used to link an individual Event (E5) to the generic description of the cultural heritage event (HC4), such as the Palio di Siena race of the present year (E5) which is about (P129) the Palio tradition (HC4). Nevertheless, it is important to define a new property to specify this special link between the HC4 Intangible Aspect and its punctual manifestations. This property is HP6 has manifestation event (event is manifestation of). Notice that, with this minimal requirement model, recurrent event series as specified in [29] may be computed, instantiated and associated to an HC4 instance if needed. The following is an example:

The Palio di Siena (HC1 Heritage Entity), is a horse race (HC4 Intangible Aspect) that is held twice each year, on 2 July and 16 August, in Siena, Italy (crm:E53 Place). Ten horses and riders (crm:E39 Actor), represent ten of the seventeen *contrade* (crm:E74 Group), or city wards.

<sup>11</sup> All the referred models are described here: <http://cidoc-crm.org/collaborations>. The NOnt ontology is described here: <https://dinarratives.eu/ontology/>.

The historical horse race Palio di Siena (HC4 Intangible Aspect) was held again (crm:E5 Event) on 17/8/2022 (crm:E52 Time-Span) after a two-year pause because of the COVID-19 pandemic. The winner was jockey Giovanni Atzeni (crm:E39 Actor).

For events and activities that are not strictly “cultural” and therefore fall outside the immediate scope of HC4 Intangible Entity, but which in any case concern, affect and remain somehow connected to the various cultural entities, the crm:E5 Event class can be used instead. These are activities like conservation, restoration, reconstruction, and natural events like earthquakes and floods. The type of events for which instances of crm:E5 are used can be specified by means of the crm:P2 which has type property of the CIDOC CRM. Open vocabularies, containing the most common descriptions of these cultural heritage activities, can be defined and released within the ontology to be used in combination with the above-mentioned crm:P2 property for a more standardized and complete description.

#### 6.4 Conditions and states

An important information about heritage assets is their condition state. In the CRM this is documented using E3 Condition State followed by P2 has type E55 Type, choosing the latter in a vocabulary of possible states. If this way of assessing the state of an asset seems too generic, a more precise solution is offered by E14 Condition Assessment. This is the activity dedicated to the evaluation of the condition, and its outcome is a report about the condition. For example, a paper titled “Three-Dimensional Creep Analyses of The Leaning Tower of Pisa” on the condition of the Pisa leaning tower was published in 1997 by Dryden and Wilson. This fact can be expressed as follows:

The “Analysis on Pisa Tower” (E14 Condition Assessment) in 1997 was carried out by (P14 carried out by) “Dryden & Wilson” (E39 Actor) in (P4 has time span) 1997 (E52 Time-Span). It is documented (P70 documents) in the document “Three-Dimensional Creep Analyses of The Leaning Tower of Pisa” (E31 Document).

#### 6.5 Stories and Storytellings

In our model, stories are considered as accounts of facts about a certain Heritage Entity, including (but not limited to) descriptions based on documents and on their interpretation. Stories are an integral aspect of the framework of a Heritage Entity (and thus, of its Digital Twin) since they contribute to the construction of its intangible part as they may make it more understandable, interesting, and attractive for the public. A story can be seen as the core of a series of facts and how they happened, as for example the story of the Knossos Palace and its discovery by Arthur Evans or the story of Falconry over the centuries. From a conceptual point of view, a story is equivalent to the concept of *narrative*, i.e., “a story as it happened in reality or in fiction”, as defined in the Narrative Ontology (NOnt) of the MINGEI project<sup>12</sup>. Since this definition is perfectly suited to our concept of story, and furthermore, given that Narrative in the Narrative Ontology is a subclass of E73 Information Object of CIDOC CRM, which provides it an additional level of formal compatibility, this class will be used in our model as it is.

A storytelling, instead, is considered in our model as the way in which the facts composing a story are actually narrated, presented and disseminated. Storytellings, in fact, comprise social and cultural activities of telling, writing and disseminating stories, for the purpose of education, cultural preservation or entertainment, both in oral form and by means of simple or sophisticated techniques aimed at making the narration of a story effective. An example of storytelling is how the history of Falconry is narrated in “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II. From a conceptual point of view, a result of storytelling can be seen as equivalent to the concept of narration, which in the Narrative Ontology “represents the narration of a narrative, i.e. an individual work that tells the events of the narrative through some form of media (text, video, audio, etc.)”. As stated in [27, 28] there can be many narrations of the same

---

<sup>12</sup> <https://www.mingei-project.eu>. Narrative Ontology documentation is available at: <https://dlnarratives.eu/ontology/>



story, focusing on different aspects of the *fabula*, or presenting events in a different order. In the Mingei Narrative Ontology, Narration is a subclass of F14 Individual Work of the FRBRoo Ontology, another extension of CIDOC CRM. Since also in this case, as for the story, we have verified the perfect conceptual and formal overlap of this class with our idea of storytelling, the Narration class will also be used in our model as it is.

## 7. Scope notes of Classes and Properties

### 7.1 Classes

#### HC1 Heritage Entity

Subclass of: crm:E77 Persistent Item

Superclass of:

Scope Note: This class comprises tangible and intangible entities of the real-world regarded as valuable because of their contribution to society, knowledge and/or culture. Instances of HC1 Heritage Entity may refer to real assets of any nature: physical, both movable and immovable, immaterial, or born digital. They can also refer to cultural events, traditions and practices, typical of the intangible heritage, and can be used to describe their features and their extent in space and time. In the case of events, we can create instances of event types. An instance of HC1 can be considered as the entry point for inferring the content of its corresponding HC2 instance, even if by using crm:P148 has component property it can also be used for denoting HC2's components.

Examples: the Knossos Palace, part of the Knossos WH archaeological site  
the Pafos Gate in Nicosia  
the "Palio di Siena"  
the Florence Historical Centre, a WH Site  
the Stonehenge Complex, a WH site  
the Bauhaus style.

Properties: HP1 has digital twin (is digital twin of): HC2 Heritage Digital Twin  
HP2 has story (is story about): nont:narrative  
crm:P70 is documented in (documents): HC6 Digital Heritage Document  
HP9 has visual representation (is visual representation of): HC7 Digital Visual Object  
crmdig:L1 was digitized by (digitized) D2 Digitization Process  
crmdig:L11 had output (was output of): HC8 3D Model<sup>13</sup>

#### HC2 Heritage Digital Twin

Subclass of: crm:E89 Propositional Object

Scope Note: The class consists of the information available in a given system and pertaining to an HC1 Heritage Entity. Every instance of HC1 Heritage Entity is linked to one instance of HC2 Heritage Digital Twin, which provides an archive of the documented history of the corresponding HC1 Heritage Entity. It includes digital representations of that Heritage Entity (e.g. 3D models, images, videos), textual descriptions (e.g. digital documents, narrations or stories), information of the effects on the related HC1 Heritage Entity of events that influenced or/and are related in any way to its state of (e.g. earthquakes, floods etc.) and of activities (e.g. restorations, conservations etc.) carried out on it.

---

<sup>13</sup> This means that an HC1 instance is related to an HC8 instance via the property path crmdig:L1/crmdig:L11.

Examples: the HDT of Pisa Leaning Tower  
the HDT of the Neptune Fountain in Bologna  
the HDT of Knossos Palace  
the HDT of the Pafos Gate in Nicosia  
the HDT of the “Palio di Siena”  
the HDT of the Florence Historical Centre

Properties: HP3 is digital twin component of (has digital twin component): HC2 Heritage Digital Twin

### **HC3 Tangible Aspect**

Subclass of: HC1 Heritage Entity  
crm:E18 Physical Thing

Scope Note: This class comprises tangible, material entities of the real-world, both movable (e.g. archaeological, artistic and cultural objects) and immovable (e.g., built heritage like monuments, buildings, cities and other complexes), regarded as valuable because of their contribution to society, knowledge and/or culture. The “tangible” term in the name of this class does not exclude that its instances also possess an intangible aspect, which is specified through the HP5 has intangible aspect property.

Examples: the Neptune Fountain in Bologna (Italy)  
the Pisa Leaning Tower, a UNESCO World Heritage (WH) Site  
the Nike of Samothrace of the Louvre Museum in Paris (France)

Properties: HP5 has intangible aspect (is intangible aspect of): HC4 Intangible Aspect  
HP7 is manifestation of (is manifested by): HC4 Intangible Aspect

### **HC4 Intangible Aspect**

Subclass of: HC1 Heritage Entity  
crm:E89 Propositional Object.

Scope Note: This class comprises cultural events, traditions and practices having particular social, historical and cultural significance, including practices and expressions, memories and oral traditions about events, things, people.

Examples: the Mediterranean diet  
Falconry  
the Rebetiko music tradition  
the “Palio di Siena”

Properties: HP6 has manifestation event (event is manifestation of): crm:E5 Event

### **HC5 Digital Representation**

Subclass of: crmdig:D1 Digital Object

Scope Note: This class comprises the digital virtual representations of an HC1 Heritage Entity such as e-texts, images, audio or video items, 3D models, etc., that are documented as single units.

Examples: the digital version of Vasari’s “Vite”

the video <https://www.youtube.com/watch?v=P1Uv4Zf5xKk>  
the Pafos Gate laser scanning 3D model

### **HC6 Digital Heritage Document**

Subclass of: HC5 Digital Representation

Scope Note: This class comprises pieces or collections of digital, non-visual documents, either born-digital or digitised from physical, real-world ones, typically containing textual or numerical information regarding an HC1 Heritage Entity and intended to become part of the related HC2 Heritage Digital Twin. Documentation of this kind may include scientific data, research results and interpretation, as well as historical and cultural information, including textual descriptions related to the nature, conditions, positioning and to the whole set of events in which the cultural entity has been involved and the actors who have participated in them.

### **HC7 Digital Visual Object**

Subclass of: HC5 Digital Representation

Scope Note: This class comprises digital visual objects, such as photos and videos, but also special imagery such as X-ray images, spectra of chemical and physical analyses, and so on, intended to become part of the HC2 Heritage Digital Twin of an HC1 Heritage Entity. Digital documentation of this kind can be born digital or digitised from physical objects (such as paper photographs, drawings and so on). Particularly relevant digital visual objects are also Virtual Reality (VR) and Augmented Reality (AR) models, other types of visual digital artefacts pertaining to a HC1 Heritage Entity. Both VR and AR models rely on 3D models of the related heritage entity, but may add or remove parts of it, or require further digital input as in AR, so they should be catalogued separately from 3D models.

Examples: The Europeana digital version of the paper picture of the Pisa Leaning Tower taken by Paolo Monti in 1960  
([https://www.europeana.eu/it/item/9200369/webclient\\_DeliveryManager\\_pid\\_6363979\\_custom\\_att\\_2\\_simple\\_viewer](https://www.europeana.eu/it/item/9200369/webclient_DeliveryManager_pid_6363979_custom_att_2_simple_viewer)).

### **HC8 3D Model**

Subclass of: HC5 Digital Representation

Scope Note: This class is used for rendering in detail the 3D model of HC1 Heritage Entity and intended as a particular crmdig:D1 Digital Object having its definite identity and resulting from operations such as digitization, acquisition, processing and other actions typical of the three-dimensional modelling world (e.g., 3D scanning, wireframe modelling and so on). The particular features of a 3D model (e.g., its type, format, resolution, size, etc.) and its relationships with the series of activities carried out for its creation and manipulation are modelled through the properties inherited from its superclass HC5, which in turn inherits from crmdig:D1 Digital Object, and through the other classes and properties of CRMdig.

Examples: The 3D model of the Neptune Fountain produced by ISTI-CNR (Pisa, Italy) as part of the documentation used for the restoration of the Neptune Fountain in Bologna (Italy).  
<https://www.cnr.it/en/focus/074-43/3d-supported-restoration-the-neptune-fountain-in-bologna>.

## 7.2 Properties

### HP1 has digital twin (is digital twin of)

Domain: HC1 Heritage Entity  
Range: HC2 Heritage Digital Twin

Scope Note: This property links an instance of HC1 Heritage Entity with an instance of its related HC2 Heritage Digital Twin in a given system.

Examples: The Pafos Gate in Nicosia, Cyprus (HC1) has digital twin (HP1) the Pafos Gate digital twin (HC2) created by Cyprus Institute.

### HP2 has story (is story about)

Domain: HC1 Heritage Entity  
Range: nont:Narrative

Scope Note: This property links an instance of HC1 Heritage Entity with an instance of a nont:Narrative that refers to it.

Examples: Falconry has story (HP2) the history of Falconry over the centuries.

### HP3 is digital twin component of (has digital twin component)

Domain: HC2 Heritage Digital Twin  
Range: HC2 Heritage Digital Twin

Scope Note: This property associates an instance of HC2 Heritage Digital Twin with another HC2 of which is component. The term 'component' here is not limited to physical or geographical relationships (see examples), but encompasses any kind of main-associated relationship.

Examples: The HC2 Digital Twin of Pafos Gate in Nicosia (Cyprus) HP3 is a digital twin component of the HC2 Digital Twin of Nicosias' City Walls.  
The HC2 Digital Twin of the "Cento Camini" Medici Villa in Artimino (Florence) HP3 is a digital twin component of the HC2 Digital Twin of the UNESCO WHS Medici Villas in Tuscany.  
The HC2 Digital Twin of Vichy is a HP3 digital twin component of the HC2 Digital Twin of the UNESCO WHS The Great Spa Towns of Europe  
The HC2 Digital Twin of the "Basilica of San Salvatore in Spoleto, Italy" is a HP3 digital twin component of the HC2 Digital Twin of "Spoleto", which is a HP3 digital twin component of the UNESCO WHS "Longobards in Italy. Places of Power"

### HP4 narrates (is narrated through)

Domain: nont:Narration  
Range: nont:Narrative

Scope Note: This property links an instance of nont:Narration with an instance of a nont:Narrative which has this narration. It is similar to the nont:hasNarration property, but is not a subproperty of crm:P148 has component.

Examples: The "*De Arte Venandi Cum Avibus*" treatise by the Holy Roman Emperor Frederick II narrates (HP4) the history of Falconry.

**HP5 has intangible aspect (is intangible aspect of)**

Domain: HC3 Tangible Aspect  
Range: HC4 Intangible Aspect

Scope Note: This property associates an instance of HC3 Tangible Aspect with its intangible aspects (HC4), i.e. the cultural, social and historical value it incorporates.

Examples: The “Theotokos of Vladimir” (HC3) icon HP5 has intangible aspect the secular veneration that is addressed to it (HC4).  
The UNESCO WHS site “Routes of Santiago de Compostela” (HC3) has intangible aspect (HP5) pilgrimage to Santiago (HC4).

**HP6 has manifestation event (event is manifestation of)**

Domain: HC4 Intangible Aspect  
Range: crm:E5 Event  
SubPropertyOf: crm:P129 is about (is subject of)

Scope Note: This property associates an instance of HC4 Intangible Aspect with the instances of the crm:E5 Event (or of the unique and specific crm:E5 Event) through which the intangible entity manifests itself in the physical world.

Examples: The Palio di Siena (HC4) has manifestation event (HP6) the historical horse race that was held in Siena on 17/8/2022 (E5)

**HP7 is manifestation of (is manifested by)**

Domain: HC3 Tangible Aspect  
Range: HC4 Intangible Aspect

Scope Note: This property associates instances of HC3 Tangible Aspect with the HC4 Intangible Aspect of which they are the manifestation in the physical world.

Examples: The set of devotional graffiti engraved on the walls of the Church of the Holy Sepulchre in Jerusalem (HC3) is the manifestation of (HP7) the pilgrimage of which the church is the final destination (HC4).

**HP8 is narrated in document (document used for narration)**

Domain: nont:Narration  
Range: crm:E31 Document

Scope Note: This property associates an instance of nont:Narration with instances of E31 Document used to implement it.

Examples: The “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II (nont:Narration) is narrated in document (HP8) the “MS. Lat. 419” manuscript , now in the library of the University of Bologna E31.

**HP9 has visual representation (is visual representation of)**

Domain: HC1 Heritage Entity  
Range: HC7 Digital Visual Object

Scope Note: This property associates an instance of HC1 Heritage Entity with instances of HC7 Digital Visual Object in which it is represented.

Examples: The Pisa Leaning Tower (HC1) has visually representation (HP9) the Europeana digital version of the paper picture of the Pisa Leaning Tower taken by Paolo Monti in 1960 ([https://www.europeana.eu/it/item/9200369/webclient\\_DeliveryManager\\_pid\\_6363979\\_custom\\_att\\_2\\_simple\\_viewer](https://www.europeana.eu/it/item/9200369/webclient_DeliveryManager_pid_6363979_custom_att_2_simple_viewer)) HC7

#### **HP10 tells about (is told by)**

Domain: nont:Narrative

Range: crm:E5 Event

Scope Note: This property is intended to identify the specific events (E5) to which a nont:Narrative relates.

Examples: The history of Falconry (nont:Narrative) tells about (HP10) the writing of “De Arte Venandi Cum Avibus” treatise by the Holy Roman Emperor Frederick II (E5).

### **8. Use cases**

In this section we develop two complete examples of HDT. In the first one, the focus is on art history and on scientific analyses carried out on a painting. The second one describes the monastery of St. John Lampadistis in Kalopanayotis, Cyprus, including the description of intangible aspects.

#### **8.1 The analyses on the portrait of Caterina Cornaro from the Leventis museum of Nicosia**

##### **8.1.1 History of the painting and results of the analyses carried out on it**

The painting was acquired by a private collector (Michael Zeippeki) who donated it to the Leventis museum. It is a 19<sup>th</sup> century portrait of Caterina Cornaro, the last queen of Cyprus [30]. Visual investigations hinted at the existence of an underlayer of paint. Further heritage science investigations (x-ray imaging, digital microscopy, multi-spectral imaging, XRF) confirmed the existence of an earlier, 16<sup>th</sup> century painting [31]. Tobias Lange, a restorer from Dresden, Germany, removed the 19<sup>th</sup> century painting and exposed the 16<sup>th</sup> century one (Figure 1). Many details changed: the crown, turban, breast cover and right shoulder size and orientation.





Figure 1. (a; b; c; d: clockwise from top left): The painting's development: a – as appeared before latest restorations; b – infra-red imaging reveals an underlayer paint with a modified pictorial composition; c – X-Ray imaging displaying in more details these differences: shoulder in a different position, a cloth covering the chest, the veil position shifted, a pointed, un-decorated crown, the turban broader and richer, the ear more to the left; d – the painting as of today, after restorations that removed the upper pictorial layer (added in the 19<sup>th</sup> century) and highlighting the 16<sup>th</sup> century original.

Art historical investigations indicate a lost Titian portrait, which was the source of inspiration for several copies, made by different artists during the 16<sup>th</sup> century and which circulated in Europe. Common to all is the shape of the pointed crown, the queen depicted as a mourning widow with characteristic clothes and a veil.

A close examination of the painting revealed some details implying relevant information about the painting. **Boldface** indicates the parts developed in the semantic example.

Table 2. Results deduced from observation of the painting and from literary references, in boldface those modelled with HDT in the example

Observation	Inference
<b>Painting canvas nailed on stretches</b>	<b>Painting cut from a larger one</b>
No signatures	A 19 <sup>th</sup> century painting overpaint on a 16 <sup>th</sup> century one, unknown artists
Old adhesive on the retro - GK9654	GK = Royal Prussian Cataloguing system
Old adhesive on the retro in German – mentioning the collectors	Edward Solly was an English collector focusing on Renaissance paintings – he donated his collection in the 19 <sup>th</sup> century to the Royal Prussian house.
Observations on a cross-section sample from the stretcher with a transmitted light microscope (technical details mentioned)	Wood structure indicates poplar ( <i>Populus</i> ), widely used in S. Europe, but not N. Europe, to make stretchers, indicating that it was probably made in Italy.
References in various books and collection inventories (Vasari's Lives, Colbertaldo's Storia di Caterina Cornaro) on paintings	Portraits of the queen by several Venetian artists, such as Titian, Giorgione, Tintoretto, Paolo Veronese, etc., done from imagination.
<b>Carlo Ridolfi writes in his Le Meraviglie dell' Arte on Titian's work "... with the same way the Queen Caterina Cornaro is depicted in widow's dress that lets the whiteness of the skin rise through the black..."</b>	<b>Indication of a now-lost painting of Titian with an iconography of "Caterina Cornaro as a widow".</b>
Marin Sanuto – chronicler and member of the Major Council of Venice mentions in his diary that upon her arrival in Venice, the queen wore "a black velvet dress, a veil on her head and Cypriote jewellery".	Hanging pearls as "Cypriote jewellery", the black dress, the almost transparent veil and the chest cover became a fashion of depicting Cypriot nobility during the 16 <sup>th</sup> century paintings in Cyprus.
Use of gesso on a tabby waved canvas, brown ochre ( <i>imprimatura</i> ) on the preparation layer, paint layer applied with fine brush strokes.	Techniques of a Venetian Renaissance painter.
Canvas of thin linen fabric in simple tabby weaves.	Produced in the Low Countries and widely used in Venice during the 16 <sup>th</sup> century.
Thin gesso layer on the canvas.	Fills the depressions in the weave.
Brown underpaint layer on top of the gesso.	Typical of 16 <sup>th</sup> century paintings.
No underdrawing beneath the painted surface following IRR and X-ray analysis.	Outlining of forms with a brush and fluid black was common to Titian's practice as evidenced on his autographed paintings.
No outlining of the canvas for a duplicating process, which would serve to duplicate the portrait to another surface.	This suggests a creation after a prototype, perhaps a repetition, worked with a free hand.
Later overpaints in various areas of the painting, such as re-positioning of the ear, change in the shape of the crown and turban, shoulders and the covering of the chest with a white fabric.	Efforts to bring the composition to the fashion of the 19 <sup>th</sup> century, as well as to centralize better the figure, since it was initially cut from a larger painting.
Overall composition of the 16 <sup>th</sup> century painting is of high quality.	As described by C. Ridolfi, this painting probably coincides with the times of Titian, but unclear by whom and is among the earliest among the queen's portraits.

Table 3. List of Heritage Science experiments.

Instruments and environmental conditions		Reason
Imaging	Ricoh WG-30 digital camera	Delineation of composition details on pictographic layers.
	Hirox KH-8700 digital microscope with a magnification range of 5×–2500× with dual illumination revolver zoom lenses.	
Spectroscopy	XRF - ARTAX-200 $\mu$ -XRF Bruker, Mo X-ray tube operated at 50 kV and 700 $\mu$ A, CCD camera with sample illumination and laser spot, a silicon drift detector with a resolution of <150 eV and a 0.65 mm collimator. Acquisition time = 90 s. Measurements taken in air atmosphere and with a Mo filter. The energy to channel calibration was done with a bronze standard, using the Cu- and Sn-K $\alpha$ lines; the Mn-K $\alpha$ line of a manganese standard for the full width at half maximum calibration. Spectra acquisition and evaluation were done with Spectra 5.3 software.	Identification of pigments used for all pictorial and preparation layers.
SEM	SEM-EDS - TM3000 Hitachi, with a Backscattered Electron (BSE) detector coupled with EDS for elemental analysis. The system was operated at 15 kV at a working distance of 8.5 mm; images were taken at magnifications of 15× to 30,000×.	Painting stratigraphy

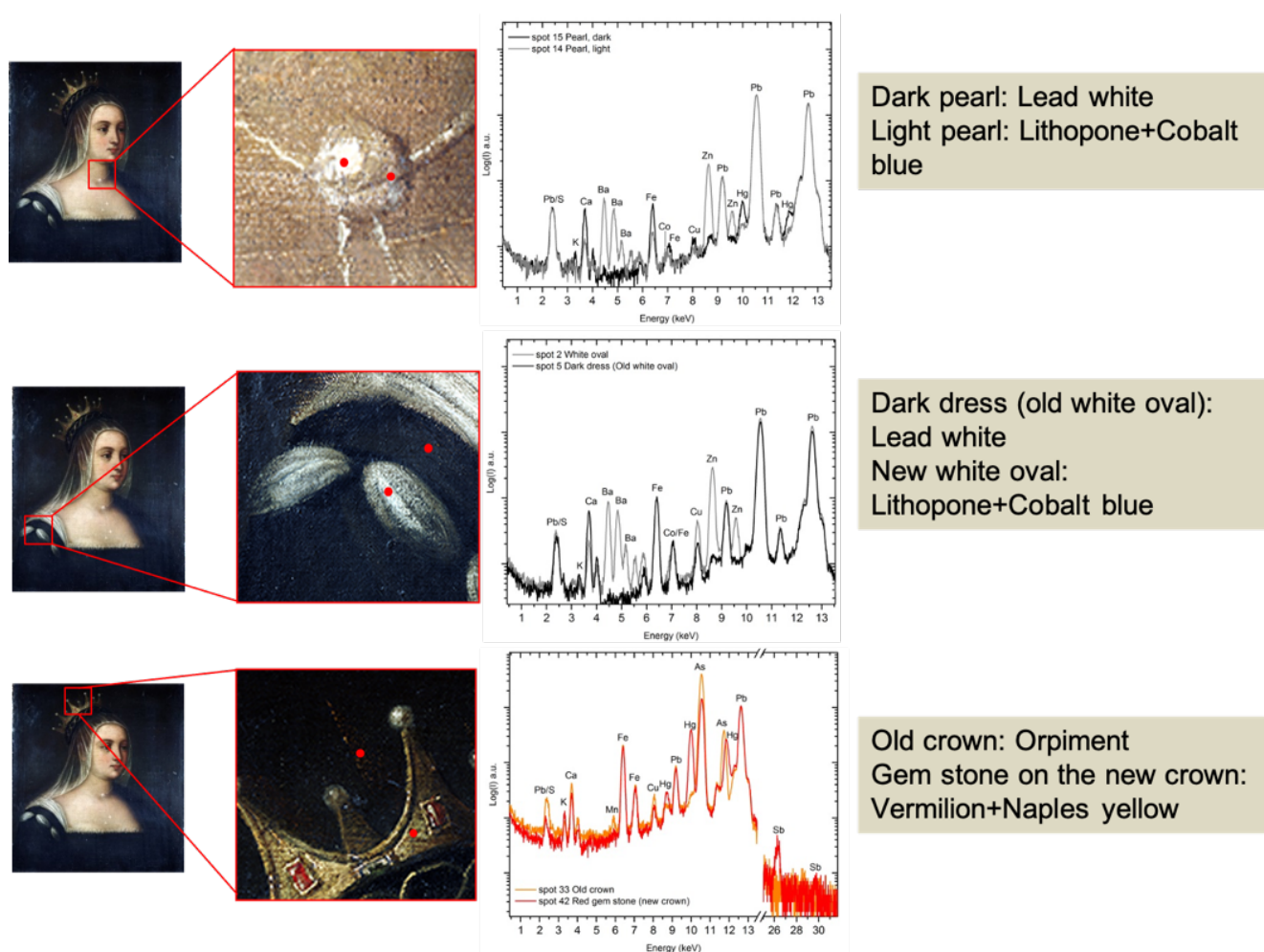


Figure 2. X-Ray fluorescence analyses of the various spots on the painting and suggested pigments used.

Table 4. Results of the analyses and conclusions based on them.

Observation		Inference
EDS	Intense Ca and S lines	Preparation layer of canvas– a thick white layer of calcium sulphate, either anhydrite ( $\text{CaSO}_4$ ) or gypsum ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ).
	Pb, Fe, Ca, Mg, Na, Al and P.	Brown preparation layer called <i>imprimatura</i> , characteristic of Renaissance paintings consisting of a mixture of lead white, earth pigments (ochre), ivory black and dolomite (?).
A thin organic layer.		Residues of varnish from the original painting.
Two uppermost transparent layers, recognizable under UV light microscopy.		Layers of varnish of the overpainting.
Staining tests on the painting cross- sections detected a proteinaceous material in a paint layer of the overpainting as well as in the Ca-containing white ground layer.		Presence of a glue binder.
XRR	Bright areas.	X-ray opaque pigments composed of heavy elements.
	Dark areas.	X-ray pigments composed of light elements.
XRF	Basic lead carbonate, intense Pb lines	<b>Lead white (<math>2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2</math>), characteristic of Renaissance paintings.</b>
	Intense Ba and Zn lines.	Lithopone (barium sulfate-zinc sulfide, $\text{BaSO}_4 \cdot \text{ZnS}$ ). Pigment used in the second half of the 19 <sup>th</sup> century (thus dating the overpaint).

Table 5. Colour Analysis with XRF

	Original paint		Overpaint
White	Lead white ( $2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$ )		A mixture of Lead white with Lithopone ( $\text{BaSO}_4 \cdot \text{ZnS}$ )
Carnation	A mixture of Lead white, Vermilion ( $\text{HgS}$ ), Ochre (iron yellow, brown) and Copper green.		
Yellow	Mainly Orpiment ( $\text{As}_2\text{S}_3$ ). Small intensity of Sn lines might be related to small additions of a yellow tin pigment, such as Lead-tin yellow ( $\text{Pb}_2\text{SnO}_4$ ). Small intensity of Ag lines might be related to admixtures of silver in natural orpiment.		Mainly Lead yellow (Litharge $\text{PbO}$ )
Red	none	Vermilion for the lighter and paler red, Vermilion with an antimony pigment such as Antimony red ( $2\text{Sb}_2\text{S}_3 \cdot \text{Sb}_2\text{O}_3$ , superposition or mixture) used to highlight the borders of the gemstones which appear brighter and darker.	
Green			Copper green
Black	Iron-Manganese black (Umber) and Copper (green) pigments		

Each statement from the tables above is supported by available datasets.



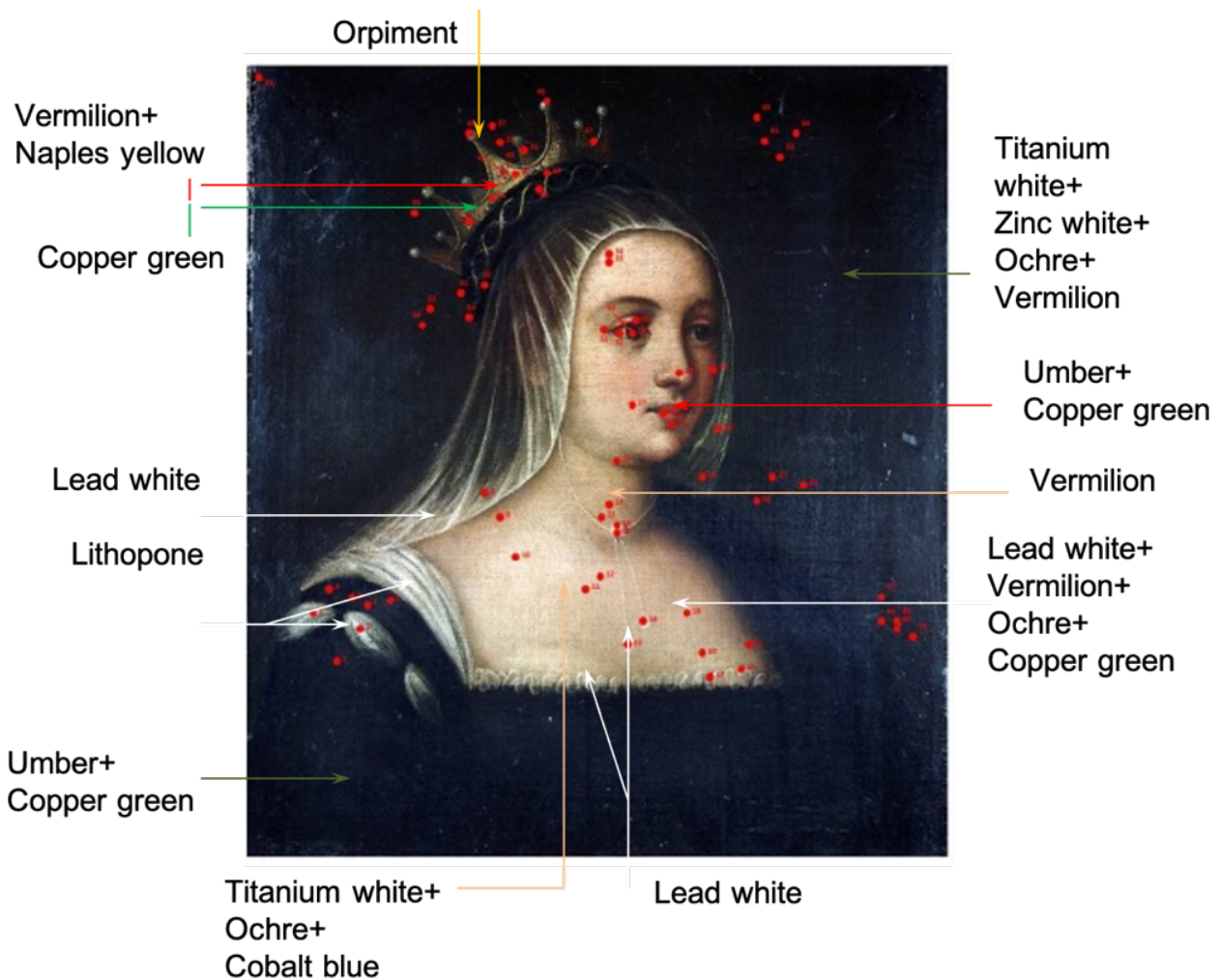


Figure 3. Distribution of pigments on the painting, demonstrating a mixture of original 16<sup>th</sup> century ones and later, 19<sup>th</sup> and 20<sup>th</sup> century pigments

### 8.1.2 Modelling Art History information

As a preliminary test, we modelled two of the results proposed above (highlighted in boldface). The first concerns a direct observation of the condition of the painting, the second concerns information derived from a book in which the painting is probably mentioned.

In particular, in the first case, some physical peculiarities are observed ( $HC1 \rightarrow O8 \rightarrow S20$ ) on the painting canvas ( $O9 \rightarrow S9$ ) which allow researchers to make hypotheses ( $J2 \rightarrow I2$ ) about past conditions and provenance of the object.

In the second case, from Carlo Ridolfi's book (E31), in which the object is mentioned ( $HC3 \rightarrow P70$ ), we read that Titian painted a portrait of Caterina Cornaro, now lost ( $J7 \rightarrow I7$ ). This leads to the hypothesis ( $J2 \rightarrow I2$ ) that this portrait could be the lost one painted by Titian.

Finally, the two  $I2$  Belief(s) resulting from the above observations (together with the others from Heritage Science described below, and others deriving from further investigations and analyses not modelled here) can then be used as a premise ( $J1$ ) to support further inferences ( $I5$ ), such as the attribution of the painting to Titian ( $I2$ ).

The full model is shown in Figure 4.

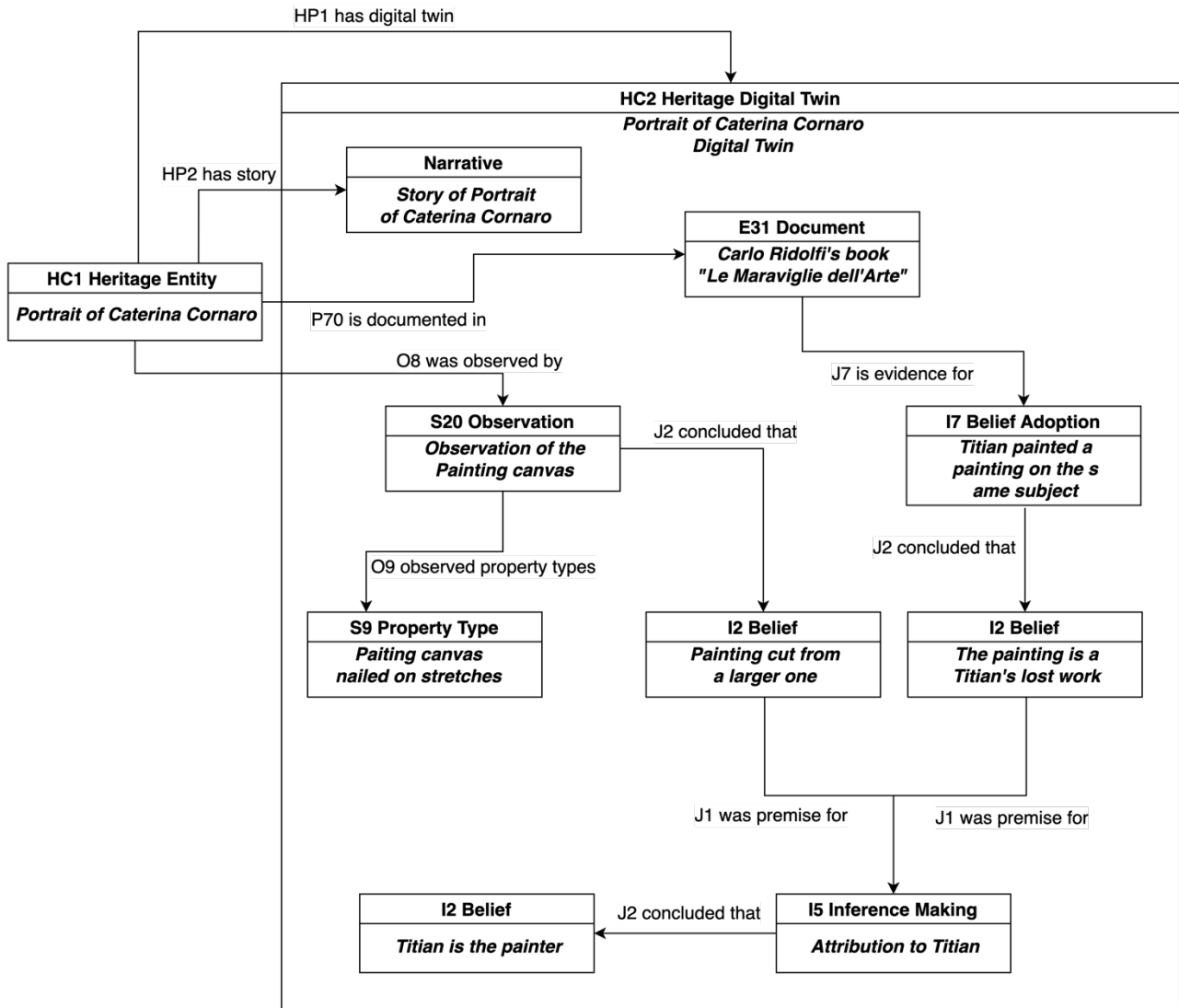


Figure 4. Modelling Art History research activities

### 8.1.3 Modelling Heritage Science results

For the Heritage Science test, we chose spectroscopy and x-ray imagery as representative examples of scientific investigations (also highlighted in boldface in this document, see above). In this case, the information derives from a spectroscopic examination of the painting (HC1 → P39 → D11), intended to identify specific pigments (P17 → E7), that was performed using a specific device (L12 → D8) equipped with additional physical accessories (P16 → E19), operated by specific software (L23 → D14) and using a series of settings and environmental parameters (L10 → D9).

The spectroscopic analysis produced a series of results (L20 → D9) which, once evaluated (O16 → S6 → O11 → S9), led the experts to some conclusions (J2) regarding the chemical elements used by the artist as characteristic of a certain historical period (I2), in this case the “Renaissance”.

The painting was also analysed by means of x-ray imaging techniques (L1 → D2) which created an x-ray image (HC7) of it, through which bright areas present on the painting are observed (O8 → S20 → S9) and are associated (J2) by the experts with the use of pigments composed of heavy elements (I2).

Also in this case, as for Art History, the final scientific results can constitute the bases (J1) for the attribution of the painting to a certain author (I5 → J2 → I2), in this case to Titian. The full model is shown in Figure 5.



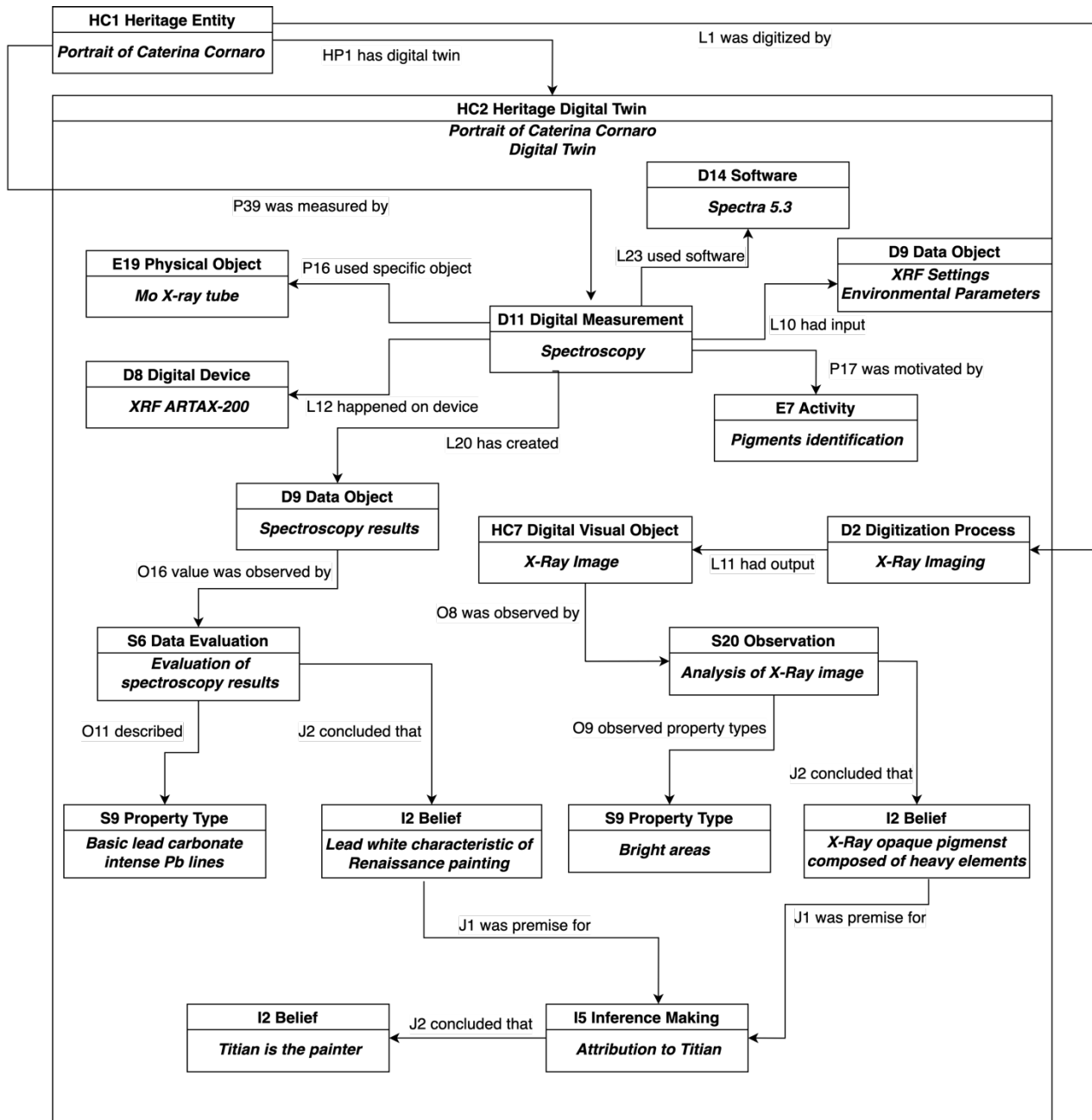


Figure 5. Modelling Heritage Science research activities and their results

## 8.2 The monastery of St. John Lampadistis in Kalopanayotis, Cyprus.

### 8.2.1 Tangible aspects

The monastery of St. John Lampadistis consists today of a group of buildings built and renovated in different periods. While its founding date is unclear, the *katholicon* (the monastery church), dedicated to Saint Herakleidios, born in the village of Lampadistis and the first bishop of Cyprus, is dated to the 11th century. Among the wall-paintings of the narthex an inscription, dated to the 15th century, describes this church as a “*katholiki*”. While written sources indicate the monastery was in use until the 19th century, afterwards it has been used as the main church of the village. Apart from the complex of the three churches under one roof, a phenomenon unique for Cyprus, there are other monastic buildings including cells, auxiliary rooms and an oil press. One of the rooms is used today to house icons from the monastery as well as other churches of the village of Kalopanagiotis [31].

The main monastery church is a domed cross-in-square structure, dated to the 11th century. In the 12th century the vaulted chapel of St. John Lampadistis was added to the north of the first church, above the tomb of the Saint. Relics of the saint, kept and displayed in a rich reliquary, and his tomb are now incorporated in the church as part of this chapel. This second chapel collapsed and was almost entirely rebuilt in the 18th century. In the middle of the 15th century a common narthex was built to the west of the two churches.

During the second half of the 15th century a vaulted chapel was added to the north of that of St. John. It became known as the 'Latin chapel' because of the assumption that it was built for the Latins (Catholics) and indicating the possible co-existence of the two rites in the same church, in line with the atmosphere of tolerance which prevailed in Cyprus after the Council of Florence (1439). Sometime between the 15th and the beginnings of the 18th century), a timber roof covered with flat hooked tiles sheltered the entire roof complex. As a result of its tripartite character, the building acquired an external image of a large building covered with a timber roof.

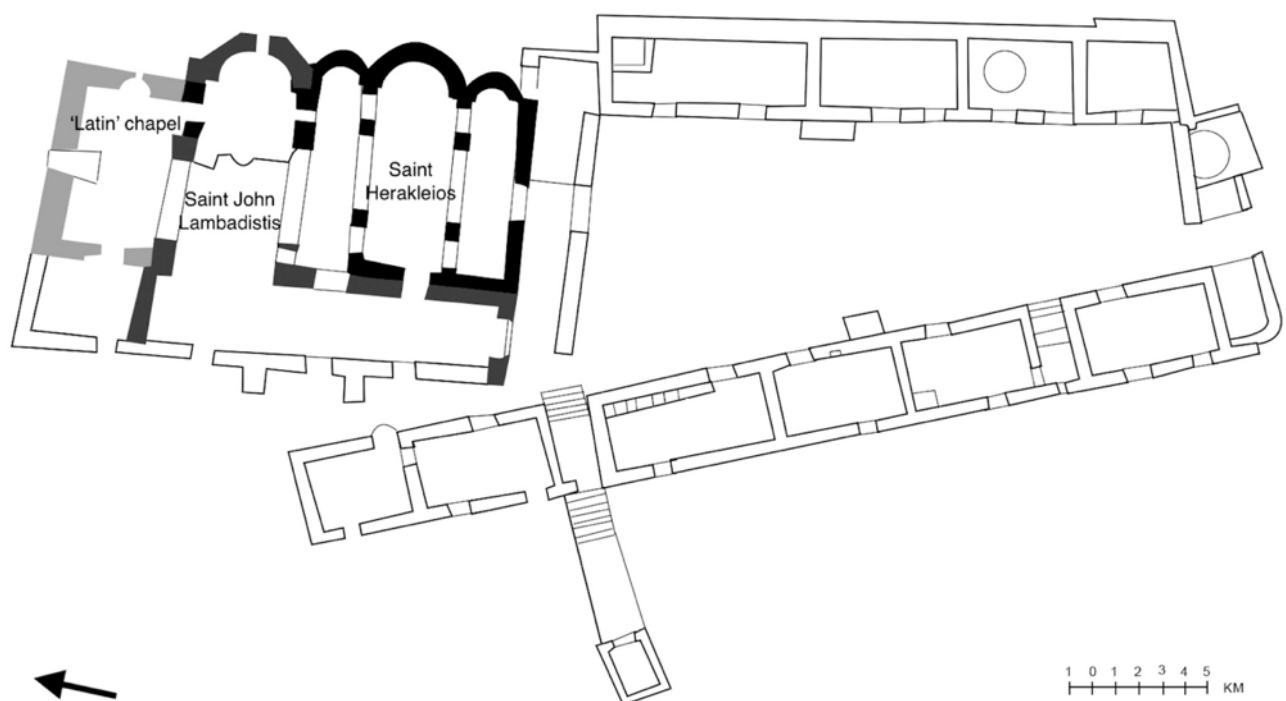


Figure 6. The plan of St. John Lampadistis complex (after [32], Figure 4)

The wall-paintings of the church are primarily from the 11th to the 15th centuries. The apse of the church, as well as some other parts, preserve fragmentary scenes dated to the 11th and 12th century. The rest of the church was painted in the 13th and 14th century. The narthex was decorated in a later period and is the work of an artist from Constantinople, who fled to Cyprus after the fall in 1453. These wall-paintings follow the trends of the Byzantine capital. The 15th century frescoes of the 'Latin' chapel display Byzantine and Italian Renaissance elements. A wooden templon screen, the oldest in Cyprus and dated to the 13th-14th century, is painted with decoration imitating coats-of-arms.

### 8.2.2 Intangible aspects

#### *The life of St. John Lampadistis*

A now lost 1640 manuscript written by a priest named Savvas from the village of Agios Theodoros Agrou and later copied by monk named Kirililos, of the Stavrovouni Monastery in 1903 narrates the life of St. John Lampadistis and the prayer said in church on his feast day on 4th of October. St. John was born in the village of Lampadistis, a now extinct village presumably located somewhere between the modern villages of Galata and Kakopetria, being the last offspring of the priest's village, Papa Kyriakos and his wife, Anna. His birthdate is unknown, probably sometime in the early 17th century. Early in his life, he was sent to learn to read and

write through the study of the Holy Scriptures, the child showing great aptitudes. At the time of his adulthood, a marriage was arranged for him, but the parents of his future wife served him poisoned fish, causing him to lose his eyesight. Now being unsuitable to marriage, St. John turned to spiritual life and spent his days in prayer. After 12 years of living as a blind person, he died. Soon people reported light beaming from his grave. Word spread and people known to be possessed came seeking John's body to pray to it and be cured. Following pressure from the people, his father agreed to open his son's tomb and found his relic's, while his heart was preserved "like a dry fig". Hence, John's remains were deposited in the chapel of St. Herakleidios at Kalopanayiotis. The many miracles attributed to John earned him the status of a saint. His father, hearing a voice telling him to build a church dedicated to his son, which he did and the church of St. John Lampadistis was built adjacent to the one of St. Herakleidios. St. John's reputation as a miracle maker expanded and the inhabitants of the island decided to view him as their guardian saint.

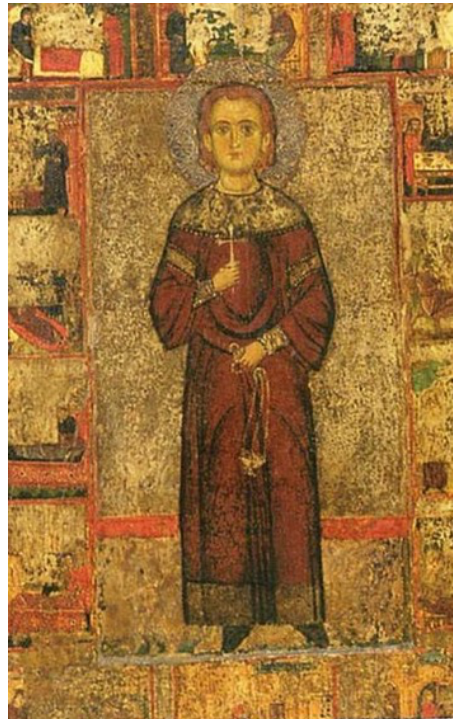


Figure 7. Vita icon depicting St. John and central episodes from his life

#### *Pilgrims and their testimonies - the Karamanlides*

The Karamanlika are a Turkish-speaking population of Greek Orthodox rite originally from Asia minor. During the 18th century a group of them, while on pilgrimage to the Holy Land, passed by the monastery of St. John Lampadistis and left several written testimonies of their passage on the church's wall close to its reliquary, the earliest being dated to 1749 and the latest to 1880, with a highest frequency between 1770 and 1780.



Figure 8. The Saint John Lampadistis reliquary, showing graffiti carved on the wall by Karamanlika pilgrims

### 8.2.3 Digital models

The 3D model of the monastery and related historical imaging documentation are available here:

<https://modelier.us.aldrin.io/models/d10a14ee-4192-45b6-b54b-8d11ecd6f70e/v2/embed/>

and a 360 panoramic walk-through is available here:

<https://dioptra.cyi.ac.cy/sites/360/lampadistis/>

part of a broader online-accessible database of the Troodos complex monasteries registered as World Heritage Sites <http://ihat.cyi.ac.cy/?q=Collection>

### 8.2.4 Ontological modelling examples

From the ontological point of view, the information about the structure of the monastery of St. John Lampadistis can be modelled by combining the classes and properties of the HDT Ontology with those of CIDOC CRM and CRMba. The monastery (HC1) is in fact made up of various buildings (P46 → B1) including the central church (*katholicon*) where the frescoes are found. Each building can in turn contain several building sections (BP1 → B2) clearly identifiable by their morphology, and of each of them it is possible to specify and describe the construction phases (P108 → E12) and the relative dating (P4 → E52). As far as the frescoes on the walls are concerned, it is possible to identify those pertaining to each section of the building (P56 → E26) so that even for them the phases of realization (P108 → E12) and the dating (P4 → E52) can be clearly distinguished. Figure 9 illustrates the case of the modelling of the Latin Chapel and its frescoes, highlighting, through the distinction between the two different production activities (E12 → P4 → E52), how the painting of this section of the building is coeval with the construction of the chapel itself (both belonging to the 15th century, in this case). The monastery is also visually represented through various digital technologies (3D models, 360 panoramic views, etc.) which can be perfectly represented through our modelling tools either with (P129 → HC8) as in Figure 9, or with (L1/L11 → HC8) when describing the digitisation process is needed.

The HDT Ontology and the CIDOC CRM ecosystem can be also employed to model the intangible aspects of the monastery, represented by its history (HP5 → HC4 and HP2 → Narrative) and linked in particular to the life of St. John Lampadistis, narrated by the priest Savvas (HP4 → Narration and HP8 → E31) in a manuscript (P128 → E18) that is now lost (P44 → E34). It is also possible to model the information related to the graffiti engraved on the walls of the monastery church by the Karamanlika pilgrims (P56 → E26), to specify their approximate dating (the 18th century in this case) (P4 → E52), and to link the images that depict them (P138 → E36).

Finally, the complex of information described above and their mutual relationships form the elements that make up the Heritage Digital Twin of the monastery of St. John Lampadistis. The HDT Ontology establishes this fundamental relationship via the property HP1 has digital twin, used to bind each Heritage Entity (HC1) to its digital representation (HC2 Heritage Digital Twin).

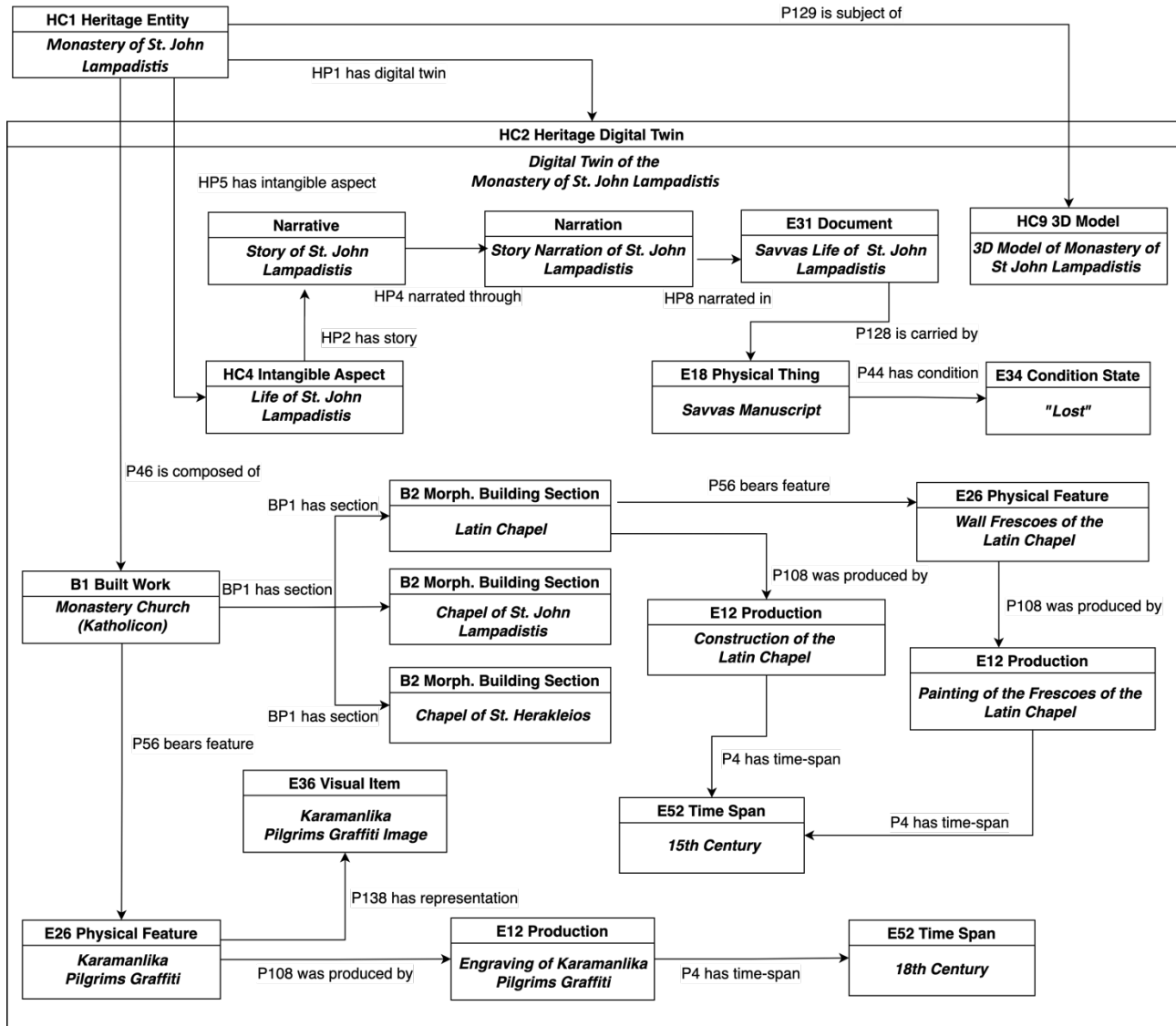


Figure 9: The ontological model of (part of) the Heritage Digital Twin of St. John Lampadistis Monastery

### 8.3 Final considerations on the examples

The patterns built by combining the different extensions of the CIDOC CRM and our new classes and properties are used to model different situations and scenarios. They can be replicated to describe all the other observations, analyses, hypotheses, and any inferences and conclusions that can be drawn from them. The conclusions in turn can serve to support new hypotheses and theories, the graph can be extended without any limitation.

Finally, as seen in the diagrams, all the semantic information acquired about the Heritage Entity and modelled through the ontology eventually becomes part of the Digital Twin (HC2), in which all the knowledge about the object is organized.

## 9. Conclusions and further work



This paper is the outcome of a research activity which is still ongoing. We decided to publish it as it is now, in our opinion a stable and coherent version, to open a discussion and improve it with the contributions of the users' community. We believe that the proposed solution reconciles the different and sometimes incompatible approaches summarily sketched in section 3 and enables an approach a la carte according to different user needs while preserving overall interoperability. Existing documentation may be easily mapped to the HDT ontology and different systems may be integrated into it, combining the existing information under the HDT overarching umbrella. Doing so will enhance the interoperability of different systems and enable the creation of a well-organized data space to support the work by researchers, heritage professionals and the public. Moreover, the HDT is ready for simulation – the second step of digital twin systems – such as answering to “what if” questions as well as causing automatic reactions to actual external events. The CRM community (CRM SIG) is currently working on a new extension for representing not only past activities but also future ones, CRMact<sup>14</sup>, that could be the way towards modelling such dynamics.

Our plans for future work include polishing the HDT ontology with the suggestions coming from the community. In parallel, we will start mapping existing schemas, such as the Europeana EDM ontology<sup>15</sup> or those based on HBIM or 3D annotations, on the HDT ontology adding specific classes and properties if needed, to ultimately create a framework where everybody can recognize their own documentation scheme but can also interoperate with those created by others.

We believe that this objective is a necessary condition to support collaborative scenarios on digital cultural heritage as those envisaged by the EU strategies on digitization.

#### Grant information

The research leading to the present paper was partially funded by the following grants:

ARIADNEplus: European Commission programme H2020 (INFRAIA-01-2018-2019) Grant no. 823914

4CH: European Commission programme H2020 (DT-TRANSFORMATIONS-20-2020) Grant no. 101004468

#### References

1. Pollard, A. M. and Bray, P. (2007) “A Bicycle Made for Two? The Integration of Scientific Techniques into Archaeological Interpretation”, *Annu. Rev. Anthropol.* 2007. 36:245–59
2. Meghini, C., Scopigno, R., Richards, J.D., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavriliis, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C. and Vlachidis, A. (2017) “ARIADNE: A Research Infrastructure for Archaeology”. *JOCCH Special Issue on Digital Infrastructure for Cultural Heritage*, Part 2 Volume 10 Issue 3, Article No. 18
3. Richards, J and Niccolucci, F. (eds.) (2019) *The ARIADNE impact*. Budapest, Archaeolingua.
4. Moscati, P. (2019) “30 anni di Archeologia e Calcolatori. tra memoria e progettualità”, *Archeologia e Calcolatori* 30, 2019, 1-12
5. Djindjian, F. (2019) “Archaeology and computers: a long story in the making of modern archaeology”, *Archeologia e Calcolatori* 30, 2019, 13-20
6. Gardin, J-C. (1979) *Pour une archéologie théorique* Paris, Hachette, 1979. English translation: *Archaeological Constructs*, Cambridge, Cambridge University Press, 1980
7. Dallas, C. (2016) “Jean-Claude Gardin on archaeological data, representation and knowledge: Implications for digital archaeology”, *Journal of Archaeological Method and Theory* 23 (1), 305-330
8. Marlet, O., Zadora-Rio, E., Buard, P.Y., Markhoff, B. and Rodier, X. (2019): “The Archeological Excavation Report of Rigny, an Example of an Interoperable Logician Publication.” *Heritage* 2019, 2(1), 761-773; <https://doi.org/10.3390/heritage2010049>
9. Zadora-Rio, E. and Galinié, H. (2020) *L'église de Rigny et ses abords. De la colonia de Saint-Martin de Tours au transfert du centre paroissial (600-1865)*. Presses universitaires de Caen
10. Dell'Unto N. and Landeschi G. (2022) *Archeological 3D GIS*. London & New York, Routledge

---

<sup>14</sup> <https://cidoc-crm.org/crmact/node/8732>

<sup>15</sup> <https://pro.europeana.eu/page/edm-documentation>

11. Bastem, S. S. and Cekmis, A. (2021) Development of historic building information modelling: a systematic literature review, *Building Research & Information*, <https://doi.org/10.1080/09613218.2021.1983754>.
12. Diara, F. and Rinaudo, F. (2021). ARK-BIM: Open-Source Cloud-Based HBIM Platform for Archaeology *Applied Sciences* 11, no. 18: 8770. <https://doi.org/10.3390/app11188770>.
13. Apollonio, F. I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A. R., Scopigno, R. and Sobrà, G. (2017) "A 3D-centered information system for the documentation of a complex restoration intervention", *Journal of Cultural Heritage*, 2017, Vol 29, 89-99
14. De Luca, L., Abergel, V., Guillem, A., Malavergne, O., Manuel, A., Néroulidis, A., Roussel, R., Rousset, M. and Tournon S. (2022) "L'écosystème numérique n-dame pour l'analyse et la mémorisation multi-dimensionnelle du chantier scientifique Notre-Dame-de-Paris", <https://hal.archives-ouvertes.fr/hal-03826931>
15. Croce, V, Caroti G., De Luca L., Jacquot K., Piemonte A. and Véron, P. (2021) "From the Semantic Point Cloud to Heritage-Building Information Modeling: A Semiautomatic Approach Exploiting Machine Learning" *Remote Sens.* 2021, 13(3), 461
16. Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021) "Digital twin paradigm: A systematic literature review." *Computers in Industry*, 130, 103469
17. European Commission (2022). *Destination Earth*. <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>
18. Jones, D., Snider, C., Nassehi, A., Jason Yon, J. and Hicks, B. (2020) "Characterising the Digital Twin: A systematic literature review" *CIRP Journal of Manufacturing Science and Technology* 229 (2020) 36-52
19. VanDerHorn, E. and Sankaran M. (2021) "Digital Twin: Generalization, characterization and implementation", *Decision Support Systems* 145 (2021) 113524
20. Qinling Qi, Tianliang Hu, Fei Tao and Nabil Anwer (2021) "Enabling technologies and tools for digital twin". *Journal of Manufacturing Systems* March 2021. <https://doi.org/10.1016/j.jmsy.2019.10.001>
21. Bolton, A., Lorraine, B., Dabson, I., Enzer, M., Evans, M, Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A. and Makri, C. (2017) *The Gemini Principles: Guiding values for the national digital twin and information management framework*. <https://doi.org/10.17863/CAM32260>
22. Shahzad, M., Shafiq, M.T., Douglas, D. and Kassem, M. (2022) "Digital Twins in Built Environments: An Investigation of the Characteristics, Applications, and Challenges." *Buildings* 2022,12,120. <https://doi.org/10.3390/buildings12020120>
23. Jouan, P., and Hallot, P. (2020) "Digital twin: Research framework to support preventive conservation policies" *ISPRS International Journal of Geo-Information*, 9(4), 228
24. Zeyu Wang, Weiqi Shi, Kiraz Akoglu, Kotoula, E., Ying Yang, and Rushmeier, H. (2018) "CHER-Ob: A Tool for Shared Analysis and Video Dissemination" *ACM J. Comput. Cult. Herit.* 11, 4, Article 18 (November 2018)
25. Baca, M. and Harpring, P (eds.) (2022) *Categories for the Description of Works of Art*, Getty Trust and College Art Association. [https://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](https://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)
26. Niccolucci, F., Felicetti, A. and Hermon, S. (2022) "Populating the Digital Space for Cultural Heritage with Heritage Digital Twins" *Data* 2022, 7(8), 105; <https://doi.org/10.3390/data7080105>
27. Bartalesi, V., Meghini, C. and Metilli, D. (2017) "A conceptualisation of narratives and its expression in the CRM" *Int. J. Metadata Semant. Ontol.* 2017, 12, 35–46, <https://doi.org/10.1504/IJMSO.2017.087692>
28. Meghini, C., Bartalesi, V., and Metilli, D. (2021) "Representing Narratives in Digital Libraries: The Narrative Ontology". *Semantic Web*, vol.12, no. 2, pp. 241 – 264, 2021
29. Carriero, V. A., Gangemi, A., Nuzzolese, A. G. and Presutti, V. (2019) "An Ontology Design Pattern for Representing Recurrent Events" *10th Workshop on Ontology Design and Patterns (WOP 2019), co-located with 18th International Semantic Web Conference (ISWC 2019)*, Auckland, New Zealand, October 27, 2019
30. E. Poyiadji-Richter (ed.) (2019) *The Secrets of a Royal Portrait*, Nicosia, Leventis Foundation



31. Hermon, S. and Bakirtzis, N. (2019) "More than Meets the Eye: Integrating Art History, Digital Technologies and Natural Sciences", in E. Poyiadji-Richter (ed.) *The Secrets of a Royal Portrait*, Nicosia, Leventis Foundation, pp. 37-46
32. Irakleous, S. and Bakirtzis, N., 2020. "Karamanlides at the monastery of Saint John Lampadistis in Kalopanagiotis: pilgrimage votives, architecture and sacred landscape" *Cahiers du Centre d'Études Chypriotes*, (50), pp. 383-405.

# Artificial Intelligence and Ontologies for Managing Heritage Digital Twins Data

Volume 10 · Issue 1 January 2025







*data*



Article

---

# Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data

---

Achille Felicetti and Franco Niccolucci



<https://doi.org/10.3390/data10010001>

## Article

# Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data

Achille Felicetti \*  and Franco Niccolucci 

VAST LAB, PIN, Piazza dell'Università 1, 59100 Prato, Italy; franco.niccolucci@gmail.com

\* Correspondence: achille.felicetti@gmail.com

**Abstract:** This study builds upon the Reactive Heritage Digital Twin paradigm established in prior research, exploring the role of artificial intelligence in expanding and enhancing its capabilities. After providing an overview of the ontological model underlying the RHDT paradigm, this paper investigates the application of AI to improve data analysis and predictive capabilities of Heritage Digital Twins in synergy with the previously defined RHDTO semantic model. The structured nature of ontologies is highlighted as essential for enabling AIs to operate transparently, minimising hallucinations and other errors that are characteristic challenges of these technologies. New classes and properties within RHDTO are introduced to represent the AI-enhanced functions. Finally, some case studies are provided to illustrate how integrating AI within the RHDT framework can contribute to enriching the understanding of cultural information through interconnected data and facilitate real-time monitoring and preservation of cultural objects.

**Keywords:** artificial intelligence; ontologies; digital twins; cultural heritage



Academic Editor: Francesco M. Donini

Received: 3 December 2024

Revised: 19 December 2024

Accepted: 24 December 2024

Published: 26 December 2024

**Citation:** Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2025**, *10*, 1. <https://doi.org/10.3390/data10010001>

**Copyright:** © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Digital twins are rapidly transforming various sectors, offering a powerful means to create digital replicas of real-world objects and systems. In their broadest sense, digital twins can be seen as virtual representations of physical systems that mirror their real-world counterparts. These representations are kept up to date through a continuous flow of information between the physical and digital realms. The concept of digital twins extends beyond a simple one-to-one mapping of physical attributes, encompassing the system's environment, processes, and their interconnections.

Digital twins are finding applications across a wide range of industries and research fields, demonstrating their versatility in representing and managing complex real-world systems. While initially prominent in industrial settings for testing and simulating the behaviour of machinery and devices, digital twin technology has expanded into diverse sectors. For instance, the European Commission has proposed the ambitious “Destination Earth” initiative [1], aiming to create a comprehensive environmental model of the Earth for forecasting and mitigating the impacts of climate change. In the realm of architecture and urban planning, digital twins, particularly within the framework of Building Information Modelling (BIM), are being used for documenting historical structures, designing new buildings, and for large-scale urban planning projects. The increasing use of sensors and the Internet of Things (IoT) has further broadened the applications of digital twins, allowing for real-time monitoring and data analysis in fields like cultural heritage preservation and restoration.

In the ever-evolving landscape of digital innovation, data stand as the lifeblood of every digital twin. As faithful digital replicas of physical systems, these digital objects are

not merely static representations but also dynamic entities that thrive on the continuous flow and integration of data. The complexity and diversity of these data necessitate a robust framework for their management, and it is here that ontologies emerge as fundamental tools. Ontologies, designed for the rational and coherent organisation of data, provide the ideal structure for managing the intricate web of information that constitutes a digital twin as they enable the classification, interlinking, and querying of data, thereby facilitating a deeper understanding and more effective utilisation of the information.

The recent advent of artificial intelligence adds a new dimension to this digital ecosystem. AI technologies offer advanced tools for data analysis, pattern recognition, and predictive modelling, enhancing the capabilities of digital twins and ontologies. This synergy between digital twins, ontologies, and AI creates a powerful combination that can constitute a new pillar of modern data management, in which digital twins provide the virtual canvas, ontologies the structured framework, and AI the analytical skills, collectively enabling unprecedented insights and informed decision-making processes. As digital twins evolve, ontologies ensure that the data remain organised and accessible while AI leverages these data to drive innovation and efficiency. Together, these elements form the foundation upon which the future of data-driven systems is built, promising a new era of technological advancement and human understanding.

In this paper, we present an exploration of the role that digital twins are assuming in the context of cultural heritage, with a particular emphasis on the critical aspect of data management. We will offer an overview of the Heritage Digital Twin model that we are defining and delve into the relationship between data, ontologies, and artificial intelligence, highlighting how these elements form the cornerstones that define the future of data-driven cultural digital twin systems, demonstrating how they are collectively advancing our ability to manage, preserve, and understand our shared cultural legacy knowledge.

## 2. Digital Twins and Cultural Heritage

Digital twins have increasingly become relevant in various research and application domains, reflecting the growing complexity of their role as digital counterparts of real-world entities and phenomena. As explored in many reviews across disciplines [2,3], the digital twin concept not only serves to capture the intricacies of digitised entities in both industrial and scientific settings but also reflects a broader philosophical engagement with how we represent and interact with reality. The term is often used loosely, sometimes as a generic descriptor for any digital replica of real-world objects, while at other times to mean a more rigorous, structured information system that interfaces with the physical world. The Destination Earth initiative [1], promoted by the European Commission and aimed at creating a complete and accurate digital twin of the planet and its environment, demonstrates that the concepts and terms of this technology are resonating even within high-profile regulatory bodies.

Originally, digital twins served the need for production optimisation, performance enhancement, product innovation, and predictive maintenance of the industrial sector. However, over time, the potential of this technology has expanded into various fields, such as environmental engineering and urban planning. Notable examples in these fields include the digital twin of Zurich, aimed at improving urban planning [4], and the digital twin of Singapore, which integrates data from multiple sources, such as sensors, satellites, and drones, to create a virtual replica of the city for urban growth management and environmental monitoring [5]. In other fields, such as architecture and construction, digital twins are revolutionising the design and building processes by enabling accurate simulations of how structures interact with their surrounding environment before they are physically constructed. This allows for the early detection of design flaws, reducing maintenance



interventions. Furthermore, scientific research has begun to harness the power of digital twins to create virtual laboratories which simulate complex natural phenomena and deliver reliable results in reduced timeframes, minimising the risks and costs associated with real-world experimentation. NASA, for instance, has long embraced digital twin technology [6] and recently has also applied it to its Mars 2020 mission, for which the digital twins of the Perseverance rover and the Ingenuity helicopter were created to simulate their physical counterparts in real-time, allowing for remote control and monitoring of their interaction with the Martian environment.

In light of these numerous and significant examples, the time seems ripe for applying this transformative technology to the realm of cultural heritage.

### 2.1. Defining Digital Twins for Cultural Heritage

The definition of what a digital twin actually is varies depending on its intended use and the context in which it is applied. A recent survey by Semeraro et al. [3] lists about 30 different definitions of digital twin, including the famous (and in some ways foundational) one formulated in 2014 by Grieves [7]. But, when considering its application in cultural heritage, a deeper reflection on methodology is required to define a model capable of ensuring an accurate and faithful representation of real-world cultural entities. It is important to note that in many cases, digital twins in other fields are created either before the real object exists, as in predictive industrial production or urban planning, or simultaneously with it, as in the case of NASA's space missions. In contrast, cultural heritage objects typically are complex entities that have existed since long before their digital reproductions and have lifespans that extend across multiple temporal and spatial dimensions. Thus, defining the nature and identity of cultural objects is a necessary precondition to formulating an appropriate paradigm for their digital representation. A "Heritage Digital Twin", as we defined it, must certainly be understood as a "duplicate" of a real cultural entity. However, it would be overly simplistic to reduce this duplication to merely physical and material aspects, overlooking the intricate network of information of which the essential nature of a cultural object consists. Cultural entities, moreover, may be immaterial, as in the case of events and cultural manifestations, such as the *Palio di Siena* [8–10] or the *Calcio Storico* in Florence [11,12]. These intangible heritage assets, which exist beyond any physical form, persist through documentation, serving as testimony to past events or as formulations for future activities.

While initial approaches to the digital twin in cultural heritage were closely tied to the notion of three-dimensional (3D) visualisations as mere "digital replicas", especially in archaeology, this has evolved significantly in recent years. In fact, early visual reconstructions, dating back to the late 20th century, were motivated by the need to reproduce the imagined pristine appearance of ruins and artefacts, replacing traditional methods such as drawings and physical casts. However, this approach has rapidly expanded to include the documentation of artefacts, monuments, and sites, leading to an extensive production of 3D models of varying quality and detail. As the field matured, it became clear that shape alone was insufficient for advanced research or practical applications. The 3D-COFORM project (2008–241) [13] had the merit of demonstrating that comprehensive documentation linked to 3D models is essential for conservation and restoration purposes and that information about materials, construction methods, and interventions can be embedded directly into the 3D models. This principle is further extended in Heritage BIM (HBIM) [14], a specialised adaptation of Building Information Modelling for historic buildings, in which information is structured according to a standard data model, the Industry Foundation Classes (IFCs) [15], aimed to integrate concepts unique to cultural heritage. Yet, both the 3D-based approach and the HBIM methodology reveal inherent limitations since they treat the 3D model as the root of all documentation,



making any additional data a mere appendage and impeding large-scale comparative analyses across different assets.

A more robust solution lies in a semantic approach, where information is organised in classes and interlinked using meaningful relationships, enabling efficient querying across various heritage objects. For example, by linking materials (such as wood) to heritage assets via properties (such as “is made of”), one can efficiently retrieve all assets constructed using wood. The core of a Cultural Digital Twin, then, is not merely its 3D visualisation but the aggregation of various forms of documentation—literary sources, photographs, archival material, scientific analyses, and more. 3D models can indeed be powerful and versatile tools for rendering the tangible aspects of cultural heritage (such as size, shape, colour, and texture of objects and monuments), but they are far less suited for representing the historical narratives, artistic values, social relevance, and other intangible elements that are crucial to fully understanding cultural objects. These aspects require expression through more advanced semantic modelling frameworks. This is where ontologies come into play, allowing for the structured organisation of knowledge in a way that goes beyond simple shape or geometry.

## 2.2. Heritage Digital Twins

According to these considerations, we defined [16–18] the concept of Heritage Digital Twin (HDT) as the complex of all available digital information concerning a real-world heritage asset, either movable (e.g., cultural objects), immovable (e.g., monuments and sites), or even intangible (e.g., cultural manifestations or traditions). This paradigm, unlike 3D-centric systems, encompasses all available digital cultural information concerning cultural objects, including visual representations, reports, documents, scientific analyses, conservation data, and historical research, a perspective that emphasises the interconnectedness of various data, moving beyond the limitations of 3D models as the primary source of information.

Moreover, the implementation of digital twins relies on several interconnected components that work together, making digital twins “reactive” to ensure the effective monitoring and preservation of cultural heritage assets. Components include sensors, i.e., physical devices strategically positioned on or near the heritage asset, aimed at measuring specific physical quantities like temperature, humidity, or vibration, and deciders for comparing the incoming sensor data against established thresholds or rules for determining whether any actions are required. Additional components in this system are the activators, which carry out real-world actions based on the decisions made by the deciders. The digital twin thus creates a reactive loop that continuously monitors and responds to changing conditions, ensuring optimal preservation of cultural heritage and whose functioning relies heavily on aggregated data.

Data, therefore, constitute the backbone of the digital twin system and serve as a comprehensive repository of knowledge about a heritage asset, supporting informed decision-making in all aspects of its management. In this perspective, our digital twin paradigm becomes instrumental in advancing data management and prescribes the use of efficient data organisation tools, like ontologies, to improve the effectiveness of heritage documentation.

## 2.3. Digital Twins and Ontologies

Over time, ontologies have proven to be extremely powerful tools when dealing with intricate and multifaceted data, such as those produced within the cultural heritage domain. They provide a structured, formal framework that enables the organisation, interlinking, and querying of diverse data types, thereby facilitating a deeper understanding and more

effective utilisation of complex information. Additionally, ontologies provide the formal description of all digital twin operations, ensuring the constant preservation of the context in which its components operate and creating a sort of real-time “digital memory” that makes it possible, at any time, to retrieve, understand and reuse all the information in a complete and accurate way. The use of standardised vocabularies further ensures the consistency and uniformity of data, an essential aspect for guaranteeing interoperability among different systems and platforms. Thanks to these characteristics and their ability to effectively orchestrate the use of aggregated documentation, ontologies can serve as the “engine” of digital twins throughout every phase of their operational life [19].

Ontologies also offer benefits in terms of the long-term preservation of digital twin information, particularly in terms of data durability and context preservation [20]. Ontological data, in fact, can be permanently stored in plain text files, which are highly durable supports and can be easily read and interpreted by humans and machines, thus becoming an ideal means for preserving over time the information and the underlying context in which digital twins operated.

We formalised all the above concepts and definitions in a couple of seminal papers [16,17], where these considerations have been explored in depth. In the same works, we have also proposed some solutions that seem particularly adequate for dealing with the scenario we have outlined. Thus, we introduced the Reactive Heritage Digital Twin Ontology (RHDTO) [18], a compatible extension of the widely used CIDOC CRM conceptual reference model [21] (henceforth “the CRM”), providing a standardised framework for organising and interlinking diverse heritage data within digital twins, enabling interoperability with other CRM-compliant systems. The first release of RHDTO was met with considerable enthusiasm from the scientific community, and the paper in which it is described [16] quickly gained significant popularity, emerging as the winner of the Data 2022 Best Paper Award. Notably, digital twins will also constitute the conceptual foundation of ECHOES, the recently started initiative launched by the European Commission to implement the future European Collaborative Cloud for Cultural Heritage (ECCCH) [22]. The RHDTO is also the backbone of the forthcoming large-scale ARTEMIS project, which will start in 2025, developing virtual restoration tools. The next Section of this paper provides a general overview of this new ontology and its various features.

### 3. The RHDTO Model

The Reactive Heritage Digital Twin Ontology (RHDTO) we defined [18] is a model developed to manage and ensure interoperability among the vast array of data that constitute the backbone of the heritage digital twin. Its development is grounded in the experience gained from significant data aggregation and integration projects, such as ARIADNE (harmonisation of archaeological data) [23–25] and 4CH (study and preservation of cultural heritage) [26]. The RHDT ontology is constructed as an extension of the CRM, a reference model for cultural heritage data, and proposes a series of entities designed to express all available cultural and scientific documentation in a standardised format. Classes such as *HC4 Intangible Aspect*, *HC5 Digital Representation*, and *HC9 Heritage Activity* that we introduced are employed to model cultural entities and their documentation semantically, while specifically designed properties are used to express the complex relationships with events, people, places, and actions that constitute their history and cultural value.

The RHDTO represents an innovative solution to manage and interconnect the broad spectrum of data that are the informational core of the digital twin of a heritage asset. The methodological choice to design it as an extension of the CRM stems from the novelty of the cultural digital twin concept and its logical structure. On the other hand, CRM is an ISO standard widely used for cultural heritage documentation, and its compatibility guarantees

wide interoperability with existing documentation. The ontology defines *ex novo* the entities necessary to describe the specific elements of this domain and to model them appropriately, maintaining consistency with the primary model (also called CRMbase to distinguish it from extensions) and its extensions used to model entities such as times, places, events, agents, and physical objects, and their mutual correlations. This synergy implements a high degree of interoperability in the documentation and analysis of cultural heritage, as it allows the alignment of the RHDTO structure with those used in other cultural heritage research domains (such as archaeology, art history, and architecture). For instance, it enables the identification and classification of architectural elements according to specific historical periods, styles, construction techniques, and materials employed and defines the diachronic destinations and uses of specific spaces or elements of a particular building.

Table 1 lists the abbreviations used for referencing the CRM model and its extensions in this paper.

**Table 1.** Abbreviations for classes and properties of CRM and extensions used in this paper.

Extension Name	Used for	Reference	Class Symbol	Property Symbol
CRM	General use	[21]	<i>E</i>	<i>P</i>
CRMdig	Digital objects	[27]	<i>D</i>	<i>L</i>
CRMsci	Scientific analyses	[28]	<i>S</i>	<i>O</i>
CRMpe	Digital Interoperability	[29]	<i>PE</i>	<i>PP</i>
RHDTO	Heritage Digital Twins	[16–18]	<i>HC</i>	<i>HP</i>

### 3.1. RHDTO General Classes

The RHDTO model is primarily conceived as a tool to capture the dual nature of cultural heritage (both tangible and intangible) and to provide, through specific classes and properties, a mechanism for dynamically documenting and analysing their mutual relationships. It is also designed to model their digital counterparts and the mechanisms that render them interoperable as interconnected elements of the cultural digital twin. The classes and properties of this ontology are identified by a code formed by the letters HC for classes and HP for properties, followed by a number.

The main class of the model, *HC1 Heritage Entity*, represents real-world entities relevant to their contribution to society, knowledge, and culture. It is a general-level class that includes both material and immaterial entities, which are modelled through its two subclasses, *HC3 Tangible Aspect* and *HC4 Intangible Aspect*. The latter is specifically intended for modelling historical and cultural events, traditions, cults, and practices typical of intangible heritage. The class *HC2 Heritage Digital Twin* represents cultural digital twins as the network of information that characterises them, effectively configuring them as informative digital replicas of real-world entities.

The ontology also provides classes used for modelling digital iconographic and multimedia representations (*HC5 Digital Representation*, *HC7 Digital Visual Object*, *HC8 3D Model*) used as additional components of heritage digital twins. Specific classes also exist to express elements of the digital documentation (*HC6 Digital Heritage Document*) that contribute to defining the identity and history of the described cultural objects. A dedicated class (*HC9 Heritage Activity*) is devoted to representing activities related to the study, scientific investigation, and digital reproduction of cultural entities. The same class can be used for any other activity relevant to the creation and management of the digital twin.

To model stories, i.e., those fragments of knowledge derived directly from documentation or oral tradition and presented in narrative form to enrich the intangible heritage of the described cultural objects, we used some classes defined in the Narrative Ontology (NOnt) [16,17,30], an ontological model developed within the MINGEI project [31] and also compatible with the CRM. The *Narrative* class of this ontology perfectly fits the concept of

story we have formulated, while the *Narration* class is used for rendering storytelling, i.e., the modalities in which the facts constituting a story are presented and shared.

The RHDTO model ensures that all instances of the used classes are always equipped with the necessary modalities to be mutually correlated. This is achieved through the use of properties inherited from the CRM, as well as through new properties specifically designed to express the complex links between cultural entities, their history, related documentation, and their digital representation. Among the new properties introduced, we have *HP1 has digital twin*, designed to indicate the relationship between cultural entities (*HC1*) and their digital twins (*HC2*); *HP5 has intangible aspect*, intended to associate material (*HC3*) and immaterial (*HC4*) aspects of cultural entities; *HP3 has story*, to link cultural entities (*HC1*) with the stories (*Narrative*) in which they are mentioned; and *HP4 is narrated through*, to indicate the storytelling modalities through which a story (*Narrative*) is told (*Narration*). This enables the creation of particularly detailed and extensive knowledge graphs that are perfectly capable of reproducing every aspect of the cultural entity for which the digital twin is built and are compatible with any other archive constructed according to the same criteria.

### 3.2. Reactive Aspects of Heritage Digital Twins

As mentioned in the previous Section, the reactive part of the digital twin is represented through a network of sensors, deciders, actuators, and other devices that are able to guarantee complex interactions with the real world. The main classes we introduce in RHDTO to model these components are *HC9 Sensor*, a specialised version of the CRMdig:D8 Digital Device, and *HC10 Decider* class, a subclass of CRMpe:PE1 Service class of the Parthenos Entities Model, another CRM-compatible ontology. As we will outline in the next sections of this paper, deciders are very peculiar components since they are responsible for analysing input from a variety of sources and generating output instructions directed toward other components or human operators.

The actions performed by deciders are modelled as events through the *HC14 Activation Event* class. This is where the decision-making process takes place, one of the actions for which the use of artificial intelligence can bring the greatest benefits to the evolution of digital twins. Activators (*HC11 Activator*), instead, serve as components that execute actions based on the instructions received from deciders.

Specific classes are provided by RHDTO to model aspects such as the placement of sensors on cultural objects or their surrounding spaces (*HP15 is positioned on*); the software through which the sensors are operated (*HP11 is operated by*); the measurements performed by them (*HC13 Sensor Measurement*); the signals (*HC12 Signal*) generated and sent by the sensors to the digital twin (*HP12 was transmitted to*). For representing human actors responsible for ensuring the safety of cultural assets, the CRM:E39 Actor class is used. This interaction underscores the importance of collaboration between the digital system and human operators in effectively managing risks.

A detailed description of all the classes and properties of the RHDTO model, along with illustrative case studies that demonstrate its effective application, is provided in [18].

## 4. Artificial Intelligence and Cultural Heritage

The way in which groundbreaking technological advancements increasingly influence our daily lives is now well-known and evident to all. In the case of artificial intelligence, although we may not be certain we can talk about a revolution, it is clear that this technology is rapidly becoming a part of our world, even in sectors seemingly distant from the futuristic scenarios where it is usually imagined, such as cultural heritage and museums. Whether it is completing an unfinished symphony by a legendary composer [32], reconstructing and understanding ancient texts written in extinct languages [33], recreating the aspect of lost

environments [34,35], identifying archaeological sites from aerial or satellite images [35,36], or simply providing guidance on how to preserve [37] or restore [38] damaged artwork, the new AI systems are proving to be powerful tools, enriching the arsenal of digital tools available to scholars in unprecedented ways.

In recent years, Europe has taken the lead in releasing a series of recommendations that demonstrate a significant commitment to regulating artificial intelligence in an ethical and responsible manner, defining, among other things, the first-ever legal framework on AI. The 2020 *White Paper on Artificial Intelligence* [39,40] laid the groundwork by outlining a balanced approach to promote excellence and trust in AI, while the recent *AI Act Regulation of the European Commission* [41] places a strong emphasis on ethics, classifies AI systems based on risk, imposes stringent requirements for high-risk systems, and prohibits uses of AIs that violate fundamental rights. However, the European Commission seemed to underestimate the benefits that AI can bring to the cultural heritage sector. On the other hand, scholars and other operators in the sector have immediately realised the importance of AI technologies at the European level and have swiftly put them into action, developing AI-based innovative solutions to enrich and preserve cultural legacy [42,43], with the awareness that it represents the living testament to the evolution of human civilisation and not just a collection of artefacts and documents. The digital transformation of cultural heritage presents an unparalleled opportunity for AI to flourish as a research domain and to accelerate advanced access to this invaluable resource [44].

#### 4.1. Artificial Intelligence for Processing Cultural Documentation

One of the fields where artificial intelligence certainly excels is in the analysis of cultural documentation, as it brings a myriad of advantages that significantly enhance the preservation, accessibility, and understanding of cultural heritage. For instance, AI technologies can automate the collection and analysis of vast amounts of data and streamline and expedite the digitisation and archiving of cultural information, thereby reducing costs and time. Additionally, they can improve accessibility through automated translation and transcription, reducing the manual work involved in traditional documentation methods while ensuring a more comprehensive capture of their content and meaning.

In our past activities, we made many attempts to replicate some of these processes, for instance, trying to extract knowledge from textual documentation using NLP and machine learning techniques [45–47]. However, these efforts were often limited by the complexity and variability of the data, as well as the need for extensive manual annotation and preprocessing needed to train the NLP tools. With the advent of AI technologies, this process has become significantly more efficient and effective since AI models can handle large volumes of textual data, automatically identifying and extracting relevant information without the need for extensive manual intervention.

By processing vast amounts of unstructured data, AI can unveil relationships and contextual insights hidden within the textual tapestry, detecting, for instance, patterns that aid in understanding the historical conditions and transformations of cultural heritage entities, offering valuable insights for the comprehension of cultural objects. AI systems, in fact, possess the remarkable ability to identify connections and patterns that may elude human cognition, thereby catalysing the creation of new and unexpected knowledge. The latter capability transcends mere data processing and challenges our traditional notions of knowledge generation, suggesting that the pursuit of understanding, if approached and guided in the right way, can become a very fruitful co-creative human-machine process in the near future. Human supervision, in fact, remains essential in all the phases of this process to guarantee the accuracy and relevance of the extracted knowledge, as well as to guide the interpretative processes in line with domain-specific expertise.



#### 4.2. Artificial Intelligence and Ontologies

As it is widely acknowledged, emerging AI technologies are not without challenges, particularly regarding data accuracy and the reliability of outcomes. In fact, the opaque nature of AIs often makes it difficult to understand the reasoning behind their outputs, often making their responses unreliable. This can pose significant challenges in cultural heritage, where trust and accountability are paramount, since these issues may lead to the misrepresentation of cultural narratives and the distortion of historical facts. Careful verification and validation are thus essential to ensure the integrity and authenticity of the knowledge that is derived or created and that is eventually integrated into the Heritage Digital Twins.

Ontologies can become essential tools for overcoming issues related to transparency, interpretability, and accountability of the information and serve as verifiable records of data and decision-making criteria used by AI systems. Since ontologies especially excel in capturing domain-specific knowledge and causal relationships in a structured format, they can provide AIs with unambiguous representations of the analysed and generated data and their context in a “language” that allows them to “reason” in formal ways and produce clean and contextually relevant outputs [48]. Thanks to these features, the semantic knowledge graphs, typically implemented through ontologies, have the potential to help reduce the risk of “hallucinations”, i.e., the typical scenario in which AIs generate unreliable or erroneous information, producing outputs not grounded in the actual data or context.

Thanks to these features, ontologies can also facilitate the establishment of XAI (Explainable Artificial Intelligence) systems, an evolution now considered necessary for making AI decisions and processes understandable and transparent, enabling users to comprehend how models work [49]. The integration of ontological knowledge not only promotes the development of more explainable and trustworthy tools but also aligns these technologies with ethical standards and human expectations, contributing to the responsible advancement of AIs [50,51].

#### 4.3. Artificial Intelligence and Heritage Digital Twins

AIs and ontologies provide even bigger advantages when employed together for implementing the Heritage Digital Twin model, a system which operates on top of semantically rich knowledge graphs. Particularly with regard to scenarios of documentation and preservation of historical sites, artefacts, and intangible heritage, the combination of these powerful technologies significantly enhances the functionality of digital twins, transforming them from static models into dynamic systems capable of enriching, simulating, predicting, and optimising real-world processes mostly in real-time and in completely new ways.

Some recent attempts to employ AI in digital twins have produced very encouraging results in different research fields, highlighting several benefits already in the present and leaving room for great future developments [52,53]. In cultural heritage, where traditional digital twins may focus on documenting and preserving physical objects or sites, AI-powered components can expand computational features by offering dynamic insights through continuous learning and adaptation. For example, an AI relying on the rich knowledge bases of the Heritage Digital Twins can guess and simulate how an ancient building may have been used, predict the long-term effects of restoration techniques on delicate artefacts, or even assist in reconstructing lost historical knowledge hidden in ancient documents. AI can also assist digital twins in the continuous monitoring and analysis of data coming from sensors and other external digital sources, helping to detect modifications in the condition of heritage objects, whether due to environmental factors, deterioration, or human intervention. Additionally, AI-driven Heritage Digital Twins can speed up and increase the efficiency of predictive analysis, allowing conservators and

cultural heritage professionals to foresee potential risks such as environmental degradation or structural damage even before they actually occur, a predictive capability that would be essential in safeguarding assets that are sensitive to climate change, pollution, and other environmental condition changes over time. Moreover, AI can assist in simulating various conservation strategies within the digital twin, imagining and implementing innovative virtual environments that would allow experts to plan and execute test interventions without endangering the original artefact or site.

The structured nature of the knowledge graph of the Heritage Digital Twins, implemented through ontologies, can further help to ensure that all the AI-driven decision-making processes remain transparent and explainable, allowing AI-empowered digital twins to provide clear motivations for all the operations performed and decisions taken. Additionally, AI can enrich and expand the digital twin knowledge graph by providing explanations along the whole decision chain, ensuring transparency and accountability also in future operations. This technological synergy is also defining a new paradigm, called Cognitive Digital Twin (CDT), that represents a significant evolution in digital twin technology, leveraging advanced AI capabilities to not only replicate physical entities but also simulate cognitive processes, potential reactions and behaviours [54].

## 5. Implementing AI-Powered Heritage Digital Twins

Building on the theoretical considerations discussed in the previous Sections, we have refined our Heritage Digital Twin model, identifying scenarios where AI could effectively contribute, in combination with the various components of the Digital Twin, to enhance its capabilities, such as the extraction and semantic encoding of information from heritage documentation, the continuous updating and enrichment of the knowledge graph, and the enhancement of analytical and predictive capabilities for identifying and preventing risk situations.

### 5.1. Artificial Intelligence and Documentation

In light of the above discussion, the most immediate and obvious way to harness the power of AI is in the analysis of historical and archival documentation included within the Heritage Digital Twin. In this scenario, AI is used to process historical documents, research papers, and even multimedia content to extract relevant data, derive dates, locations, people, and information about objects and events, and structure them according to the classes and properties of the RHDTG. AI can also handle multilingual documents, using language translation models to process texts in various languages and derive information from them. This capability is particularly useful in cultural heritage, where documentation may span different languages and historical periods. The extracted data can then be used to populate the knowledge graph of the Heritage Digital Twin, enriching its informational and historical context.

AI can also be used to foster the continuous updating of the knowledge graph as new documentation becomes available. For instance, AI algorithms can be set to periodically scan for new sources, extract information, and update the knowledge base, ensuring that the Heritage Digital Twin remains up-to-date and widely comprehensive over time. This automated process significantly reduces the manual effort required for knowledge curation and management, making it more efficient and scalable.

The enriched knowledge graph can, in turn, be used to advance the training of the AI, further enhancing its performance, allowing the refinement of its understanding and improving its ability to retrieve relevant information. This makes the AI increasingly “informed” about the most relevant facts concerning the cultural entity reproduced by the Digital Twin and its historical and cultural context.



### 5.2. Artificial Intelligence and Decider Components

The use of AI can also have a significant impact in enhancing the reactive components of the Heritage Digital Twins. In a previous work [18], we have defined the Decider as an intelligent component able to process incoming data from sensors and information stored in the digital twin knowledge base to detect potential risks and execute predefined algorithms or apply decision-making rules to determine appropriate actions to preserve cultural heritage entities. In this scenario, AI technologies can improve the efficiency of this essential component [55], especially in scenarios where real-time decisions are required. In particular, AI would allow the Decider to refine and enhance its decisional process, allowing for faster and more efficient responses to critical issues detected by the digital twin. AI tools, in fact, excel in quickly processing vast amounts of data, identifying complex patterns, and making predictions based on evolving information. Additionally, AI can efficiently combine and analyse data coming from various sources, including historical data, real-time sensor data, and scientific data, enabling more comprehensive and accurate decisions based on a wide range of inputs.

The AI can also capture the decisions made by the Decider and the context in which they were made by providing a comprehensive view of past actions and their outcomes in order to create in the knowledge graph historical records of single decisions or decision chains that can be used to improve future decision-making processes. Moreover, the predictive analytics typical of AI improve the capabilities of the digital twin to anticipate future trends and potential risks, allowing the Decider to adopt a proactive rather than reactive approach, thereby helping to prevent issues before they arise.

To optimise these AI features, the system can be additionally trained to recognise aspects critical to conservation and preservation, for instance, through the analysis of reports, protocols and recommendations produced by experts in conservation and restoration. This targeted training helps the AI focus on critical elements and make more informed decisions. Once trained, the AI can be further tested through simulated risk scenarios, allowing researchers to assess and refine its decision-making responses under controlled conditions. The responses can then be integrated back into the training process to further refine and improve the AI decision abilities. This iterative approach ensures that the AI is not only capable of making accurate decisions but also continuously improves its performance based on new data and expert feedback.

### 5.3. Embedding Physics-Based Models

As we already underlined, the efficiency of a Heritage Digital Twin lies in its ability to continuously update its information, even if, in fact, this may not be feasible for every aspect of the reproduced objects under all conditions. Indeed, while the information on cultural objects available today is very extensive, relying solely on acquired data for complex systems to work properly can prove difficult and, in many cases, unreliable. For instance, in the preservation of historical buildings, the internal condition of a centuries-old wall cannot be constantly assessed by breaking it open, and instead, inferences must be made from limited sensor data. Similarly, in analysing ancient artefacts, conservators must rely on sparse, noisy, and indirect observations to infer internal conditions and material composition [56].

A possible solution for addressing this challenge could lie in the integration, directly in the AI, of predictive physics-based models, which encode the governing laws of nature, with scalable methods for enhancing data simulation, optimisation and decision-making [57]. This would allow AI-driven deciders to provide consistent and accurate decisions based on predefined criteria and algorithms, reducing the risk of human error and bias and ensuring that decisions are made objectively and reliably.

The use of ontologies, which provide the semantic mechanisms to describe this type of information and orient AI reasoning, allows for filling the gap in the data coming from observation by means of inferences derived from the analysis of the physical and mathematical laws that commonly regulate natural processes. This approach, encompassing the field of computational science, aims to predict complex phenomena such as the deterioration of historical buildings, the response of ancient artefacts to different conservation treatments, and the structural integrity of archaeological sites under various environmental conditions.

## 6. Extending the RHDTO Model

To document all the proposed scenarios of employing AI in support of the Heritage Digital Twin, we extended the RHDTO to model the various components involved and capture their complex interactions. In particular, we introduced a new class, *HC15 AI Component*, representing the AI component within the digital twin system and describing its role in automating and enhancing various processes performed, such as data analysis, decision-making, and predictive modelling. These processes are represented by the new *HC16 Simulation and Prediction* class so that they can be recorded in the digital twin's knowledge base to be examined and reused by AI in future scenarios. Below is a tentative definition of these classes:

- **HC15 AI Component** (Subclass of: CRMdig:D14 Software).

This class is intended to model artificial intelligence systems integrated into the Heritage Digital Twin. AI components provide advanced capabilities for extracting and semantically encoding information from heritage documentation, continuously updating and enriching the digital twin's knowledge base. Additionally, instances of this class perform tasks such as data analysis, decision-making, and pattern recognition driven by algorithms or machine learning models, enhancing the analytical and predictive capabilities of digital twins to provide deeper insights into cultural entities and their historical and cultural context. Unlike conventional software, instances of *HC15* simulate intelligent behaviour, learning from data and adapting to new conditions while making informed decisions. AI components are continuously trained using data stored in the knowledge graph, real-time data coming from sensors, and scientific documentation and protocols produced by experts, thus refining their responses through iterative feedback to prioritise key preservation factors. They also leverage physics-based models to infer conditions that cannot be directly observed, such as material degradation of cultural objects;

- **HC16 Simulation or Prediction** (Subclass of: CRM:E5 Event).

This class is used to represent events of analysis, simulation, and prediction performed by the AI Component within the digital twin system, providing a structured way to document the AI's analytical activities. Simulations and predictions are critical components of the digital twin system, as they allow for the anticipation of future trends, the identification of potential issues, and the optimisation of decision-making processes. By performing simulations, the AI can test various scenarios and evaluate their impacts without the need for real-world experimentation. This is particularly useful in scenarios where real-time decisions are critical, such as emergency response or dynamic system optimisation. Predictions, on the other hand, enable the AI to forecast future states or events based on historical data and current conditions, providing valuable insights for proactive maintenance and conservation efforts.

To enhance the functionality and transparency of the AI Component, we also defined some new properties essential for describing the nature and functionalities of the AI Component and the results of the operations performed by it in ontological (and thus, formal) terms. The new properties include the following:

- **HP16 uses algorithm** (Domain: *HC15 AI Component*; Range: *CRMdig:D14 Software*). This property describes the specific algorithm or model employed by the AI Component within the digital twin system to process documentation efficiently, implementing its capability to draw meaningful conclusions from the analysed data. By recording this information, the property emphasises the methodological foundation of the AI component's functionality, underscoring its reliance on algorithmic approaches to achieve its objectives;
- **HP17 was trained using** (Domain: *HC15 AI Component*; Range: *CRM:E31 Document*). This property specifies the training documentation, protocols or datasets that were utilised to train the AI component, indicating that the AI has undergone a training process fundamental for enabling it to learn patterns, make predictions, and improve its performance in tasks related to cultural heritage documentation. This property is also essential for identifying the source and nature of the training data, which is critical for understanding the capabilities and limitations of the AI;
- **HP18 analysed** (Domain: *HC15 AI Component*; Range: *HC2 Heritage Digital Twin*). A property that indicates the relationship between the AI Component and the documentation or data being examined. This property indicates that the AI has performed a thorough examination of the specified documentation, extracting relevant information, patterns, and insights. The analysis may involve various techniques such as natural language processing, data mining, or pattern recognition to interpret and contextualise the data. The property is also essential for keeping precise and transparent records of the extracted knowledge provenance;
- **HP19 extracted knowledge** (Domain: *HC15 AI Component*; Range: *HC2 Heritage Digital Twin*). This property refers to the information or insights that are derived from data processing performed by AI components within the context of digital twins. This property captures the significant historical data that AI algorithms analyse and interpret, transforming raw information into structured graphs. The extracted knowledge can encompass various types of data, including but not limited to historical facts, relationships, and context, which are essential for creating a comprehensive semantic graph;
- **HP20 performed** (Domain: *HC15 AI Component*; Range: *HC16 Simulation or Prediction*). This property is used to establish a direct link between an AI Component and a specific Simulation or Prediction event, exactly indicating the analysis, simulation or prediction performed by an AI Component, thus ensuring traceability, accountability, and transparency in all AI's operations;
- **HP21 relies on** (Domain: *HC10 Decider*; Range: *HC15 AI Component*). This property can be used to model the connection between the Decider and the AI Component, indicating that the Decider has utilised the AI Component's support in its decision-making processes. The property is fundamental for documenting the symbiotic relationship between the two components, underscoring the role of AI in augmenting and enhancing the decision-making abilities of the Decider. Additionally, this property is particularly valuable for understanding the context and rationale behind the Decider's decisions, enabling users to validate and interpret its actions effectively.

Examples of how these classes and properties can be used to semantically model AI-empowered Heritage Digital Twins are provided in Section 7.

#### *Why Artificial Intelligences Cannot Be (E39) Actors (and Probably Never Should)*

Those readers familiar with ontologies and the CRM world may have noticed that the names of some of the entities we have defined, such as *HC16 Simulation or Prediction*

or *HP20 performed*, show great similarities with those of classes and properties already existing in the CRM ecosystem, and may wonder why these classes were not used directly in RHDT0. This is because these entities often refer to concepts orbiting in the sphere of the CRM:*E7 Activity* class, namely those events specifically carried out by instances of CRM:*E39 Actor* in CRM. For example, the CRMsci:*S7 Simulation or Prediction* class is placed in the Activities branch (E7) of the CRM, and therefore, among those “actions intentionally carried out by (P14) instances of *E39 Actor*”, as stated in the E7 scope note.

Even the CRMdig:*D7 Digital Machine Event* class, which at first glance might seem optimal for representing the reasoning typically performed by AIs, is described in its scope note to “comprise events that happen on physical digital devices following a human activity that intentionally caused its immediate or delayed initiation and results in the creation of a new instance of D1 Digital Object on behalf of the human actor”.

The question of whether an AI should be classified as a CRM:*E39 Actor* in an ontological framework such as CRM raises significant philosophical and conceptual issues. At the heart of the matter is the distinction between human and machine agency. In the CRM, the *E39 Actor* class encompasses entities capable of intentional actions, i.e., humans, organisations, and groups that act with purpose, self-awareness, and often within complex socio-cultural contexts. In contrast, AIs, despite their increasing sophistication, operate without true intentionality. AI components, no matter how advanced, remain tools executing algorithms based on rules and models designed by human agents to perform tasks or simulate decision-making processes. While AI can mimic certain aspects of human cognition and agency, such as decision-making or adaptive learning, it lacks the intrinsic properties of autonomy, moral responsibility, and consciousness that define actors in a human sense.

Thus, to classify AI as a CRM:*E39 Actor* would suggest a level of equivalency with human beings or institutions, which raises concerns about anthropomorphising technology. Such a classification could generate a misleading understanding of AI’s capabilities, implying that AI might possess the self-determined agency and the social roles traditionally ascribed to human actors. While AI can perform actions, these are fundamentally reactive, governed by predefined logic or learning models. Its “decisions” are calculative rather than reflective, lacking the contextual and ethical reasoning that characterises human decision-making. Furthermore, from an ontological perspective, Actors in the CRM are embedded in cultural, historical, and social systems in ways that an AI component cannot be. AI operates within a technical framework, and while it can engage with cultural heritage data, it does so as an instrumental tool, not as an agent with cultural or ethical background.

Another reason why AI cannot be considered an *E39 Actor(s)* lies in the nature of human decision-making activity. Unlike AI, which relies on algorithms and data processing, humans often make instinctive decisions guided by intuition or “good feelings” that sometimes turn out to be correct, even without consciously knowing why. A well-known example is found in chess, where grandmasters like Garry Kasparov have spoken of making incredibly complex decisions in mere seconds based on intuition rather than detailed analysis. This kind of decision-making draws on a blend of experience and tacit knowledge, something AI cannot replicate.

Therefore, it looks philosophically and ontologically more appropriate to us to place AI Components in our ontology within the category of CRMdig:*D14 Software*, acknowledging its role as an advanced digital tool and, at the same time, to define *HC16 Simulation or Prediction* as a direct subclass of CRM:*E7 Event*. This preserves the integrity of the distinction between human and machine agency, recognising the instrumental power of AI without attributing to it qualities it does not possess. By doing so, we maintain a clear boundary

between human actors and AI components, which, while powerful, remain, in essence, computational tools programmed to serve human purposes.

## 7. Case Study: Arnolfo Di Cambio's Tower in Florence

To showcase the applicability of our model, and especially of the new AI classes and properties defined, we selected the Tower of Arnolfo di Cambio, one of the most iconic monuments of the city of Florence (Italy). This is the prominent clock tower of *Palazzo Vecchio*, the historic building now housing the Florence city council, designed and created in the late 13th century by Arnolfo di Cambio, a notable Italian architect and sculptor, to serve as a symbol of the city's civic power [58,59].

In ontological terms, the Tower, being a monument and thus a physical, cultural object, can be represented in RHDTO by means of the *HC1 Heritage Entity* class. The *HC2 Digital Twin* class can be instantiated to define the Heritage Digital Twin of the Tower within the semantic space of our system.

Historically, the Tower has been a site of public announcements and a prison, showcasing its multifaceted role in Florence's medieval and Renaissance history. The Tower's significance in the history and cultural imagination of Florence is further underscored by extensive historical and archival documentation, including historical maps, city records, and construction documents, that provide insights into the Tower's origins, modifications, and enduring significance to the city over time. Additionally, the Tower is extensively represented in various artistic media, including 3D models, paintings, prints, and photographs, depicting it in its urban context and showcasing its architectural grandeur and central role in the Florentine skyline (see Figure 1).



**Figure 1.** Palazzo Vecchio and the Arnolfo Tower in Florence.

### 7.1. Extracting Knowledge from Documentation

All the above documentation can be integrated into the Heritage Digital Twin, both in digital format (i.e., by digitising and incorporating it into the HDT system) and as a semantic representation of its nature and content in the knowledge graph by means of the advanced analytical capabilities deployed through the use of AI algorithms, thus

enhancing the informational richness and historical significance of the Heritage Digital Twin information layer.

An example of how this mechanism could work is to show how the AI Component can be used to derive relevant information from historical documentation concerning the monument. The Tower, in fact, is mentioned in many important historical sources concerning the history of Florence, including, among others, the *Istorie Fiorentine* by Giovanni Villani [60], an author who directly witnessed the historical circumstances surrounding its construction during the late medieval period. Additionally, Giorgio Vasari, in his *Vite* [61], discussed Arnolfo's work and highlighted the technical and artistic aspects of his Tower, offering a Renaissance perspective on its architectural significance.

In ontological terms, the CRM:E31 *Document* class can be declared to represent these documents within the semantic space of our system. Artificial intelligence can be employed to “read” these works, analysing and processing the texts to extract relevant content. Extracted historical information can then be “transcribed” into fragments of semantic knowledge using the classes and properties of the RHDTO and the CRM ecosystem, thereby contributing new data to the history of the cultural object within the Digital Twin's knowledge base. The extraction process itself can be documented through the new *HC15 AI Component* class and its properties, namely, *HP18 analysed* and *HP19 extracted knowledge*, which also provides details on the provenance of the newly generated information fragments. Properties such as *CRM:P70 is documented in*, appropriately instantiated by the AI, can also be employed to maintain a consistent relationship between the derived knowledge and its source.

From a technical point of view, the way in which the semantic fragments created by the AI enrich the knowledge graph is rooted in the technique of *named graphs*, which provides a powerful mechanism for organising and representing data. By employing this technique, AI components can generate specific contexts or subsets of data within the broader graph structure, labelling it with appropriate names, thus allowing for more structured interpretations and analyses of their content. In fact, each named graph is a “semantic box” that encapsulates a distinct layer of information and whose provenance can always be precisely traced, enabling the AI to deal with insights that are both relevant and contextually aware. This approach not only enriches the overall knowledge representation but also facilitates dynamic interactions and relationships among various sections of the information stored within the knowledge graph, ultimately enhancing the system's capacity for intelligent reasoning.

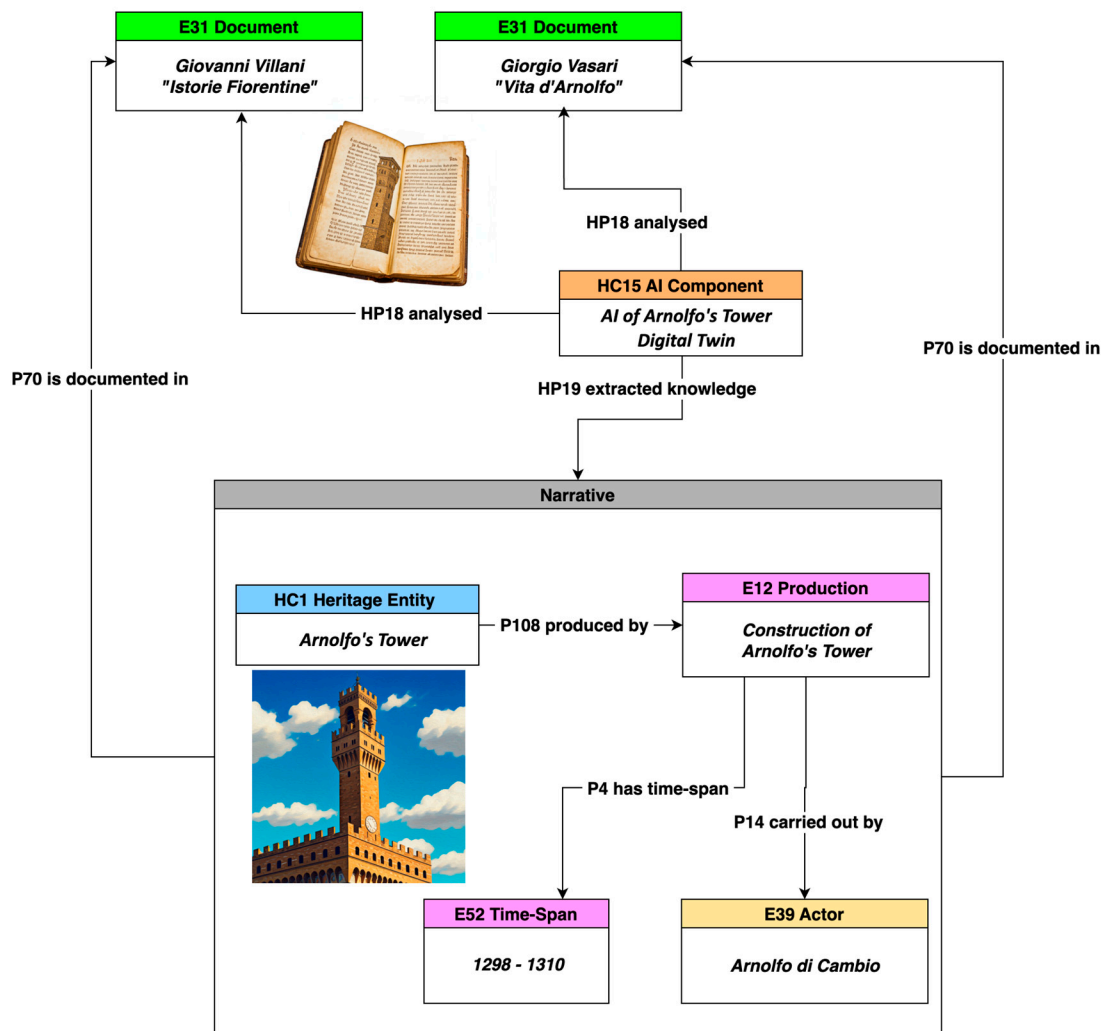
Figure 2 illustrates an example of how this process can take place by showing the extraction and the semantic encoding performed by AI of information relating to the time span of the construction of the Tower, performed between 1298 and 1310 by Arnolfo, as reported by Villani and Vasari. The named graph in which this knowledge is encapsulated is rendered, instantiating the *Narrative* class of the NOnt model.

To evaluate the potential of this process, we conducted preliminary experimental tests on selected small excerpts from the chosen works. We utilised open-source pre-trained AI models available on the Hugging Face platform [62] (especially the LLaMA 3.2 model [63]) to analyse the texts, extract key entities (such as people, objects, and events) and identify semantic relationships. All experiments were performed following only an initial, minimal fine-tuning of the models, achieved by providing the AIs with a minimal set of historical documentation and the specifications of our ontology.

Among the various challenges encountered during this process was the fact that the text was written in ancient Italian, which often diverged significantly from contemporary language and posed considerable difficulties for automated text analysis. Additionally, terminological and semantic ambiguities, particularly in relation to unclear temporal



contexts of historical events, necessitated a hybrid AI-human approach to achieve effective disambiguation and contextualisation. Nevertheless, despite the exploratory nature of this approach, the outcomes were significant. For example, the AI successfully identified the construction period of Arnolfo di Cambio's Tower (1298–1310) and its association with the places and the key historical figures mentioned in the text.



**Figure 2.** The knowledge extraction process is modelled using the RHDTO model.

The tested AI models, fed with the classes and properties specifications of RHDTO and the other models of the CIDOC CRM ecosystem, were further employed to encode the extracted results into a structured semantic format. This automatic process ensured that the knowledge fragments retrieved from the texts were consistently represented according to our established ontological framework. Although the precision achieved during encoding was generally acceptable, minimal human intervention was still required also in this case, particularly to verify the accuracy of representations involving complex or ambiguous data. Nonetheless, the outcomes are encouraging, indicating, even at this preliminary stage of our research, that AI tools hold significant potential for semantic codification of cultural heritage contexts.

The process of knowledge extraction described above primarily focuses on using AI in textual documentation. However, as advancements in AI technologies continue to evolve, this capability will extend to images, 3D models, maps, audio-visuals and other forms of



documentation, making this approach (and thus, this conceptual model) applicable to other possible use cases.

### 7.2. Artificial Intelligence and the Reactive Components of Heritage Digital Twins

To illustrate how AI can assist the digital twin reactive components, and in particular the Decider, in processing signals from sensors and making informed decisions on the actions to take to ensure the safety of cultural objects, we consider a recent paper concerning the seismic monitoring of Arnolfo's Tower, reporting an activity that was set up by the researchers from the Italian National Institute of Geophysics and Volcanology (INGV) and the Department of Architecture at the University of Florence and is still ongoing [64]. The aim of this seismic study is to deepen the understanding of the Tower's dynamic characteristics and evaluate its response to various vibration sources. To this end, two seismic stations were installed at the base and top of the Tower, enabling the identification of its primary modal frequencies using free vibration analysis. During the investigation period considered in the study, the monitoring system detected a heightened sensitivity to low-frequency seismic waves that align with the Tower's natural frequencies of oscillation, suggesting that the Tower may be more affected by distant seismic events than by nearby ones despite the latter being clearly perceptible within the city. The study also highlighted the impact of "cultural noise", i.e., vibrations caused by human activities, on the Tower's high-frequency oscillations. These sources include the ringing of the Tower's bell, the movement of visitors, and general urban activity. The researchers plan to extend the monitoring network by adding more sensors within the Tower and the neighbouring *Palazzo Vecchio*. This expansion will facilitate a more detailed analysis of the modal shapes and the dynamic interactions between the structures. Additionally, further investigation is needed to clarify the differences in the Tower's responses to nearby versus distant earthquakes, which could have important implications for assessing the seismic vulnerability of historic structures. This continuous monitoring approach would provide crucial data to the Heritage Digital Twin of Arnolfo's Tower for refining numerical models and improving the accuracy of structural response predictions, as demonstrated by other relevant studies in this field [65]. It would also allow for the assessment of the impacts of both natural and anthropogenic vibrations on the long-term integrity of the Tower, as well as other precious insights that can guide future conservation and preservation strategies for these and other similar heritage structures.

The contribution of AI in this context could be significant to analyse and classify seismic events according to their intensity and distance, thereby identifying those that may pose a greater threat to the stability of the Tower while distinguishing them from anthropogenic vibrations, such as those generated by large crowds of tourists visiting the *Palazzo Vecchio* or the underlying *Piazza della Signoria*. To enhance its effectiveness, scientific reports, including the one detailing this study, could be used to train the AI, equipping it with an understanding of the specific challenges and risks associated with the Tower's structure and condition. This approach would be akin to having the AI "study" to become a "digital seismologist", i.e., an "expert" capable of analysing and distinguishing relevant data and operating more effectively based on specialised knowledge. Naturally, all the processes performed and decisions made by the Decider and the AI will be encoded using the RHDTO model, thereby becoming an integral part of the Heritage Digital Twin's knowledge base. This provenance information will allow experts to evaluate the quality of the decisions made, correct any inaccuracies, and further refine the AI training.

To semantically represent this scenario, we once again leverage the CRM ecosystem and the existing entities of RHDTO. Additionally, we deploy the newly defined entities of our ontology. To describe the seismic sensors installed on the Tower, we instantiate the *HC9*

*Sensor* class. The placement of each of the sensors on the monument (*HC1 Heritage Entity*) is modelled through the *HC15 is positioned on the property*. *HC13 Sensor Measurement*, in combination with the properties *CRMsci:O24 measured* and *CRMdig:L12 happened on device*, can be instantiated to model the events monitored by each sensor, such as seismic movements and the “cultural noise” generated by human actions, including tourist movements, bell ringing, urban traffic, local events, and daily operational tasks. The events themselves are modelled using the *CRM:E5 Event* class. The resulting seismic signals are represented by the *HC12 Signal*, and their transmission to the Digital Twin of Arnolfo’s Tower (*HC2*) is modelled through the *HP12 was transmitted to the property*. The integrated monitoring system receiving these signals is represented by instances of the *HC10 Decider* class.

The interaction between the Decider (*HC10*) and the AI Component of the Heritage Digital Twin, rendered by instantiating the new *HC15 AI Component* class, implements a system designed to acquire the transmitted values, analyse the Tower’s physical and environmental conditions, integrate information from other sensors, and enrich the Heritage Digital Twin’s Knowledge Base. The role of AI in conducting simulations and predictive tasks is rendered through the property *HP20 performed*, while the operations themselves are modelled through instances of the *HC16 Simulation or Prediction* class. The interaction between the Decider and the AI Component for enhanced decision-making is captured by the *HP21 relies on* property.

Finally, the process by which the AI analyses the digital twin’s Knowledge Base, critical for accessing and assessing relevant information for informed decision-making, is represented by the *HP18 analysed* property, while the *HP17 was trained using* property is employed to specify the scientific papers and other resources used for training the AI, thereby enhancing its capabilities and “expertise” in the field of seismology.

The ontological representation of all these processes is illustrated in Figure 3.

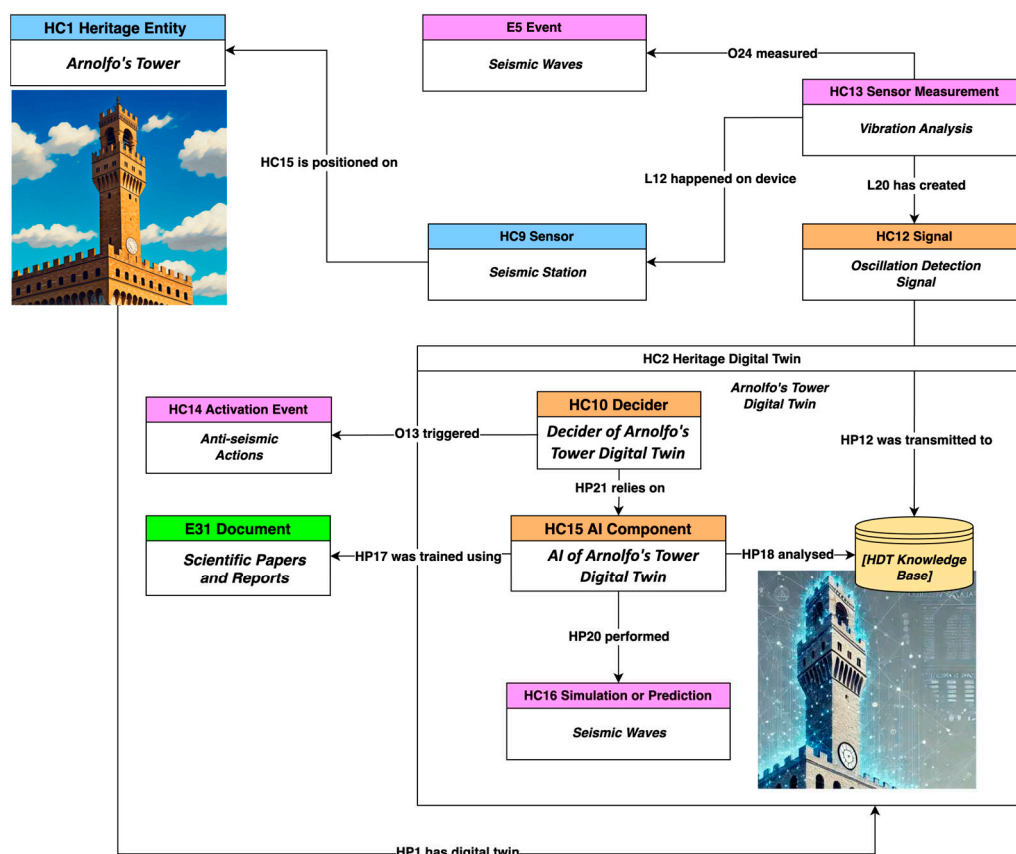


Figure 3. Ontological representation of AI and reactive components of the Heritage Digital Twin.

As in the case of knowledge extraction, both the decisions made by the decider and the contributions of the AI—such as the specific insights and analyses it provided—are captured in a named graph aimed at enriching the overall knowledge graph by documenting the decision-making process, the rationale behind the actions taken, the relevant environmental data, and the AI’s analytical contributions. This facilitates future reference, analysis, and potential learning for similar situations.

## 8. Conclusions

This paper is a first step in showing how the integration of artificial intelligence in Heritage Digital Twins, based on extensive semantic knowledge graphs and empowered by ontologies, already represents a significant advancement in the field of cultural heritage data management. These three components, when well-orchestrated, offer unprecedented capabilities for the preservation, analysis, and decision-making processes related to cultural objects. However, the successful integration of these elements requires careful consideration to avoid the pitfalls of the well-known “Three-Body Problem” [66], with the complexity and potentially chaotic consequences that can arise from the interaction of multiple dynamic systems.

What we illustrate in this paper is only a glimpse into the transformative potential of AI within the digital twin landscape, where these advanced tools reveal new dimensions for enhancing documentation, analysis, and interpretation. We stand at the dawn of a new technological era that promises to reshape our understanding of interaction, intelligence, and the very fabric of knowledge itself. Yet, it is inevitable that AI applications will, in time, extend far beyond these initial boundaries, unlocking possibilities across diverse domains, both within and outside the digital twin framework.

In parallel, another significant paradigm shift is already underway. While many of today’s most prominent AI tools are semicommercial, proprietary products, i.e., powerful yet closed systems that limit adaptability and transparency, we are also witnessing the rise of very promising open-source alternatives, providing resources that not only democratise access to advanced AI capabilities but also offer enhanced transparency and customisation. As this new era unfolds, it invites us to engage thoughtfully with the possibilities and responsibilities embedded in these evolving tools.

Given the encouraging results obtained with minimal training and little effort in the tests conducted on existing AI systems carried out during this work, we plan to further expand our active exploration of these emerging technologies in future activities. This includes identifying the latest AI tools and subjecting them to rigorous testing as they become available, ensuring a seamless integration with our evolving methodologies. Our efforts will focus on further benchmarking these tools against cultural heritage and scientific datasets alongside the information embedded in our knowledge bases, rigorously scrutinising their performance in navigating both structured and unstructured data domains and assessing their capacity to handle increasingly sophisticated decision-making tasks.

In parallel, we will train a custom open-source AI model to reflect our specific methodologies and intellectual approaches and achieve seamless integration into the overall framework of the Heritage Digital Twin ecosystem. We will also pioneer advanced training techniques based on semantically encoded information to guarantee that the model is uniquely tailored, rigorously refined, and capable of addressing the numerous challenges required by the Heritage Digital Twin lifecycle.

At the same time, we will deepen our use of ontologies to govern and refine the power of these tools, ensuring that AI-driven processes align with structured and meaningful frameworks. We firmly believe that in a landscape where the complexity of algorithms can render decisions inscrutable, only ensuring traceability can transform the opaque nature

of AI into a more comprehensible narrative, enhancing accountability and fostering a collaborative dialogue between human intuition and artificial reasoning. We also hold the view that the use of AI always necessitates the involvement of human experts who are called to provide guidance, validation, and supervision. Only human oversight is able to ensure that the decisions made by the AI are grounded in reality, that the data stored in the knowledge graphs are accurate and remain relevant, and that the Heritage Digital Twins, today and in the future, remain faithful representations of cultural entities, bridging the gap between machine representation and human understanding.

By framing data within a coherent ontological context, we elevate the AI's output from mere black-box results to a rich tapestry of interrelated concepts and insights, paramount for the implementation of effective Heritage Digital Twins. This transparency fosters a deeper trust in the AI's outcomes, as scholars can discern not only the results but also the underlying logic that informed them. Only in this way can AI become a participant in a shared epistemological journey, where the interplay of logic and data is made visible, paving the way for more informed and responsible decisions and choices. This not only increases the reliability of AI outputs but also aligns with the EU's vision of an AI landscape that prioritises human values, safeguards against biases, and promotes social well-being, serving as a vital step toward ensuring that AI technologies are developed and deployed in a manner that is ethical, responsible, and worthy of public trust.

**Author Contributions:** All authors have equally contributed to the present paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created for this study.

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

## References

1. European Commission, Destination Earth. Available online: <https://destination-earth.eu> (accessed on 21 October 2024).
2. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A Systematic Literature Review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [CrossRef]
3. Semeraro, C.; Lezoche, M.; Panetto, H.; Dassisti, M. Digital Twin Paradigm: A Systematic Literature Review. *Comput. Ind.* **2021**, *130*, 103469. [CrossRef]
4. Schrotter, G.; Hürzeler, C. The Digital Twin of the City of Zurich for Urban Planning. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 99–112. [CrossRef]
5. Caprari, G.; Castelli, G.; Montuori, M.; Camardelli, M.; Malvezzi, R. Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives. *Sustainability* **2022**, *14*, 6263. [CrossRef]
6. Li, L.; Aslam, S.; Wileman, A.; Perinpanayagam, S. Digital Twin in Aerospace Industry: A Gentle Introduction. *IEEE Access* **2022**, *10*, 9543–9562. [CrossRef]
7. Grieves, M. Digital Twin: Manufacturing Excellence Through Virtual Factory Replication. *White Pap.* **2014**, *1*, 1–7.
8. Drechsler, W. The Contrade, the Palio and the *Ben Comune*: Lessons from Siena. *Trames* **2006**, *10*, 99–125. [CrossRef]
9. Palio di Siena. Available online: [https://en.wikipedia.org/wiki/Palio\\_di\\_Siena](https://en.wikipedia.org/wiki/Palio_di_Siena) (accessed on 21 November 2024).
10. Paradis, T.W. *Living the Palio: A Story of Community and Public Life in Siena, Italy*; Paradis: Indianapolis, IN, USA, 2019.
11. Artusi, L. *Calcio Fiorentino. History, Art and Memoirs of the Historical Game. From Its Origins to the Present Day*; Scribo Edizioni: Florence, Italy, 2016.
12. Calcio Storico Fiorentino. Available online: [https://en.wikipedia.org/wiki/Calcio\\_storico\\_fiorentino](https://en.wikipedia.org/wiki/Calcio_storico_fiorentino) (accessed on 21 November 2024).
13. Niccolucci, F. Technologies, Standards and Business Models for the Formation of Virtual Collections of 3D Replicas of Museum Objects: The 3D-COFORM Project. In Proceedings of the 2010 IST-Africa Conference, Durban, South Africa, 19–21 May 2010; pp. 1–8.

14. Lovell, L.J.; Davies, R.J.; Hunt, D.V.L. The Application of Historic Building Information Modelling (HBIM) to Cultural Heritage: A Review. *Heritage* **2023**, *6*, 6691–6717. [CrossRef]
15. Industry Foundation Classes (IFC). Available online: <https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/> (accessed on 21 November 2024).
16. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [CrossRef]
17. 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. *arXiv* **2023**, arXiv:2302.07138. [CrossRef]
18. Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. [CrossRef] [PubMed]
19. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a Semantic Construction Digital Twin: Directions for Future Research. *Autom. Constr.* **2020**, *114*, 103179. [CrossRef]
20. Albers, L.; Große, P.; Wagner, S. Semantic Data-Modeling and Long-Term Interpretability of Cultural Heritage Data—Three Case Studies. In *Digital Cultural Heritage*; Kremers, H., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 239–253. ISBN 978-3-030-15198-0.
21. CIDOC CRM. Available online: <https://cidoc-crm.org/> (accessed on 21 November 2024).
22. ECHOES—European Cloud for Heritage OpEn Science. Available online: <https://www.echoes-ecch.eu> (accessed on 21 November 2024).
23. Richards, J.D.; Niccolucci, F. (Eds.) *The Ariadne Impact*; Archeolingua: Budapest, Hungary, 2019; ISBN 978-615-5766-31-2. [CrossRef]
24. Aspöck, E.; Richards, J. (Eds.) *International Data Aggregation for Archaeological Research and Heritage Management: The ARIADNE Experience*; Internet Archaeology, Special Issue 64; University of York: York, UK, 2023. Available online: <https://intarch.ac.uk/journal/issue64/index.html> (accessed on 21 November 2024).
25. ARIADNE Research Infrastructure. Available online: <https://www.ariadne-research-infrastructure.eu> (accessed on 21 November 2024).
26. Competence Centre for the Conservation of Cultural (4CH). Available online: <https://www.4ch-project.eu> (accessed on 21 November 2024).
27. Doerr, M.; Stead, S.; Theodoridou, M. Definition of the CRMdig. An Extension of CIDOC-CRM to Support Provenance Metadata. 2022. Available online: <https://cidoc-crm.org/sites/default/files/CRMdigv4.0.pdf> (accessed on 21 November 2024).
28. Kritsotaki, A.; Hiebel, G.; Theodoridou, M.; Doerr, M.; Rousakis, Y. Definition of the CRMsci. 2017. Available online: <https://cidoc-crm.org/crmsci/ModelVersion/version-1.2.3> (accessed on 21 November 2024).
29. Bruseker, B.; Doerr, M.; Theodoridou, M. Report on the Common Semantic Framework. PARTHENOS Project, 2017. Available online: [https://www.parthenos-project.eu/Download/Deliverables/D5.1\\_Common\\_Semantic\\_Framework\\_Appendices.pdf](https://www.parthenos-project.eu/Download/Deliverables/D5.1_Common_Semantic_Framework_Appendices.pdf) (accessed on 21 November 2024).
30. Meghini, C.; Bartalesi, V.; Metilli, D. Representing Narratives in Digital Libraries: The Narrative Ontology. *Semant. Web* **2021**, *12*, 241–264. Available online: <https://www.semantic-web-journal.net/content/representing-narratives-digital-libraries-narrative-ontology> (accessed on 21 November 2024). [CrossRef]
31. The MINGEI Project. Available online: <https://www.mingei-project.eu/> (accessed on 21 November 2024).
32. Brandt, A.K. Beethoven’s Ninth and AI’s Tenth: A Comparison of Human and Computational Creativity. *J. Creat.* **2023**, *33*, 100068. [CrossRef]
33. Rizk, R.; Rizk, D.; Rizk, F.; Kumar, A. A Hybrid Capsule Network-Based Deep Learning Framework for Deciphering Ancient Scripts with Scarce Annotations: A Case Study on Phoenician Epigraphy. In Proceedings of the 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 9–11 August 2021; IEEE: Lansing, MI, USA; pp. 617–620. [CrossRef]
34. Magnani, M.; Clindaniel, J. Artificial Intelligence and Archaeological Illustration. *Adv. Archaeol. Pract.* **2023**, *11*, 452–460. [CrossRef]
35. Cobb, P.J. Large Language Models and Generative AI, Oh My!: Archaeology in the Time of ChatGPT, Midjourney, and Beyond. *Adv. Archaeol. Pract.* **2023**, *11*, 363–369. [CrossRef]
36. Argyrou, A.; Agapiou, A. A Review of Artificial Intelligence and Remote Sensing for Archaeological Research. *Remote Sens.* **2022**, *14*, 6000. [CrossRef]
37. Mishra, M.; Barman, T.; Ramana, G.V. Artificial Intelligence-Based Visual Inspection System for Structural Health Monitoring of Cultural Heritage. *J. Civ. Struct. Health Monit.* **2024**, *14*, 103–120. [CrossRef]
38. Cao, J.; Zhang, Z.; Zhao, A.; Cui, H.; Zhang, Q. Ancient Mural Restoration Based on a Modified Generative Adversarial Network. *Herit. Sci.* **2020**, *8*, 7. [CrossRef]



39. European Commission (Ed.) *White Paper on Artificial Intelligence. A European Approach to Excellence and Trust*; European Commission: Brussels, Belgium, 2020.
40. Nikolinakos, N.T. A European Approach to Excellence and Trust: The 2020 White Paper on Artificial Intelligence. In *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies—The AI Act*; Law, Governance and Technology Series; Springer International Publishing: Cham, Switzerland, 2023; Volume 53, pp. 211–280, ISBN 978-3-031-27952-2.
41. European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Creating Harmonized Rules on Artificial Intelligence and Amending Regulations (Artificial Intelligence Regulation)*; European Parliament: Brussels, Belgium, 2024. Available online: <http://data.europa.eu/eli/reg/2024/1689/oj> (accessed on 21 November 2024).
42. Pasikowska-Schnass, M.; Young-Shin, L. *Artificial intelligence in the Context of Cultural Heritage and Museums: Complex Challenges and New Opportunities*; Technical Report PE 747.120; European Parliamentary Research Service: Brussels, Belgium, 2023. Available online: [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2023\)747120](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)747120) (accessed on 21 November 2024).
43. Münster, S.; Maiwald, F.; Di Lenardo, I.; Henriksson, J.; Isaac, A.; Graf, M.M.; Beck, C.; Oomen, J. Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe. *Heritage* **2024**, *7*, 794–816. [\[CrossRef\]](#)
44. Directorate-General for Communications Networks; Content and Technology; Izsak, K.; Terrier, A.; Kreutzer, S.; Strähle, T.; Roche, C.; Moretto, M.; Sorensen, S.; Hartung, M.; et al. *Opportunities and Challenges of Artificial Intelligence Technologies for the Cultural and Creative Sectors*; Publications Office of the European Union: Luxembourg, 2022. Available online: <https://op.europa.eu/en/publication-detail/-/publication/359880c1-a4dc-11ec-83e1-01aa75ed71a1/language-en> (accessed on 21 November 2024).
45. Felicetti, A.; Williams, D.; Galluccio, I.; Tudhope, D.; Niccolucci, F. NLP Tools for Knowledge Extraction from Italian Archaeological Free Text. In *Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) Held Jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, San Francisco, CA, USA, 26–30 October 2018; IEEE: San Francisco, CA, USA; pp. 1–8. [\[CrossRef\]](#)
46. Bombini, A.; Alkhansa, A.; Cappelli, L.; Felicetti, A.; Giacomini, F.; Costantini, A. A Cloud-Native Web Application for Assisted Metadata Generation and Retrieval: THESPIAN-NER. *Appl. Sci.* **2022**, *12*, 12910. [\[CrossRef\]](#)
47. Bombini, A.; Castelli, L.; Felicetti, A.; Niccolucci, F.; Reccia, A.; Taccetti, F. Towards the Creation of AI-Powered Queries Using Transfer Learning on NLP Model—The THESPIAN-NER Experience. In *Image Analysis and Processing. ICIAP 2022 Workshops*; Mazzeo, P.L., Frontoni, E., Sclaroff, S., Distant, C., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2022; Volume 13374, pp. 266–277, ISBN 978-3-031-13323-7. [\[CrossRef\]](#)
48. Tzitzikas, Y.; Mountantonakis, M.; Fafalios, P.; Marketakis, Y. CIDOC-CRM and Machine Learning: A Survey and Future Research. *Heritage* **2022**, *5*, 1612–1636. [\[CrossRef\]](#)
49. Longo, L.; Brcic, M.; Cabitza, F.; Choi, J.; Confalonieri, R.; Ser, J.D.; Guidotti, R.; Hayashi, Y.; Herrera, F.; Holzinger, A.; et al. Explainable Artificial Intelligence (XAI) 2.0: A Manifesto of Open Challenges and Interdisciplinary Research Directions. *Inf. Fusion* **2024**, *106*, 102301. [\[CrossRef\]](#)
50. Pansoni, S.; Tiribelli, S.; Paolanti, M.; Di Stefano, F.; Frontoni, E.; Malinverni, E.S.; Giovanola, B. Artificial intelligence and cultural heritage: Design and assessment of an ethical framework. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 1149–1155. [\[CrossRef\]](#)
51. Gîrbacia, F. An Analysis of Research Trends for Using Artificial Intelligence in Cultural Heritage. *Electronics* **2024**, *13*, 3738. [\[CrossRef\]](#)
52. Jayasinghe, S.C.; Mahmoodian, M.; Sidiq, A.; Nanayakkara, T.M.; Alavi, A.; Mazaheri, S.; Shahrivar, F.; Sun, Z.; Setunge, S. Innovative Digital Twin with Artificial Neural Networks for Real-Time Monitoring of Structural Response: A Port Structure Case Study. *Ocean Eng.* **2024**, *312*, 119187. [\[CrossRef\]](#)
53. Zayed, S.M.; Attiya, G.M.; El-Sayed, A.; Hemdan, E.E.-D. A Review Study on Digital Twins with Artificial Intelligence and Internet of Things: Concepts, Opportunities, Challenges, Tools and Future Scope. *Multimed. Tools Appl.* **2023**, *82*, 47081–47107. [\[CrossRef\]](#)
54. D’Amico, R.D.; Sarkar, A.; Karray, M.H.; Addepalli, S.; Erkoyuncu, J.A. Knowledge Transfer in Digital Twins: The Methodology to Develop Cognitive Digital Twins. *CIRP J. Manuf. Sci. Technol.* **2024**, *52*, 366–385. [\[CrossRef\]](#)
55. Seng, K.P.; Ang, L.M.; Ngharamike, E. Artificial Intelligence Internet of Things: A New Paradigm of Distributed Sensor Networks. *Int. J. Distrib. Sens. Netw.* **2022**, *18*, 155014772110628. [\[CrossRef\]](#)
56. Emmert-Streib, F. Defining a Digital Twin: A Data Science-Based Unification. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1036–1054. [\[CrossRef\]](#)
57. Ritto, T.G.; Rochinha, F.A. Digital Twin, Physics-Based Model, and Machine Learning Applied to Damage Detection in Structures. *Mech. Syst. Signal Process.* **2021**, *155*, 107614. [\[CrossRef\]](#)
58. Sinibaldi, G. *The Palazzo Vecchio in Florence*; La Libreria dello Stato: Rome, Italy, 1950.
59. Palazzo Vecchio in Florence. Available online: [https://en.wikipedia.org/wiki/Palazzo\\_Vecchio](https://en.wikipedia.org/wiki/Palazzo_Vecchio) (accessed on 21 November 2024).



60. Villani, G. *Villani's Chronicle; Being Selections from the First Nine Books of the Croniche Fiorentine*; Selfe, R.E.T., Wicksteed, P.H., Eds.; Archibald Constable & Co.: London, UK, 1906. First published 1322.
61. Vasari, G. *The Lives of the Painters, Sculptors and Architects*; Gaunt, W., Ed.; Dent: London, UK, 1963. First published 1550.
62. Hugging Face Community. Available online: <https://huggingface.co> (accessed on 16 December 2024).
63. Llama Open-Source AI models. Available online: <https://www.llama.com> (accessed on 16 December 2024).
64. Azzara, R.M.; Tanganelli, M.; Trovatielli, F.; Vettori, N. Results from the Seismometric Continuous Monitoring of an Ancient Bell Tower: The Arnolfo Tower, Palazzo Della Signoria, Florence, Italy. In *Structural Analysis of Historical Constructions*; Endo, Y., Hanazato, T., Eds.; RILEM Book Series; Springer Nature Switzerland: Cham, Switzerland, 2024; Volume 47, pp. 165–178. ISBN 978-3-031-39602-1. [\[CrossRef\]](#)
65. Dal Zilio, L.; Giardini, D.; Carbonell, R.; Wiemer, S. Harnessing the Potential of Digital Twins in Seismology. *Nat. Rev. Earth Environ.* **2023**, *4*, 510–512. [\[CrossRef\]](#)
66. Barrow-Green, J. The Three-Body Problem. In *The Princeton Companion to Mathematics*; Gowers, T., Barrow-Green, J., Leader, I., Eds.; Princeton University Press: Princeton, NJ, USA, 2008; pp. 726–728.

**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.



# Dig *Italia*

Anno XIX, Numero 2 - **2024**

Rivista del digitale nei beni culturali

ICCU-ROMA



ICCU

Istituto centrale per il catalogo unico  
delle biblioteche italiane e per le informazioni bibliografiche  
[www.iccu.sbn.it](http://www.iccu.sbn.it)

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista  
è ammessa con obbligo di citazione

«DigItalia», rivista del digitale nei beni culturali, è una rivista *peer reviewed*  
e segue il codice etico delle pubblicazioni

I contributi possono essere proposti alla rivista dagli autori o su iniziativa del Comitato Scientifico. Gli articoli vengono sottoposti al Comitato di Redazione per un primo accertamento sulla corrispondenza con i campi di ricerca della rivista. I contributi delle sezioni Saggi e Progetti vengono indirizzati in forma anonima ad almeno uno studioso di comprovata competenza sui temi affrontati. I revisori fanno pervenire i loro giudizi alla redazione (favorevole alla pubblicazione, favorevole con modifiche/miglioramenti, non favorevole). Se il giudizio finale è positivo, viene comunicata agli autori l'accettazione del contributo, insieme ad eventuali indicazioni suggerite dai valutatori, di cui si garantisce comunque l'anonimato.

*In copertina:*

L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.C. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicinabile all'Apollo di Leochares (IV sec. a.C.)

**Direttore Fondatore**

Marco Paoli

**Direttore Responsabile**

Giuliano Genetasio

**Comitato di Redazione**

Amalia Maria Amendola

Valentina Atturo

Laura Borsi

Elisabetta Castro

Silvana de Capua

Maria Cristina Mataloni

Valentina Menesatti

Lucia Negrini

Federica Olivotto

Claudia Pecoraro

Angelo Restaino

Elisa Sciotti

Alice Semboloni

Maria Lucia Violo

**Grafica & Impaginazione**

MLA&Partner - Roberta Micchi

**Produzione e Stampa**

Istituto Poligrafico e Zecca dello Stato S.p.A.

Roma

**Editore**

ICCU

Istituto centrale per il catalogo unico  
delle biblioteche italiane

e per le informazioni bibliografiche

Viale Castro Pretorio, 105

00185 Roma

T +39 06 49.210.425

email: [ic-cu.digitalia@cultura.gov.it](mailto:ic-cu.digitalia@cultura.gov.it)

<https://digitalia.cultura.gov.it>



UNICAM

### **Comitato Scientifico**

Stefano Allegrezza  
Enrico Pio Ardolino  
Giovanni Bergamin  
Julian Bogdani  
Dimitri Brunetti  
Flavia Bruni  
Simonetta Buttò  
Elisabetta Caldelli  
Rosaria Campioni  
Laura Ciano  
Gianfranco Crupi  
Andrea De Pasquale  
Pierluigi Feliciati  
Giuliano Genetasio  
Marina Giannetto  
Maria Guercio  
Mauro Guerrini

Klaus Kempf  
Maurizio Messina  
Laura Moro  
Maria Teresa Natale  
Marco Paoli  
Don Valerio Pennasso  
Marco Pizzo  
Paola Puglisi  
Roberto Raieli  
Gino Roncaglia  
Maria Letizia Sebastiani  
Giovanni Solimine  
Laura Tallandini  
Anna Maria Tammaro  
Antonella Trombone  
Chiara Veninata  
Paul Gabriele Weston



# Il futuro digitale di 4CH: il gemello digitale dei beni culturali

«DigItalia» 2-2024

DOI: 10.36181/digitalia-00105

**Franco Niccolucci – Achille Felicetti**

*Polo Universitario di Prato (PIN)*

*Il presente articolo descrive il modello di gestione dei dati sviluppato nel progetto europeo 4CH e la creazione della base della conoscenza per il progetto. Questa si basa sul concetto di gemello digitale in un modo innovativo, cioè come complesso dei dati relativi ad un bene, sia grafici o visuali come immagini o modelli 3D sia testuali o documentali, e delle loro interrelazioni. Si tratta di un approccio innovativo che consente di gestirne la documentazione in modo unitario, indipendentemente dal formato dei dati e dalla loro provenienza.*

## Introduzione

Come descritto in un precedente articolo su questa rivista<sup>1</sup>, il progetto europeo 4CH<sup>2</sup> ha avuto come obiettivo la progettazione del centro di competenza europeo per la conservazione del patrimonio culturale. Conclusosi a dicembre 2023, 4CH ha prodotto una serie di documenti in cui sono stati analizzati tutti gli aspetti necessari alla creazione di tale Centro, da quelli istituzionali e organizzativi a quelli economici e finanziari e a quelli tecnologici. Confermando una visione “confederale” del Centro di Competenza, cioè basata su una molteplicità di Centri nazionali o regionali relativamente autonomi, coordinati da un Centro europeo incaricato inoltre di fornire servizi comuni, sono in corso di attuazione iniziative per la creazione di soggetti giuridici autonomi in grado di assumere direttamente questo compito a livello europeo e italiano. A livello europeo si prevede la creazione di una entità giuridica di diritto belga senza fine di lucro e dotata della necessaria autonomia giuridica e operativa. La legislazione belga è infatti una delle poche in Europa a consentire la creazione di un ente di questo tipo<sup>3</sup>, con sede legale in Belgio ma operativo a livello internazionale e per questo ampiamente utilizzato da molti consorzi europei per la continuazione delle attività di progetto. Per l'Italia si sta invece considerando una tipologia flessibile, prevista dal nostro ordinamento, a cui potranno aderire sia i partner italiani

<sup>1</sup> Franco Niccolucci, *4CH: un progetto per sviluppare le applicazioni e le competenze digitali per la gestione del patrimonio culturale*, «DigItalia. Rivista del digitale nei beni culturali», 17 (2022), n. 2, p. 161-167. <<https://doi.org/10.36181/digitalia-00045>>. L'articolo illustra la progettazione e la creazione di un centro di competenza per la conservazione dei beni culturali basato sui modelli 3D di siti ed edifici storici, in risposta a un bando della Commissione Europea che collegava tale Centro ai principi esposti nella Raccomandazione 1970/2021 sulla creazione di un “Data Space” per i beni culturali, attualmente in corso di costituzione da parte di un consorzio guidato da Europeana.

<sup>2</sup> 4CH – *Competence Centre for the Conservation of Cultural Heritage* è stato un progetto finanziato all'interno del programma Horizon2020 della Commissione Europea con il Grant 101004468, operativo nel periodo dal 1/1/2021 al 31/12/2023.

<sup>3</sup> Si tratta dell' *Association Internationale Sans But Lucratif* (AISBL).

del progetto che altri enti interessati al raggiungimento degli obiettivi progettuali. Queste attività istituzionali sono in corso di attuazione da parte del coordinatore di 4CH, Francesco Taccetti, per conto dell'Istituto nazionale di fisica nucleare (INFN).

Per quanto riguarda la tecnologia 3D e le sue applicazioni ai beni culturali, coordinata nel progetto 4CH da Roberto di Giulio per conto di INCEPTION, uno spin-off dell'Università di Ferrara, è stato di recente pubblicato il bando per la creazione di un centro europeo di competenza "virtuale" (cioè non basato su un'infrastruttura fisica) per il 3D, che dovrà supportare le istituzioni del patrimonio culturale nell'utilizzo di questa tecnologia, collaborando con il centro di competenza di 4CH. Recentemente, il progetto vincente è risultato *3D-4CH - Online Competence Centre in 3D for Cultural Heritage*, coordinato da INCEPTION srl, uno dei coordinatori di 4CH, che inizierà le attività nel febbraio 2025. L'obiettivo di 3D-4CH è la predisposizione di standard, tecnologie e servizi per le applicazioni ai beni culturali e la formazione degli operatori del settore, in collaborazione con altre iniziative collegate alla digitalizzazione e alla creazione di modelli tridimensionali per i beni culturali.

Fra i servizi centrali che 4CH dovrà assicurare restano quindi da definire gli aspetti che riguardano la gestione dei dati, un altro pilastro del progetto originario, che gli autori del presente contributo hanno coordinato durante la sua realizzazione. All'interno di questo aspetto puramente digitale possiamo distinguere un primo gruppo di contenuti che riguardano una serie di servizi offerti dal centro in modo virtuale agli operatori e alle istituzioni culturali. Si tratta in sostanza di una biblioteca digitale che contiene i riferimenti a prodotti interni o esterni utili a questa comunità, come ad esempio strumenti di formazione prodotti all'interno del progetto o da terzi, manuali, documentazione, e così via, sia in formato testuale che multimediale. Da un punto di vista informatico l'organizzazione di questi dati è relativamente semplice: si tratta di un catalogo, organizzato per parole chiave, collegato a un archivio interno dove sono depositati i relativi contenuti oppure con link ad archivi esterni. Resta quindi da progettare e realizzare la gestione dei dati veri e propri, ovvero quella che è stata denominata la *Base della Conoscenza*, indicata con l'acronimo *KB* corrispondente alle iniziali del termine inglese corrispondente<sup>4</sup>.

## 1. La Base della Conoscenza e il suo contenuto

Come in molti altri settori della ricerca, anche nel campo della conservazione e restauro dei beni culturali la mole di dati prodotti dalle attività di operatori e ricercatori è in continua crescita. Se l'utilizzo di tecnologie digitali produce dati e risultati facilmente condivisibili, almeno in via teorica, il loro deposito in sistemi isolati l'uno dall'altro, veri e propri silos informativi, ne rende spesso l'accesso problematico e di fatto irrealizzabile. Si tratta invece di un'esigenza fortemente sentita, come dimostrano esperienze analoghe in settori affini. Per esempio, il sistema ARIADNE<sup>5</sup>, anch'esso prodotto di un progetto europeo e attualmente gestito da una AISBL costituita per questo scopo, si è rivelato di grande potenzialità nel campo della ricerca archeologica. Con una disponibilità di circa quattro milioni di contenuti, il catalogo di ARIADNE si è rivelato di

<sup>4</sup> Una *Base della Conoscenza* è internazionalmente denominata con il termine tecnico *Knowledge Base*. Questo in qualche modo richiama il concetto più familiare di *Data Base*, da cui differisce sostanzialmente in quanto utilizza una tecnologia più avanzata, più flessibile e più aderente ai contenuti che sarà sinteticamente descritta nel seguito dell'articolo. Per questo motivo si utilizzerà l'acronimo *KB* per indicare la Base della Conoscenza.

<sup>5</sup> Sono stati pubblicati numerosi articoli sul progetto ARIADNE, fra cui: Sara Di Giorgio – Paola Ronzino, *ARIADNE, l'infrastruttura europea per l'archeologia*, «DigItalia. Rivista del digitale nei beni culturali», 11 (2017), n. 1/2, p. 124-135; Franco Niccolucci, *ARIADNEplus: l'avventura continua*, «DigItalia. Rivista del digitale nei beni culturali», 15 (2020), n. 2, p. 88-95. <<https://doi.org/10.36181/digitalia-00016>>.

estrema utilità come testimoniano<sup>6</sup> ricercatori di ogni parte d'Europa e numerose nuove istituzioni che chiedono di esservi incluse.

È ragionevole quindi ritenere che un sistema di dati relativi alla conservazione e restauro avrebbe lo stesso successo. Progetti in tal senso sono ad esempio in corso in Belgio e nel Regno Unito, finanziati dai rispettivi governi. La KB di 4CH intende quindi agire in modo proattivo ed essere pronta a collegare e indicizzare questi sviluppi. Allo stesso tempo, la ricerca sull'organizzazione ottimale della KB può contribuire a migliorare l'organizzazione dei dati stessi oltre che a strutturarne in modo efficiente il catalogo. In effetti, il lavoro sulla KB di 4CH ha anche contribuito a sviluppare aspetti teorici che riteniamo molto rilevanti per quanto riguarda i dati stessi. L'implementazione del sistema di ricerca è stata fornita dal CNAF<sup>7</sup>, il centro di calcolo dell'INFN. Non entreremo qui nella descrizione tecnica di tale sistema, che è riportata in dettaglio nella documentazione<sup>8</sup> del progetto, per due motivi. Il primo è determinato dall'impostazione del presente contributo, che intende presentare gli aspetti strutturali della KB nell'organizzazione dei contenuti, anziché quelli strettamente tecnici. Il secondo consiste nel fatto che la strumentazione tecnologica su cui un'implementazione è basata è in continuo progresso e quindi nuovi e più efficienti motori di ricerca possono essere utilizzati senza alterare le caratteristiche del sistema, anzi migliorandole, né cambiare i suoi contenuti. Ad esempio, il motore del sistema ARIADNE è stato sostituito in maniera indolore dopo alcuni anni di funzionamento con uno più efficiente senza nessun cambiamento evidente nell'interfaccia utente, se non una più veloce e più efficiente esecuzione. Basterà quindi accennare che il sistema è basato su un database *NoSQL*, cioè non strutturato in tabelle rigide ma invece organizzato secondo relazioni flessibili fra le entità che rappresentano le informazioni.

Il sistema risultante è stato valutato molto positivamente dagli esperti nominati dalla Commissione Europea in sede di revisione del progetto, benché tuttora sviluppato solo a livello di prototipo.

## 2. La Base della Conoscenza di 4CH e i modelli 3D

Benché lo scopo del progetto riguardasse programmaticamente, come da bando, i monumenti, gli edifici storici e i siti, cioè il patrimonio culturale materiale e inamovibile, ci si è presto resi conto che tale limitazione era immotivata, illogica e disfunzionale. Non esiste infatti un bene culturale che non abbia una componente immateriale e non c'è motivo di fare distinzioni fra una statua collocata in un museo, mobile anche se con difficoltà, e una posta in una piazza o in una fontana, invece inamovibile. Anche privilegiare in modo eccessivo la tecnologia tridimensionale e farne il fulcro della documentazione appariva riduttivo, ed era forse condizionato da una tendenza presente nella ricerca, influenzata dall'indubbia potenzialità dei modelli tridimensionali. Le prime applicazioni di questa tecnica ai beni culturali nascono negli anni Novanta del secolo scorso in campo archeologico, spesso uno dei più sensibili all'innovazione tecnologica digitale, e sono utilizzate per narrare i risultati della ricerca in un modo di forte impatto e di facile comprensione da parte del pubblico, sostituendo ad esempio nei musei i plastici e i diorami già uti-

<sup>6</sup> La pubblicazione più recente a questo proposito è: *International Data Aggregation for Archaeological Research and Heritage Management: the ARIADNE experience*, ed. by Edeltraud Aspöck, Julian D. Richards, «Internet Archaeology», 64 (2023), <<https://intarch.ac.uk/journal/issue64/index.html>>.

<sup>7</sup> L'acronimo CNAF sta per Centro Nazionale Analisi Fotogrammi, un nome un po' anacronistico che risale al 1962 ma che oggi designa il più potente centro di calcolo italiano e uno dei maggiori in Europa.

<sup>8</sup> La documentazione tecnica è descritta in uno dei *deliverable* del progetto: 4CH Project, *Deliverable 3.3 - Final report on services and tools*, 2023, p. 11-21, <<https://doi.org/10.5281/zenodo.11204151>>.

lizzati nelle ricostruzioni del passato. Il campo di applicazione si estende progressivamente agli oggetti e agli edifici, documentandone la forma ora anche con finalità di gestione, di analisi e di conservazione. Ci si rende presto conto però che specialmente per quest'ultima è necessario aggiungere informazioni di altro genere, relative ad esempio ai materiali, al grado di conservazione o di degrado, agli interventi di restauro e così via. Il modello 3D diventa così anche il supporto su cui posizionare queste informazioni aggiuntive, come nel caso del restauro della fontana del Nettuno di Bologna<sup>9</sup> in cui vengono individuate le aree della statua da restaurare e se ne documentano le condizioni "appendendole" al suo modello 3D.

Questo metodo, di grande efficacia visiva e documentativa, presenta un rischio intrinseco, già evidenziato anni fa nel progetto europeo 3D-COFORM<sup>10</sup>: nei necessari processi di trasformazione della nuvola di punti si possono modificare o addirittura eliminare parti di modello (un procedimento chiamato decimazione) per ridurne il "peso" digitale e quindi inavvertitamente modificare o perdere delle informazioni associate alle regioni dell'oggetto digitale modificate o eliminate. Occorre quindi un'estrema cautela nella manipolazione del modello 3D.

Se in questi casi il modello 3D consiste in una nuvola di punti ottenuta mediante tecniche di scansione laser o fotogrammetria, in campo architettonico prende invece piede una metodologia mutuata dalla progettazione. I modelli 3D in questo caso sono ottenuti in CAD<sup>11</sup> e le altre informazioni vengono aggiunte utilizzando un sistema derivato dalla progettazione di nuovi edifici e denominato BIM (*Building Information Modeling*), che prevede la possibilità di unire alle informazioni grafiche di progetto quelle relative ai materiali da utilizzare, ai servizi presenti nelle pareti, alla struttura dell'edificio e altre ancora. Il BIM è codificato in uno standard ISO e possiede software specializzato per il suo utilizzo. Esso nasce quindi per documentare edifici nuovi, e richiede un adattamento e informazioni aggiuntive quando invece s'intende documentare quelli storici; non si presta inoltre ad oggetti come quelli presenti in un museo. Occorrono comunque delle aggiunte, che danno luogo a un sistema chiamato HBIM, dove H sta per *Heritage*, patrimonio culturale. Tuttavia, tali estensioni non-standard dello standard BIM risultano difficili da gestire con il software esistente.

Ma al di là di queste problematiche – tutte le soluzioni digitali ne presentano diverse – c'è un aspetto che rende questi approcci inadatti a una base della conoscenza come quella di 4CH: essendo focalizzati su un singolo oggetto, quello rappresentato dal modello 3D, essi guardano all'albero e non vedono la foresta. Fuor di metafora, essi non consentono una visione d'insieme del patrimonio culturale, ad esempio per individuare gli oggetti fatti con lo stesso materiale, oppure quelli di una determinata epoca o soggetti a uno specifico intervento di conservazione. In altri termini, non è possibile (o è troppo complicato) determinare ad esempio tutti i monumenti equestri prodotti in un materiale soggetto a degrado per via delle piogge acide; non si può stabilire quali interventi di restauro conservativo siano stati compiuti con successo su dipinti prodotti usando determinati materiali; e così via. Il modello 3D è poi irrilevante (o forse addirittura inesistente) per documentare beni immateriali come il canto a tenore sardo, la transumanza o la falconeria, tutti inseriti nella lista UNESCO del patrimonio immateriale dell'umanità. Si è reso

<sup>9</sup> Fabrizio Ivan Apollonio et al., *A 3D-centered information system for the documentation of a complex restoration intervention*, «Journal of Cultural Heritage», 29 (2017), p. 89-99.

<sup>10</sup> 3D-COFORM è un progetto sviluppato negli anni 2008-2012 e dedicato alla creazione di collezioni digitali di modelli 3D di beni culturali. Informazioni sul progetto e sui suoi risultati si possono reperire sul sito della Commissione Europea: <<https://cordis.europa.eu/project/id/231809>>.

<sup>11</sup> Anche se questo acronimo è di uso frequente, ricordiamo che il suo significato è *Computer-Aided Design*, cioè la parte grafica della progettazione supportata da computer mediante software specializzato.

quindi necessario sviluppare un approccio diverso alla documentazione del patrimonio culturale, basato su un punto di vista olistico che integri forma (se esistente), sostanza, condizione e interventi deliberati o accidentali su di esso. Il paradigma che integra tutti questi aspetti e potenzialmente altri ancora è il concetto di gemello digitale, anche noto come *digital twin*<sup>12</sup>.

### 3. I gemelli digitali nella documentazione del patrimonio culturale

Il concetto di *digital twin*, cioè di gemello digitale, è ampiamente diffuso in campo industriale. Esso designa un complesso di dati e processi digitali che rappresentano e simulano il comportamento di un apparecchio collocato nel mondo reale. È quindi composto da un insieme di dati che definiscono e descrivono complessivamente il modello digitale, e da connessioni con il mondo reale che permettono di simulare le reazioni dell'oggetto reale rappresentato dal suo gemello digitale rispetto ad eventi che accadono (o possono accadere) nel mondo reale. Tralasciando momentaneamente l'aspetto di simulazione, il concetto di gemello digitale corrisponde esattamente alle nostre esigenze: un oggetto digitale composto da tutte le informazioni pertinenti a un bene culturale, collegate da relazioni che ne indicano le reciproche corrispondenze e interdipendenze. Quindi il gemello digitale di un edificio storico sarà composto dal modello della sua forma, come un modello 3D, ma anche dalla documentazione dei suoi materiali costruttivi, delle relazioni fra i suoi interni, da informazioni storico-architettoniche, dalla storia del suo utilizzo, dalla documentazione degli interventi di restauro e così via, compreso il suo valore civico presente e passato. Non è necessario che tutte queste informazioni siano collocate nello stesso archivio digitale, in quanto il sistema di documentazione può aggregare e catalogare informazioni accessibili in rete anche se poste in archivi informatici diversi. Per esempio, si potrebbe trattare di uno (o più) modelli 3D depositati in sistemi diversi a seguito di attività di digitalizzazione compiute separatamente; di fotografie e filmati; di documentazione ufficiale quale quella contenuta in schede ministeriali; di articoli scientifici di ricerca storico-architettonica pubblicati in riviste; di rapporti di interventi di restauro conservati presso gli istituti che li hanno effettuati; di documentazione disparata sull'uso dell'edificio in passato e nel presente; e di storie relative all'edificio stesso, che contribuiscono a crearne il valore culturale per la sua comunità di riferimento e in generale per tutti. Il complesso di tutte queste informazioni relative a un bene culturale è definito come il *gemello digitale culturale* di tale bene, in inglese *Heritage Digital Twin* (in sigla: HDT). Il gemello digitale culturale di un bene comprende quindi tutte le informazioni disponibili per il bene in questione e le loro interrelazioni. Il sistema che organizza tali dati, che prende com'è noto il nome tecnico di *ontologia*, cioè la «rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse»<sup>13</sup>, è stato denominato HDTO, dove O sta per ontologia. Quindi HDTO è formata dai concetti generali necessari e pertinenti alla documentazione dei beni culturali (tecnicamente denominati *classi*) e dalle loro interrelazioni (chiamate

<sup>12</sup> La bibliografia sui *digital twin* e le loro applicazioni in campo industriale è sterminata. Forse la prima applicazione nel campo del patrimonio culturale si può far risalire al progetto *Gemini* (2018-2022) curato dal *Centre for Digital Built Britain* dell'Università di Cambridge, che ha sviluppato una serie di principi alla base di una progettazione del costruito finalizzata al bene pubblico e alla condivisione. Anche se il progetto si è concluso si possono trovare informazioni sul sito ad esso dedicato: <<https://www.cdbb.cam.ac.uk/what-we-did/national-digital-twin-programme>>. Per quanto invece riguarda le applicazioni di questo concetto ai beni culturali si può fare riferimento agli articoli sul tema citati nella nota 15. Anche se di contenuto tecnico, queste pubblicazioni illustrano la necessità e l'uso del concetto di *digital twin* con una serie di esempi relativi al patrimonio culturale rivolti agli operatori del settore. I contenuti di questi articoli sono comunque riassunti nel seguito del presente contributo.

<sup>13</sup> La definizione è tratta da Wikipedia, <[https://it.wikipedia.org/wiki/Ontologia\\_\(informatica\)](https://it.wikipedia.org/wiki/Ontologia_(informatica))>.

proprietà). Per ciascun oggetto culturale, ogni informazione ad esso relativa apparterrà a una classe, di cui le singole informazioni saranno esempi – o, come si dice tecnicamente, *istanze*<sup>14</sup>.

Vale la pena di osservare che questo approccio riduce la centralità del modello 3D nell'organizzazione dei dati, rendendolo solo uno degli elementi, anche se fra i più importanti, che descrivono l'oggetto.

Per garantire l'interoperabilità di questa organizzazione dei dati con quelle esistenti, l'ontologia HDTO è compatibile con lo standard per la documentazione dei beni culturali che è, com'è noto, il CIDOC *Conceptual Reference Model* (CRM). HDTO è infatti un'estensione compatibile del CRM, di cui specializza alcuni concetti, utilizzando anche altre estensioni, già esistenti, dello stesso CRM, come quella destinata agli oggetti digitali in generale, denominata CRMdig; quella per le analisi scientifiche, CRMsci; e altre ancora.

Nel progetto 4CH, la creazione dell'ontologia HDTO è stato il primo passo per organizzare i dati e aggregare le informazioni provenienti da fonti diverse. Nella base della conoscenza (KB) di 4CH organizzata secondo tale ontologia non sono contenuti quindi i dati, ma solo i metadati, con il riferimento agli archivi in cui i dati sono depositati. Una ricerca impostata sulla KB da un utente sarà così effettuata sui metadati che permetteranno la selezione dei dati rilevanti e, a richiesta, l'accesso ai dati originari.

Questa ontologia è stata accolta molto favorevolmente dalla comunità scientifica, tanto che il primo articolo in cui veniva descritta<sup>15</sup>, anche se in modo da sviluppare ulteriormente, ha raggiunto rapidamente una significativa popolarità ed ha ricevuto un premio come il migliore nella sua categoria. Importanti centri di restauro si sono dichiarati disponibili a fornire documentazione da usare come test per il sistema. Ulteriori sviluppi dell'ontologia sono in corso, ad esempio, per prendere in considerazione la presenza di sensori che rilevino in modo automatico la condizione dei beni documentati nel sistema e l'eventuale presenza di situazioni di rischio per la loro conservazione<sup>16</sup>.

L'utilizzo dell'ontologia HDTO rende la nostra Base della Conoscenza (KB) interoperabile con tutti i sistemi basati sul CRM o comunque con esso compatibili, permettendone quindi l'integrazione diretta nella KB attraverso i metadati. Il contenuto della KB è gestito attraverso un database NoSQL, cioè non strutturato in tabelle ma basato invece sulle classi definite nell'ontologia e sulle relazioni fra di esse, sviluppando così un sistema estremamente articolato e ricco di collegamenti interni. Per esempio, edifici storici posti nella stessa città sono collegati dal luogo comune in cui si trovano; ciascuno di essi sarà poi collegato ad altri edifici dello stesso stile architettonico oppure con la stessa funzione (chiesa, edificio civico, e così via). Un altro possibile collegamento riguarda il materiale di costruzione, e collega a ciascuno di questi tutti gli edifici che lo utilizzano. Sono invece specifici di ogni bene le informazioni che lo riguardano direttamente, come il nome e quelle visive: foto, modelli 3D, video e altro ancora. Ricompare qui il modello tridimensionale, an-

<sup>14</sup> Il termine *istanza* con questo significato è assente dai dizionari della lingua italiana. Si tratta infatti di un anglicismo tecnico abbastanza recente, calco della parola *instance* che in inglese significa esempio.

<sup>15</sup> Franco Niccolucci – Achille Felicetti – Sorin Hermon, *Populating the Digital Space for Cultural Heritage with Heritage Digital Twins*, «Data», 7 (2022), <<https://doi.org/10.3390/data7080105>>. Questo articolo è stato accolto con grande interesse dalla comunità scientifica e premiato come il migliore pubblicato sul tema nel 2022. Più di recente, è stato pubblicato: Franco Niccolucci – Achille Felicetti, *Digital Twin Sensors in Cultural Heritage Ontology Applications*, «Sensors», 24 (2024), <<https://doi.org/10.3390/s24123978>>, che illustra l'utilizzo di sensori nel campo dei beni culturali.

<sup>16</sup> Questo tipo di sistemi prendono il nome di *Internet of Cultural Things* (IoCT), applicazione ai beni culturali del concetto di IoT (*Internet of Things*), dove sensori dialogano fra loro attraverso la rete e svolgono in modo automatico operazioni sugli oggetti presenti nel sistema. Esempi tipici sono le applicazioni di domotica e quelle di sicurezza.



#### 4. Un esempio di applicazione: il Battistero di San Giovanni a Firenze

The diagram illustrates the relationships between various digital heritage resources, organized into two main horizontal sections. The top section is centered around the **Battistero di Firenze** (Florence Baptistery), represented by an image of the building. The bottom section is centered around the **Porta Est del Battistero** (East Gate of the Baptistery), represented by an image of the door.

**Top Section (Battistero di Firenze):**

- HC8 3D Model** (Modello 3D del Battistero (Esterno) su Sketchfab) and **HC8 3D Model** (Modello 3D del Battistero (Interno) su Sketchfab) are connected to **HC3 Tangible Aspect** (Battistero di Firenze) via **HP9 has digital representation**.
- HC3 Tangible Aspect** (Battistero di Firenze) is connected to **HC6 Heritage Digital Document** (Fiori, G. et al (2016) "I mosaici del Battistero Fiorentino: Indagini archeologiche e restauro" Ravenna Museiv. Conservazione e restauro del mosaico antico e contemporaneo) via **P70 is documented in**.
- HC3 Tangible Aspect** (Battistero di Firenze) is connected to **HC2 Heritage Digital Twin** (Digital Twin del Battistero) via **HP1 has digital twin**.
- HC3 Tangible Aspect** (Battistero di Firenze) is connected to **ES3 Place** (Piazza San Giovanni Firenze) via **P56 has current location**.
- HC3 Tangible Aspect** (Battistero di Firenze) is connected to **HC6 Heritage Digital Document** (Bartoli, G., Betti, M., Monchetti, S. (2017) Modellazione numerica ed analisi strutturale del Battistero di San Giovanni a Firenze. Mandragora) via **P70 is documented in**.
- HC3 Tangible Aspect** (Battistero di Firenze) is connected to **Narrative** (Versi di Dante sul Battistero di Fienze) via **HP2 has story**.
- HC6 Heritage Digital Document** (Bartoli, G., Betti, M., Monchetti, S. (2017) Modellazione numerica ed analisi strutturale del Battistero di San Giovanni a Firenze. Mandragora) is connected to **HC3 Tangible Aspect** (Battistero di Firenze) via **P46 is composed of**.
- Narrative** (Versi di Dante sul Battistero di Fienze) is connected to **HC3 Tangible Aspect** (Battistero di Firenze) via **HP4 narrated through**.
- HC2 Heritage Digital Twin** (Digital Twin del Battistero) is connected to **HC3 Tangible Aspect** (Battistero di Firenze) via **HP9 has digital twin component**.

**Bottom Section (Porta Est del Battistero):**

- HC6 Heritage Digital Document** (Vasari, Le vite, 1588) is connected to **HC3 Tangible Aspect** (Porta Est del Battistero) via **P79 is documented in**.
- HC7 Digital Visual Object** (Foto Porta Est del Battistero su Wikimedia) is connected to **HC3 Tangible Aspect** (Porta Est del Battistero) via **HP9 has digital representation**.
- HC3 Tangible Aspect** (Porta Est del Battistero) is connected to **ES3 Place** (Museo Opera del Duomo Firenze) via **P55 has current location**.
- HC3 Tangible Aspect** (Porta Est del Battistero) is connected to **HC2 Heritage Digital Twin** (Digital Twin della Porta Est del Battistero) via **HP1 has digital twin**.
- HC3 Tangible Aspect** (Porta Est del Battistero) is connected to **Narrative** (Frases di Michelangelo sul nome della Porta Est "Porta del Paradiso") via **HP2 has story**.
- HC3 Tangible Aspect** (Porta Est del Battistero) is connected to **HC6 Heritage Digital Document** (Alighieri, D. La Divina Commedia, Inferno, XIX canto) via **HP6 narrated in**.
- HC2 Heritage Digital Twin** (Digital Twin della Porta Est del Battistero) is connected to **HC3 Tangible Aspect** (Porta Est del Battistero) via **HP9 has digital twin component**.

123

fia relativa, gli aspetti letterari (la citazione di Dante nella *Divina Commedia*) e la particolarità di questo edificio di contenere un bene descritto separatamente e dotato di un proprio gemello digitale (non riportato in dettaglio nell'esempio), cioè la Porta est, anche nota come Porta del Paradiso perché così chiamata da Michelangelo, realizzata da Lorenzo Ghiberti. Del gemello digitale fanno parte anche documentazioni visive, fra cui un modello 3D.

Nella descrizione precedente ci siamo limitati a indicare alcuni esempi del contenuto della KB relativi al Battistero di San Giovanni per dare un'indicazione della funzionalità del sistema. Si potrebbero naturalmente aggiungere moltissime altre informazioni, per esempio quelle relative al restauro delle formelle della "Porta del Paradiso" danneggiate dall'alluvione del 1966, oppure quelle sulla conservazione dei mosaici oggetto di un recente intervento di conservazione, e altre ancora. L'efficienza del sistema si basa sulla sua capacità di aggregare informazioni da fonti disparate (purché disponibili online in formato digitale) e di stabilire così collegamenti fra contenuti digitali diversi ma tutti pertinenti allo stesso bene, il Battistero. Questa operazione ha l'ulteriore vantaggio di non duplicare i contenuti, di cui sono solo raccolti i metadati.

## 5. Prospettive per il futuro

Per arrivare a un sistema effettivamente operativo e utilizzabile saranno necessari numerosi passaggi. Il primo richiede la collaborazione di ricercatori, operatori culturali e istituzioni che posseggono i dati necessari. Se abbiamo infatti già a disposizione dei dati reali di test, va però ulteriormente sviluppata la collaborazione con questi soggetti, in Italia e all'estero, per ampliare la base da aggregare nella KB. Queste collaborazioni permetteranno di raffinare e precisare ulteriormente il nostro approccio semantico a fronte di problematiche reali.

Il secondo riguarda assicurare le risorse organizzative e finanziarie per supportare il lavoro. Anche su questo aspetto il gruppo di ricerca che ha lavorato in 4CH sta attivamente ricercando soluzioni.

Uno dei problemi, anzi forse il problema che si presenta è la presenza e la disponibilità di dati e metadati adeguati. Anche assumendo che la documentazione sia conservata in archivi potenzialmente accessibili per l'aggregazione, ci si può aspettare che spesso i metadati disponibili siano modesti, essendo probabile che tali archivi siano stati concepiti più come depositi statici che come oggetto di ricerche strutturate. La creazione di metadati apparentemente non indispensabile è un'attività noiosa, che richiede tempo e non produce effetti tangibili nell'immediato. Nella fase iniziale ci siamo quindi limitati a compilarli manualmente per testare il funzionamento del sistema. Per una fase successiva di funzionamento a regime sono allo studio sistemi automatici che possono essere attivati contestualmente alla creazione del documento, oppure utilizzati su documenti già esistenti, basati sull'intelligenza artificiale. Un esperimento iniziale<sup>17</sup> di qualche tempo fa su un piccolo insieme di documentazione archeologica ha permesso di verificare la fattibilità di questo procedimento, estraendo automaticamente da rapporti di scavo archeologico le parole rilevanti (ad esempio anfora, ceramica, e così via) e permettendo così l'indicizzazione automatica del testo analizzato.

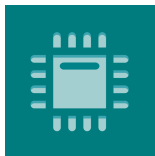
Questo procedimento di arricchimento automatico dei metadati è destinato a migliorare rapidamente in efficienza e nei tempi di "apprendimento" grazie ai progressi recenti in questo settore.

<sup>17</sup> Achille Felicetti – Daniel Williams – Ilenia Galluccio – David Tudhope – Franco Niccolucci, *NLP Tools for Knowledge Extraction from Italian Archaeological Free Text*, in: *3rd Digital Heritage International Congress (DigitalHERITAGE) & 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, San Francisco, CA, USA, 2018, p. 1-8, DOI: <<https://doi.org/10.1109/DigitalHeritage.2018.8810001>>.

Si tratta di evitare le cosiddette “allucinazioni”, cioè risultati clamorosamente errati derivanti dall'utilizzo di informazioni non pertinenti. Ciò è consentito da alcuni modelli di intelligenza artificiale che, a differenza di quelli più noti, non pescano a caso fra una grande quantità di informazioni ma invece utilizzano solo un insieme predefinito di partenza. Si tratta comunque di un'attività di ricerca da impostare per il futuro. Sarà poi necessario sviluppare una descrizione sistematica dei protocolli di analisi scientifica, producendo thesauri e dizionari controllati da affiancare a quelli largamente diffusi per le informazioni storico-artistiche quale, ad esempio, lo Art & Architecture Thesaurus (AAT) sviluppato dal Getty Institute.

*The present article describes the data management model developed in the 4CH European project and the creation of the project knowledge base. This is based on the digital twin concept in an innovative way, namely as the collection of data pertaining to an asset, both graphical or visual data, such as images or 3D models, and textual or documental data, as well as their interrelations. It is an innovative approach enabling the management of the documentation in a unitary way, independently of the data format and their provenance.*





*sensors*



Article

---

# Digital Twin Sensors in Cultural Heritage Ontology Applications

---

Franco Niccolucci and Achille Felicetti



<https://doi.org/10.3390/s24123978>

## Article

# Digital Twin Sensors in Cultural Heritage Ontology Applications

Franco Niccolucci \*  and Achille Felicetti 

VAST Lab, PIN, Piazza dell'Università 1, 59100 Prato, Italy; achille.felicetti@gmail.com

\* Correspondence: franco.niccolucci@gmail.com

**Abstract:** This paper concerns the extension of the Heritage Digital Twin Ontology introduced in previous research to describe the reactivity of digital twins used for cultural heritage documentation by including the semantic description of sensors and activators and all of the process of interacting with the real world. After analysing previous work on the use of digital twins in cultural heritage, a summary description of the Heritage Digital Twin Ontology is provided, and the existing applications of digital twins to cultural heritage are overviewed, with references to reviews summarising the large production of scientific contributions on the topic. Then, a novel ontology named the Reactive Digital Twin Ontology is described, in which sensors, activators, and the decision processes are also semantically described, turning the previous synchronic approach to cultural heritage documentation into a diachronic one. Some case studies exemplify this theory.

**Keywords:** digital twins; cultural heritage documentation; ontologies

## 1. Introduction

Digital twins are increasingly used in all research and application fields.

As discussed in recent reviews on the use of this term in different domains [1], this concept captures the complexity of digital equivalents of the real world both in industrial and research applications. It is sometimes present as a generic term to designate a digitised model of real artefacts or, more specifically, a structured information system about reality. Also, regulatory authorities are using this terminology. Notable examples are offered by the European Commission, which has introduced the idea of the Digital Twin of the Earth [2] to indicate a global environmental model but is also using this notion rather loosely in its research programmes, calling for research projects that deal with “digital twins” in different domains to simulate research activities and conduct them in a digital framework.

The idea of introducing digital twins in land planning and built environment activities dates back to the Cambridge National Digital Twin project, which stated general principles for their application in this domain, the Gemini Principles [3].

As concerns cultural heritage, the term digital twin made its appearance in reference to digitised 3D models of heritage artefacts as a synonym of digital replica, suggesting an original literary rather than technical use of the term. The first applications of 3D visualisation in cultural heritage date back to the last decade of the 20th century, especially in archaeology, and were motivated by the need to illustrate the supposed pristine appearance of remains, replacing drawings and maquettes previously used for this purpose with visual reconstructions of the past. A similar approach rapidly extended to document the shape of artefacts, monuments, and sites and has now produced a huge amount of heritage 3D models of extremely different quality and detail. Looking at heritage assets in 3D rather than by 2D images has become a straightforward practice, although still needing more work, as demonstrated by the EU Recommendation on this topic [4] calling for activities incentivising the creation of new ones.

In time, it has become apparent that taking into account only the shape of heritage assets is not sufficient for advanced model use, both in research and in practice. Any serious



**Citation:** Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. <https://doi.org/10.3390/s24123978>

Academic Editor: Yi Qin

Received: 22 May 2024

Revised: 9 June 2024

Accepted: 13 June 2024

Published: 19 June 2024



**Copyright:** © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).



use requires background documentation which needs to be stored, linked to the 3D model, and made available for inspection.

Work on this topic started within the 3D-COFORM project (2008–2012) [5–7] and continued with various significant contributions, especially for heritage conservation and restoration [8,9]. In some of these applications, the additional information concerning, for example, materials used in different parts or conservation interventions has been attached as a text to the 3D model or to some of its regions. Although availing a completely different 3D technology, similar applications have been developed adopting a CAD approach to the creation of the 3D model. In this case, the heritage-related documentation is managed by extending an existing information system, called BIM (Built Information Modelling), widely used in the building industry to specify information about materials, services, and processes concerning the construction of a *new* built asset. BIM is semantically defined by the IFCs (Industry Foundation Classes), which have been extended in various way to incorporate concepts typical of cultural heritage. The resulting application, called HBIM where H stands for Heritage, has a widespread use especially for documenting monuments and historic buildings. A 2023 paper [10] surveys a very large number of such applications. Within HBIM models, digital twins have been envisaged as a way of introducing heritage-related concepts [11,12]. However, the data management software managing HBIM suffers from the limits of the data management software available for BIM, which privileges the analysis of individual assets modelled but is much less efficient in investigations across different buildings modelled in this way, for example those built with the same material (wood, bricks, etc.).

Both such approaches suffer from their “original sin”, i.e., having a 3D model as the root of the documentation tree, where any other information is just a leaf. Thus, information structures pertain to an individual heritage asset that makes further data processing a cumbersome task: comparisons among different assets, e.g., to search for construction materials across non-visual documentation, is impractical, as no advanced database technology can assist in this job because the data are not properly organised. On the contrary, well-structured heritage documentation organizes data in *classes* of which the information pertaining to an individual asset is an instance. *Classes* are interrelated by *properties*. For example, “wood” would be an instance of the class “Construction Material”, linked to the class “Heritage Asset” by a property “is made of”. This enables us to search for all assets with wood as the construction material and listing the corresponding instances, i.e., all of the buildings made of wood. Such a semantic approach enables the creation of large digital documentation systems which allow us to manage knowledge about heritage assets and the activities carried out on them, especially conservation and preservation ones. In this way, a huge amount of research results can be made available and searchable by scientists and practitioners. This availability complies with the FAIR principles as it not only supports findability and access but also makes them interoperable via the compliance to standards and ultimately allows for their reuse. The purpose of the present paper is to show how the digital twin paradigm is suitable to contribute to this objective and improve the effectiveness of data organisation in the heritage domain.

*Classes* and *properties* are organised in ontologies, and for cultural heritage, the standard one is the CIDOC Conceptual Reference Model (CRM), ISO 21127 [13]. CIDOC stands for the committee which originally started the CRM definition, now managed by a Special Interest Group. Besides organising the information related to cultural heritage, the CRM presents an important feature, *extensibility*, i.e., the possibility of specialising concepts, i.e., the *classes*, and relationships, i.e., the *properties*, to fit better to the data description in a particular application subdomain. The resulting models are interoperable with any other one compatible with the CRM by establishing a correspondence (a *mapping*) between couples of corresponding *classes* (and *properties*) in each model, either directly to each other when feasible, or via the CRM class/property which is a common superclass/super-property of both. Thus, the CRM base—the universal set of heritage-related concepts—has developed several extensions, such as CRMarchaeo for archaeological excavation, CRMsci

for scientific observations, CRMdig for digital activities and components, and so on, offering a complete framework to organize knowledge about heritage. It also encompasses concepts related to intangible heritage by means of *classes* already proposed for narration. Both *classes* and *properties* are identified by a literal mark common to all in the ontology, followed by a progressive number; for example, E followed by a number (e.g., E1) denotes the CRM *classes* and P followed by a number (e.g., P9) is used for *properties*. It must be noted that extensions to the CRM benefit from inheriting the robustness of this reference model, including, among others, the consistency of the ontology which derives directly from being an extension of the standard one. For example, a building might be, at the same time, an asset and a place, but the risk of partial overlap of the two *classes* has already been resolved within the CRM and the solution extends to all its extensions, such as ours. Therefore, concerning the construction of the RHDTO and the control of its internal consistency, we have not relied on any specific tool such as Protégé or OntoStudio, but rather on the conceptual tools and methodologies provided by the CIDOC CRM. In particular, by aligning the RHDTO with the CIDOC CRM, we were able to carry out conceptual checks at each step of the development and for each new entity introduced to ensure that the model remains consistent, coherent, and interoperable with the CIDOC CRM framework and with other ontologies and systems that use the CIDOC CRM as their core model. To achieve this alignment, we have followed a systematic and iterative process of *mapping* and refinement, in which we have identified and investigated the relevant *classes* and *properties* in the CRM and other related ontologies and defined new *classes* and *properties* in the RHDTO only when actually required. We have extensively used CIDOC CRM scope notes, examples, and formal definitions to ensure that the RHDTO's *classes* and *properties* are well structured and consistently linked with any other entity in the CRM ecosystem.

In the development of the RHDTO model, we have exclusively focused on the conceptual design of the ontology without restricting our investigation to its implementation in a specific tool or language. This is because the RHDTO is intended to be a high-level ontology that provides a comprehensive framework for modelling the dynamic and complex nature of cultural heritage entities and risk management operations and, as such, is designed to be agnostic to any particular tool or format. Of course, once the general conceptualisation is formalised, the ontology can be implemented by means of different technologies. For example, the Heritage Digital Twin Ontology (HDTO), which we already released and which we are extending here, has already been expressed in the RDFS and the data have been encoded using the Turtle syntax, this being a particularly suitable compact and human-readable format for the RDF to describe the objects of our domain and to be queried and reasoned over by using standard Semantic Web technologies, such as SPARQL and OWL reasoners. This is what we plan to do with the RHDT extension as well, using the RDFS or any other suitable format, when the model is tested and evaluated and has reached the necessary stability and maturity to be released and used.

Regarding the potential computability of the ontology, during the development phase, we have verified the completeness and expressivity of the extension by ensuring that it provides all of the features for modelling the multifaceted nature of all cultural heritage and risk management entities, while at the same time keeping the introduced *classes* and *properties* to the minimum necessary number. However, the actual computability of the RHDTO, in terms of its performance evaluation, adaptability to services, and scalability, requires further investigation and testing during the implementation phase.

In the present paper, as well as in the previous ones, we aim to describe the logical structure of the ontology as a solution to the digital documentation needs of the heritage sector. A rich and extensive semantic structure may improve cultural heritage documentation, which is indeed possible, although ineffective, also with much simpler systems, but the richness and flexibility of more advanced ones, such as the CRM and its extensions—including the present one—allow us to cover the required documentation extension. A rich documentation allows us to use the data for many digital applications, from valorisation to the focus

of our work, conservation, and preservation. A digital twin holistic paradigm appears to us as the most effective framework for this purpose.

The above-mentioned critical points of 3D-based systems are among the reasons why in 2022 and 2023, we proposed a novel ontology for heritage information as a CRM extension. It is based on the Heritage Digital Twin (HDT), a holistic approach to heritage information where the 3D graphical component is just one element [14,15]. In this ontology, named HDTO, all of the documentation is organised according to a compatible extension of the basic CRM model, and thus, information from other documentation systems organised according to the CRM can be straightforwardly imported into it. Moreover, the HDTO enables us to document the intangible component of tangible heritage assets, and in the particular case of intangible heritage, such as the one enlisted in the UNESCO list [16], it is the only existing one providing complete coverage; for intangible heritage, the visual documentation may consist of video or audio recordings or may even be totally absent. Three-dimensional models have also been used to document aspects of intangible heritage (see [17] for a review of such applications), but they are usually less suitable as the starting point of the related documentation which includes stories, traditions, and other immaterial content is very often unrelated to the shape. However, we will thoroughly explore the complexity of documenting intangible heritage in future work.

The HDTO has been used to set up the cloud-based knowledge base (KB) created in 4CH [18], an EU-funded project designing a Competence Centre for the Conservation of Cultural Heritage. Documentation in the 4CH KB includes the relevant information about heritage assets, from the visual and 3D data to the results of scientific analyses, conservation activities, and historical research. The description on the 4CH KB implementation is fully documented in the above-mentioned technical report, which also addresses aspects concerning its scalability to very large amounts of data and their processing. As described in the related report [18], the system has been implemented on the INFN-CNAF super-computer based in Bologna, Italy, planned to become *Leonardo*, one of the most powerful ones in Europe. This machine can effectively allow for fast data processing regardless of their size and amount by providing computational resources in real time as required. The applicability of real-time calculations in the knowledge base depends on several factors, such as the size and complexity of ontology-encoded information, the volume and velocity of the data, the computational resources and algorithms available, and the specific requirements and constraints of the services and applications. We also recognize that the knowledge base may contain a large and diverse number of subclasses and instances, which may pose challenges and trade-offs in terms of the accuracy, completeness, and timeliness of the calculations. In this regard, we have designed the RHDT Ontology to be highly expressive while also minimising redundancy and complexity, which should help to improve the efficiency of data processing. To address these challenges, we also plan to adopt a multi-strategy approach at the implementation stage, which combines different techniques and tools for ontology engineering, data management, and reasoning, such as modularisation, approximation, sampling, and parallelisation. Thus, we believe that real-time calculations are feasible for the types of data and decisions required by the knowledge base extended according to the RHDT Ontology, such as the detection and diagnosis of critical events, the prediction and simulation of future scenarios, and the optimisation and adaptation of the system's behaviour.

On the other hand, data aggregation from heritage institutions is a process that may require time to establish cooperation agreements that go beyond the digital technicalities, such as those concerning access rights, ownership of the data, and other legal aspects. In our opinion, preparing the technical framework—as our work on the ontology aims to do—is however a necessary step in this direction. It also allows us to open a discussion with heritage institutions and professionals, the data providers, which may lead to fine-tuning and improving the data organisation framework.

Organising heritage data management is *per se* important, especially in the present time in which the data overflow risks are becoming unmanageable. It also has important impacts

on research and management. Aggregating heritage-related documentation, possibly stored in different repositories, within a holistic approach as the one we propose may have different uses. For example, it may enable us to improve the valorisation of heritage assets and the enjoyment by the public by providing a one-stop shop for all of the content required to build scientifically based explanations, also availing approaches such as those pertaining to virtual, augmented, and extended reality. However, this is not our goal. Our work focuses on the conservation, preservation, and when necessary, restoration of heritage assets. In this regard, a rich documentation is absolutely required, and a wide knowledge base allows for comparisons with similar conditions and interventions on other assets. Indeed, heritage is affected by many different risks, both individual—i.e., related to individual assets—and systemic. The latter include, for example, the effects of catastrophic events such as floods, fires, landslides, and earthquakes, as well as human actions such as vandalism, crime, and war. *Sensors* and IoCT play a crucial role in documenting the availability of automated reactions to conservation risks, and thus, their introduction in the ontology is a significant step forwards in the semantic description of heritage assets.

While the previous version of the HDTO provided the first step in organising the documentation of heritage assets into a static system, with the present improved version, we aim to move towards a dynamic connection between the real thing and its digital replica. The RHDTTO can therefore provide several benefits since it establishes a consistent framework for representing heterogeneous information and ensure that data are captured and stored in a format ready to be used by a wide range of services and applications. In particular, the ontology is fundamental for integrating data from a variety of sources, including historical and material records, scientific studies, and *sensors*, as well as data on the general architecture of the system and its functioning, information with different structures and formats. The ontology is thus fundamental to create the complex knowledge graph necessary to make these data interoperable, an operation not feasible when using databases or other non-semantic tools. Since the RHDTTO is designed as a system with different components and data, the ontology is therefore essential to orchestrate this heterogeneous mass of information and allow the system to work at its best.

The HDTO does not (yet) consider the dynamic and interactive aspects connecting a digital twin to reality; timewise, it is static, just incorporating in each documentation item a timestamp that may be used to reconstruct the diachronic evolution of the asset. In both of our previous papers mentioned above, the authors acknowledged that this approach was just a first step in the definition of a full-fledged digital twin, still lacking the modelling of interactions with reality-based information.

In the present paper, and in other forthcoming related ones, we will try to move beyond this limitation and to develop an extended documentation system where interaction with reality may be documented and put to work in a continuous interchange of information and process activation between the real world and its equivalent in the digital universe. Such extension is motivated by the need to reflect either one of the features embedded in the digital twin concept, its reactivity to inputs coming from reality and its capacity to produce corresponding outputs and real-world actions. Besides this theoretical necessity, we believe that a reactive heritage digital twin may better model heritage assets which are immersed in a continuous changing landscape, are affected by phenomena happening in it, and with their own changes, contribute to compose its evolution. The new semantic model proposed here extends but in no way supersedes the previous version, which conserves its usefulness in the many cases in which this dynamic perspective is not required. The HDTO is, anyway, the substrate on which the digitally simulated reactions take place, and the new ontology named the Reactive Heritage Digital Twin Ontology (RHDTTO) is an extension incorporating all of the previously defined *classes* and *properties*.

Due to the complexity of such model extension, we will proceed by taking steps with separate contributions, the present one being dedicated to *sensors* and *activators*. This stepwise approach will also allow for a more focused discussion within the research community, hopefully producing improvements to be incorporated in future versions of

the ontology. We will often use very simple examples to keep the reader's attention on the ontological modelling rather than on the technical aspects of the sensor and to facilitate the model's explanation to non-technical heritage professionals who play a very important role in use-inspired [19] research in this domain, but of course, the ontology works also in more complicated cases, as will be illustrated in the sequel.

Finally, and to conclude this long introduction, we will not address implementation aspects in this paper. For the existing knowledge base, the previous version of the ontology was written in the RDF and implemented as described in the already mentioned technical report [18]. The implementation was preceded by the logical description of the ontology [14,15] which was very favourably received by the heritage community. The implementation was then reviewed by a commission of experts and considered an excellent result of the related funded project. We will follow the same approach with the extended ontology presented here, stimulating a discussion and receiving suggestions for improvement as far as the novel dynamic approach to heritage documentation presented here is concerned. Then, we will test its effectiveness on case studies (the example presented in this paper may be one of them), in parallel with collecting data from the producers/owners, i.e., cultural heritage institutions, finally implementing the system in the RDF as required.

## 2. A Summary of the HDTO

To keep this paper self-contained, we summarize below the main features of the HDTO. For full details, please refer to the above-mentioned papers [14,15] introducing the concept of the Heritage Digital Twin (HDT).

According to [20], "a digital twin is a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between the physical and virtual system". Another definition in [21] describes the digital twin as made across five dimensions: physical entities, virtual models, services, data, and connections.

Both our papers [14,15] acknowledged that what they proposed was only half way in the work needed to develop a full-fledged concept of digital twins related to cultural heritage, as they addressed only three of the above-mentioned aspects of the digital twin: physical entities (renamed as real-world entities to also encompass intangible aspects), virtual models, and data, while services and connections were temporarily set apart to be considered in future work, of which the present paper is an initial part. This led to a formal semantic definition of the HDT and of the related ontology, the HDTO. The main reason for this partial move was to start setting up an overarching data infrastructure where all of the parts of the virtual model include the relationships to a well-organised data system. The HDTO is thus conceived as an extension of the CRM, enabling semantic interoperability with all of the (many) existing information systems for cultural heritage documentation which adopt this standard or compatible systems, such as the EDM (Europeana Data Model) [22] and its forthcoming extension expected to be used in the data space for cultural heritage. Our above-mentioned papers were well received by the research community, having been extensively quoted in the literature and the first one [14] being the recipient of the *Data* 2022 Best Paper Award.

The most important innovation proposed in the HDTO is to define the concept of the Heritage Digital Twin (HDT) in a holistic perspective, i.e., as *the complex of all available digital information concerning a (real world) heritage asset*, either movable (e.g., museum exhibits), immovable (e.g., monuments and sites), or even intangible (e.g., traditions). In the resulting data organisation, all of the data, such as reports, documents, datasets, and visual representations (2D, 3D, or 4D), or any other related digital data, are linked with each other within the HDT by appropriate properties. Thus, the HDT becomes a full-fledged digital alter ego of real-world heritage assets.

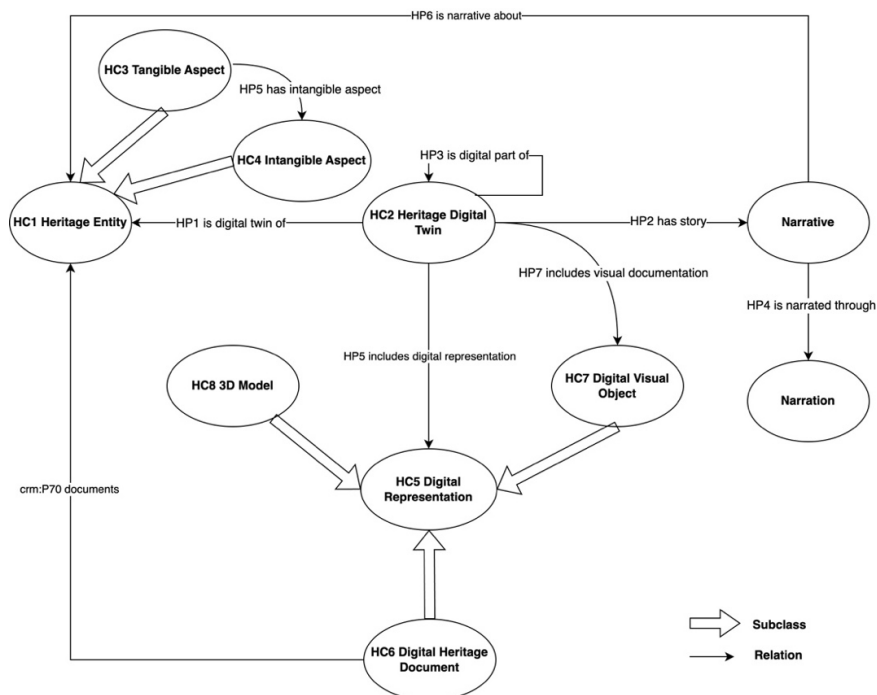
As presented in the two papers [14,15] mentioned above, the HDTO aims to encapsulate the dual aspects of cultural heritage, both tangible and intangible, and the cultural events in which cultural entities are involved throughout their history, offering specific

classes and properties for the dynamic documentation and analysis of their interactions. Moreover, the HDTO is pivotal in structurally organising data to document digital twin systems, knowledge bases, and other similar operational platforms. Being machine-readable and actionable, the HDTO functions as an internal “language” of the digital twin system, ensuring smooth interaction among its various digital components. The HDTO classes describe the more common high-profile entities, such as places, agents, physical objects, events, and temporal entities, thereby ensuring high-level interoperability in multiple domains. New classes and properties are defined to describe more specific concepts such as digital twins, cultural entities, 3D models, and others, having no exact match in the CRM ecosystem. To facilitate the distinguishing of classes and properties in the text, henceforth, we will indicate their names and symbols in *italic boldface*.

The root class of the HDTO is *HC1 Heritage Entity*, comprising tangible and intangible entities of the real world regarded as valuable because of their contribution to society, knowledge, and/or culture. *HC1* is a subclass of *E1 CRM Entity*, the root class of the CRM. Instances of *HC1 Heritage Entity* may refer to real assets of any nature: physical, both movable and immovable, immaterial, or born digital. Subclasses of *HC1 Heritage Entity* model the material and immaterial aspects that a cultural entity can have. In particular, the *HC3 Tangible Aspect* class models the tangible, material component of entities of the real-world, both movable (e.g., archaeological, artistic, and cultural objects) and immovable (e.g., built heritage as monuments, buildings, cities, and other complexes). *HC4 Intangible Aspect* comprises instead cultural events, traditions, and practices having particular social, historical, and cultural significance, including practices and expressions, memories, and oral traditions about events, things, and people. They can be independent or related to some physical object. Other classes in the HDTO model the digital documentation related to heritage entities, encompassing 3D models (rendered through the dedicated *HC8 3D Model* class), images, other audio–visual mediums (modelled using the *HC7 Digital Visual Object* class), as well as textual information made available in digital format (represented by means of the *HC6 Digital Heritage Document* class). Stories and storytelling, intended, respectively, as accounts of facts about cultural entities and the way in which these facts are narrated, are modelled using the narrative and narration classes provided by the Narrative Ontology (NOnt), another CRM compatible ontology for to the description of texts and their content [23]. The use of the NOnt, as well as other compatible models such as CRMsci and CRMdig, demonstrates the capability of the HDTO to integrate semantic tools defined in other contexts to represent the different facets of a digital twin. However, no classes or properties of the NOnt are used in this paper, so we refer to [15] for definitions and examples.

Heritage Digital Twins of cultural entities are rendered in the HDTO by means of the class *HC2 Digital Twin* used to organize and connect the digital information available in a given system and pertaining to an *HC1 Heritage Entity*, including digital representations and information of the effects of events that influenced or/and are related in any way to its state and activities (e.g., restorations, conservations, etc.) carried out on it. The *HC2* class semantically represents the way in which the Heritage Digital Twin is implemented and the building blocks through which it operates. In addition to those inherited from the CRM, the HDTO also provides other properties, such as *HP1 is digital twin of*, linking a heritage entity (*HC1*) with its digital twin (*HC2*). In general, the HDTO properties are mainly intended to semantically interconnect the various pieces of information modelled through the classes. Thus, they are elements used to implement the knowledge graph of the Heritage Digital Twin. The main classes of the HDTO and the properties linking them are shown in the following diagram (Figure 1). The figure outlines the semantic structure of the HDTO, which is fully described and illustrated with diagrams in [14,15].





**Figure 1.** A semantic diagram of the HDTO, showing its main classes and properties.

An additional advantage of the introduction of the HDT derives from the already mentioned modelling of historical buildings using BIM (Building Information Modelling), based on the Industry Foundation Classes (IFCs), and the HBIM (Heritage BIM) based on them. The extensible nature of the HDTO allows us to incorporate BIM and such extensions into it via mappings from/to the IFCs and the proposed HBIM extensions to the HDTO.

### 3. Sensors and Activators

In a simple definition, a *sensor* is a device that produces an output signal for the purpose of detecting a physical phenomenon. More precisely, a *sensor* is a device, module, machine, or subsystem that detects events or changes in its environment and sends the information to other electronics, frequently a computer processor. Thus, a *sensor* measures one (or more) physical quantity and sends a signal based on the measured value(s) to an *activator*, which automatically activates some action if required according to built-in criteria. Very simple everyday examples are thermostats, which combine a temperature *sensor* and an *activator* switching off or on an air conditioner if the temperature is within or outside a pre-selected range. The automated action may be very simple, such as the switching on/off of a warning red light in a car fuel gauge or an anti-theft alarm which detects the interruption of an electrical circuit due to the opening of a house front door and activates an alarm siren. The latter case may also consist of a movement detector (the *sensor*) that sends a signal to the siren *activator*. This second example is a bit more sophisticated as it may take into account other aspects, for example the volume of the intrusive object, and decide if it is a thief or just your cat wandering around during its human's absence.

In any case, the model of the overall action consists of several steps:

- A *sensor*: measuring one (or more) physical quantity—in the above examples, they, respectively, are the temperature, the quantity of fuel in the car tank, and the amount of infrared radiation for an anti-theft movement *sensor*—and forwarding the measure(s) to the *decider*.
- A *decider*: comparing the *sensor* measurement with pre-set decision rules and deciding if some action needs to be activated. In the above examples, for the thermostat, the decision is based on a threshold, a limit temperature, and the required action is switching the air conditioning on/off; for the fuel gauge, the threshold is a minimum



acceptable quantity of it, and the action is switching on the red warning light in the car dashboard; and for the anti-theft system, the action involves determining if the size of the intruder according to the infrared size is comparable with a human or is smaller, such as a cat. If the *sensor* signal falls in the alarm range, the *decider* sends an activation signal to the *activator*. The decision rules will often be much more complex and imply different actions according to the values measured by the *sensor*.

- An *activator*: When authorised by the *decider*, it commands actions in the real world, e.g., switching some device on or off, such as the air conditioning system, the car red warning light, and the alarm siren. The action may be programmed to result in something more complex than a simple “Do”; in this case, we may logically split such a complex *activator* into different elementary activators. For example, an anti-theft *activator* may command different actions such as “activate the siren” and “call police”; thus, it corresponds to two elementary activators.

In simple cases, the *decider* is included either in the *sensor* or in the *activator*, frequently in the latter as in the above examples. It is however a logical functionality different from the device, and thus, we will keep it separate from both.

The logic of the *sensor–decider–activator* system is often hard-coded in the electronics of the devices, especially for simple ones, but there is also a strong tendency towards soft-coding to support its integration with other smart devices and enable remote control via wi-fi and the internet. This has led to an increase in the purely digital component, which facilitates the inclusion of this system in a digital twin framework. In any case, both the *sensor* and the *activator* are placed in the real world, so they do not belong to the digital twin but should be included in the overall semantic description of the whole system. The decision process may be very simple, as in the above examples, or very complicated when several factors must be taken into account, requiring processing and possibly, in a forthcoming future, an AI-based decision process.

#### 4. The Internet of Cultural Things (IoCT)

The Internet of Cultural Things, i.e., the use of the so-called Internet-of-Things (IoT) in cultural heritage applications, has been a subject of study and research in the last 10 years.

The concept of the Internet of Cultural Things (IoCT) refers to the application of Internet of Things (IoT) technologies in the context of cultural heritage and cultural assets (see [24] for a recent survey on the topic). It involves the integration of sensors, devices, and information systems within cultural contexts to monitor, preserve, study, and enhance artworks, archaeological sites, museums, monuments, and other elements of cultural heritage. In the development of digital twins in the cultural heritage sector, IoCT plays a fundamental role. In the cultural context, digital twins have been extensively used to create a digital replica of an artwork, an archaeological site, or a museum environment, allowing cultural operators to explore, preserve, and share cultural heritage in innovative ways. Sensors range from the very simple ones previously described to a network of devices measuring physical quantities and exchanging the related information with each other. Both sensors and activators may be directly linked to the digital twin or be connected in a network in which they are the nodes.

The orchestration of such IoCT nodes may envisage different ways in which devices and sensors within an IoCT network communicate with each other and with data processing systems, allowing devices and sensors within an IoCT network to effectively and efficiently collaborate to collect, transmit, simulate, and process relevant cultural data, thereby contributing to the development of digital twins and innovation in the cultural heritage sector.

Among the many studies concerning IoCT and digital twin applications in the cultural heritage domain, [25] reviews the most important applications, mainly focusing on the valorisation of cultural heritage and to assist visitors and promote tourism. Very simple examples of such applications are the sensors activating a video when somebody enters a museum room or, in general, starting an automated reaction to human presence to make

the heritage experience more interactive, enjoyable, and participative for visitors. More complex ones collect measures from various devices, connected in an articulated system as described above, and use these to make decisions on actions to be performed by one or more activators.

Digital twin applications to the conservation of monuments are discussed in general in [26], while [27] describes the use of sensors to monitor the conservation conditions of historic buildings, in this case a church in Matera's "Sassi" considered in its complex environment. Applications to movable artefacts are less frequent; among them are the one described in [28] in relation to a famous violin and the study of palaeolithic lithics presented in [29]. A semantic approach to digital twins is introduced in [30], which discusses the ontologies for the description of sensors and IoT in preventive conservation. Other related papers are [31–34].

## 5. The Reactive Heritage Digital Twin Ontology (RHDTO)

### 5.1. Introductory Notes

In this section, we introduce the new entities of the RHDT aimed at modelling the reactive elements of the digital twin, such as sensors, services, and data processing engines. As mentioned earlier, our ontology builds upon the CRM ecosystem and incorporates elements from the CRMdig, CRMsci, and CRMpe extensions. Concepts and logics from various existing ontologies, developed to model sensors and related phenomena, have been investigated to fine-tune our conceptual tools. Models such OntoSensor [31] and SensorML [32,33] provide a rich and expressive framework for representing sensors, their properties, and their relationships to other entities, such as observed properties, features of interest, and observation processes. Nevertheless, our primary goal is not only to model the phenomena related to digital objects, their nature, and their interactions, but also to develop a semantic tool for making Reactive Digital Twins operational and achieving integration and interoperability between the different types of data they store and manage.

By leveraging the CRM ontology and extending it with new classes and properties where necessary, we aim to create a comprehensive ontology that can effectively represent the dynamic and complex nature of cultural heritage and risk management, and support decision-making and risk assessment in real-world scenarios. The CRM, with its robust framework for modelling events, is particularly well suited for capturing the dynamic aspects of the Reactive Heritage Digital Twins and representing the complexity typical of this domain, including state changes detected by sensors and other devices, component interactions, and processes. Leveraging this event-centred approach allowed us to better focus on system behaviours and predict future outcomes by identifying common patterns and trends within the system's operations.

Throughout the development process, we have thus prioritised the reuse of existing classes and properties from the CRM ecosystem, only introducing new ones when necessary to accurately represent the intended semantics of the described entities. This approach strikes a balance between ensuring interoperability and maintaining precision in describing the functionalities and components of the system, while also keeping the ontology compact and easy to understand and apply. Therefore, the resulting RHDT Ontology only introduces classes and properties to describe the specific components and events involved in the various operations performed by the system, the interaction between devices and services, and the communication with human operators.

Although our ontology is specifically designed for cultural heritage, many of the new concepts we have introduced, such as the devices, sensors, activators, data, and signals, may have broader applicability beyond this domain. In fact, we have taken care to define these entities in general terms to make them consistent with relevant models developed in other fields, such as OntoSensor and SensorML, and to facilitate interoperability and reuse in other domains.

Table 1 reports the symbols used for classes and property names in the CRM and its extensions used in the present paper. For example, in the CRM, the class symbol is E and

the property symbol is P. Symbols are combined with a number to identify a specific class or property, e.g., *E1* or *P2*.

**Table 1.** The CRM extensions used in this paper.

Extension Name	Used for	Reference	Class Symbol	Property Symbol
CRM	General use	[13]	<i>E</i>	<i>P</i>
CRMdig	Digital objects	[35]	<i>D</i>	<i>L</i>
CRMsci	Scientific analyses	[36]	<i>S</i>	<i>O</i>
CRMpe	Interoperability framework	[37]	<i>PE</i>	<i>PP</i>
HDTO	Heritage Digital Twins	[14,15]	<i>HC</i>	<i>HP</i>

Also, in the following sections, we will use the convention of denoting the semantic elements in the text by *italics–boldface*.

Another extension of the CRM, proposed in [38], describes a model for the results of scientific analyses on heritage assets to be used for the data concerning them. It is not used in the present paper, but its classes and properties are relevant in applications to support the description of the condition states of artefacts, the experiments carried out to evaluate them, and their results.

## 5.2. Sensors

Sensors are a central element of the Reactive Digital Twin. In our ontological view, a sensor can be defined in general terms as a digital device placed on physical objects or in specific locations intended to measure and collect data about them, process it, and transmit it to the digital twin system for analysis and further processing. A sensor can measure various physical or environmental properties, such as temperature, pressure, humidity, light, sound, position, velocity, or acceleration, and can be of different types, such as analogic or digital, wired or wireless, active or passive. Building upon this general definition, we introduced the new *HC9 Sensor* class, specialising the *D8 Digital Device* of CRMdig, a general class aimed at describing instances of material items capable of processing or producing digital data.

The positioning of a sensor on a cultural object can be modelled using the new *HP15 is positioned on* property, allowing for the representation of the spatial relationship established between the object to be monitored (*HC3 Tangible Aspect*) and the sensor placed on it. For sensors positioned in a space adjacent to the cultural object, we can use the *P55 has current location* property of the CRM, connecting a sensor to the instance of *E53 Place* where the monitored object is currently located to indicate that both of these physical entities share the same space.

During its operational lifecycle, a sensor can assume a particular status representing its condition at a given time, which can be modelled using the *E3 Condition State* class of the CRM. Instances of this class enable the description of sensor states (such as on/off, operational/non-operational) and operation modes, reporting any faults or errors, as well as any maintenance or calibration activities performed on it. Additionally, sensors are typically operated by different kinds of software, tailored for gathering and processing acquired data, generating signals, and executing specific functions such as configuration, calibration, monitoring, and diagnostics. To represent this operational software, we employ the *D14 Software* class of CRMdig that was chosen due to its comprehensive semantic features allowing for the modelling of all of the software components of the Reactive Digital Twin. While the *D14* class already provides the necessary features to model the software components of any device, we deemed it important to explicitly represent the relationship between the sensor and its controlling software. For this reason, a specific *HP11 is operated by* property was created to indicate the close interconnection existing between the sensor and the software that controls it. This also enables the link between the software, the measurement operations performed by the sensor, and the digital signals generated from them.

### 5.3. Measurements and Signals

In a CRM perspective, the measurement operations performed by sensors can be modelled as events. A sensor measurement event, in fact, involves the collection of data identifying various modifications of conditions on the object or in the environment under examination, within a certain spatiotemporal interval. In accordance with these features, we have defined the new **HC13 Sensor Measurement** class as a subclass of the **S21 Measurement** class of the CRMsci ontology to describe the measurement events performed by the sensors connected to Reactive Digital Twins. The **L12 happened on device** property of CRMsci is used to specify the sensor (**HC9**) on which the measurement (**HC13**) took place. Measured events can instead be modelled by means of the CRM **E5 Event**, a class particularly well suited for describing generic events of various types. The measurement (**HC13**) and the measured (**E5**) events are ontologically related through the **O24 measured** property of CRMsci. Instances of the CRM **E55 Type** class can be assigned to instances of **HC13** to detail, among the other things, the detected types of risk conditions related to the cultural object, and subsequently trigger the generation of the corresponding signals. This assignment is usually carried out by means of the **L17 measured thing of type** property of CRMsci.

Generated signals can be modelled as digital objects since they codify a measurement taken by a sensor to be transmitted, under certain conditions, to the system for further processing or analysis. Typically, a signal is a piece of software encoded in a formal language (e.g., XML, JavaScript, etc.) and generated in response to a specific event, such as a change in sensor reading, a threshold crossing, or a specific timer. It may contain various information, including sensor identifiers, timestamps, recorded values, measurement units, and details about condition assessment quality. In our model, instances of signals are represented by means of the new **HC12 Signal** class, a specialisation of the **D9 Data Object** class of the CRMdig ontology. Signals (**H12**) generated in the presence of a potential risk detected by sensors can be linked to the measurement event that generated them through the **L20 has created** property of CRMdig. Once generated, signals are transmitted to the Reactive Digital Twin and specifically to dedicated intelligent digital agents (deciders) running as part of the Reactive Digital Twin system. The transmission of signals to deciders is recorded via the new **HP12 was transmitted to** property.

### 5.4. Deciders

The RHDTO provides a new **HC10 Decider** class, specialising the scope of **PE1 Service** of the CRMpe ontology, to describe in detail the deciders and their features. Deciders can receive and analyse input data from various sources, such as sensors, activators, or other services, and query the digital twin knowledge base to acquire knowledge concerning the linked cultural objects and their potential risks. They further process all of this information using algorithms, rules, or models and generate output instructions, such as commands, feedback, or status updates, to be sent to other components of the digital twin, to activators, or to other external devices. Moreover, deciders can also send alerts to human operators in various forms, such as email, SMS, push notifications, and visual or audible alerts, depending on the nature of the detected issue.

Ontologically, the actions performed by deciders after the decision-making process is complete can be modelled as events. Thus, we have designed a specific **HC14 Activation Event** class, a subclass of the CRM **E5 Event**, to represent them. The **O13 triggered** property of CRMsci is particularly useful for linking activation events to the deciders by which they were triggered. Activation events can prompt specific actions by digital agents, such as adjusting a valve or activating a pump, and/or alert human operators, informing them of the necessary countermeasures to be taken in response to the detected potential risks. Both operations are modelled by means of the new properties **HP14 alerted**, pointing to instances of the **E39 Actor** class, and **HP13 activated**, linking activation events with activators.

### 5.5. Activators and Human Interaction

Activators are other fundamental components of Reactive Digital Twins. In our RHDT Ontology, they are described as devices that enact actions upon physical objects or processes based on the instructions received by the deciders operating within the Reactive Digital Twin system. They can be of mechanical, electrical, digital, hydraulic, or pneumatic types, and can execute a range of actions, including movement, adjustment, control, or regulation. Activators may operate in distinct modes, including automatic, manual, or semi-automatic, and can be governed by various software systems, such as firmware, drivers, or applications. Moreover, they offer various degrees of precision, accuracy, and responsiveness, contingent upon their design, fabrication, and maintenance processes.

Activators interact with the Reactive Digital Twin through multiple channels, including sensors, APIs, or communication protocols, enabling the reception of commands, feedback, or status updates from the system. Given their specialised nature, we have defined a new **HC11 Activator** class, a subclass of **D8 Digital Device** of the CRMdig ontology to describe these peculiar devices. As in the case of sensors, activators may be operated by software (again encoded through the **D14 Software** class of CRMdig) intended for interpreting the commands and feedback coming from the Reactive Digital Twin system and controlling the activator's actions accordingly.

As already anticipated, the **E39 Actor** CRM class is used to represent people and/or institutions responsible for the safety and security of cultural entities and their environment. These are the people who (may) receive alerts from deciders in order to take appropriate actions in the case of risky events. As we said, the **HP14 alerted** property is used to associate an instance of **HC14 Activation Event** with an instance of **E39 Actor**, indicating the alerts sent to the specified individuals or groups. This property is of particular interest since it describes the modalities of interaction and collaboration between the digital system and the real world, making particularly evident, also on the ontological level, the importance of the human component for the effective prevention and management of dangerous situations.

A schematic overview of the RHDT Ontology, its entities, and their semantic relationships is illustrated in Figure 2.

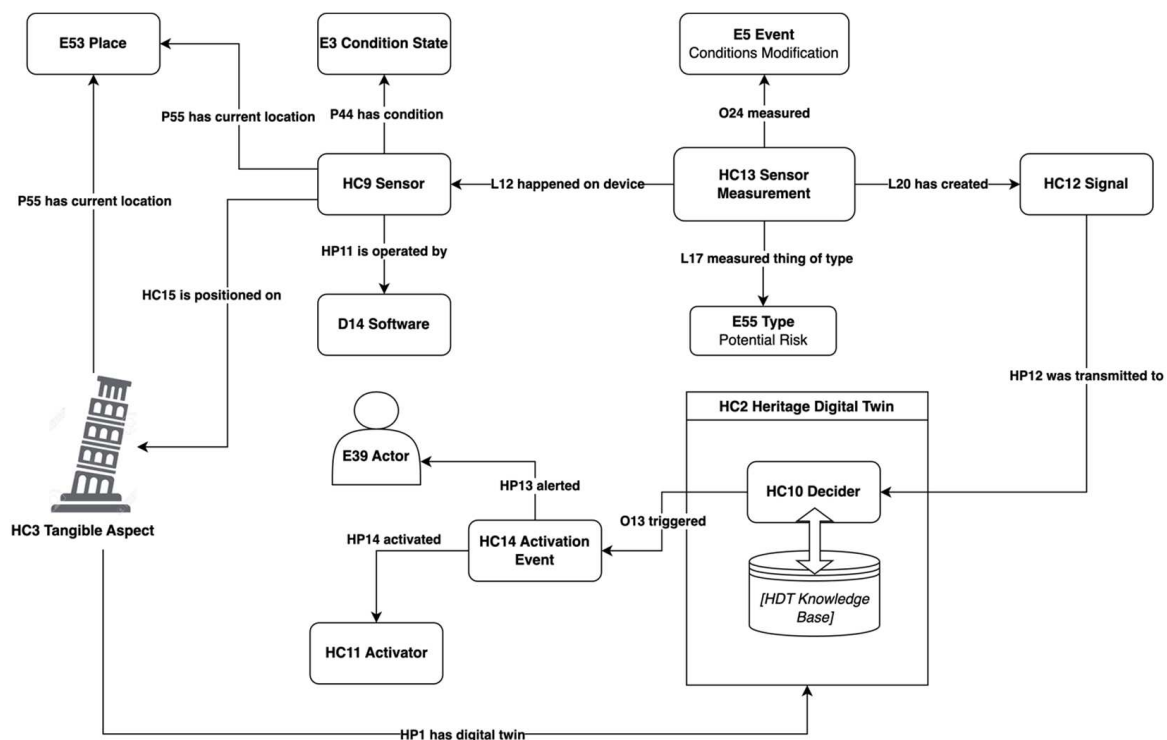


Figure 2. Classes and properties of RHDT Ontology.

## 6. A Summary of the New Classes and Properties

The new classes and properties are listed below with their scope note, i.e., a short description.

### 6.1. Classes

#### 6.1.1. HC9 Sensor

Subclass of *D8 Digital Device* of CRMdig Ontology

This class comprises specialised devices utilised for monitoring and evaluating the conditions of artifacts, structures, or environments of historical significance. These devices are designed to detect changes in parameters such as temperature, humidity, vibration, or light exposure which may impact the conservation of cultural assets. Operated by dedicated software, sensors can exist in various states, including active, standby, or alert. For example, a temperature sensor may indicate fluctuations that could endanger delicate artifacts like manuscripts or artworks. When certain predefined thresholds are surpassed, the sensor enters an alert state, signalling potential risks and prompting appropriate intervention measures.

#### 6.1.2. HC10 Decider

Subclass of *PE1 Service* of CRMpe Ontology

This class comprises software components responsible for receiving signals from sensors and autonomously making decisions aimed at safeguarding cultural objects. Acting as an intelligent service within the conservation systems of the digital twin, deciders process incoming data from sensors and information stored in the digital twin knowledge base to detect factors such as environmental conditions, artifact vulnerability, and conservation protocols. Based on this analysis, it executes predefined algorithms or decision-making rules to determine appropriate actions for preserving cultural heritage. These actions could include adjusting environmental controls, activating protective measures, or triggering alerts to conservators or relevant personnel when necessary.

#### 6.1.3. HC11 Activator

Subclass of *D8 Digital Device* of CRMdig Ontology

This class comprises digital devices responsible for executing actions determined by the decider of the digital twin system to safeguard cultural objects. Serving as a crucial link between decision-making and practical implementation, the activator translates directives from the decider into tangible interventions aimed at mitigating risks or optimising conservation conditions. Interventions may encompass a variety of mechanisms, including the activation of mechanical, electronic, pneumatic, or hydraulic systems, among others. Activators ensure the timely and effective execution of proactive conservation strategies, contributing to the long-term preservation of cultural heritage assets.

#### 6.1.4. HC12 Signal

Subclass of *D9 Data Object* of CRMdig Ontology

This class is used to model particular data objects generated by sensors and to document specific detected conditions. Signals are typically transmitted to the digital twin, where they are processed and analysed by the system's algorithms to generate insights into the condition and conservation needs of cultural heritage. Signals may be encoded in a specific data formatting language, facilitating efficient transmission, storage, and analysis. Utilising standardised formats ensures interoperability and compatibility among different sensor systems and conservation platforms, including the digital twin knowledge base, enabling the seamless integration of data from diverse sources for comprehensive conservation management. The incorporation and analysis of encoded signals enables the digital twin to support informed decision-making and proactive preservation strategies for cultural heritage assets.

#### 6.1.5. HC13 Sensor Measurement

Subclass of **S21 Measurement** of CRMsci Ontology

This class comprises specific measurement events in which a sensor detects and quantifies a specific parameter or condition relevant to the monitoring and conservation of cultural assets. Events of this kind occur when a sensor registers changes in parameters such as temperature, humidity, light exposure, or vibration, capturing information that reflects the environmental conditions surrounding cultural objects or structures. Sensor measurement is thus an essential feature for continuously assessing the time and circumstances in which conditions of risk may affect cultural heritage, providing valuable insights into factors that may impact the preservation of artifacts or sites over time.

#### 6.1.6. HC14 Activation Event

Subclass of **E5 Event** of CRM Ontology

This class serves to model actions performed by an activator to initiate specific interventions or alerts aimed at safeguarding cultural assets. Activation events occur when an activator executes directives received from the digital twin's decider, triggering actions such as activating climate control systems, deploying protective enclosures, or alerting personnel through various communication channels, including email notifications, SMS, and other similar ones.

#### 6.2. Properties

**HP11 was operated by the following:**

Domain: **HC9 Sensor**;

Range: **D14 Software**.

This property links an instance of **HC9 Sensor** to the instances of **D14 Software** that operate it. The software is usually a piece of code running on the sensor and responsible for controlling and managing it, for instance by configuring its settings, collecting and processing its data, and generating signals based on its measurements.

**HP12 was transmitted to the following:**

Domain: **HC11 Signal**;

Range: **HC9 Decider**.

This property associates the instances of **HC11 Signal** with the instances of **HC9 Decider**, indicating that a certain signal has been transmitted to the digital twin decider for processing. The property can be used to model the flow of data from sensors to the deciders and can be useful for tracking the status of signals and ensuring that they are properly processed by the decider. The property can also be used to model the relationship between signals and the specific decider services that process them, allowing for more fine-grained analysis and optimisation of the digital twin system.

**HP13 activated the following:**

Domain: **HC14 Activation Event**;

Range: **HC11 Activator**.

This property associates the instances of **HC14 Activation Event** with the instances of **HC11 Activator**, indicating the specific digital device activated by the digital twin system based on the decision made by the decider component regarding the actions to be taken according to the detected risk.

**HP14 alerted the following:**

Domain: **HC14 Activation Event**;

Range: **E39 Actor**.

This property associates the instances of **HC14 Activation Event** with the instances of **E39 Actor**, indicating the action of alerting human personnel. The property can be used to model the communication between the digital twin and human operators, following the



decision taken by the decider component, and to describe the modalities of collaboration and decision-making between the digital system and the real world.

*HP15 was positioned on the following:*

Domain: *HC9 Sensor*;

Range: *HC3 Tangible Aspect*.

This property is used to model the spatial relationship between instances of HC9 Sensor and the physical object on which the sensors are located. The property is fundamental to document the specific case in which a sensor is physically placed on or attached to a cultural heritage physical object rather than being simply placed nearby or in the same environment as the monitored object.

## 7. Example: Giovanni Pisano's Pulpit in Pistoia, Italy

To exemplify the documentation of a system using sensors, we will consider a recent paper concerning the pulpit in the church of Sant'Andrea in Pistoia (Italy), a medieval masterwork by the Italian sculptor Giovanni Pisano described in [39]. The physical system includes sophisticated sensors and takes into account previous work by a worldclass restoration centre based in Florence, the *Opificio delle Pietre Dure* (OPD), as described in the above-mentioned paper. The array of sensors installed for the purpose of gathering environmental and dynamic data encompassed various types, among which were humidity sensors situated on the wall in the side nave adjacent to the pulpit, and uniaxial accelerometers positioned atop the pulpit and on the ground near its base. This specific configuration facilitated meticulous monitoring of both the environmental area surrounding the pulpit and any dynamic oscillations occurring on its surface.

To demonstrate the functionality of our ontology in representing this scenario, we specifically focus on these two types of sensors, illustrating how they can be semantically described using our classes and properties and integrated into the overall description of the installed monitoring system.

The semantic modelling starts by noting that Giovanni Pisano's pulpit is a monument and thus a physical cultural object that can be represented by instantiating the *HC3 Tangible Aspect* class of our RHDT Ontology. The church of Sant'Andrea in Pistoia (Italy) can be represented by an instance of the *E53 Place* class of the CRM, while the fact that the pulpit is housed in this church can be rendered through the *P55 has current location* property of the same model. Specific identifiers for both of these heritage entities can be defined and/or derived using for example the global identifiers provided by Wikidata, such as the <https://www.wikidata.org/wiki/Q3925522> (accessed on 20 May 2024) URI identifying the pulpit, or the <https://www.wikidata.org/wiki/Q1148335> (accessed on 20 May 2024) URI identifying Sant'Andrea's church in the Wikidata system. The *HC2 Digital Twin* class can be instantiated to define the Reactive Digital Twin of the pulpit in the semantic space of our model.

As previously noted, instances of the *HC9 Sensor* class could be employed to represent the diverse types of the sensors used to monitor the monument. We focus our example on two instances of this class used to represent the temperature and humidity sensor installed on the walls of the church and the uniaxial accelerator sensor positioned on top of the pulpit. Regarding their specific position, the distinct placements of these two sensors are rendered either through the CRM property *P55 has current location*, used to indicate the positioning of the temperature and humidity sensor on the wall of the church, and the property *HC15 is positioned on*, used to specify the placement of the uniaxial accelerometer sensor directly on the surface of the monument. The *HP11 is operated by* property is utilised to establish the connection between the sensors and the corresponding instances of the CRMdig *D14 Software* class, employed to represent the software operating them.

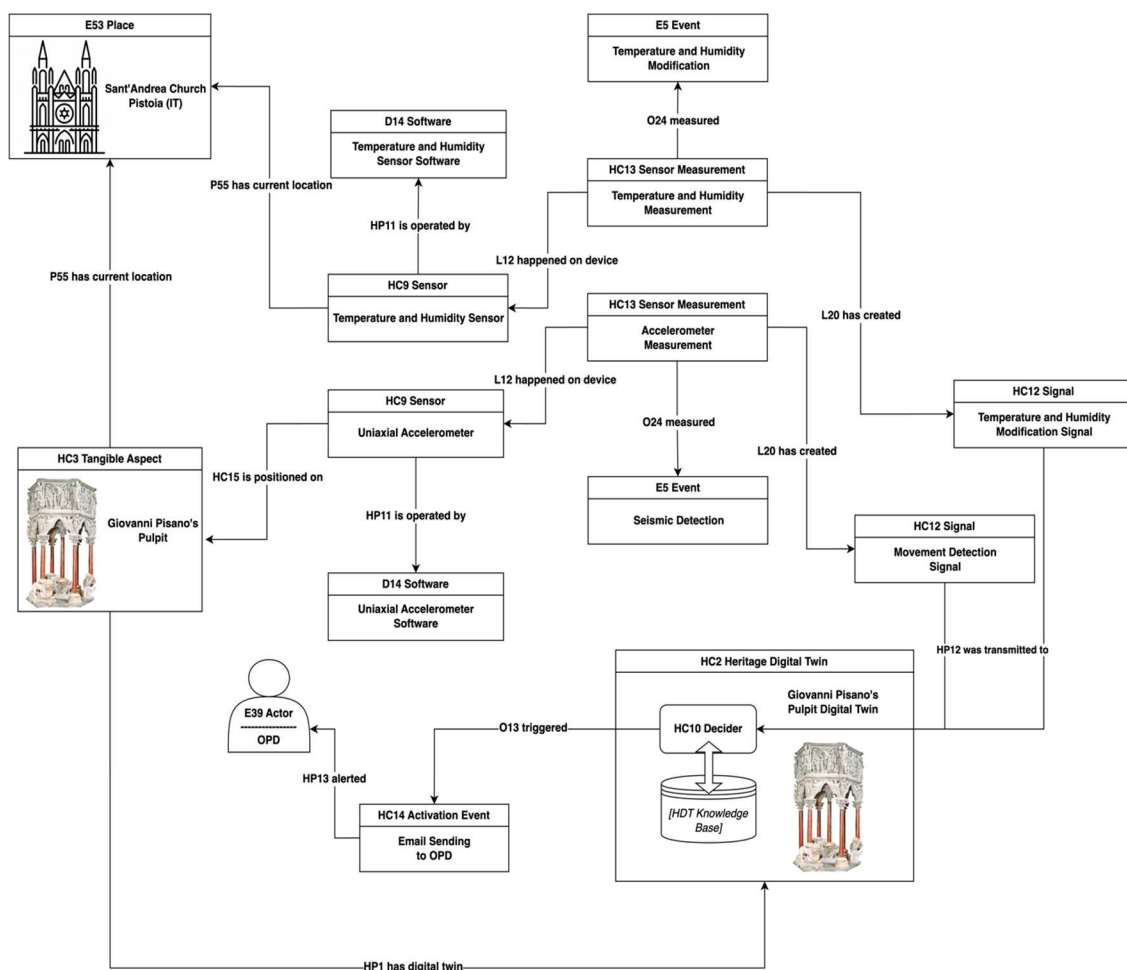
An associated instance of the *HC13 Sensor Measurement* class is defined to specify, through the CRMdig *O24 measured* property, the type of event monitored by each sensor, and specifically, the seismic movements detected through variations in acceleration by the uniaxial accelerometer, and changes in the temperature and humidity parameters of

the church recorded by the temperature and humidity sensor. Both of these events are modelled by means of the *E5 Event* class of the CRM.

The signals resulting from these measurements are represented by using the *HC12 Signal* class and linked to the event that generates them via the *L20 has created* property of CRMdig. The signal transmission to the digital twin of the pulpit (*HC2*) is encoded through the *HP12 was transmitted to* property and the integrated monitoring system that receives them by instances of the *HC10 Decider* class. This is the system designed to acquire the transmitted values, analyse the various physical and environmental conditions of the monument, extend the information coming from other sensors, and enrich and interact with the digital twin's knowledge base.

The sensor system of Giovanni Pisano's pulpit does not include activators, but just for the sake of exemplifying how the system would be described when including an activator, we have added a hypothetic second part, not present in the actual system. In this supposed system, a decider (*HC10*) evaluates the variations in some measures from standard values and if they exceed a threshold, it sends an email to the OPD to intervene. This process can be semantically modelled by means of the *O13 triggered* CRMdig property and *HC14 Activation Event* representing the action of email transmission by the system to the OPD competent office (*E39 Actor*).

Figure 3 represents the knowledge graph of the example. It illustrates the semantics of this example in a simplified way to show an application of the RHDTO to a typical monitoring scenario for a heritage asset.



**Figure 3.** A semantic graph of the Giovanni Pisano's pulpit monitoring system using the RHDTO Ontology.

## 8. Conclusions and Future Work

In the present paper, we have addressed the introduction of sensors and activators in the RHDTO to describe the reactions of a system modelled as the digital twin of a heritage asset to external real-world events. The impact of such events is analysed in terms of measures of one or more physical quantities made by sensors and analysed by deciders which command actions (or possibly no action) to activators.

This model does not cover all of the possible reactions of the digital twin to external events. They may be described by information collected by external systems and forwarded to the digital twin system, such as weather forecasts or earth science information and then combined with other data from the decider, for example with the geographic location of the asset, ultimately leading to actions by activators in a way similar to the sensor–decider–activator chain but with the sensor replaced by the external information. A rather famous example is the MOSE system in Venice, used to limit the effects of flooding caused by high tide and strong northern winds on the monuments of the famous city by rising mechanical barriers at the lagoon mouths on the Adriatic Sea. MOSE bases its action on forecasting high tides, which have a well-known regularity, and wind strength and direction with weather forecasts. Although questioned in light of some anticipated environmental side effects, MOSE has so far succeeded in avoiding the effects of *acqua alta* (literally, “high water”, i.e., high tide) on the city monuments.

Finally, another possibility is the simulation of external events, digitally modelled in the system and described in the digital twin, to be analysed by the decider to produce as the output the simulated consequences, possibly also including the simulated effects of corrective actions by activators. For example, one might ask the system what are the effects of extremely hot summers on an art gallery and how are these mitigated by air conditioning. All such simulations will need scientific models behind the scenes and the related processing programmes, which may be stored in a computer system and documented with all of the other information in the corresponding Reactive Digital Twin.

**Author Contributions:** All authors have equally contributed to the present paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data created.

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

## References

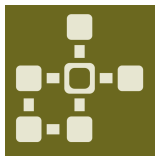
1. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [CrossRef]
2. European Commission, Destination Earth. Available online: <https://digital-strategy.ec.europa.eu/en/policies/destination-earth> (accessed on 1 May 2024).
3. Bolton, A.; Lorraine, B.; Dabson, I.; Enzer, M.; Evans, M.; Fenemore, T.; Harradence, F.; Keaney, E.; Kemp, A.; Luck, A.; et al. The Gemini Principles: Guiding Values for the National Digital Twin and Information Management Framework, 2018. Available online: <https://www.cdbb.cam.ac.uk/system/files/documents/TheGeminiPrinciples.pdf> (accessed on 1 May 2024).
4. European Commission. Commission Recommendation of 10.11.2021 on a Common European Data Space for Cultural Heritage. Official Journal of the European Union, 2021. Available online: <http://data.europa.eu/eli/reco/2021/1970/oj> (accessed on 1 May 2024).
5. 3D-COFORM. Available online: <https://cordis.europa.eu/project/id/231809> (accessed on 1 May 2024).
6. Rodriguez Echavarria, K.; Theodoridou, M.; Georgis, C.; Arnold, D.; Doerr, M.; Stork, A.; Peña Serna, S. Semantically rich 3D documentation for the preservation of tangible heritage. In Proceedings of the 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, Brighton, UK, 19–21 November 2012; Arnold, D., Kaminski, J., Niccolucci, F., Stork, A., Eds.; pp. 41–48.

7. Pena Serna, S.; Schmedt, H.; Ritz, M.; Stork, A. Interactive Semantic Enrichment of 3D Cultural Heritage Collections. In Proceedings of the 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST, Brighton, UK, 19–21 November 2012; Arnold, D., Kaminski, J., Niccolucci, F., Stork, A., Eds.; 2012; pp. 33–40.
8. Apollonio, F.I.; Basilissi, V.; Callieri, M.; Dellepiane, M.; Gaiani, M.; Ponchio, F.; Rizzo, F.; Rubino, A.R.; Scopigno, R.; Sobrà, G. A 3D-centered information system for the documentation of a complex restoration intervention. *J. Cult. Herit.* **2017**, *29*, 89–99. [\[CrossRef\]](#)
9. Gros, A.; Guillem, A.; De Luca, L.; Baillieu, E.; Duvocelle, B.; Malavergne, O.; Leroux, L.; Zimmer, T. Faceting the post-disaster built heritage reconstruction process within the digital twin framework for Notre-Dame de Paris. *Sci. Rep.* **2023**, *13*, 5981. [\[CrossRef\]](#)
10. Lovell, L.J.; Davies, R.J.; Hunt, D.V.L. The Application of Historic Building Information Modelling (HBIM) to Cultural Heritage: A Review. *Heritage* **2023**, *6*, 6691–6717. [\[CrossRef\]](#)
11. Jouan, P.; Hallot, P. Digital twin: A HBIM-based methodology to support preventive conservation of historic assets through heritage significance awareness. *ISPRS Arch.* **2019**, *42*, 609–615. [\[CrossRef\]](#)
12. Jouan, P.; Hallot, P. Digital Twin: Research Framework to Support Preventive Conservation Policies. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 228. [\[CrossRef\]](#)
13. CIDOC CRM. Available online: <https://cidoc-crm.org/> (accessed on 1 May 2024).
14. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Digital Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [\[CrossRef\]](#)
15. 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. *arXiv* **2023**, arXiv:2302.07138. [\[CrossRef\]](#)
16. UNESCO List of Intangible Cultural Heritage. Available online: <https://ich.unesco.org/en/lists> (accessed on 1 May 2024).
17. Skublewska-Paszkowska, M.; Milosz, M.; Powroznik, P.; Lukasik, E. 3D technologies for intangible cultural heritage preservation—Literature review for selected databases. *Herit. Sci.* **2022**, *10*, 3. [\[CrossRef\]](#)
18. 4CH Project. *Deliverable 3.3—Final Report on Services and Tools*; Zenodo: Cham, Switzerland, 2023; pp. 11–21. [\[CrossRef\]](#)
19. Stokes, D.E. *Pasteur’s Quadrant—Basic Science and Technological Innovation*; Brookings Institution Press: Washington, DC, USA, 1997; p. 196.
20. VanDerHorn, E.; Sankaran, M. Digital Twin: Generalization, characterization and implementation. *Decis. Support Syst.* **2021**, *114*, 113524. [\[CrossRef\]](#)
21. Qi, Q.; Tao, F.; Hu, T.; Anwer, N.; Liu, A.; Wei, Y.; Wang, L.; Nee, A.Y.C. Enabling technologies and tools for digital twin. *J. Manuf. Syst.* **2021**, *58*, 3–21. [\[CrossRef\]](#)
22. Definition of the Europeana Data Model v5.2.8. Available online: [https://pro.europeana.eu/files/Europeana\\_Professional/Share\\_your\\_data/Technical\\_requirements/EDM\\_Documentation/EDM\\_Definition\\_v5.2.8\\_102017.pdf](https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf) (accessed on 1 May 2024).
23. Meghini, C.; Bartalesi, V.; Metilli, D. Representing narratives in digital libraries: The narrative ontology. *Semant. Web* **2021**, *12*, 241–264. [\[CrossRef\]](#)
24. Liang, X.; Liu, F.; Wang, L.; Zheng, B.; Sun, Y. Internet of Cultural Things: Current Research, Challenges and Opportunities. *Comput. Mater. Contin.* **2023**, *74*, 469–488. [\[CrossRef\]](#)
25. Li, S.; Du, K.; Xu, W.; Cao, Y. The Applications of Internet of Things in Architectural Heritage Preservation. In Proceedings of the 11th International Conference on Information Systems and Computing Technology (ISCTech), Qingdao, China, 30 July–1 August 2023; pp. 469–474. [\[CrossRef\]](#)
26. Astorga González, E.M.; Municio, E.; Noriega Alemán, M.; Marquez-Barja, J.M. Cultural Heritage and Internet of Things. In Proceedings of the GoodTechs ’20: Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good, Antwerp, Belgium, 14–16 September 2020; pp. 248–251. [\[CrossRef\]](#)
27. Lerario, A.; Varasano, A. An IoT Smart Infrastructure for S. Domenico Church in Matera’s “Sassi”: A Multiscale Perspective to Built Heritage Conservation. *Sustainability* **2020**, *12*, 6553. [\[CrossRef\]](#)
28. Manfriani, C.; Gualdani, G.; Goli, G.; Carlson, B.; Certo, A.R.; Mazzanti, P.; Fioravanti, M. The Contribution of IoT to the Implementation of Preventive Conservation According to European Standards: The Case Study of the “Cannone” Violin and Its Historical Copy. *Sustainability* **2021**, *13*, 1900. [\[CrossRef\]](#)
29. Marulli, M.R.; Sorrentino, G.; Menna, F.; Paggi, M. Digital twin models of replicative ground stones: Insight into simulating usage of Upper Palaeolithic tools. *Sci. Rep.* **2023**, *13*, 18298. [\[CrossRef\]](#)
30. Michalakakis, K.; Moraitou, E.; Aliprantis, J.; Caridakis, G. Semantic Representation and Internet of Things in Cultural Heritage Preventive Conservation. In Proceedings of the International Conference on Cultural Informatics, Communication & Media 2019-CICMS 2019, Mytilene, Greece, 13–15 June 2019. [\[CrossRef\]](#)
31. Russomanno, D.J.; Kothari, C.R.; Thomas, O.A. Building a Sensor Ontology: A Practical Approach Leveraging ISO and OGC Models. In *IC-AI*; Hamid, R., Joshua, A.R., Eds.; CSREA Press: Las Vegas, NV, USA, 2005; pp. 637–643.
32. Barnaghi, P.; Meissner, S.; Presser, M.; Moessner, K. Sense and Sensability: Semantic Data Modelling for Sensor Networks. *Proc. ICT Mob. Summit* **2009**, 2009, 1–9.
33. SensorML. Available online: <https://www.ogc.org/standard/sensorml/> (accessed on 1 May 2024).
34. Peralta, L.R.; de Brito, L.P.L.; Gouveia, B.T.; de Sousa, D.G.; Alves, C.D.S. Automatic monitoring and control of museums environment based on wireless sensor networks. *Electron. J. Struct. Eng.* **2010**, *1*, 12–34. [\[CrossRef\]](#)

35. Doerr, M.; Stead, S.; Theodoridou, M. Definition of the CRMdig. An Extension of CIDOC-CRM to Support Provenance Metadata. 2022. Available online: <https://www.cidoc-crm.org/crmvig/sites/default/files/CRMdigv4.0.pdf> (accessed on 1 May 2024).
36. Kritsotaki, A.; Hiebel, G.; Theodoridou, M.; Doerr, M.; Rousakis, Y. Definition of the CRMsci. 2017. Available online: <https://cidoc-crm.org/crmsci/ModelVersion/version-1.2.3> (accessed on 1 May 2024).
37. Bruseker, B.; Doerr, M.; Theodoridou, M. Report on the Common Semantic Framework. PARTHENOS Project 2017. Available online: [https://www.parthenos-project.eu/Download/Deliverables/D5.1\\_Common\\_Semantic\\_Framework\\_Appendices.pdf](https://www.parthenos-project.eu/Download/Deliverables/D5.1_Common_Semantic_Framework_Appendices.pdf) (accessed on 1 May 2024).
38. Niccolucci, F.; Felicetti, A. A CIDOC CRM-based Model for the Documentation of Heritage Sciences. In Proceedings of the 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), San Francisco, CA, USA, 26–30 October 2018; pp. 1–6. [CrossRef]
39. Marafini, F.; Betti, M.; Bartoli, Z.G.; Casarin, F.; Marchesini, F.; Barontini, A.; Mendes, N. Static and Dynamic Monitoring of Giovanni Pisano's Pulpit in Pistoia (Italy). In *SAHC 2023, RILEM Bookseries 47*; Endo, Y., Hanazato, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2024; pp. 197–210. [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.





*computers*

IMPACT  
FACTOR  
**4.2**

CITESCORE  
**7.5**

Article

---

# Digital Twins and the Stendhal Syndrome

---

Franco Niccolucci and Achille Felicetti



<https://doi.org/10.3390/computers14040136>



## Article

# Digital Twins and the Stendhal Syndrome

Franco Niccolucci \*  and Achille Felicetti \* 

Fondazione PIN, Piazza dell'Università, 59100 Prato, Italy

\* Correspondence: franco.niccolucci@gmail.com (F.N.); achille.felicetti@gmail.com (A.F.)

**Abstract:** The “Stendhal Syndrome” mentioned in the title refers to the first (early 19th century) documented perception of the role of intangible aspects in characterising cultural heritage. This paper addresses the semantic organisation of data concerning the digital documentation of cultural heritage, considering its intangible dimension in the framework of Digital Twins. The intangible component was one of the aspects motivating the need of setting up the Heritage Digital Twin (HDT) ontology and its extensions, published in a series of papers since early 2023. In this paper, we analyse how places, persons, and things may give value to a heritage asset, being linked to and supporting its intrinsic cultural significance. This development stems from the consideration of heritage studies and research carried out by scholars and organisations such as UNESCO and ICOMOS, which underline the paramount role of the intangible component in defining heritage assets. The paper then expands the previous semantic structure of the Heritage Digital Twin ontology as concerns the intangible aspects of a heritage asset, extending the HDT concepts by defining new classes and properties related to its intangible component. These are discussed in various cases concerning places, monuments, objects, and persons, and fully developed in examples.

**Keywords:** heritage digital twins; intangible cultural heritage; data organisation and semantics



Academic Editor: Paolo Bellavista

Received: 24 February 2025

Revised: 31 March 2025

Accepted: 3 April 2025

Published: 6 April 2025

**Citation:** Niccolucci, F.; Felicetti, A. Digital Twins and the Stendhal Syndrome. *Computers* **2025**, *14*, 136. <https://doi.org/10.3390/computers14040136>

**Copyright:** © 2025 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) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The “Stendhal Syndrome” mentioned in the title is the medical name given to a psychosomatic condition affecting some individuals when they visit locations of exceptional beauty or great historic importance. It is the same deep emotion that most educated people feel, but at a lesser degree, when visiting heritage masterpieces or historic places. The name derives from a report by Stendhal, pseudonym of the 19th century French writer Marie-Henri Beyle, the author, among others, of the famous novels *Le Rouge et le Noir* (1830) and *La Chartreuse de Parme* (1839). Reporting about a travel in Italy made in 1817 in his book *Rome, Naples et Florence* (1826) [1], he states that when visiting the Church of Santa Croce in Florence, he was deeply touched by the funerary monuments of so many great men, such as the artist Michelangelo, the astronomer Galileo Galilei, the opera composer Gioacchino Rossini, and many other famous Italians. Several of these tombs are not exceptional artworks, but the buried persons give them a special value: so, in this case it is history that makes heritage. Stendhal reports that he was not only impressed by such historical presence: “*I was in a sort of ecstasy, from the idea of being in Florence, close to the great men whose tombs I had seen. Absorbed in the contemplation of sublime beauty. . . I reached the point where one encounters celestial sensations. . . Everything spoke so vividly to my soul. . . I had palpitations of the heart. . . Life was drained from me. I walked with the fear of falling*”. The impression gave him physical symptoms, which are nowadays called Stendhal syndrome.

Stendhal quotes a contemporary poem by the Italian poet Ugo Foscolo who noted the exceptional value of the presence of the memory of so many notable persons in the Santa Croce church.

The poetic perception of the importance of the intangible background of monuments and sites, as described above, was not immediately recognised as important by heritage professionals, especially in Europe. Their attention focused on the tangible aspects, which were analysed and classified according to styles and artistic or architectural movements. Thus, the tombs mentioned by Stendhal were studied according to their shape and appearance, and not to the person who was buried there. Stendhal's reaction was considered as a syndrome, something between a medical/psychological condition and the extreme sensibility of a poet. Thus, the style classification of heritage assets became a scholar's approach to study tangible objects. This statement does not mean that this approach was negative: it was just partial and revealed as insufficient especially when applied to heritage outside Europe. Nevertheless, it led to important achievements as the *Athens Charter* (1931) and the *Venice Charter* (1964) [2], produced by ICOMOS (International Council on Monuments and Sites), which dictated important principles for the conservation and restoration of heritage property. However, the consideration that heritage assets, whatever their nature, consist of the inseparable combination of tangible aspects with intangible ones emerged soon. It is well described in a keynote lecture significantly titled "*The interdependency of the tangible and intangible cultural heritage*" given by Mounir Bouchenaki, former Director General of ICCROM and Director of the UNESCO World Heritage Centre, at the ICOMOS 14th General Assembly and Scientific Symposium held in Zimbabwe in October 2023 [3]. The lecture may be summarised in its final statement: "*Both the tangible and the intangible heritage rely on each other when it comes to understanding the meaning and importance of each*". The *Australia ICOMOS Burra Charter* [4], first adopted in 1979 and revised in 2013, had already addressed similar issues, extending the *Venice Charter* by providing standard guidelines for heritage conservation practice of places, which may have tangible and intangible dimensions. Conservation of places involves their cultural significance, i.e., the aesthetic, historic, scientific, social, or spiritual values for past, present, or future generations. Finally, the report *Innovation in Cultural Heritage Research* [5], published by the EU DG Research and Innovation in 2018 and illustrating the foundations of the 2018 European Year of Cultural Heritage, expressed in detail the same concept, applying them to the digitisation policy of the European Union.

The above-mentioned Burra Charter [4] starts with a definition of places as follows: "Place means a geographically defined area. It may include elements, objects, spaces and views. Place may have tangible and intangible dimensions". Places have cultural significance, i.e., "aesthetic, historic, scientific, social or spiritual value for past, present or future generations", which is "embodied in the place itself, its fabric, setting, use, associations, meanings, records, related places and related objects". This statement redefines Stendhal's sentiment as the instinctive recognition of the cultural significance of the Santa Croce tombs rather than a psychological disease (a syndrome): in sum, it is the perception of the inseparable relationship between the tangible and intangible dimensions of heritage assets. It is not surprising that the Australia ICOMOS Charter proposed this extension of the previous concept of heritage asset to places: the mountain Uluru, a UNESCO World Heritage site also known with its English name Ayers Rock, is a sacred place for the Pitjantjatjara, the aboriginal people of the area, but it shows no features of what a perspective limited to tangible aspects would recognise as a heritage asset.

In the following sections, we will show how the above considerations substantially impact the digitalisation of cultural heritage, i.e., the documentation of cultural heritage by digital means aimed at its study and research, conservation, restoration, and valorisation

towards a wide public. While this consideration is acknowledged—at least instinctively—by heritage scholars and professionals, it is less common in communicating heritage. Ignoring it may lead to disrespectful tourism practices and ultimately to overtourism, refusal of visitors by the indigenous population, and damages to heritage unconsciously caused by unaware tourists, who sometimes confuse heritage places with theme parks. Also heritage data collections often disregard this intangible dimension.

In the present paper we will show how this perspective can be captured by a Digital Twin approach in the framework of digitalised cultural heritage and stored into its data. This is not only an improvement in the semantic precision of data characterisation, but also the answer to a precise need as manifested in the above-mentioned documents: without it, the Digital Twin would lack an important component of its structure.

## 2. Digital Twins Today

Digital Twins (DT) made their appearance several years ago as a “living model” for NASA. Since then, they have come a long way, becoming nowadays a popular simulation model in many industrial sectors: Google Scholar mentions 23,000 related papers published since 2024, more than a thousand of which have appeared in January 2025. A recent book [6] synthetically describes different DT applications, while a recent survey paper [7] surveys and summarises industrial applications contributions. It defines a DT as “*comprised of three components*”:

- (i) *Physical twin: A real-world entity (living/non-living) such as part/product, machine, process, organisation, or human, etc.*
- (ii) *Digital twin: The digital representation of the physical twin with the capability to mimic/mirror its physical counterpart in real time.*
- (iii) *Linking mechanism: The bidirectional flow of data between the two which operates automatically in real-time”.*

With this definition, DTs continuously interact with the real-world, receiving inputs about the current status of the physical twin, and returning outputs about actions to be undertaken in the real-world or expected status of the physical twin. Thus, the focus is on simulation and action: the Digital Twin is designed to reproduce as faithfully as possible the relevant aspects of the physical twin, the real-world object it simulates. Information about other aspects may be disregarded: for example, the Digital Twin of a car to be used for autonomous/driverless driving may ignore aspects of the physical twin, which are irrelevant for the intended purpose, such as the fabric of the seats, the car colour, the owner’s gender and first name, and so on: the intended simulation guides the choice of the details to be incorporated in the Digital Twin and consequently its data structure.

On the other hand, DTs currently used for cultural heritage almost always refer to the physical component of the heritage asset. Three-dimensional models are the basis for such DTs, and in quite a few cases, they are considered the DT, with additional annotations attached to it [8]. In this approach, the DTs are built on point-cloud models produced by 3D scans. Another research thread avails of HBIM, the heritage extension of BIM (Build Information Modelling) [9], a documentation system created to assist the design of new buildings and the planning of related works. HBIM is based on a CAD procedure to reproduce the object shape and on a database containing the related information such as materials, services, and so on. An extensive and thorough analysis of the use of Digital Twins for built cultural heritage conservation is contained in [10].

In the authors’ opinion, both such approaches have shortcomings. In the “augmented objects” approach, information is attached to specific points or regions of the 3D model. While there is no restriction on the attachments, and thus they might incorporate any kind of information including immaterial aspects (although in practice this never happens),

there are two major issues: the connection between the “attachments” and the 3D model is unstable, as it may change when transforming the model, for example, by reducing its size via decimation or creating a different mesh via point interpolation. The second issue is that cross-referencing the information across different assets is (almost) impossible: for example, an attachment stating “this part of the object is made in wood” would refer only to a specific object, the one modelled. Searching for all objects with a part made in wood is impossible or at least is not envisaged as relevant because the focus of the investigation is on a single heritage object, the one modelled in 3D. Regarding HBIM models, extending to heritage-related data the original BIM database, designed to assist in the construction of new buildings is cumbersome and, in most cases, impossible. Thus, the related information is severely limited and, so far, never includes, even tentatively, any reference to the intangible component of the asset. Moreover, it works properly only for buildings, and its applicability to movable objects, e.g., statues, or to spaces as intended in the Burra Charter, is questionable and actually never attempted.

It is to overcome these critical aspects that we introduced in previous papers [11,12] what we called the Heritage Digital Twin (HDT) as the holistic assemblage of all the information pertaining to a heritage asset, intangible elements included. For the sake of completeness, the next section includes a short summary of the main features of the HDT ontology. The HDT is a compatible extension of the CRM (Conceptual Reference Model) ISO 2117:2023 standard [13] used for the documentation of cultural heritage, maintaining semantic interoperability with the heritage documentation systems that adhere to that standard—in practice all the good ones.

If data should flow bidirectionally between the Digital Twin and the physical one, the HDT has been extended [14] to the Reactive Heritage Digital Twin (RHDT), which incorporates in its schema *sensors*, receiving inputs from the real-world in real time, and *activators*, which push actions in the real-world. A very simple example consists of sensors receiving input from IoT devices, for example, a temperature measurer, which activates a fire extinguisher according to predefined rules. The activation is triggered by a dedicated *decider*, a DT component that compares the measured temperature as transmitted to it by the sensor with pre-determined thresholds and orders the fire extinguisher to start action if the temperature is beyond a specific value. The *decider* may also avail of AI-based procedures to take complex decisions [15].

### 3. An Overview of the Heritage Digital Twin (HDT) Ontology and of Its Extension to Reactive HDT

The Heritage Digital Twin (HDT) ontology is designed to address the critical aspects outlined above. As a compatible extension of the CRM, the model is intended to represent a complete solution for managing and interconnecting the broad spectrum of data that forms the informational core of a Heritage Digital Twin. The ontology provides a high degree of interoperability in cultural heritage documentation and analysis, allowing for alignment with other research domains such as archaeology, art history, and architecture. It introduces a series of entities designed to express all available cultural and scientific documentation in a standardised format. Furthermore, it captures the dual nature of cultural heritage, encompassing both tangible and intangible aspects, and provides a mechanism for dynamically documenting and analysing their relationships. The HDT ontology is fully described in [11,12]. In the following, as is usual for ontologies, classes are identified by one or more capital letters followed by a number and a name; a similar approach is used for properties, identified by one or more capital letters followed by a number. The HDT and its extensions use the letters HC for the classes and HP for the properties.

Key classes in the model include *HC1 Heritage Entity* and *HC3 Tangible Aspect*, used to describe real-world entities, and an *HC2 Heritage Digital Twin* class meant to represent cultural Digital Twins as informative digital replicas. Other classes, such as *HC4 Intangible Aspect*, *HC5 Digital Representation*, and *HC9 Heritage Activity*, are employed to semantically model the intangible side of cultural entities and their documentation, along with specifically designed properties to articulate complex relationships with the places, people, objects, and events that constitute their historical and cultural significance. The HDT ontology also includes classes for modelling digital iconographic and multimedia representations, digital documentation elements, and activities related to the study and digital reproduction of cultural entities. It also provides features to represent the information relating to stories and narrations derived from documentation or oral tradition.

The HDT model was subsequently extended to also deal with the reactive part of the Digital Twins, consisting of the network of sensors, deciders, actuators, and other IoCT devices that ensure complex interactions with the real-world. Key classes introduced in the resulting R(eactive)HDT ontology model [14] include *HC9 Sensor*, *HC10 Decider*, and *HC11 Activator*, designed to model the IoCT capabilities of recording conditions, analysing input from various sources, and generating output instructions to trigger informed actions. In addition, the RHDT ontology provides specific classes to model aspects such as the placement of sensors, the software operating them, the measurements they perform, the signals generated, and the people responsible for ensuring the safety of cultural assets. Finally, a further extension to model AI components was analysed in [15].

Within the framework of the RHDT ontology, the representation of intangible heritage represents an innovative aspect that allows the model to move beyond the mere cataloguing of physical attributes to embrace the sphere of cultural practices, historical events, and traditions. To this end, the *HC4 Intangible Aspect* class serves as a cornerstone, a semantic space designed to capture the very essence of intangible heritage, allowing it to reside within the Digital Twin as a fundamental and interwoven component of a comprehensive representation. The class is conceived as a template for the documentation of traditions and events, a generic representation that embodies the essence of cultural manifestations, since it acknowledges the continuity and adaptability of intangible heritage and its capacity to evolve while retaining its core identity. The connection between the tangible and intangible is then thoughtfully articulated through the *HP5 has intangible aspect* property, a bridge that semantically links the physical dimension of a heritage asset to the intangible values it embodies. This property is crucial for understanding the full cultural context of a tangible object, informing the Digital Twin that its relevance is not solely determined by its physical properties, but also by the traditions, stories, and practices it represents.

Further elaborating on this connection, the *HP7 is manifestation of* property identifies a reciprocal link to denote the instances in which an intangible aspect is made manifest. This interplay between specific events and their more abstract counterparts is further refined by the introduction of the property *HP6 has manifestation event* property, which allows for a more direct connection between the abstract representation of an intangible aspect and its concrete expressions in time and space. The ontology also expands to the realm of narratives through the use of the Narrative Ontology (NOnt) introduced in [16], whose classes are employed to represent stories as components of intangible heritage. These are not just records of past events but are understood as dynamic constructions contributing to the social and cultural significance of a heritage entity. This approach makes the RHDTO capable of highlighting the ways in which these narratives are expressed and transmitted, recognising the importance of different modes of communication in preserving and conveying cultural knowledge. It ensures that intangible aspects are not treated as mere addenda but are semantically interwoven with the tangible dimensions of cultural



heritage, allowing for a more comprehensive and interconnected representation of the cultural assets within the Digital Twin.

#### 4. The Concepts of Place and Space in the CRM Ontology

As previously noted, the concept of place as defined in the Burra Charter is of fundamental importance for cultural heritage information. At first sight, “place” seems a very simple concept and is considered as such in the CRM [13] standard. The CRM uses the letter E for the classes and P for the properties. All the CRM definitions and concepts are documented in dedicated sections of the CRM web site [13], which can be searched for the item concerned.

The CRM defines the class E53 Place as “extents in space, in particular on the surface of the Earth, in the pure sense of physics: independent from temporal phenomena and matter”. It works well in statements such as “The Leaning Tower is located in Pisa”, where Pisa is the place. There are however limitations to a fair use, as shown by the following—somehow humorous—example. The famous Windsor Castle, a royal residence, is located at Windsor, in the English county of Berkshire. Its construction dates back to the 11th century. So, its place is “County of Berkshire”. The Heathrow airport is located in the London Borough of Hillingdon, a borough of the Greater London. It started operations in 1929 as a small airfield. Its place is therefore “Borough of Hillingdon”. The two locations have a common border, and the distance between the castle and the airport is actually about 9 miles (15 km) by road, and 6.7 miles (10.7 km) if measured as the crow flies. An ignorant (or perhaps just humorous) tourist asked on the web “why they built the Windsor Castle so close to the Heathrow airport”. The question was not unfounded, because the concept of place does not bring the notion of time and the two places, the Windsor Castle one and the Heathrow Airport one, are very close to each other. To resolve this issue (and perhaps also more serious ones) is why the CRM introduced the more precise class E92 Spacetime Volume, which comprises four-dimensional point sets (volumes) in physical spacetime, i.e., the extent of a material phenomenon in space and time. If the location of the above two spaces (Windsor Castle and Heathrow airport) is considered as a spacetime volume, the Windsor one extends in the space occupied by the castle and in the timeframe from circa 1200 AD onwards, while the Heathrow one extends in the space occupied by the airport and in the period from 1929 AD onwards. In both cases, the spaces occupied vary in time, for example, as concerns Heathrow, where the growth of the airport increased the space occupied by it at precise dates; however, the two locations remained close to each other only since the beginning of existence of the Heathrow airport in 1929. Thus, the concept of spacetime volume greatly improves the precision of the information and avoids misconceptions as the one mentioned in the above joke, or in other more serious ones, for example, when referring to “Rome” without specifying if it is intended the Roman age Rome, the medieval one, or the present time capital of Italy: the spacetime volume resolves this ambiguity.

One might ask if such spacetime volume concept may also describe the concept of place as defined in the Burra Charter. As already mentioned, the Burra *Place* is defined as *a geographically defined area, with tangible and intangible dimensions, having cultural significance, i.e., aesthetic, historic, scientific, social or spiritual value*: none of these elements is referred in the CRM concept of spacetime volume, which incorporates only measurable aspects such as space and time, both somehow material as they belong to physics. A solution is proposed in the next section.

#### 5. Heritage Place as a Conceptual Object

Besides being insufficient to manage the Burra Charter concept of (heritage) place, the CRM approach to places also differs from the one adopted in gazetteers, notably in the

*Pleiades* gazetteer. *Pleiades* [17,18] is probably the most extensive and complete gazetteer of ancient places, mainly conceived for historians and archaeologists. It has extensive coverage for the Greek and Roman world, the Ancient Near Eastern, Byzantine, Celtic, and Early Medieval geography. *Pleiades* distinguishes between *places*, which are constructed by human experience, and *locations*, i.e., *concrete*, objective spatial entities, i.e., a portion of space on the Earth. This differs from the CRM definition of *E53 Place*, i.e., “an extent in space, in particular on the surface of the Earth, in the pure sense of physics” (which rather corresponds to the *Pleiades* location), as well as from the already mentioned CRM definition of *E92 Spacetime Volume*. *E53* also does not fit with the Burra definition of place as mentioned above. Thus, a new class should be introduced to capture the cultural and historical significance of places.

This distinction reflects the dualism between two foundational categories of spatial objects, defined as follows. Smith [18–22] introduced the distinction of two foundational types of boundaries of physical entities, on which a top-level distinction between spatial entities is based on *bona fide boundaries*, i.e., *natural* or mind-independent boundaries, which are physical boundaries in the objects themselves that exist independently from human perception, and correspond to the CRM approach; and *fiat boundaries*, i.e., *artificial* or mind-dependent boundaries, which are non-physical boundaries, that depend on human decision or considerations and are the products of mental activities, i.e., the approach of gazetteers and heritage-related space definitions such as the Burra Charter.

However, the CRM has a class which in our opinion may be suitable to address the semantic issues contained in the Burra definition of place: the class *E28 Conceptual Object*, which comprises “non-material products of our minds and other human produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication (. . .) They exist as long as they can be found on at least one carrier or in at least one human memory. Their existence ends when the last carrier and the last memory are lost”. Moreover, this *E28* has a subclass, the *E89 Propositional Object*, a refinement of it, defined as comprising “immaterial items, including but not limited to stories, plots, procedural prescriptions, algorithms, laws of physics or images that are, or represent in some sense, sets of propositions about real or imaginary things and that are documented as single units or serve as topic of discourse”. Such refined class improves the precision of the classification as it excludes other concepts such as *Type* and *Appellation* that also belong to *E28*. Availing of this *E89*, we define a new subclass of *E89*, called *HC15 Conceptual Space*, as “the complex of information referring to an extent of space and the related aesthetic, historic, scientific, social or spiritual value for past, present or future generations embodied in the place itself, its fabric, setting, use, associations, meanings, records, related places and related objects”. It may be noted that this definition copies almost verbatim the Burra Charter definition of cultural significance and of place. It is unfortunate that the term “place” has a meaning in the CRM as in the definition of the class *E53 Place*, which concerns “extents in space, in particular on the surface of the Earth, in the pure sense of physics”, while elsewhere “place” refers to a much wider concept as in the Burra Chapter. Thus, to avoid misunderstandings when operating within the CRM semantic framework, we prefer to use the term “space” in the new class name.

This newly defined class better represents the Burra Chapter definition of places and the *Pleiades* concept of place. It may refer to an ample variety of cases: imaginary “places” such as Atlantis or Peter Pan’s Neverland, as well as places set up by human feelings but linked to a physical space. The latter is the case of Uluru/Ayers Rock, where a whole community refers to it as a sacred space, and it applies to many other circumstances, as shown in the examples described below. A conceptual space may also have a *location*, a spacetime volume linked to the conceptual place, for example, in the Uluru/Ayers Rock case, the geographical region comprising the mountain and the time span extending from



the beginning of the cults related to the place until today. The (bona fide) borders of the location may be fuzzy, which is perfectly acceptable according to the CRM scope note (i.e., the definition) of spacetime volume.

In sum, to fit with the needs of cultural heritage documentation, we think that it is necessary to introduce a new spatial concept, the conceptual space defined above. It has fiat borders, i.e., borders defined by human decisions or considerations and are the products of mental activities. Accordingly, this new conceptual space is a specialisation (a subclass) of the CRM E89 Propositional Object.

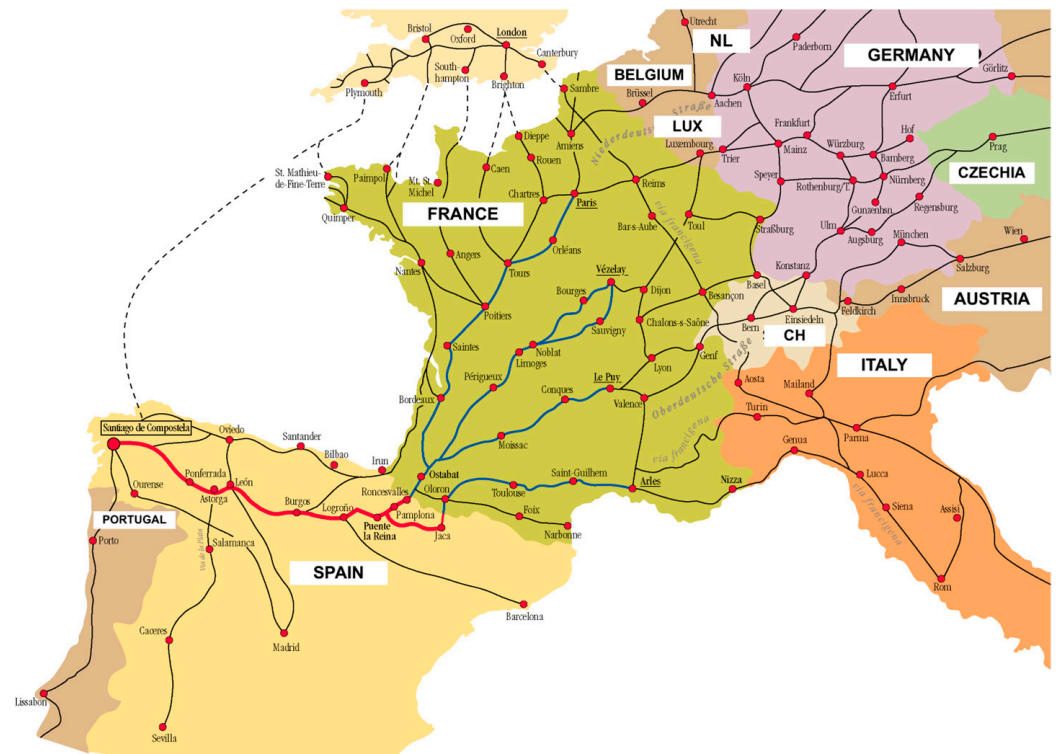
Examples show that conceptual spaces belong also to the cultural heritage of regions like Europe, where the tangible component has had an overwhelming importance until recently.

The conceptual space models the cause of Stendhal syndrome. This feeling derives by regarding the Santa Croce church as the complex of the considerations about the great humans buried in it, rather than the physical building, which is instead the location of such conceptual space. This approach further explains the title of the present paper.

Another clear example is provided by the *Heritage Routes* promoted by the Council of Europe [23,24], which show a “journey through space and time” demonstrating “a shared and living cultural heritage” among different countries and cultures of Europe. The subjects of such routes range from music (the *Mozart Ways*) to agriculture (the *Route of the Olive Tree* and the *Iter vitis*), arts and crafts (*Ceramics*, *Art Nouveau*, *Impressionisms*), artists and scientists (*Leonardo da Vinci*, *Le Corbusier*, *Alvar Aalto*), and more.

We consider one of the most famous of these routes, the pilgrimage itinerary of the “*Camino de Santiago*”, the *Santiago de Compostela Pilgrim Routes* [24], also known as Saint James Way, as it well expresses the above-mentioned mixture of tangible and intangible aspects. It concerns the Christian pilgrimage to the tomb of Saint James in Santiago de Compostela (Spain), which started in the Middle Ages and is still alive today. The Spanish and French parts of the *Camino* are also included on the UNESCO World Heritage List. Completing the pilgrimage guarantees a plenary indulgence to pilgrims. As a result of this pilgrimage, a rich heritage was formed in time. Tangible heritage such as places of worship, hospitals, accommodation facilities, bridges, as well as non-tangible heritage in the form of myths, legends, and songs are present along the Santiago Routes. The religious aspect is also part of this intangible component. The *Camino* location is rather fuzzy as there is no specific and detailed indication on a precise route to reach Santiago: the pilgrimage’s route is rather a category of the spirit rather than a place in the CRM sense. Of course, the general definition of the pilgrimage must abstract from individual choices and feelings, and document opportunities or typical activities along way. If such activities may be carried out at specific locations or buildings (a church or chapel, for example), these should be defined individually as heritage assets, possibly worth owning their own Digital Twin. In conclusion, the *Camino* comprises the pilgrimage itinerary description, the physical route, the pilgrims’ religious beliefs, how they carry out the pilgrimage, and what they do during it. The *Camino* develops mostly in Western Europe through the Netherlands, Belgium, France, and Spain, with a Portuguese extension, although there are also places in Europe, e.g., in Poland, Latvia, and Italy, which refer to the devotion to the saint. It is worth mentioning that a single, official version of the *Camino*’s route does not exist. The paths have changed over centuries, and the precise itinerary depends on historical sources, regional interpretations, and even personal choices made by pilgrims. Various organisations, including UNESCO, the already mentioned Council of Europe, and local associations such as the European Federation of Saint James Way [25], provide descriptions and maps of the routes, each with its own perspective and level of detail. For example, UNESCO highlights the French and Spanish routes recognised as World Heritage

Sites, while other institutions like the Council of Europe and the European Federation of Saint James Way, emphasise the *Camino's* cultural and historical significance as a European Cultural Route or focus on providing detailed guides and practical resources for pilgrims. Figure 1 aims at providing a general overview of such routes.



**Figure 1.** Pilgrimage itineraries of the *Camino de Santiago* in Central and Western Europe. The itineraries in blue (France) and red (Spain) indicate the part of the *Camino*, which are included in the UNESCO World Heritage list. Credits: map by Manfred Zentgraf, Volkach, Germany, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=1377930> (accessed on 31 January 2025).

This example shows some relevant aspects of the analysis of the “place” general concept. They show that there are places which are not simply “a geographic concept”, as they take their heritage value from beliefs shared by a community. The granularity of this background community, and its importance, are largely discretionary.

As already mentioned, a conceptual space may refer to some extent in space on the Earth defined with a more or less precise definition, its location. The *Camino de Santiago* concerns a broadly defined part of Europe which has fuzzy borders and came to existence in the Middle Ages. Similarly, Uluru/Ayers Rock is a mountain with relatively indetermined borders in Australia and became a sacred place when humans started inhabiting Australia. So, both refer to a spacetime volume. We will represent this aspect with the property *HP15 is spatiotemporally defined*, with *HC15 Conceptual Space* as domain and *E92 Spacetime Volume* as range.

The intangible component of a heritage asset may include other aspects rather than a specific spatial location. While there is always a conceptual element behind it, its reification may concern an object, for example, a religious icon, the relics of a saint, a national flag, a set of memorabilia (“objects that are collected because they are connected with a person or event that is thought to be very interesting” according to the Cambridge Dictionary definition), and so on. In such cases, the object is not relevant per se, but for what it symbolically represents. We will use for this kind of objects the class *HC16 Conceptual Thing* to include both the tangible and intangible component. A conceptual thing is linked to

the object it refers to by the property *HP16 has manifestation*, with domain *HC16 Conceptual Thing* and range *E18 Physical Thing*.

In other cases, what embodies the cultural value is a living being or a group thereof. It may be a person who lived in the past or one who is still alive, or a group. This case will be modelled with the class *HC17 Conceptual Agent*. This concept was introduced in the UNESCO definition of intangible heritage [26]. The property which represents this relationship is defined as *HP17 is embodied in*, with the class *E39 Actor* as the range: Actor, in the CRM context may represent a person or a group (respectively, *E21 Person* and *E74 Group*). Note that the three cases illustrated above, i.e., conceptual space, conceptual thing, and conceptual agent, may refer to the same conceptual object. For example, the *Camino de Santiago* can be regarded as a conceptual space (the itinerary), a conceptual thing (the complex of the heritage physical assets along the way and at destination, for example, the Saint's relics), and a conceptual agent, the saint himself.

## 6. The Extended HDT Model

In the HDT vision, a Heritage Digital Twin is mainly composed of all the available information about the real-world cultural entity it aims to reproduce. In this perspective, documentation is the main source through which the intangible aspect of the replicated cultural entity is witnessed within the HDT. Documentation, thus, needs to be properly addressed to capture the complexity and deep significance of its intangible sphere. To this aim, in the ontology, we managed to improve the features provided by the existing *HC4 Intangible Aspect* class to capture detail about intangible spaces, objects, and people that hold significant cultural and symbolic value in an improved semantic form. In turn, we defined new ways for linking these intangible aspects to their real-world counterparts.

### 6.1. Conceptual Spaces

As already outlined, conceptual spaces exist within documentation as objects of discourse living in a non-spatial and atemporal dimension, while obviously retaining the potential to refer to real places having actual manifestation in space and time. From an ontological point of view, conceptual spaces can be better approximated as propositional objects in the sense of the CRM, i.e., as immaterial sets of propositions about real or imaginary things that, similarly to the CRM conceptual objects, are documented or communicated between persons, and continue to exist on multiple carriers simultaneously as long as they can be found on at least one carrier, or in at least one human memory. We therefore propose a new class for modelling conceptual spaces, tentatively defined as follows:

*HC15 Conceptual Space*

Subclass of: *E89 Propositional Object*

This class comprises non-material, intangible entities that represent places or locations as they exist within the realm of discourse, memory, and cultural practices. They are created by individuals or communities and become objects of discourse about their identity, function, and historical implications. Unlike physical spaces, which have a tangible and geographically defined presence, conceptual spaces have their roots in the HDT documentation, which may include written documents, oral traditions, artistic representations, and collective memories. They typically hold significant cultural and symbolic value, contributing to the collective identity and heritage of a community.

As an example of conceptual space, we can think to the already mentioned sacred space of Ayers Rock, designated out of reverence by the indigenous community and existing within the realm of collective memories and oral traditions (as witnessed, for example, in stories and audiovisual or media documentation) rather than solely in its physical form. The sacredness of this conceptual place is maintained through cultural practices and ceases

to hold symbolic significance if these memories fade, even if the actual place to which this intangible value refers, continues to exist in space and time, together with the physical object (the rock) located there. The Uluru/Ayers Rock and the Heritage Routes examples offer a different set of examples for the conceptual space class.

To describe the relations that may exist between a conceptual space and its actual corresponding spatial manifestation in the real-world, we also introduced a new *HC15* property, defined as follows:

*HP15 is spatiotemporally defined by*

Domain: *HC15 Conceptual Space*

Range: *E92 Spacetime Volume*

This property establishes a relationship between a *conceptual space (HC15)* and the real-world spatiotemporal volume (*E92 Spacetime Volume*) to which it is directly related. In particular, it specifies that a conceptual space, while existing in an intangible, conceptual dimension within the realm of discourse and memory, could have a corresponding manifestation as physical place and timeframe that can be defined in terms of a spacetime volume. This property emphasises the dual nature of spaces, which can exist both as non-material products of the mind and as physical places that can be geographically and historically contextualised.

As an example, Uluru/Ayers Rock *HP15 is spatiotemporally defined by* a portion of space represented in Figure 2. For this example, defining a temporal range may be more difficult as it refers to oral sources. Most studies date back the aboriginal religious beliefs to 10,000 BCE. Properties such as *P70 is documented in* and *P129 is subject of* can be used to establish relationships between a conceptual space (*HC15*) and the documentation (*E31 Document*) in which it is described, mentioned, or otherwise reported within a specific textual, visual, or multimedia source, thus highlighting the role of documentation in sustaining and transmitting its cultural and historical value.



**Figure 2.** Map of Uluru-Ayers Rock area (picture from OpenStreetMap. CC-BY-SA 2.0. <https://openstreetmap.org/copyright> accessed on 31 January 2025).

Documentation about Uluru/Ayers Rock is more difficult to exemplify, as it resides in the minds and beliefs of the local native population. Instead, the Heritage Routes may rely on a large set of documents, from religious to heritage related ones, the latter being connections to other related heritage assets.

## 6.2. Conceptual Objects and People

As previously noted, cultural objects and people can also show an intangible dimension, going beyond their physical nature. As in the case of conceptual spaces, these conceptual entities exist primarily within the realm of discourse and are sustained through various forms of documentation. Therefore, they need to be defined with new specific classes and properties to describe and associate them with the corresponding real-world entities they refer to. For instance, physical objects that are the focus of specific cults or historical relevance (e.g., orthodox icons, relics of saints, objects belonging to relevant historical figures) possess a conceptual side deriving from mentions or descriptions of them in written or oral documentation. Similarly, historical figures celebrated or remembered through documents, narratives, and artistic representations should be distinguished from the related real historical persons who lived and actively participated in historical events. In this perspective, the *E18 Physical Thing* and *E39 Actor* classes of the CRM seem inadequate to model entities of this kind, since they refer exclusively to real objects and people rather than conceptual or symbolic representations. Also in this case, we propose to model these intangible entities by means of subclasses of *E89 Propositional Object*, as we did for the conceptual spaces defined above. In particular, we propose the following two new classes:

### *HC16 Conceptual Thing*

Subclass of: *E89 Propositional Object*

This class encompasses entities that acquire cultural, symbolic, or intellectual significance independently of their material form. These include conceptual representations of real-world objects or legendary artefacts expressed through documentation. Unlike physical objects, their identity is shaped by discourse, representation, and transmission across different media and traditions, and are instantiated through written accounts, oral traditions, iconographic depictions, or digital representations, ensuring their persistence within cultural memory. Their identity remains intact even if their material manifestations are lost or altered, provided they continue to be referenced, interpreted, or reconstructed within the cultural discourse.

An example of this class is the complex of funerary monuments in the Santa Croce church in Florence, not considered only as artistic artefacts, but as the subject of a discourse about notable Italians.

### *HC17 Conceptual Agent*

Subclass of: *E89 Propositional Object*

This class refers to historical, legendary, or fictional individuals whose identity and significance transcend their physical existence. These figures are constituted through collective memory, literary narratives, artistic representation, and ideological constructs rather than direct biological lineage or material evidence. Conceptual persons may include deified heroes, mythical founders, or fictional protagonists, as well as documented representations of historical individuals when expressed within specific narratives or symbolic frameworks. Their continuity relies on being referenced within traditions, texts, or artistic expressions, allowing their legacy to persist even in the absence of tangible proof of their existence.

The figure of Stendhal exemplifies very well the nature of what a conceptual agent could be, existing simultaneously as a historical figure and a literary construct. While the man Stendhal lived and wrote, the conceptual Stendhal persists through the centuries,



shaped by literary tradition, cultural memory, ideological reinterpretation, and the famous Syndrome named after him.

To articulate the relationships of conceptual things and conceptual agents with their real-world counterparts, the new, extended version of the RHDT ontology includes a set of new properties. Their scope notes are described below.

*HP16 has manifestation*

Domain: *HC16 Conceptual Thing*

Range: *E18 Physical Thing*

This property establishes a relationship between a conceptual thing (*HC16*) and the physical thing (*E18*) that serves as its tangible manifestation. It can be used to associate conceptual entities, such as legendary artefacts, symbolic objects, or documented representations of physical things, with their corresponding material instances. This linkage enables a structured understanding of how conceptual entities interact with the physical world, indicating that while conceptual things exist primarily in discourse and documentation, it may have one or more physical embodiments that materialise its symbolic, cultural, or intellectual significance.

*HP17 is embodied as*

Domain: *HC17 Conceptual Agent*

Range: *E39 Actor*

This property establishes a relationship between a conceptual agent (*HC17*) and an Actor (*E39*) who is recognised as its real-world counterpart. The property enables the documentation of relationships between conceptual agents and historical figures, such as cases where legendary characters are linked to real individuals or where cultural traditions associate a symbolic identity with a known person, thus facilitating the study of historical reinterpretations, attributions, and the evolution of conceptual agents in collective memory.

### 6.3. Empowering the *HC4 Intangible Aspect Class*

As we stated at the beginning of this section, the *HC4 Intangible Aspect* class is a key component of the RHDT ontology and plays a central role in capturing the non-material dimension of cultural heritage, incorporating it in its digital representations. In this regard, *HC4* functions as an overarching category that incorporates specific types of conceptual entities, such as spaces, objects, and people, that exist within documentation and collective memory, ensuring that they are adequately represented as living elements of discourse. To express this ontological relationship, we introduce the *HP18 incorporates property*, defined as follows:

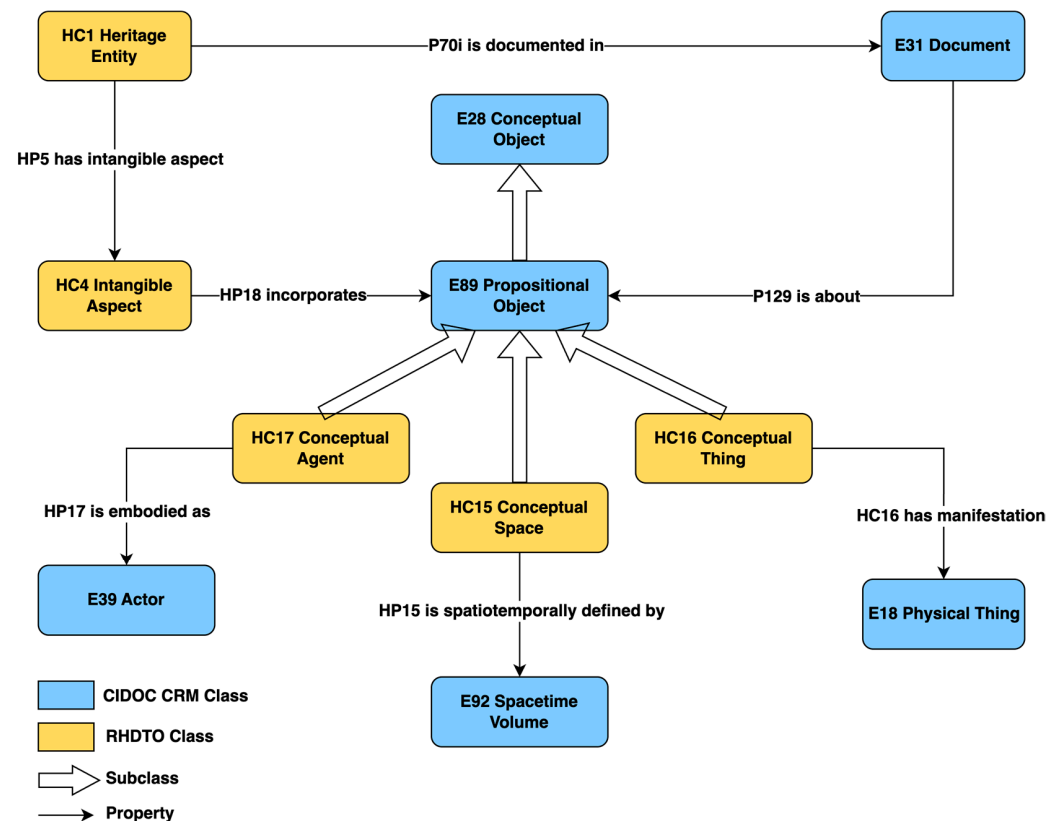
*HP18 incorporates*

Domain: *HC4 Intangible Aspect*

Range: *E89 Propositional Object* (includes *HC15*, *HC16*, *HC17* as a superclass of such classes)

This property establishes a relationship between an intangible aspect (*HC4*) and one or more propositional objects (*E89*) that contribute to its conceptual structure. In particular, it provides a semantic way to describe how cultural narratives, symbolic representations, and conceptualised historical entities contribute to the formation of intangible aspects of cultural heritage. This property indicates that an intangible aspect, as represented in documentation and cultural memory, can subsume or encompass specific conceptual entities such as conceptual spaces, conceptual things, and conceptual agents, effectively linking them to broader intangible heritage dimensions.

The new classes and properties described in this section are represented in the diagram in Figure 3.



**Figure 3.** Semantic representation of new RHDTO intangible heritage classes.

#### 6.4. An Example: Stendhal in Santa Croce

In this section, we will develop the semantic representation of the example of the Santa Croce church and Stendhal syndrome mentioned above. The visit of Stendhal to the Basilica of Santa Croce in Florence, which we cited in the opening of this paper, provides a vivid illustration of the interplay between physical heritage and its conceptual extensions in narrative and cultural memory. The story, reported in his book *Rome, Naples et Florence*, has in fact the power to transform this historical monument into a conceptual space (HC15), existing independently of its tangible architectural structure (HC3 *Tangible Aspect*). Santa Croce thus becomes an object of discourse shaped by the literature (E31 *Document*), which serves as carriers of its intangible significance. If we consider that in addition to Stendhal's work, there are obviously many other literary, historical, and artistic documents in which it is possible to recognise the same conceptual space of Santa Croce, it is easy to realise how this conceptual element becomes a central node of its intangible aspect, representing an aggregating node within the documentation of the Digital Twin (the HDT) of this famous monument. We mention also the fact that Stendhal's feelings of reverence were mentioned in the previous literary work *I Sepolcri* (*Sepulchres*) by the Italian poet Ugo Foscolo [27], some verses of which are also verbatim quoted in Stendhal's text.

Ontologically, the relationship between the conceptual and the spatiotemporal manifestation of the church can be formalised through the property *HP15 is spatiotemporally defined by*, linking the conceptual space of Santa Croce to its real-world spacetime volume (E92), i.e., its physical presence in Florence, existing over a defined period of time (see Figure 4).

Within this space, Stendhal saw and described the tombs of Michelangelo (see Figure 5), Galileo, and other celebrated individuals. These tombs are obviously physical entities (E18), made of marble and still visible today. But in the act of literary representation, they assume a second identity, becoming conceptual things (HC16) linked to the material monuments,



but persisting beyond their physical form. The property *HP16 has manifestation* captures this duality, linking the conceptual thing (*HC16*) of Michelangelo's and Galileo's funerary monuments to the actual physical objects (*E18*) located in Santa Croce. This illustrates how material objects can be reinterpreted, documented, and reimagined across different media, to reinforce their conceptual dimension.



**Figure 4.** The Santa Croce church in Florence (credit: By Diego Delso, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=125634027>) (accessed on 31 January 2025).



**Figure 5.** The tomb of Michelangelo in Santa Croce (credit: By Diego Delso, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=125633750> accessed on 31 January 2025).

The figures commemorated within these tombs undergo a similar transformation. Michelangelo and Galileo were historical persons (*E39 Actor*) who lived, acted, and influenced their world. However, as they are reintroduced in Stendhal's account, they take on a different ontological status, i.e., the intangible value of the memory, an aspect that can be modelled by means of the *HC17 Conceptual Agent* class. Indeed, this is precisely the memory of the great men of the past that triggered in Stendhal his famous syndrome. The property *HP17 is embodied as* allows to link the conceptual representations of Michelangelo and Galileo to their historical counterparts.

As already mentioned, another important aspect of this ontological modelling is represented by Stendhal himself. As a real historical person (*E39 Actor*), he physically visited Santa Croce and recorded his impressions. However, within his own narrative, he also created an intangible version of himself, a conceptual agent (*HC17*) who narratively experienced and interpreted Santa Croce, filtering reality through subjective reflection. The distinction between author and narrator highlights the role of documentation in shaping how places, people, and objects are understood, generating the intangible value of sites and monuments, as in this case occurs to Santa Croce.

Finally, the church, the tombs, and the historical figures all exist as intangible aspects (*HC4*) of the cultural heritage entity (*HC3*), relating the physical world to its conceptual representations. The property *HP18 incorporates* expresses this connection, showing how the intangible aspect (*HC4*) of Santa Croce encompasses its conceptual spaces, things, and agents, preserving its significance in the context of the documentation of the Digital Twin of the church, in which it is described not only as a merely physical structure but, highlighting its immaterial aspects, as a conceptual entity that functions as a repository of historical values, artistic memory, and cultural significance.

The diagram in Figure 6 illustrates with a graph the ontological representation of the concepts expressed above by means of the classes and properties of the RHDT ontology model. To keep the diagram readable by a human, only some of the relationships and classes are shown as examples of a more complex structure: this problem does not exist, of course, when the semantic graph is loaded into a computer using, for example, RDF encoding.

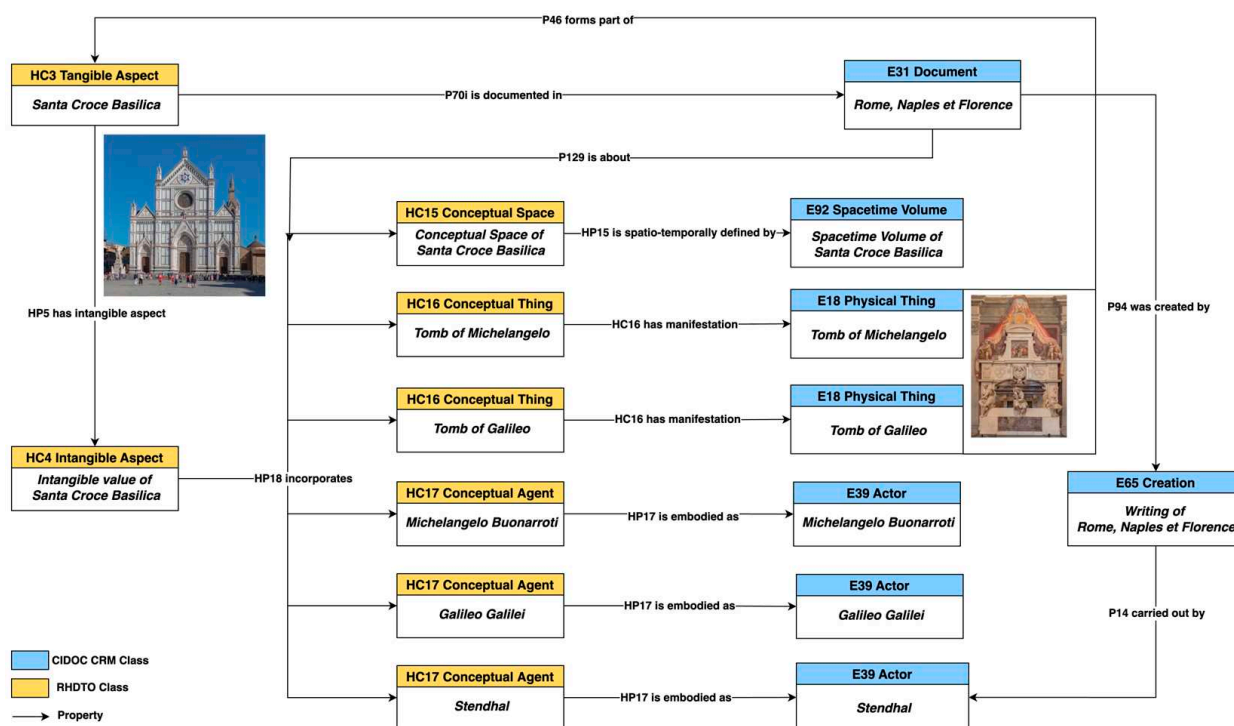


Figure 6. RHDTO representation of Stendhal's visit to Santa Croce.

## 7. Discussion and Further Work

In this paper, we have analysed several examples of places that must actually be considered as conceptual objects, such as Uluru and the Santiago Routes. Further work will analyse more examples of HC15 Conceptual Space with reference to relevant concepts used in cultural heritage practice. One significant example could be the concept of landscape as defined in the Landscape Convention [28], which combines physical, cultural, living, and human elements reflecting a common identity.

Regarding the above-mentioned Santa Croce church, there is a clear tangible component which can be modelled as such using the RHDT ontology. Historical elements such as its Gothic character may also be included as additional knowledge about its architectural features. Its Digital Twin could include information to monitor the church elements such as columns, capitals, paintings, and so on, availing of the RHDT (Reactive HDT) ontology [14,15].

On the other hand, Santa Croce owns also an intangible dimension with many components. One is of course the religious component, being a Catholic church. Its relationship with the Franciscan order is related to this aspect. The causes of Stendhal syndrome add a further dimension to its intangible component. All these intangible aspects may be properly documented in the RHDT model using the HC15 Conceptual Space class and the other related classes. The visitor's awareness level is of course subjective: visitors must understand what a church is in the Catholic religion, and probably also the importance of Saint Francis and his reforms. To a deeper understanding of the cultural values related to Santa Croce, which caused Stendhal sentiments, knowledge about the famous people buried in the church sepulchres is moreover required. Properly documenting them in the knowledge base built on RHDT (and its extensions) is a step in this direction.

RHDT also supports historical research and the conservation of the monument by connecting this church (as well as other monuments and artefacts, of course) to the knowledge of conservation activities carried out on other artefacts somewhere else. This topic is the subject of the EU-funded research project ARTEMIS [29], just started on 1st January 2025, which relies on a knowledge base driven by the RHDT ontology and aims at exploring the feasibility of "virtual" artwork and monument restoration founded on the data stored in its knowledge base. In ARTEMIS, the authors of the present paper are in charge of setting up such semantic infrastructure, implementing the RHDT concepts in practice.

Finally, the information contained in RHDT may also support valorisation, being the starting point to develop stories for visitors and providing access to material to compose visual explanations, with the additional advantage that the communication content would be compliant with principles such as the *London Charter* [30–32], which requires that every computer visualisation of cultural heritage be based on scientific bases. We have not explored this opportunity so far, and plan to do so in future research.

The above considerations suggest future work to improve the fit of the semantic description to the needs of heritage documentation. It will be necessary to improve the instruments required to organise space-based information, starting in the present paper with the notion of conceptual space. Since most of the immaterial concepts are based on narration and beliefs, these two areas are a clear target for a further expansion of the RHDT ontology. The CRM framework and its extensions provide the context in which such expansion may be designed in a compatible and interoperable way, demonstrating the power of digitalisation for the comprehension, documentation, study, conservation, and valorisation of cultural heritage. We do hope that the Heritage Digital Twin approach will support these achievements.

**Author Contributions:** Both authors have equally contributed to this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data is contained within the article.

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

## References

1. Stendhal, B. *Rome, Naples et Florence*; Delaunay Libraire, Palas Royal: Paris, France, 1826.
2. The Venice Charter. 1964. Available online: <https://www.icomos.org/en/participer/179-articles-en-francais/ressources/charters-and-standards/157-thevenice-charter> (accessed on 31 January 2025).
3. Bouchenaki, M. The interdependency of the tangible and intangible cultural heritage. In Proceedings of the Keynote lecture at the ICOMOS 14th General Assembly and Scientific Symposium, Victoria Falls, Zimbabwe, 27–31 October 2023; Available online: [https://openarchive.icomos.org/468/1/2\\_-\\_Allocution\\_Bouchenaki](https://openarchive.icomos.org/468/1/2_-_Allocution_Bouchenaki) (accessed on 31 January 2025).
4. The Burra Charter. 2013. Available online: <https://australia.icomos.org/wp-content/uploads/The-Burra-Charter-2013-Adopted-31.10.2013.pdf> (accessed on 31 January 2025).
5. European Commission: Directorate-General for Research and Innovation; Sonkoly, G.; Vahtikari, T. Innovation in Cultural Heritage Research—For an Integrated European Research Policy Publications Office. 2018. Available online: <https://data.europa.eu/doi/10.2777/673069> (accessed on 31 January 2025).
6. Crespi, N.; Drobot, A.T.; Minerva, R. (Eds.) *The Digital Twin*; Springer: Cham, Switzerland, 2023. [CrossRef]
7. Singh, M.; Srivastava, R.; Fuenmayor, E.; Kuts, V.; Qiao, Y.; Murray, N.; Devine, D. Applications of Digital Twin across Industries: A Review. *Appl. Sci.* **2022**, *12*, 5727. [CrossRef]
8. Messaoudi, T.; Véron, P.; Halin, G.; De Luca, L. An ontological model for the reality-based 3D annotation of heritage building conservation state. *J. Cult. Herit.* **2018**, *29*, 100–112. [CrossRef]
9. Lovell, L.J.; Davies, R.J.; Hunt, D.V.L. The Application of Historic Building Information Modelling (HBIM) to Cultural Heritage: A Review. *Heritage* **2023**, *6*, 6691–6717. [CrossRef]
10. Vuoto, A.; Funari, M.F.; Lourenço, P.B. Shaping Digital Twin Concept for Built Cultural Heritage Conservation: A Systematic Literature Review. *Int. J. Archit. Herit.* **2023**, *18*, 1762–1795. [CrossRef]
11. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Digital Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [CrossRef]
12. 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. *arXiv* **2023**, arXiv:2302.07138. [CrossRef]
13. Available online: <https://cidoc-crm.org/> (accessed on 31 January 2025).
14. Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. [CrossRef] [PubMed]
15. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2025**, *10*, 1. [CrossRef]
16. Meghini, C.; Bartalesi, V.; Metilli, D. Representing Narratives in Digital Libraries: The Narrative Ontology. *Semant. Web* **2021**, *12*, 241–264. Available online: <https://www.semantic-web-journal.net/system/files/swj2574.pdf> (accessed on 31 January 2025). [CrossRef]
17. Available online: <https://pleiades.stoa.org/> (accessed on 31 January 2025).
18. Smith, B. Fiat objects. In *Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology*, 11th European Conference on Artificial Intelligence. Amsterdam: European Coordinating Committee for Artificial Intelligence, Amsterdam, The Netherlands, 8–12 August 1997; Guarino, N., Pribbenow, S., Viue, L., Eds.; European Coordinating Committee for Artificial Intelligence: Luxembourg, 1994; pp. 15–23.
19. Smith, B. On Drawing Lines on a Map. In *Spatial Information Theory: Proceedings in COSIT '95*; Frank, A.U., Kuhn, W., Mark, D.M., Eds.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 475–484.
20. Smith, B.; Varzi, A.C. Fiat and Bona Fide Boundaries: Towards an Ontology of Spatially Extended Objects. In *Spatial Information Theory: A Theoretical Basis for GIS*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 103–119.
21. Smith, B.; Varzi, A.C. Fiat and Bona Fide Boundaries. *Philos. Phenomenol. Res.* **2000**, *60*, 401–420. [CrossRef]
22. Smith, B. Fiat Objects. *Topoi* **2001**, *20*, 131–148.
23. Available online: <https://www.coe.int/> (accessed on 31 January 2025).
24. Available online: <https://www.coe.int/en/web/cultural-routes/the-santiago-de-compostela-pilgrim-routes> (accessed on 31 January 2025).



25. Available online: <https://saintjamesway.eu> (accessed on 31 January 2025).
26. Available online: <https://www.unesco.org/en/intangible-cultural-heritage> (accessed on 31 January 2025).
27. Foscolo, U. (1807) *I Sepolcri*. English Translation: *Sepulchres, Parallel Text Translation from Italian* by J.G. Nichols; Alma Classics: London, UK, 2015.
28. Council of Europe Landscape Convention as Amended by the 2016 Protocol. Available online: <https://rm.coe.int/16807b6bc7> (accessed on 31 January 2025).
29. Available online: <https://www.artemis-project.eu> (accessed on 31 January 2025).
30. Available online: <https://www.london-charter.org/> (accessed on 31 January 2025).
31. Beacham, R.; Denard, H.; Niccolucci, F. An Introduction to the London Charter. In *The e-Volution of Information Communication Technology in Cultural Heritage: Where Hi-Tech Touches the Past: Risks and Challenges for the 21st Century. Short Papers from the Joint Event*; Ioannides, M., Arnold, D., Niccolucci, F., Mania, K., Eds.; CIPA/VAST/EG/EuroMed; Archaeolingua: Budapest, Hungary, 2006; Available online: [https://www.london-charter.org/media/files/beacham-denard-niccolucci\\_intro.pdf](https://www.london-charter.org/media/files/beacham-denard-niccolucci_intro.pdf) (accessed on 31 January 2025).
32. Denard, H. A New Introduction to the London Charter. In *Paradata and Transparency in Virtual Heritage*; Bentkowska-Kafel, A., Baker, D., Denard, H., Eds.; Digital Research in the Arts and Humanities Series; Ashgate: Farnham, UK, 2012; pp. 57–71. [[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.



Article

---

# Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins

---

Achille Felicetti, Aida Himmiche and Miriana Somenzi

## Special Issue

Digitalization of Cultural Heritage with Artificial Intelligence: Machine Learning and Deep Learning Solutions

Edited by

Dr. Marcello La Guardia, Dr. Valentina Bonora and Dr. Andrea Masiero





## Article

# Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins

Achille Felicetti <sup>1,\*</sup> , Aida Himmiche <sup>2</sup>  and Miriana Somenzi <sup>2</sup> <sup>1</sup> VAST LAB, PIN, Piazza dell'Università 1, 59100 Prato, Italy<sup>2</sup> ARIADNE Research Infrastructure AISBL, Rue de Grand-Bigard 14, Berchen-Sainte-Agathe, 1082 Bruxelles, Belgium; aidahimmiche@gmail.com (A.H.); somenzi.miriana@gmail.com (M.S.)

\* Correspondence: achille.felicetti@gmail.com

## Abstract

This paper explores the integration of Artificial Intelligence and semantic technologies to support the creation of intelligent Heritage Digital Twins, digital constructs capable of representing, interpreting, and reasoning over cultural data. This study focuses on transforming the often fragmented and unstructured documentation produced in cultural heritage into coherent Knowledge Graphs aligned with internationally recognised standards and ontologies. Two complementary AI-assisted workflows are proposed: one for extracting and formalising structured knowledge from heritage science reports and another for enhancing AI models through the integration of curated ontological knowledge. The experiments demonstrate how this synergy facilitates both the retrieval and the reuse of complex information while ensuring interpretability and semantic consistency. Beyond technical efficacy, this paper also addresses the ethical implications of AI use in cultural heritage, with particular attention to transparency, bias mitigation, and meaningful representation of diverse narratives. The results highlight the importance of a reflexive and ethically grounded deployment of AI, where knowledge extraction and machine learning are guided by structured ontologies and human oversight, to ensure conceptual rigour and respect for cultural complexity.



Academic Editor: Jose María Alvarez Rodríguez

Received: 19 July 2025

Revised: 30 August 2025

Accepted: 11 September 2025

Published: 15 September 2025

**Citation:** Felicetti, A.; Himmiche, A.; Somenzi, M. Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins. *Appl. Sci.* **2025**, *15*, 10061. <https://doi.org/10.3390/app151810061>

**Copyright:** © 2025 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) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** artificial intelligence; ontologies; knowledge graphs; digital twins; cultural heritage

## 1. Introduction

It is said that Helen of Troy, the most beautiful woman in the world according to Greek mythology, never set foot on the shores of Asia Minor. In a dramatic and unsettling version of the myth, told by Euripides and echoed in later traditions, it was not the real Helen whom Paris carried off to Troy. Instead, a simulacrum, an *eidolon*, a perfect likeness devoid of substance, was sent by the gods to take her place (this version of the myth is attested in Apollodorus, *Epitome* III. 5; Euripides, *Electra*, 128 and *Helen* 31 ff.; Servius on Virgil's *Aeneid*, I. 655 and II. 595; Stesichorus, quoted by Tzetzes, *On Lycophron* 113). While the true Helen waited in silence in Egypt, perhaps unaware of the war unfolding in her name, it was her image that ignited desire, deceived kings, and plunged the world into its most storied conflict [1]. The *eidolon* was not a lie but a form with agency: it acted, provoked, and transformed the course of history. Though lacking life of its own, it possessed real effectiveness.

In this figure of the operative double, of an image that does not merely represent but intervenes, there is a striking anticipation of the modern notion of digital twin. Such an

entity is designed not just to mirror the original but to accompany or replace it in action, to operate in its stead, often with greater reach and precision. A digital twin, to fulfil this role, must do more than resemble. Like Helen's phantom, it must be endowed with structured intelligence and a "grammar" that sustains its coherence. The story of Helen's *eidolon* speaks to an ancient intuition: that a likeness, when crafted with precision and animated with purpose, can operate in the world with consequences equal to, or even surpassing, those of its original. This mythic logic finds a contemporary counterpart in the digital age, where cultural entities are increasingly represented by what might be called *simulacra*, i.e., constructed doubles whose very name shares etymological roots with the verb *to simulate*. These digital twins are not inert copies but dynamic, structured representations designed to engage with the world through processes of analysis, interaction, and prediction.

In the realm of cultural heritage, the digital twin has evolved far beyond the notion of a static visualisation or isolated model. It now constitutes a complex informational organism, integrating diverse layers of documentation: historical texts, scientific analyses, sensory data, and curatorial knowledge. To preserve coherence across this multiplicity of sources, ontologies provide the foundational structure to build the memory of the digital twin by serving as its semantic architecture and allowing knowledge to be expressed formally, queried effectively, and shared across systems and institutions.

Yet, structure alone cannot suffice. Just as Helen's *eidolon* required a divine breath to become operative, digital twins must be endowed with intelligence in order to act meaningfully. Artificial Intelligence, when guided by ontological frameworks, enables digital twins to transcend passive representation. Through the semantic interpretation of textual and visual sources, the detection of patterns in environmental data, and the simulation of complex behaviours, AI seems ideal to become the interpretive engine of the twin. In this framework, natural language processing, entity recognition, semantic inference, and predictive modelling are no longer isolated functionalities but become part of a coherent system of knowledge and memory production.

A distinctive innovative aspect of this study is the systematic combination of Artificial Intelligence and ontologies in the construction of digital twins. Previous approaches, especially in cultural heritage, have often relied on AI without semantic guarantees, or on ontological models without operational integration. The framework presented here links the two dimensions inseparably: ontologies act as the semantic backbone that ensures transparency, coherence, and interoperability, while AI functions as the cognitive engine that extracts, enriches, and interprets knowledge from the complex body of documentation that sustains the digital twin, including analytical reports, curatorial records, and other heritage data. It is precisely in the synergy between ontologies and AI that the true potential of digital twins can unfold. Their mutual interaction constitutes the epistemological core of a new generation of digital twins in which ontologies ensure that AI models are not only efficient but also transparent, interpretable, and anchored in shared conceptual frameworks. In return, AI can enrich semantic information by identifying latent connections, disambiguating meanings, and populating knowledge graphs with contextually valid information. The result is a self-expanding semantic system, capable of both reasoning and remembering, a system that learns, explains, and operates in alignment with the epistemic and ethical values of cultural heritage research.

This paper is structured as follows: Section 2 outlines the historical evolution of digital twins, particularly those enhanced by AI, from their origins in engineering and aerospace to their gradual redefinition within the cultural heritage domain. Section 3 introduces the Heritage Digital Twin and its associated ontology, designed to implement cognitively enriched, conceptually expressive constructs. Section 4 explores the uses of new Artificial Intelligence technologies to extract knowledge from cultural heritage

documentation and encode it according to the semantics of ontological models widely used in this domain to build the knowledge graphs of Heritage Digital Twins. Section 5 investigates how knowledge graphs can be used to enrich the semantic toolbox of Artificial Intelligences, transforming them into agents capable of reasoning on large and intricate amounts of data and answering complex scientific questions formulated in natural language. Section 6 presents a case study where the entire pipeline is tested on a set of scientific reports from the ARTEMIS project documentation to identify and distil information used to populate the knowledge graphs that serve as the engine of the Heritage Digital Twins and is used to enrich and reinforce AI components, enhancing their ability to support advanced functionalities such as natural language interfaces, intelligent querying, and adaptive interaction with users. Following the discussion in Section 7, which highlights the potential of these technologies while also highlighting the inherent risks in their application, particularly when not subject to rigorous human oversight, this paper moves towards Section 8, offering some conclusive notes and future perspectives.

## 2. Digital Twins: From Conceptual Models to Operational Systems

The relevance of digital twins has steadily expanded across disciplines, reflecting their growing role as sophisticated digital counterparts of real-world entities and systems [2,3]. While often loosely applied to any virtual replica, the term also designates complex, structured information systems capable of interfacing with the physical world. Initially developed to enhance industrial performance, optimise processes, and anticipate failures, digital twins have since evolved to support various domains for their ability to improve planning decisions [4], especially by integrating multi-source data [5].

The origins of digital twins are often traced to NASA's simulation technologies in the 1960s, which allowed engineers to replicate space conditions virtually—an epistemological necessity in the face of inaccessibility [6]. Although the term “digital twin” formally emerged only in 2010 [7], the conceptual lineage is deeper and more complex, as is often the case with terms that retrospectively colonise prior practices. The applicability of digital twins has expanded dramatically since then. In manufacturing, for instance, digital twins underpin Industry 4.0 and its successor paradigms, enabling predictive maintenance, life-cycle optimisation, and human–machine collaboration [8]; and in the construction sector, digital twins facilitate accurate simulation of design scenarios before physical implementation, reducing costs and enabling early detection of critical flaws. This evolution is most visible in large-scale projects. Urban applications, perhaps the most ambitious in scale and implication, leverage digital twins to model smart cities, integrating traffic, land use, energy consumption, and citizen feedback into coherent, semantically grounded infrastructures. A digital twin of the city of Zurich is being developed to support sustainable urban planning [4]; and the new Singapore's city-scale twin integrates data from sensors, satellites, and drones for growth management and environmental monitoring [9]. Moreover, the European Commission's Destination Earth initiative envisions a planetary-scale digital twin to simulate climate dynamics, incorporating not only physical systems but also social and behavioural data [10]. This human-in-the-loop paradigm reflects a growing recognition that digital twins are not neutral mirrors but value-laden constructs. They encode assumptions about relevance, causality, and intervention and thus require governance frameworks that ensure transparency, ethical integrity, and stakeholder accountability [11].

When transposed to the cultural heritage domain, however, digital twins demand a more nuanced methodological framework. Unlike industrial or urban applications, where twins are created before or alongside their physical counterparts, heritage assets typically predate their digital representations by centuries and embody multiple temporal, material, and cultural layers. Consequently, their digital twins must account for both tangible and

intangible dimensions, incorporating expressions of human creativity, identity, collective traditions, performances, and community practices that carry historical, artistic, social, and symbolic significance [12–14]. Within this perspective, digital twin technology has increasingly been tested across diverse areas of cultural heritage, producing significant though still heterogeneous results. In archaeology, digital twins have been used to document and virtually reconstruct sites and artefacts, providing new opportunities for exploring contexts that are otherwise fragmented or inaccessible [15]. The museum sector has adopted similar strategies for collections management and preventive conservation [16], with major institutions such as the Victoria & Albert Museum and the Louvre experimenting with large-scale digitisation to safeguard fragile artefacts and to expand access to audiences worldwide [17,18]. Monuments and architectural heritage have also been at the forefront of experimentation: the digital twin of Notre-Dame Cathedral, for instance, now functions as both a record of the building before the devastating fire of 2019 and a central resource for its ongoing restoration [19]. Comparable approaches have been applied to historic buildings, where parametric twins integrate sensor data to monitor indoor climate conditions and support conservation planning [20], as demonstrated by the cases of Löfstad Castle in Sweden [21] and Sant'Andrea's church in Pistoia, Italy [14].

At the scale of urban and landscape heritage, projects such as the Great Wall Resource Management Information System [22] have generated detailed digital archives of UNESCO World Heritage sites across continents, demonstrating how digital twin technologies can be applied to cultural landscapes of vast territorial extension. Finally, in the field of restoration, digital twins enriched with AI-powered inpainting and 3D reconstruction have enabled the simulation of conservation treatments and the digital recovery of degraded artefacts, offering conservators a dynamic decision-support tool before interventions on the original object.

These examples demonstrate that, in most cases, applications of digital twin technology in cultural heritage remain predominantly tied to 3D visualisations, echoing late twentieth-century efforts to virtually reconstruct monuments and artefacts. As the model is evolving, the digital twin is gradually ceasing to be merely a static reflection, becoming a conceptual framework open to reinterpretation of operational and behavioural dimensions across disciplinary borders.

### *2.1. Digital Twins and Artificial Intelligence*

Artificial Intelligence has increasingly emerged as a transformative force in the cultural heritage domain, offering powerful methods for analysing, interpreting, and even recreating the past. Its capacity to process vast datasets, recognise patterns, and generate new content has enriched the arsenal of digital tools available to scholars, conservators, and institutions in unprecedented ways. Applications span an impressively wide spectrum. Generative systems have been used to complete unfinished compositions by legendary composers, providing novel insights into artistic style and creative processes [23]. In parallel, advances in natural language processing have enabled the reconstruction and interpretation of ancient texts written in extinct or poorly understood languages, contributing to philology, epigraphy, and the study of lost literatures [24]. In archaeology and palaeoenvironmental studies, AI-driven reconstructions have recreated the aspect of long-lost environments and past landscapes, allowing researchers to simulate ecological conditions, visualise historical settings, and contextualise human activity [25,26].

Conservation and restoration represent another key area where AI has proven invaluable. In art history, recent advances have explored the innovative application of AI in the reconstruction and preservation of historical paintings, sculptures, ceramics, and other artefacts, offering new solutions to the challenges of traditional restoration [27]. The

same technologies have been applied to questions of provenance, helping to trace the circulation of cultural objects across time and space through automated analysis of stylistic or material features [28]. Systems capable of detecting deterioration patterns in artworks and monuments provide guidance on preventive conservation strategies [29,30], while generative and inpainting algorithms have been successfully employed to support the virtual restoration of damaged paintings, sculptures, ceramics, and manuscripts [31]. These developments not only facilitate better-informed interventions but also open the possibility of testing conservation hypotheses in digital environments before their implementation on the original artefacts.

The incorporation of Artificial Intelligence into digital twin technologies is ushering in a significant conceptual and operational shift. No longer confined to functioning as elaborate digital counterparts of physical entities, digital twins are now evolving into systems endowed with interpretive, adaptive, and semantically structured intelligence, capable not just of replicating but also of understanding and responding to the complexities of their real-world counterparts. This transformation marks a notable epistemological threshold, recasting the digital twin as an active agent of interpretation rather than a passive conduit of representation. It is AI that propels this transition, furnishing digital twins with mechanisms of learning, inference, and autonomous decision-making [8]. As such, digital twins are no longer mere digital projections but are emerging as computational systems capable of analysis, simulation, and strategic recommendation [32]. Evidence of this transformation is manifesting across a broad range of domains. In manufacturing, for example, AI-enhanced digital twins support predictive maintenance, improve workflow design, and enable real-time responsiveness by interpreting sensory inputs and operational histories. Factories thus begin to resemble adaptive environments, governed not solely by pre-programmed rules but by evolving patterns of interaction and feedback [33]. In urban systems, the convergence of AI and digital twins is transforming how cities are modelled, understood, and governed. These intelligent frameworks are being applied to simulate environmental and infrastructural conditions—ranging from carbon emissions and heat distribution to traffic and pedestrian behaviour. As a result, municipalities are empowered to enact more nuanced and anticipatory interventions [34,35].

Importantly, the symbiosis between AI and digital twins is extending its reach beyond industrial or scientific contexts, entering realms such as cultural heritage with promising, often experimental, applications. One particularly innovative avenue involves the reconstruction of lost or undocumented heritage through AI-based image generation. Here, AI is employed to synthesise visualisations derived from oral testimonies and collective memory, complementing tangible-based HBIM methods [36].

In heritage-related scenarios more broadly, AI integration into digital twins is already proving beneficial for documentation, conservation, and public access. AI-driven twins are being employed to improve crisis management by integrating multi-sensor data and supporting real-time decision-making, with clear implications for the protection of heritage assets in emergency situations [37]. In cultural tourism contexts, the combination of digital twins and AI enables predictive monitoring, preventive conservation, and immersive storytelling, enhancing both preservation and visitor engagement [38]. At the same time, AI-enabled digital twins are increasingly able to reshape the relationship with artworks and heritage environments by allowing replication, remote access, and large-scale data integration [17]. More advanced approaches are exploring semantically grounded twins enhanced with memory-enabled AI agents, capable of reasoning over time and providing transparent, value-aligned support for long-term preservation planning [39]. Automation is expediting data collection, facilitating large-scale digitization, and enhancing accessibility via automated transcription and translation tools. At the same time, AI is processing



real-time sensor data to detect threats and recommend countermeasures, increasing the system's responsiveness to environmental risks or structural degradation. In this way, conservation scenarios can be virtually simulated, allowing interventions to be evaluated without exposing heritage objects to physical risk [14].

Despite these advances, the implementation of AI within digital twin ecosystems remains uneven and often constrained. Many applications persist in isolated pilot stages, lacking the infrastructural robustness or institutional momentum to scale effectively [40]. The interdependence of software, hardware, and human expertise highlights the need for deployment strategies that go beyond technical engineering to embrace socio-technical coordination [41]. Also, on a conceptual level, the integration of AI into the epistemic fabric of digital twins remains underdeveloped. While machine learning models excel in identifying correlations and predicting outcomes, they often fall short in contributing to the construction of coherent knowledge systems [42]. Nevertheless, progress continues steadily, and the expectation is that AI will increasingly endow digital twins with the capacity to fulfil their epistemic and operational potential. Yet this trajectory raises critical concerns about transparency and information quality. The internal logic of many AI systems remains opaque, making it difficult to trace how conclusions are reached—especially problematic in domains requiring interpretive subtlety, such as cultural heritage, urban planning, or medicine. Here, data in isolation is insufficient; meaningful engagement demands systems that can not only reason but also explain. An overreliance on autonomous inference without appropriate mechanisms for interpretability risks introducing errors, distortions, or even ethical blind spots. This situation prompts a foundational question: how far can epistemic authority be entrusted to systems incapable of articulating their own reasoning? Especially now that decisions once reserved for human deliberation start to be automated by systems that remain inaccessible to scrutiny, ensuring intelligibility and accountability is thus not a secondary concern but a prerequisite for ethical and sustainable implementation [11].

## 2.2. Ontologies: Illuminating Artificial Intelligence

In response to these fundamental questions, emerging approaches are foregrounding the importance of structured semantic architectures as a prerequisite for meaningful AI deployment. Rather than relying solely on data quantity or statistical modelling, these strategies emphasise the need for carefully curated knowledge frameworks, defining concepts, relationships, and ontological commitments, as foundational to advanced reasoning and explanation [43]. Ontologies, in this regard, are serving a dual role: they are enabling more nuanced and context-sensitive AI operations and also acting as safeguards against the risks of misinterpretation or hallucination [44]. Additionally, ontologies are employed in the development of explainable machine learning pipelines, addressing crucial aspects like feature categorisation and metadata description, which are fundamental for the transparency of the machine learning process [45]. By embedding formal logic and expert-curated structure within digital twin systems, ontologies help ensure that AI outputs remain anchored in coherent and verifiable knowledge domains. This allows intelligent systems not merely to function but to explain, interrogate, and evolve. In doing so, they contribute to the transformation of digital twins into epistemic infrastructures: semantic, reflexive, and dynamically coupled with the evolving landscape of human understanding [46]. This role of ontologies as a foundation for trustworthy AI is a crucial aspect, since they are able to provide more structured and explicit knowledge that enables AI systems to better articulate their reasoning. This extends beyond mere performance, addressing ethical concerns such as bias mitigation and transparency enhancement.



### 2.3. Towards the Cognitive Heritage Digital Twins

The Heritage Digital Twin currently being designed and developed [12–14,47] is shaped as a point of confluence; a space where these diverse technologies do not merely coexist but actively intertwine. It is precisely at this intersection that ontologies, Artificial Intelligence, and semantic infrastructures converge to generate a new kind of cultural intelligence. As evidenced in Section 2.1, typical applications of digital twins in heritage contexts largely focus on high-resolution 3D models, often derived from laser scanning or photogrammetry. These visual representations, while technically sophisticated, are semantically shallow and remain inert unless complemented by rich contextual meta-data and interpretive frameworks, offering a simulacrum in the visual sense but lacking interpretive depth.

Rather than reducing cultural heritage to its visual or geometric likeness, the paradigm of Heritage Digital Twin embraces the full constellation of digital documentation that relates to real-world cultural entities, whether movable, like artworks and artefacts; immovable, like monuments or architectural sites; or even intangible, such as rituals, traditions, and other expressions of cultural memory. This paradigm shifts the focus away from purely 3D-centred representations, proposing instead a holistic digital ecosystem that brings together visual depictions, textual accounts, scientific analyses, conservation records, and historical interpretations. It is a vision grounded in the belief that meaning arises not from isolated forms but from the interplay of all the fragments through which heritage is remembered, transmitted, studied, and understood.

Cultural entities, in fact, are not governed solely by physics but essentially by memory and cannot be fully described through geometry or physics alone since they carry meanings, contexts, and symbolisms that resist reduction to quantitative attributes. As a result, the replication of a heritage entity does not have to rest on form alone but requires a cognitive and semantic infrastructure and the ability to articulate not only what an object is but what it means, represents, and has undergone across time. As such, the digital twin in this domain demands not only a new technical infrastructure but a rethinking of its very ontological assumptions.

At the heart of any meaningful digital representation of cultural heritage lies a conceptual operation as much as a technical one: the deliberate choice to model reality through shared ontological commitments. Cultural entities are not inert objects, they are constructed through layers of description, interpretation and transmission. If their digital counterparts are to achieve more than a superficial resemblance, they must be capable of reflecting this diachronic and interpretive richness. Ontologies provide the necessary framework for this task: a rigorous grammar through which cultural knowledge may be structured, preserved, and subjected to critical inquiry. By expressing information as entities and relationships, ontologies enable a deeper representation of the cultural object's life history, its transformations, and the meanings it acquires across time. In this framework, Artificial Intelligence can also reveal its full potential as a catalytic instrument capable of navigating, extracting, and reassembling meaning from the vast, fragmented archives of human knowledge and at the same time ingesting and digesting this knowledge to enhance its capabilities and serve as a privileged tool for exploring and interrogating the information embedded in the digital twin.

The culmination of this trajectory is thus the development of a new construct that unites the expressive power of ontologies with the interpretive capabilities of Artificial Intelligence in order to represent, reason about, and interact with cultural entities. Far from being a mere digital copy of its shape, the Heritage Digital Twin becomes an evolving semantic organism, an animated simulacrum (like Helen's *eidolon*) which not only has its external resemblance but reproduces the cultural object in its entirety. In its most advanced

form, the Heritage Digital Twin can become an agent (i.e., a cognitive entity) capable of dialogue, inference, and response within a meaningful digital continuum.

In this perspective, the present study proposes some innovative methodological aspects, building upon and extending existing practices. A distinctive contribution lies in the systematic combination of artificial intelligence and ontologies for the construction of Heritage Digital Twins (HDTs). While prior work in cultural heritage has often approached these technologies in isolation—relying either on AI without robust semantic guarantees or on ontological models lacking operational integration, as illustrated above—the proposed framework aims at unifying these two dimensions. Specifically, widely validated ontologies in cultural heritage documentation serve as the semantic backbone, ensuring the transparency, coherence, and interoperability of the data model. Concurrently, AI operates as the cognitive engine, dynamically extracting, enriching, and interpreting knowledge from the heterogeneous body of information that sustains the digital twin. This approach not only leverages the power of AI to process vast amounts of data—from analytical reports and curatorial records to conservation data—but also embeds this process within a structured, semantically rigorous framework. This ensures that the insights generated by the AI are verifiable and contextually relevant, a key advancement over previous approaches.

On this basis, this research pursues two main objectives: (i) to demonstrate how the proposed methodology facilitates the construction of Heritage Digital Twins by systematically examining and integrating heterogeneous documentation into a coherent and semantically grounded knowledge base and (ii) to show how this combined paradigm enhances the reactivity and cognitive potential of Heritage Digital Twins, enabling them to respond more effectively to analytical, conservation, and interpretative needs. The feasibility of these objectives is illustrated through two complementary case studies, aimed at illustrating the practical applicability of the proposed methodology: on one hand, ontologies and AI engage with the epistemic “archaeology” of cultural information, retrieving, structuring, and semantically articulating the dispersed and often latent knowledge embedded in visual and textual heritage records. On the other, they explore how this extracted knowledge, once formalised within a coherent ontological framework, can be reintroduced into AI systems to enhance their capacity for informed reasoning, contextual sensitivity, and traceable inference. A full account of the complete creation of a Cognitive Heritage Digital Twin would require a much larger study, given the scale and complexity of the task, encompassing vast amounts of documentation, data integration, and long-term monitoring. Far from exhausting the spectrum of possibilities opened by these technologies, this study exemplifies one of their most immediately impactful trajectories: the retrieval, organisation, and strategic deployment of knowledge embedded in cultural heritage documentation. The experiments conducted suggest that the procedures used could constitute a valuable contribution to anyone intending to address the same issues, especially in the cultural heritage sector where this type of experimentation is particularly needed.

### 3. Digital Twins and the Modelling of Knowledge

This paper builds upon the theoretical foundations introduced in our previous work on the Reactive Heritage Digital Twin and its associated ontology, offering a more operational perspective. It presents the design and implementation of AI-assisted semantic pipelines for the extraction, structuring, and reasoning over cultural information. By combining ontologies with intelligent processing components, it explores how to construct digital twins that are not only informative but also sentient in their logic and responsive in their behaviour. More specifically, this paper examines two complementary uses of Artificial Intelligence within this framework, situated at opposite ends of the semantic pipeline. On one side, AI is deployed for the extraction of knowledge from unstructured textual sources, in particular,

heritage science reports, and its subsequent alignment with ontological structures, resulting in the enrichment of the digital twin's semantic graph with scientific information. On the other hand, the populated graph itself becomes a resource for training AI systems, enabling them to perform tasks such as querying semantic information through natural language or assisting in the extraction of new content. This virtuous cycle of knowledge generation and reinforcement creates a recursive architecture in which semantic data and intelligent agents co-evolve, gradually enhancing the digital twin's cognitive capacity. The goal is to show how this technological convergence can support the development of truly intelligent Heritage Digital Twins: systems that do not merely reflect culture but actively participate in its documentation, understanding, and preservation.

The conceptual architecture upon which the entire system is constructed is the Reactive Heritage Digital Twin Ontology (RHDTO), an extension blossomed from the solid foundations of the CIDOC CRM [48] as a means to organise, interconnect, and semantically sustain the heterogeneous data that forms the informational substrate of cultural digital twins. Among the various ontological frameworks developed for the heritage domain, the CIDOC Conceptual Reference Model (CRM) is an internationally recognised standard (ISO 21127:2023 [49]). More than a model, CIDOC CRM constitutes an ecosystem that includes domain-specific extensions, such as CRMsci for scientific observation [50], CRMdig for digital provenance [51], and CRMhs for heritage science [52], that offer a nuanced vocabulary capable of accommodating the multifaceted nature of cultural data. The adoption of CIDOC CRM offers a new path forward, disclosing the possibility to describe not just the form of an object but the historical events, actors, techniques, and materials that constitute its biography. It is this ontological fluency, this capacity to speak the shared language of heritage knowledge, that renders the CIDOC CRM ecosystem a cornerstone of semantic continuity across the heritage information landscape.

In some previous works, the structural components of the RHDTO have been articulated, delineating its main dimensions: the *documentary* one, which captures the breadth of scientific, historical, and multimedia documentation [12,13]; the *reactive* one, which models the operational mechanisms for real-world interaction through sensors, deciders, and actuators [14]; and the *intelligent* one, in which Artificial Intelligence is integrated to interpret, enrich, and activate knowledge [53]. To operationalise these dimensions, the RHDTO introduces a set of classes and properties designed to model the multifaceted nature of heritage entities and their digital counterparts with semantic precision. At the core lies the class *HC1 Heritage Entity*, a general abstraction encompassing both tangible (*HC3 Tangible Aspect*) and intangible (*HC4 Intangible Aspect*) dimensions of cultural heritage. The digital twin itself is represented by *HC2 Heritage Digital Twin*, conceived as a structured network of informational components, including digital reproductions, documentation, and the activities underpinning their creation and management. Notably, Ref. [53] illustrates how AI can serve as the cognitive engine of the digital twin, supporting analytical reasoning and decision-making within key components such as the Decider and the Actuator, thereby enhancing the system's ability to respond to real-world conditions in a meaningful and context-aware manner. Each of these facets are modelled by means of dedicated classes, such as *HC10 Decider* and *HC15 AI Component*, and contribute to the evolving ecology of digital twins. Building on this, the subsequent sections describe several ways in which AI can be practically used for this purpose, offering an overview of the most recent methodologies and technologies available.

## 4. From Fragments to Knowledge

### 4.1. Natural Language Processing of Textual Documentation

NLP focuses on enabling machines to understand, interpret, and generate human language. As simple as this may sound from an agnostic point of view, handling human language is one of the most difficult tasks a machine can undertake. Unlike programming languages, which are based on mathematical rules and are explicitly designed to be machine-understandable, natural language is not as easily interpreted or represented in numerical form.

Over the past decade, the field of NLP has undergone a major transformation with the advent of deep learning and, more recently, Large Language Models (LLMs). These models are based on the Transformer architecture, which introduced the so-called attention mechanism:

*“Given a set of vector values and a vector query, attention is a technique to compute a weighted sum of the values, dependent on the query.” [54]*

In particular, LLMs use a specialised version called self-attention.

In traditional attention mechanisms, each query maps exactly to one key–value pair, while in self-attention each query matches each key to varying degrees and then the returned result is a sum of values weighted by the query–key match. This mechanism allows the model to assign importance to different tokens (which are often single words) within a sequence, enabling it to focus on the most relevant parts of the input when generating output.

In addition to this, LLMs are trained on vast, multilingual corpora [55] and have demonstrated impressive capabilities across tasks [56] such as machine translation, question answering, summarisation, and semantic search. Their ability to perform complex tasks such as transfer learning, a machine learning technique in which the knowledge learned from a task is re-used to boost performance on a related task, few-shot generalisation [57], and contextual language modelling [58] has opened new frontiers in computational language understanding, particularly in complex or previously underexplored text domains [59].

One such domain is cultural heritage [60], where textual artefacts may include ancient manuscripts, epigraphic inscriptions, archival inventories, scanned PDFs of scientific analyses, and handwritten letters [61]. These materials offer rich but highly challenging content for computational analysis, often exhibiting historical language variants, inconsistent orthography, multilingual passages, and physical degradation, all of which complicate standard NLP workflows.

Recent efforts by computer scientists have focused on adapting LLMs and developing domain-specific NLP pipelines to address these challenges. Techniques such as fine-tuning on specialised corpora, integrating Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) outputs [62], and combining textual data with metadata or image features [63] are increasingly used to support tasks like:

- named entity recognition
- semantic enrichment
- machine-assisted transcription
- cross-document linking

As a result, state-of-the-art NLP is becoming an essential component in the digital preservation, accessibility, and scholarly analysis of cultural heritage texts.

#### 4.2. LLMs for Identification and Semantic Extraction of Relevant Entities and Relationships

A critical challenge in the digital processing of cultural heritage texts lies in the identification and extraction of entities and relationships, a foundational step for building structured knowledge, supporting semantic search, and enabling automated reasoning.

Traditionally, this task has been addressed through Named Entity Recognition (NER) pipelines, which label predefined categories such as “person”, “location”, “organisation”, or “date” in unstructured text. Over the last decade, state-of-the-art NER systems have been greatly enhanced by transformer-based models like BERT and RoBERTa, showing high accuracy on well-structured, contemporary texts. However, these systems often struggle when applied to historical documents [64], multilingual corpora, or domain-specific ontologies, where named entities are ambiguous, infrequent, incompletely annotated, or entangled in complex relational structures (even though there are some successfully trained version of some specific domains, such as ArchaeoBERT [65]). In the context of cultural heritage, standard NER often proves insufficient. Texts, such as archival inventories, epigraphs, marginalia, or curatorial notes, frequently reference entities in indirect, partial, or obsolete forms [66]. Furthermore, the relevant concepts often extend beyond standard categories to include historical events, material properties, artistic styles, ritual roles, or geopolitical shifts, many of which are not well covered by off-the-shelf NER models [67].

These limitations pose significant barriers to aligning extracted information with structured ontologies or linking it to a Knowledge Graph.

Recent research has increasingly highlighted the intersection of LLMs and KGs, pointing to a paradigm shift in knowledge representation. On one side, LLMs are used to augment KGs by supporting knowledge extraction, construction, and refinement. On the other, KGs are used to augment LLMs in tasks such as training, prompt learning, or knowledge grounding [68,69]. These works underline a move towards hybrid models and pipelines that integrate both explicit ontological knowledge and the parametric knowledge encoded within LLMs.

KG construction itself is a complex process that requires the integration of information both from structured and unstructured data. Traditional pipelines treat these sources in isolation, struggling with the heterogeneity and noise typical of cultural heritage data. LLMs, by contrast, are trained across diverse sources and exhibit strong performance on tasks such as knowledge extraction [70], entity resolution [71], and relation alignment.

An LLM-based approach is employed in this experiment, enabling flexible, context-aware entity and relation extraction in the cultural heritage domain, eliminating reliance on pre-trained NER schemas.

Instead of classifying entities into fixed types, the LLM is prompted to identify relevant semantic units and their relationships in natural language, guided by domain-specific ontology definitions. This enables extraction and alignment of non-standard or composite entities and their mapping to structured entities in a customised Cultural Heritage Knowledge Graph.

This prompt-based methodology further supports the extraction of complex relationships, including part-whole, temporal, provenance, and scientific analysis relations, which are essential for modelling cultural heritage information in ontological formats.

By bypassing traditional NER in favour of in-context reasoning with LLMs, multiple tasks, such as entity disambiguation, coreference resolution, attribute assignment, and relation typing, are integrated into a single unified pipeline.

While LLM fine-tuning methods can leverage textual information from individual entities, formal semantics are often underutilised. Beyond template-based fine-tuning, additional LLM techniques, such as prompt learning and instruction tuning, present promising avenues to enrich semantic understanding [72,73]. Moreover, cultural heritage applications

often rely on cross-domain knowledge, highlighting the importance of ontology alignment, so the identification of cross-ontology mappings among concepts, instances, and properties with equivalence, integration, or membership relationships can occur.

Traditional alignment systems, such as LogMap [74], rely heavily on lexical matching and symbolic reasoning, but this is often not enough. Exploiting textual meta-information through LLMs offers a promising strategy for ontology alignment: even though pretrained models like BERT have been fine-tuned for this task [75], state-of-the-art LLMs are yet to be explored in the context. Our experiments show that this LLM-driven strategy significantly improves recall and semantic accuracy in entity-linking tasks when compared to traditional NER-based pipelines, particularly when dealing with fragmented historical records or low-resource languages. Moreover, the use of an open-source, locally deployable model ensures data privacy, customisability, and scalability, meeting the needs of cultural institutions and digital humanities projects that require transparent, interpretable, and adaptable AI tools.

Transitioning extracted information into an ontological encoding makes it unambiguous, facilitating integration into the Cultural Heritage Knowledge Graph and subsequent incorporation into Heritage Digital Twins, thus supporting richer and more semantically coherent digital representations of cultural assets.

#### *4.3. Building Semantic Knowledge Graphs from Extracted Information*

The process of encoding NLP-extracted data to ontological structures is never a mere act of correspondence; it is a conceptual translation that requires both technical rigour and hermeneutic sensitivity. Information drawn from textual descriptions, archival records, scientific datasets, or visual analysis carries within it layers of context, intention, and uncertainty. To encode such data into an ontological framework is to interpret it through a lens that seeks formal clarity without erasing epistemic nuance.

One of the major strengths of the RHDTO model lies in its ability to function as a connective tissue between datasets that originate from heterogeneous contexts, a conceptual harmonisation that permits the assembly of coherent knowledge graphs without sacrificing domain specificity. RHDTO provides the necessary paraphernalia for such knowledge transformation and transcoding, turning textual evidence and visual data into computable knowledge while preserving their contextual richness. This is particularly relevant in domains like heritage science, where the use of these conceptual tools allows for a fine-grained articulation of experimental procedures and datasets, as will be illustrated in Section 6. This semantic articulation reshapes the cognitive architecture through which heritage information is organised, shared, and understood, transforming the ontological framework of the knowledge graph in a philosophical intermediary between language and logic, between human interpretation and machine understanding [76]. This way, the digital twin is endowed with a cognitive dimension. It no longer merely stores or displays data but interprets and internalises it within an ontological order that mirrors, albeit partially and provisionally, the complex web of human understanding.

Artificial Intelligence, particularly in its recent semantic incarnations, also plays an increasingly central role in this mapping process. AI systems can retrieve entities, classify relationships, and extract candidate assertions from unstructured sources and convert them in ontology-grounded semantic triples. This is one of the cases in which the synergy between AI and ontologies reveals its full potential, the former accelerating the identification of candidate knowledge, the latter ensuring that such knowledge is meaningfully situated within a coherent and reusable conceptual space. Yet the true work of semantic alignment remains interpretive, demanding expert judgement in choosing the appropriate classes, properties, and event structures through which to articulate extracted content within the framework of the knowledge graph.



## 5. From Knowledge to Intelligence

### 5.1. Using Knowledge Graphs to Build Knowledge-Enriched AI Agents

Once knowledge crystallises in the form of semantic graphs, it can be used to nurture and enhance AIs. The complexity of cultural heritage data demands the use of intelligent systems that can reason over both explicit content and the nuanced context behind it. Knowledge graphs, for this purpose, have emerged as a powerful tool to build knowledge-enriched Artificial Intelligence systems. Information is structured to preserve its rich semantics and remains machine-interpretable, which enables AI systems to achieve context-aware understanding and inference instead of limiting themselves to surface-level pattern recognition [77,78].

Cultural heritage tells a story, and as stories are composed of interrelated entities, they must be represented in such fashion. A knowledge base with its nodes and edges perfectly captures this web of elements, grounded in formal ontologies [78,79], and CIDOC CRM and its extensions play a crucial role in offering a shared conceptual schema that aligns historical data with contemporary semantic technologies [78,80,81].

In recent years, there has been a growing emphasis on reinforcing the reasoning capabilities of LLMs by integrating them with external knowledge structures, particularly knowledge graphs. While LLMs such as GPT, BERT, and Mistral [82–84] have demonstrated remarkable abilities in generating fluent and plausible language, they remain prone to hallucinations, generating incorrect or misleading content especially in high-stakes domains like healthcare or science [85].

These failures often stem from limitations in their training objective, typically maximising the log-likelihood of the next token without robust grounding in factual correctness.

To address this, researchers are increasingly integrating LLMs with knowledge graphs, which encode structured, semantically rich, and traceable factual knowledge [86].

Unlike opaque neural representations, knowledge graphs enable models to reason over explicit entities and relations, align responses to verifiable facts, and offer interpretability via provenance metadata [87,88].

This integration allows models not only to retrieve precise domain-specific information but also to construct explainable inference chains, thus improving transparency and trust. Moreover, the inclusion of knowledge graphs supports granular and contextually relevant knowledge retrieval, outperforming naive augmentation with random or noisy data [89]. By aligning entities and relationships from LLM outputs with ontologies and knowledge bases, systems can avoid shallow text pattern recognition and instead achieve conceptual generalisation and symbolic reasoning.

This approach also opens the door to continuous learning pipelines, where AI outputs are validated against, or enriched by, structured data and subsequently reintroduced into the model in the form of feedback loops. These loops allow AI systems to iteratively refine their knowledge representations, adjust for emerging information, and correct previously observed errors. Combined with recent advances in multimodal and temporal input integration—e.g., linking structured data with images or spatial metadata [68]—the use of knowledge graphs is becoming central to the development of knowledge-enriched, context-aware AI systems that are both accurate and adaptable. As structured data continues to permeate all domains of life, designing robust mechanisms to represent and integrate this data into LLMs remains a crucial and active research frontier.

### 5.2. Embedding Cultural Heritage Knowledge Graphs

To date, much of the literature has concentrated on the construction of knowledge graphs, focusing on entity recognition, ontology alignment, and data integration with ontologies such as CIDOC CRM [90,91]. More recently, however, researchers have begun to

investigate the potential of graph representation learning as a means of exploiting these graphs for retrieval, enrichment, and semantic discovery.

A notable example is the CIDOC2VEC approach [92], which develops embeddings from CIDOC-CRM-compliant graphs. The method employs relative sentence walks through the KG to generate sequences of entities, which are then embedded using standard distributional techniques from natural language processing. These embeddings enable similarity-based retrieval and recommendation of cultural heritage objects across collections. CIDOC2VEC represents a significant advancement because it demonstrates how embeddings can facilitate semantic interoperability across institutions, allowing users to discover related artefacts even in the absence of explicitly asserted links.

Nevertheless, a critical limitation of CIDOC2VEC is its entity-centered, relation-agnostic design. By treating walks as linear sequences of entities, the approach largely ignores the semantic role of relations defined in CIDOC CRM, such as the distinction between *P108 has produced* (linking production events to artefacts) and *P50 has current keeper* (linking artefacts to institutions). While this simplification improves scalability, it results in embeddings that capture contextual co-occurrence but lack sensitivity to the multi-relational structure that is central to CRM's expressiveness. Consequently, retrieval is biased toward entity similarity rather than relation-aware reasoning, limiting its ability to support complex queries (e.g., retrieving artefacts created in the same workshop but preserved in different sites).

In contrast, recent advances in graph neural networks (GNNs), and particularly composition-based graph convolutional models such as CompGCN [93], explicitly model both entities and relations during the embedding process. By applying relation-specific transformations during message passing, CompGCN embeddings encode not only which entities are related but also how they are related, producing richer and more semantically faithful representations. In mainstream KG domains (e.g., Freebase, DBpedia), this relation-aware modelling has been shown to significantly improve link prediction and semantic querying [93,94].

Yet, from the review carried out in this research work, relation-aware graph convolutional models have not been applied in the querying phase of cultural heritage KGs, particularly those structured with CIDOC CRM. Existing approaches remain entity-focused, prioritizing similarity-based retrieval (as in CIDOC2VEC) or classification tasks (e.g., GCN-Boost [95]) rather than supporting relation-aware querying and ranking. This presents a clear research gap.

This study addresses said gap by exploring the use of CompGCN embeddings to enhance querying in the RHDTO grounded knowledge graph. Rather than focusing exclusively on graph construction or classification, a proposal is extended to apply embeddings post-construction in the querying phase, enabling relation-sensitive similarity scoring across entities and events. This contributes a novel paradigm for semantic search and exploratory browsing, bridging the rigour of ontology-driven reasoning with the flexibility of embedding-based retrieval.

### 5.3. Access to Heritage Data in Natural Language

As the scale, scope, and complexity of data increase, the need to simplify its querying methods does so equally. Past a time where only select individuals could access certain information thanks to their expertise in the topic and research skills, information is now at everyone's reach, a half-thought and a click away. Not only archaeologists, scientists, or other specialists may wish to browse the knowledge base, so in the context of Heritage Digital Twins, traditional querying methods would rather prove obsolete in the face of approaches utilising natural language as a medium [96].

If a user expresses their inquiry in their own language, the translation layer—natural language understanding (NLU)—bridges the gap between human intention and structured semantic content [97]. The query gains a certain level of interpretability, relaxing constraints and allowing for fuzzy reasoning during retrieval across both structural and semantic levels [98]. Not only does this present an advantage in query guarantee (retrieving at least one result) but also provides users enough feedback to adjust and shape their queries and best attract the answers they seek within the intricate knowledge graph.

At the core of this capability is a combination of tasks including semantic parsing, entity linking, intent classification, and relationship inference to decompose and interpret the natural language queries [99]. Thanks to recent advances in prompt-based large language models, using separately trained components for each of these tasks is no longer a requirement, as LLMs offer an integrated, zero, or few-shot approach to parsing natural language [100]. In the context of cultural heritage, where user queries are often imprecise, explanatory, and laden with historical, geographical, or culturally specific terms, properly instructed LLMs can produce a variety of structured outputs [77]. The LLM prompt construction includes the general context of the query to focus the interpretation, indications on the what and how of information extraction, and any additional enrichment LLMs should apply to their output. As such, the latter can be in the form of data triples—the building blocks of a graph structure [98].

Intuitively, these triples would be converted into SPARQL fragments to filter the knowledge graph data by mentioned entities, and the results would be layered according to the specific query relationships. The SPARQL query, then, may only be constructed after correct assignment of ontology labels and properties to the processed query elements—a task which would either require the involvement of an expert for labelling and validation or the implementation of an additional layer of entity and relationship classification using an ontology-trained AI model [101]. However, the usage of multiple hierarchical ontologies, one more complex than the other, renders this task more complicated than not, and while the latter option is more viable, the rigidity of SPARQL querying remains a risk; in the case of misclassification, the query will yield no results although they might exist [102].

To address this challenge, a hybrid query execution is adopted that bypasses strict SPARQL-based querying and instead leverages graph neural networks and knowledge embeddings, allowing the system to execute natural language queries in a semantically flexible and structurally tolerant way, specifically via Composition-based Graph Convolutional Networks (CompGCNs).

Following the NLU stage, the nodes represent candidate entities, types, or conceptual elements (e.g., “cathedral”, “bronze”, “Charles II”), and edges represent inferred relations or semantic dependencies (e.g., “located\_in”, “issued\_by”, “made\_of”). The constructed query graph represents the user’s intent directly, without prematurely enforcing alignment to any specific ontology schema, while capturing its semantic and relational structure.

For Heritage Digital Twins, similarity extends beyond the simple attribute matching but rather enters the world of semantic, structural, and contextual equivalence. It is rare—if at all possible—for artefacts, historical and archaeological remains, or intangible heritage to be identical. Instead, they show overlap in historical context or provenance, form, function or symbolism [103], all types of data that are hosted and structured in the knowledge graph according to the RHDTO.

For tangible heritage, this work investigates similarity computation at the following levels, for example, to identify:

- pieces that belong to the same object,
- different objects that were excavated at the same site but from different locations,
- different objects that were crafted in the same workshop,

- different objects made of material originating from the same place; inferred from the data of scientific analysis reports,
- different objects made by the same artist.

This demonstrates the great advantage of similarity computation in querying, since thus far in existing works, most of these results would not be retrieved by a basic querying system alone. In fact, some of these objectives require reliance on visual input [104] such as images or 3D models, but by using intelligent AI models, it may be possible in most cases to identify such similarities just based on metadata stored within the knowledge graph [105].

## 6. Case Study: Enriching Digital Twins with Heritage Science Data

As a testbed for exploring the integration and synergy between AI techniques and ontological modelling, some reports generated by analysis activities in the heritage science field were selected, part of the documentation of the ARTEMIS initiative [106].

### 6.1. Application of AI-Driven Semantic Pipelines

A modular AI-based pipeline was designed to extract structured semantic knowledge from unstructured scientific texts belonging to the cultural heritage domain, with a focus on modern preservation and restoration techniques such as, but not limited to, X-Ray Fluorescence (XRF) and 3D modelling techniques.

The system is particularly tailored to support the enrichment and population of an ontology-driven knowledge graph designed to be an efficient base for the construction of Heritage Digital Twins, but its architecture is flexible enough to support other knowledge-driven applications.

The defined methodology combines several state-of-the-art open-source technologies, including Natural Language Processing, Large Language Models, semantic similarity search, and ontology-based reasoning.

Textual data was primarily sourced from the ARTEMIS initiative documentation, consisting of complex PDF files with multi-column layouts, embedded images, tables, and special characters. The preprocessing and chunking pipeline consisted of the following steps:

- PDF parsing and image extraction: text and images extraction has been performed with PyMuPDF [107], a Python (Python 3.12.7, PyMuPDF 1.25.5) library able to handle even complex textual layout. Images from PDFs were also extracted and kept separately for further experiments.
- Text preprocessing: SpaCy [108] library has been used to perform tokenisation, sentence segmentation, and syntactic parsing to reduce noise and generally improve the handling of textual data by machines.
- Sentence-aware chunking: to enable efficient processing by Transformer-based LLMs, the texts were then segmented into smaller, semantically coherent units called “chunks”. Sentences were aggregated sequentially until the token limit (in this experiment the limit was 300 token) was reached to avoid splitting syntactically or semantically dependent clauses. This size has been chosen to balance the model’s context window constraints with our need to have meaningful textual context. This segmentation allows the model to focus on localised contexts, ensuring that the extracted knowledge is specific and grounded in the source material.
- Text normalisation: lowercased all text, removed punctuation, standardised whitespace, and ensured consistent character encoding for compatibility with embedding models.

Once the text chunks were generated and normalised, the pipeline employed a combination of embedding-based semantic analysis and instruction-tuned language modelling to

perform ontology term extraction and knowledge graph population. Candidate ontology terms extracted from each chunk were first encoded into dense vector representations using the SentenceTransformer model all-MiniLM-L6-v2 [109]. A pairwise cosine similarity matrix was then computed across all term embeddings to identify semantically similar terms. Terms which exceeded a chosen similarity threshold of 0.85 were clustered together, effectively collapsing synonyms and variant spellings into unified conceptual entities. The pairwise cosine similarity matrix is given by:

$$S = D^{-1}XX^{\top}D^{-1}, \quad (1)$$

where

$$D = \text{diag}(\|\mathbf{x}_1\|, \|\mathbf{x}_2\|, \dots, \|\mathbf{x}_n\|). \quad (2)$$

This semantic deduplication approach allowed for a more robust normalization than conventional string-matching methods, ensuring terminological consistency across the corpus.

Term extraction was then performed using Mistral 7B Instruct [110], an instruction-tuned LLM selected for its strong performance and open-source transparency, which guarantees the reproducibility of its outputs.

The prompting strategy was explicitly designed to extract only the entities and concepts relevant to the cultural heritage domain.

Each chunk of text was provided to the model with instructions to identify modern scientific analysis techniques, materials, instruments, and experimental procedures while excluding any narrative or explanatory content.

Prompts were crafted to elicit concise, list-form outputs, containing only the terms of interest. By carefully structuring these prompts, the model could perform zero-shot extraction of domain-specific ontology terms without relying on supervised training data, ensuring that each output was both focused and reproducible.

To ensure the reliability of the extracted terms, a human-in-the-loop validation step was implemented, allowing domain experts to manually review and select relevant terms. This hybrid approach balanced automated efficiency with expert oversight, mitigating the risk of hallucinated or irrelevant outputs while maintaining the accuracy of the ontology term set.

A key innovation of this pipeline lies in the retrieval of semantic context for each validated term: while processes like Term Extraction or Named Entity Recognition successfully provide a list of relevant concepts, their meaning is often best understood in relation to the text from which they originate, and this is especially true in our case, where the data needs to be mapped following domain ontologies rules.

To capture this context, both the refined ontology terms and the original text chunks were embedded using the same SentenceTransformers model, and a FAISS similarity index was constructed over the chunk embeddings.

FAISS [111] is an open-source library for efficient similarity search and clustering of dense vectors, optimised for high-dimensional data, such as embeddings from neural networks. The way it works is by returning *k nearest neighbors of a query vector*, typically with their indices and distances/similarity scores.

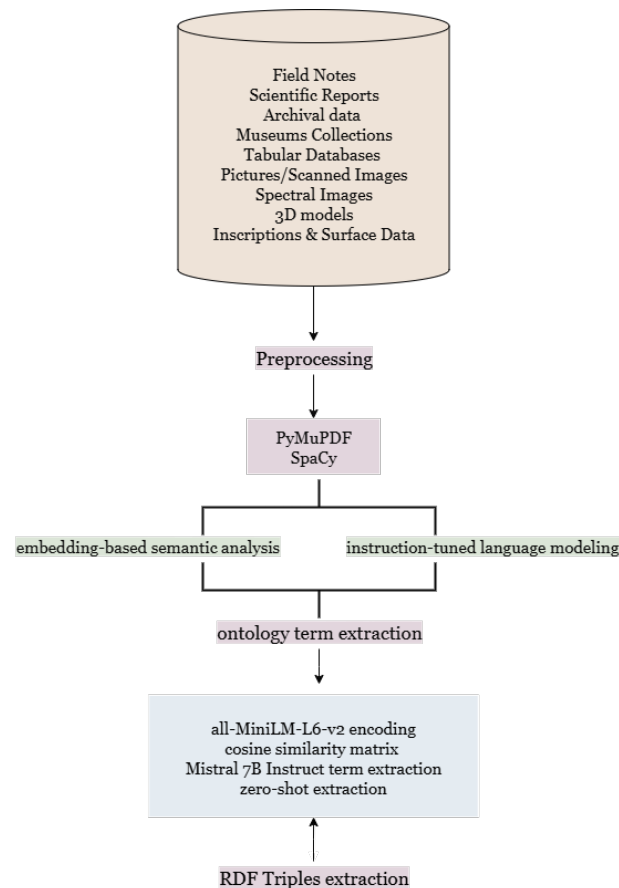
With the relevant contextual chunks retrieved, Mistral 7B Instruct was employed a second time to generate RDF triples.

The model was prompted to extract concise semantic triples aligned both structurally and semantically with existing cultural heritage ontologies. This resulted in a set of triples that formally represents relationships between terms while preserving traceability to the source text.

The final outputs of this AI-assisted pipeline include:

- a validated list of ontology-relevant terms, stored in JSON format,
- a collection of RDF triples, structured and domain ontology-aligned,
- a mapping between each term and its original context, ensuring interpretability and traceability.

Figure 1 is a diagram describing graphically the aforementioned pipeline:



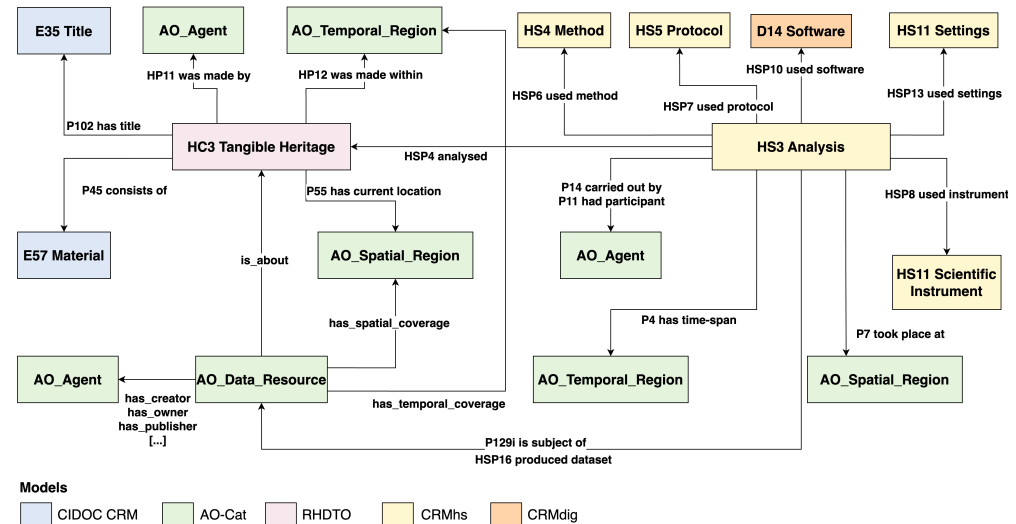
**Figure 1.** Process diagram pt.1: automatic ontology term and RDF triples extraction from cultural heritage reports.

The underlying conceptual model chosen for this effort is CRMhs [52], an ontological component of the CIDOC CRM ecosystem developed to represent the documentation and operational structure of scientific analyses carried out on cultural heritage objects. As an extension of the CIDOC CRM, CRMhs integrates seamlessly with RHDTO and gives it the ability to formalise the domain of heritage science by introducing specific classes and properties that describe scientific activities (such as *HS3 Analysis*), instruments (e.g., *HS11 Scientific Instrument*), datasets, sampling procedures, and their relationships to both the physical object under investigation and the broader research context. It supports the representation of tangible and intangible entities involved in the scientific process, such as samples, areas of interest, methods employed, and results obtained. The model also accounts for the configuration of complex devices, their software components, and even the parameters used during analysis. Moreover, since the ARTEMIS ontological architecture is intentionally modular, reflecting the diversity and complexity of the domains it seeks to represent, the project also employs the AO-Cat ontology [112] developed within the ARIADNE initiative [113] to describe datasets and their specific characteristics, including provenance, access conditions, and content typologies. Our encoding experiments, also



using AO-Cat to encode these entities, have fully demonstrated how these models are able to interact perfectly with each other, allowing the construction of semantic graphs that, despite being internally heterogeneous, constantly maintain a solid semantic coherence.

A general overview of the ontological module designed by ARTEMIS to model scientific data is shown in Figure 2.



**Figure 2.** Classes and properties used by ARTEMIS to model heritage science information.

By anchoring AI outputs to a domain-specific ontology, the system ensures that extracted data is both semantically coherent and interoperable. This ontological framework thus provides the structure for an AI-assisted information extraction workflow aimed at transforming heritage science reports into semantically enriched data. Using the categories defined by this semantic ecosystem as guidance, natural language processing tools are applied to identify and annotate references to analytical techniques, measurements, instruments, and related entities. The extracted knowledge is then converted into structured triples, for instance, in Turtle format, encoded using the grammar provided by the ontological model employed and ingested in the knowledge graph.

The following examples show some results of the pipeline tested in [114] about the use of Macro X-ray Fluorescence Scanning on a Raffaello's Portrait of Leo X. Starting from the original PDF, the JSON file containing the triples was successfully and automatically extracted, encoded in Turtle format, and ingested into the ARTEMIS knowledge base to visualise the graph. A fragment of the Turtle encoding (below) shows the ontological representation of Raffaello's painting as a cultural object by means of the *HC3* class of the RHDT Ontology.

```

artemis:LeoX_Portrait_Study_Object a rhdto:HC3_Tangible_Entity ;
  rdfs:label "Portrait of Leo X" ;
  owl:sameAs <http://www.wikidata.org/entity/Q5597> ;
  crm:P102_has_title "Portrait of Leo X with Cardinals" ;
  rhdto:HP11_was_made_by artemis:Artist_Raffaello ;
  crm:P45_consists_of "Oil on wood" ;
  rhdto:HP12_was_made_within artemis:LeoX_Portrait_Period ;
  crm:P55_has_current_location artemis:LeoX_Portrait_Place .

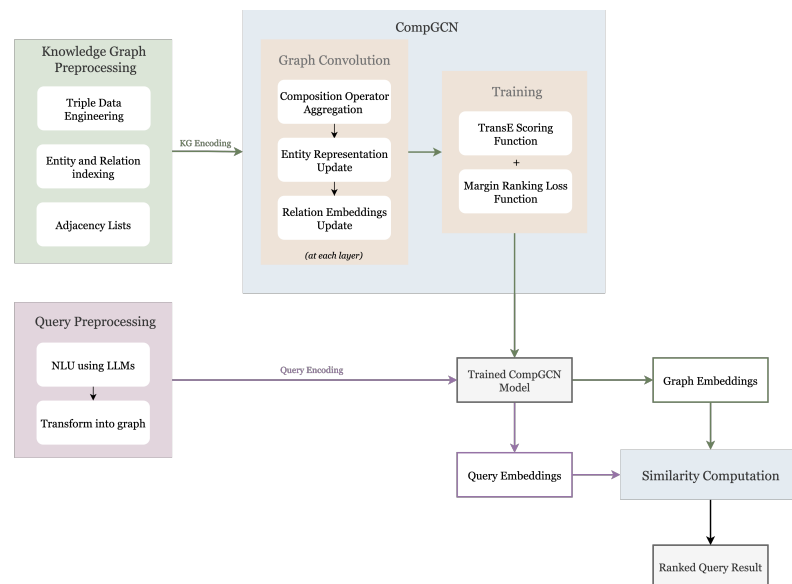
```

The following Turtle fragment, instead, shows the encoding of the analysis event conducted by scholars of the Italian Institute of Nuclear Physics (INFN) on the same cultural object, the way in which CRMhs properties are used to specify the instrumentation



## 6.2. Populating and Querying the ARTEMIS Knowledge Graph

The following subsections describe the second phase of the experiment pertaining to the querying of the knowledge graph, for which the steps are illustrated in the process diagram in Figure 4.



**Figure 4.** Process diagram pt.2: querying the KG in natural language using CompGCN embeddings.

Once the HDT knowledge graph has been built and the cultural heritage data is semantically structured, the next step is implementing the querying system, which allows us to access it. This process involves a few steps, beginning with an additional preprocessing stage. This time, we are not preprocessing textual documents but rather preparing the stored data for artificial intelligence model training by indexing each entity and relation with a unique ID and engineering the graph data as described next. The ontology-based graph structure resulting from the preceding work maintains the information in a parametrised way, linking all attributes through property edges, including the node types (e.g., by means of the *P2 has type* property of CIDOC CRM). It is good practice for data archival, consistency, and reusability; however, it poses particular challenges for an AI model, as then the nodes do not encapsulate enough meaningful information to distinguish them from others. To address this, the knowledge graph is restructured so that each node in the graph corresponds to an instance of an RHDTTO class. Therefore, the type of each entity is embedded in its node as a static feature, along with any human-readable label, title, or description it is linked to, for all of which simple embeddings were created using SentenceBERT. An example of the resulting node in JSON format is the following:

```

{ "id": "obj_0001",
  "label": "Mona Lisa",
  "type": "E22_Man-Made_Object",
  "features": {
    "label_embedding": [0.23, 0.78, ...],
    "type_embedding": [0.12, -0.09, ...] }

```

The next step of the process is to feed the new knowledge graph to an AI model and begin training. Though, grounding the knowledge graph in the rich RHDTTO makes it too complex for standard neural networks to capture its nuanced relational structure of data. The selected model, CompGCN, explicitly models both nodes and relations jointly. For cultural heritage data where each connection has a semantic meaning that depends on the

nature of the relationship, it is critical to be able to make a difference between—for example, whether a person *created* an object, *participated in* an event, or *curated* an exhibition [93]. CompGCN looks at each of these relations as its own learnable element and trains to learn their differences. It also accounts for the directionality of relationships; in the knowledge graph, many relationships are inherently directional and semantically loaded, as there is an asymmetry in reading them (e.g., “was influenced by” and “influenced”). The network learns this by identifying the roles of the relations and the flow of information between nodes [93].

Implementation-wise, the graph is transformed into triple format or adjacency lists, and each relation is modelled explicitly with its inverse counterpart. CompGCN then uses a message-passing mechanism that jointly embeds nodes and relations to spread information throughout the graph during training. The model uses composition operators on the embedding vectors to integrate relation semantics into the neighbourhood aggregation, which allows the network to model logic and semantics of the paths traversed additionally to their proximity. Node embeddings are updated at each graph convolution layer according to the unique semantic composition of the connecting relation and the characteristics of their neighbours [93].

### 6.3. Message Passing in CompGCN for Cultural Heritage Knowledge Graphs

In CompGCN, the message passing mechanism extends traditional GCNs by explicitly incorporating relation semantics [93,94]. For each triple  $(h, r, t)$ , the neighbour embedding  $\mathbf{t}^{(k)}$  is composed with the relation embedding  $\mathbf{r}^{(k)}$  using a composition operator  $\phi$ , e.g.,

$$m_{h \leftarrow (t,r)}^{(k)} = \phi(\mathbf{t}^{(k)}, \mathbf{r}^{(k)}), \quad (3)$$

where  $\phi$  can be addition, multiplication, or circular correlation. Each entity representation is then updated by aggregating over its adjacency list:

$$\mathbf{h}^{(k+1)} = \sigma \left( \sum_{(r,t) \in \text{Adj}(h)} W_r^{(k)} \cdot \phi(\mathbf{t}^{(k)}, \mathbf{r}^{(k)}) + W_0^{(k)} \mathbf{h}^{(k)} \right), \quad (4)$$

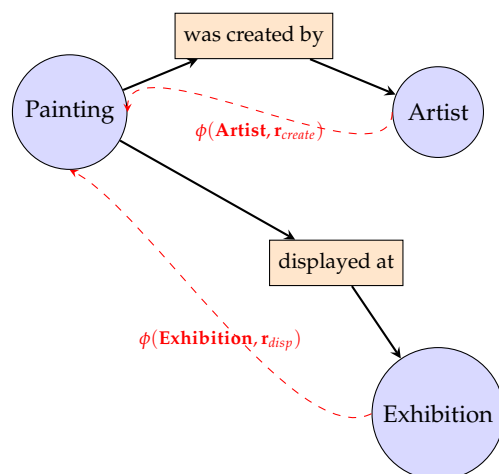
where  $W_r^{(k)}$  are relation-specific weight matrices and  $\sigma$  is a non-linear activation. Unlike earlier relational GCNs, CompGCN also updates relation embeddings:

$$\mathbf{r}^{(k+1)} = W_r^{(k)} \mathbf{r}^{(k)}. \quad (5)$$

This joint update allows both entities and relations to evolve during training, which is particularly relevant in cultural heritage graphs where relations (*was produced by*, *participated in*, *located at*) carry rich semantics. For example, the embedding of an entity such as a *Vase* is updated not only from its neighbouring *Potter* (*was produced by*) but also from its participation in a *Firing Event* [115]. After three layers, entity and relation embeddings are evaluated with a TransE scoring function [116]:

$$f(h, r, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2, \quad (6)$$

and optimised using a margin ranking loss to ensure valid triples score higher than corrupted ones. This allows embeddings to capture both contextual and relational meaning, enabling similarity-based querying of heritage data. The following schema in Figure 5 illustrates this aggregation.

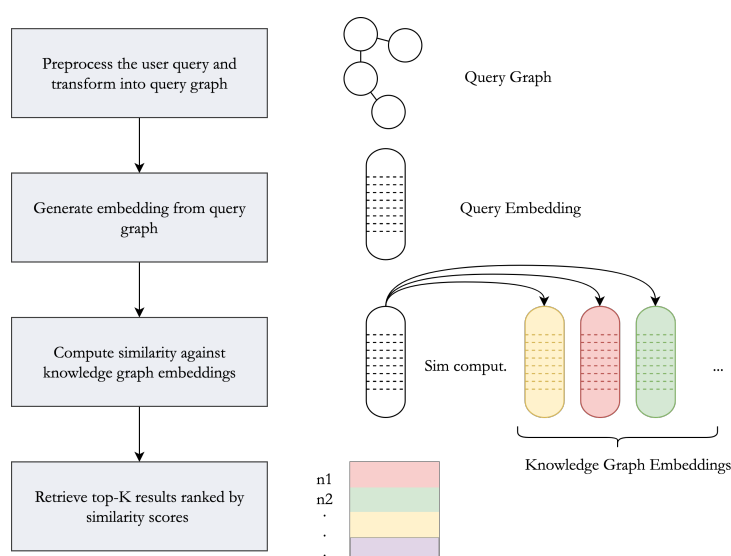


**Figure 5.** Schematic of CompGCN message passing in cultural heritage knowledge graphs. Example: *Painting* aggregates messages from *Artist* and *Exhibition*. Dashed red arrows indicate composed messages being aggregated into the entity embeddings.

#### 6.4. Similarity Computation Between the Query and KG Embeddings

Our system uses vector embeddings to compute similarity between cultural entities, taking into consideration temporal alignment, relational proximity, and typological affinities. These computations involve a combination of two measures between the embedded graph structures: the cosine similarity, which works by calculating the cosine angle between two vector representations (though it is generally used for high dimensions dense vector comparisons), and the contrastive distance learning, which improves the clustering of semantically similar subgraphs and disperses the unrelated ones [117,118].

After passing three layers of graph convolution, the node embeddings encode rich enough structural and semantic information to perform tasks such as similarity computation: once the user query is transformed into a graph structure with the LLM, it is passed through the same CompGCN encoder to generate a comparable embedding in the same latent space as that of the knowledge graph. Cosine similarity is then computed between this query embedding and candidate entity or subgraph embeddings from the knowledge graph to find semantically relevant matches even in the absence of exact alignment (the candidates are scored in an interval of 0 to 1; 1 signifying high similarity). Figure 6 visualises this process into a simplified diagram.



**Figure 6.** Diagram of the querying process using CompGCN embeddings and similarity computation.

### 6.5. Design and Implementation of the Prompt-Based Interface

Heritage Digital Twins are powerful archives, and alongside the extensive effort to improve the methods by which they are queried logically, these improvements must also be made accessible to users. The chosen approach aims to allow natural language as a querying medium, and as such, a prompt-based user interface was developed instead of the traditional, rigid hierarchy of filters or drop-down UIs. The concept is that of a chat environment for users to interact with the knowledge graph in a conversational manner. A key design consideration has been quality feedback and transparency. The model response must be synthesized and clear, allowing the user to understand how their prompt was parsed and interpreted. This is most valuable when the user's query is ambiguous or lacking specificity. For example, a user query saying "Find me Greek figures from the classical period" would yield the response "Finding Greek historical personalities from the classical period. Would you like me to expand my search to mythological entities or sculptures and statues?".

This points to another important element; iterative refinement. Following the model's reasoning during the query would let the user, therefore, correct or refine the system's interpretation of their intent. This requires the maintenance of a persistent conversation context that the system can always refer back to during the validity of dynamic session state, stored in an in-memory session store with a unique identifier passed with each message, after each of which the prompt context and history are updated and re-evaluated during the LLM convocations. The implementation of this feature further guarantees accessibility across expertise levels, as even unstructured queries can be evolved into their most effective form through a co-constructed exploration [119]. A well-rounded response is also often appreciated by non-experts if it provides additional information that might entice further interests in the topic at hand. For this, the LLM is also tasked to include adjacent context, or broader data summaries, informed by the knowledge graph into its response. This can be synthesised from the subgraphs scoring slightly lower in similarity against the query graph and should be clearly specified as such [119].

### 6.6. Presenting Complex Query Results in User-Comprehensible Formats

After a successful query, a sorted list of high-scoring entity or subgraph candidates for the given user prompt is obtained, and the task is then presented to the user in a human-readable layout. First, the LLM interprets the result data, searching for ways to highlight important information, categorise the entries, or provide contextual summaries where better suited; then, it returns it to the user in narrative format, as seen in Figure 7.

When dealing with large knowledge graphs such as the HDT's, it is crucial to avoid clutter, filter through data, and present only the salient information initially. Progressive disclosure balances usability with transparency for the sake of both casual exploration and scholarly validation. Our current implementation of the UI presents the system's response in a narrative format describing in plain text its findings, structured, listed, and categorised as needed by the LLM for better digestion of the information. Text-based responses are suitable for any type of queries, from precise search to exploratory inquiries. They can be concise and direct but can also weave together complex temporal and geographic information, benefiting from explanation and nuance to paint a holistic picture for the user [120]. Not to suggest that text-based narration is the only format supported in chat-UIs; on the contrary, an important advantage is that they are extensible. In the context of HDTs, more structured or even visual elements would render the response more impactful, and thus the interface can be upgraded to include 3D model displays, images and illustrations, geographic and temporal maps, tables, diagrams, and raw graph structures depending on



the user’s needs. Ultimately, presenting complex query results is not just a matter of UI feature design but a mediation between semantic accuracy and human cognition.

What would you like to know?

Find information about coins minted in Famagusta

Send

💡 HDT’s Response:

Famagusta, Cyprus.

### Coins of the Lusignan Kingdom of Cyprus (1192-1489 AD)

During the Crusades, Famagusta was part of the Lusignan Kingdom of Cyprus. The Lusignans issued gold coins known as Gros and silver coins known as Deniers. Below are some examples:

#### Gold Gros (Cyprus, 1376-1401 AD)

The obverse of the gold gros features a bust of Saint Christopher holding the Christ child on his shoulders. The reverse shows the Lusignan royal arms with three fleurs-de-lis and two lions rampant.

#### Silver Denier (Cyprus, 1376-1401 AD)

The obverse of the silver denier depicts a bust of Saint Catherine of Alexandria with a palm branch and a wheel in her hands. The reverse displays the Lusignan royal arms as on the gold gros.

### Coins of the Venetian Republic (1489-1571 AD)

After the fall of Famagusta to the Ottomans in 1571, it became part of the Venetian Republic. The Venetians issued coins with various designs for different regions under their control. Below are some examples specific to Famagusta:

#### Silver Grosso (Famagosta, 1492-1503 AD)

**Figure 7.** Snapshot of the functional chat-UI demonstrating a user prompt and the query results in narrative format.

## 7. Discussion: Taming the Wild Beast

The hybrid querying approach described above clearly outperforms strict SPARQL translation in several important dimensions. First, it offers greater flexibility when dealing with user queries that are incomplete, underspecified, or expressed in natural language, such as “old buildings with mythological carvings”, where formal ontology mapping would struggle or fail entirely [121]. The system provides meaningful results even in the absence of precise predicates or labels by interpreting semantic intent and mapping it onto the knowledge graph structure instead of depending on exact term matching. Secondly, explainability is supported by the architecture. The query graph can be visualised, annotated, and even interactively refined by creating an intermediate graph representation for every query. Because it makes clear which entities and relations are being taken into consideration [102], this transparency aids users in understanding how their prompts are interpreted. Third, the method allows for gradual improvement. The system can mimic a human-like dialogue by asking clarifying follow-up questions, such as whether “Hellenistic” refers to a time period or an artistic style, in order to gradually refine the search instead of returning a null result because of strict schema constraints [77]. Moreover, the system empowers prompt engineering and context enrichment. A prompt like “Find ancient settlements near rivers active during early agriculture” can be interpreted by the LLM and enriched with relevant temporal ontologies and spatial reasoning constructs

then structured into a graph-aware query. This graph is then embedded by CompGCN, which jointly encodes nodes and relations to support matching based on structure and context [93].

### 7.1. The Indispensable Role of Human Expertise in AI-Assisted Processes

Despite the growing capabilities of AI models to process, interpret, and retrieve information from both structured and unstructured data, human expertise remains a key part of any system that engages with rich cultural knowledge [101,122].

In AI-assisted querying over cultural heritage knowledge graphs, human input is not just an addition; it is essential for the reliability, interpretability, and contextual relevance of the whole system [101]. AI models, including prompt-based large language models (LLMs) and graph neural networks like CompGCN, are fundamentally statistical and generalising systems. Their ability to infer structure, relevance, and semantic similarity is limited by the quality and coverage of their training data and the ontological framing in the knowledge graph. Cultural heritage data is inherently varied, deeply contextual, and often only partially curated. This field includes ambiguities, historical gaps, and interpretive uncertainties that cannot fully be resolved by algorithms alone [101]. Human expertise is vital for curating representative datasets, designing prompt templates, aligning graph structures with domain ontologies like CIDOC CRM, and interpreting model outputs with historical context in mind [101,122]. Moreover, domain experts are crucial for assessing and improving the outputs of AI models. For example, interpreting a user query like “religious sites with serpent imagery” may involve understanding cultural motifs, symbolic meanings, and changes in iconography over time that are not clearly included in the ontology. In these situations, an expert is needed to confirm the system’s graph-based matches, identify semantically inappropriate connections, and help refine the query or results [101]. This ongoing feedback not only enhances the user experience but also supports the long-term improvement and strength of the AI model, especially when the feedback is included in training processes [122].

In the case of unavailable human expert validation, human input is still not disregarded. Cyclical processes in which the predictions or outputs of an AI model are evaluated by both humans and automated systems are called feedback loops [101,122] and are developed in AI systems to involve the user through feedback actions regarding AI results and suggestions. This feedback is turned into structured input that can enhance the underlying knowledge graph or fine-tune the model while also being used to evaluate performance metrics. In accepting, rejecting, or editing said recommendations, there is valuable human insight that the system translates into learning signals, and this information is captured as labeled training data.

The backend of the semantic pipeline can include modules for continuous or active learning, which allows the AI model to be retrained on new datasets that consider user input. Active learning techniques can help prioritise the most impactful or unclear cases for human review, making expert time more effective [101,122].

Error analysis is an important part of feedback loops. Developers can identify model blind spots or weaknesses in the training data by investigating systematic errors, such as frequent incorrect classifications or concepts that are not linked. In addition to model retraining, these insights can guide improvements to user instructions, ontology design, and data curation practices [101].

Additionally, explainability, which is vital for building trust in AI systems, relies on human interpretation. While neural architectures can generate relevance scores or attention weights, only experts can turn those signals into valid domain interpretations, such as explaining why a particular Hellenistic inscription appeared in response to a vaguely

formulated mythological query [123]. Their knowledge helps bridge the gap between the model's behaviour and human cultural understanding. Ultimately, the relationship between human expertise and machine intelligence creates a hybrid form of intelligence that is particularly effective in cultural heritage contexts. In this scenario, machines can enhance human reach and recall across extensive data collections, while humans provide interpretation, critique, and contextual relevance. This partnership ensures that AI tools do not simply replicate patterns but also contribute meaningfully to the advancement of knowledge, honoring both academic rigour and cultural sensitivity [124,125].

### 7.2. *Ethical Implications and the Pursuit of Transparent AI in Cultural Heritage*

As artificial intelligence systems play a larger role in accessing and interpreting cultural heritage data, important ethical issues need to be addressed, ensuring that these technologies uphold values such as transparency, accountability, and cultural sensitivity. Unlike traditional datasets, cultural heritage knowledge bases include narratives, identities, and power dynamics that reflect human histories and ways of knowing. Therefore, using AI in this area poses risks that go beyond technical bias, leading to issues such as historiographic distortion, cultural erasure, and unfair knowledge representation [126]. One major ethical challenge comes from the lack of clarity in current AI models, especially large language models and graph-based neural systems like CompGCN. While these systems can generate or retrieve rich context responses, they operate like “black boxes.” This makes it hard for users and curators to trace where a specific output comes from or to understand how a query match is made [127,128]. In cultural heritage applications, this lack of transparency is particularly concerning. Users may unknowingly receive results influenced by hidden biases in the data, such as Eurocentric narratives, underrepresentation of indigenous knowledge systems, or overly simplistic event interpretations outlined in systems like CIDOC CRM [126,129]. To reduce these risks, AI interfaces in this field should focus on transparency. This means developing interpretable structures, like graph-based representations used in query transformations, that users can visualise and annotate. Keeping a clear record of how data transforms from natural language input to query graph, and then to retrieved subgraphs, allows scholars and curators to review the model's behavior and ensure accuracy [127]. Additionally, creating feedback systems that let users question or adjust AI outputs is crucial for maintaining human oversight and encouraging discussion [127]. Another key ethical issue is supporting diverse perspectives. Cultural heritage is not just a fixed set of facts; it is a dynamic field full of interpretation, often featuring contested meanings and various community viewpoints. AI models should be taught and assessed with awareness of these complexities. This involves using inclusive training materials, enabling multilingual and multicultural querying options, and making sure minority narratives are not oversimplified or ignored by dominant patterns [130,131]. Transparency in how models are developed, including clear documentation of data sources, design choices, and evaluation methods, helps prevent accidental cultural oversimplification [126,132]. Finally, using AI in this area must align with legal and ethical standards concerning data ownership and management. Cultural artefacts, especially those obtained through colonial means or held in sacred regard, require careful governance when they are digitized and processed by algorithms. Transparent and explainable AI should not only reveal its inner workings but also honor the rights and values of the communities whose heritage it engages with [126,133].

## 8. Conclusions and Future Work

The experiments described in this paper offer promising results, particularly in their capacity to simplify human tasks and provide tools that accelerate the extraction, organisation,

and use of cultural knowledge. What emerges is not simply a more efficient pathway from text to data but a recursive system in which knowledge begets cognition, and cognition in turn refines the very tools through which knowledge is acquired. This methodological loop, anchored in ontological rigour and empowered by machine learning, reconfigures both the means and the ends of heritage data processing, transforming interpretation into interaction.

These outcomes represent a fundamental step towards the realisation of a fully developed Cognitive Heritage Digital Twin. As noted previously, providing a full account of the creation of an entire Cognitive Heritage Digital Twin would require a much larger study given the immense scale and complexity involved. The objective of this paper, therefore, was to exemplify some of the most immediately impactful trajectories currently emerging, centred on the use of artificial intelligence and ontologies for the retrieval, organisation, and strategic deployment of knowledge embedded in cultural heritage documentation. The presented AI-assisted semantic pipeline has been experimentally validated as a robust and efficient framework for knowledge extraction, laying the groundwork for more advanced functionalities. Future developments will aim to refine and extend the system by enhancing the accuracy of entity disambiguation, incorporating multimodal inputs such as images and tabular data, and introducing feedback loops that allow curators and domain experts to validate and iteratively improve the extracted knowledge while also enhancing the models being used.

As cultural heritage datasets span a broad spectrum of visual media, which is often heterogeneous, physically degraded, or embedded within ambiguous symbolic systems, it is important to also take those into account. Some promising zero-shot classification testing has already been conducted on images using open-source Vision Language Models like OPENClip [134] to automate the extraction of analysis types and semantic meaning to fuel Digital Twins and Knowledge Bases. This line of research will be further investigated to fine-tune the AI models and boost recognition of domain-specific figure types. Future plans will also aim to streamline the integration between natural language queries and graph-based responses, enabling more fluid and transparent interactions with the underlying knowledge base. These enhancements will not only increase the robustness of the extraction process but will further consolidate the cognitive capacities of Heritage Digital Twin, aligning their growth with the evolving needs of cultural heritage research. This entails not only refining the underlying semantic infrastructures but also designing more intuitive, transparent, and context-aware interfaces that allow scholars and practitioners to explore, interrogate, and contribute to the knowledge base with greater ease. By empowering both the machine and the human interlocutor, a more dialogic relationship with cultural data is envisioned, in which digital twins are no longer static repositories but dynamic environments for discovery, reasoning, and collaborative interpretation.

Particular attention will also be devoted to the applications of the AI–ontology tandem in the development of decisional components of Heritage Digital Twins, enabling them to react meaningfully to sensor-derived inputs and to activate appropriate, context-aware responses. Beyond decision-making and reactive behaviour, further functionalities will be explored, such as predictive modelling, simulation of possible conservation scenarios, and projection of future states of cultural assets under varying environmental or human-induced conditions. These developments will draw on the inferential and connective power of semantic graphs, which provide the epistemic infrastructure necessary to support such complex reasoning processes. This approach echoes the insight of Heraclitus, who warned that “much learning does not teach understanding” (Fragment B 40.). It is a powerful reminder, now as then, that true knowledge arises not from the mere accumulation of data but from its structuring, contextualisation, and meaningful integration. Our work

constantly strives to answer this call: to move beyond erudition towards insight; beyond cataloguing towards comprehension.

Naturally, all future enhancements to the pipeline and to the cognitive architecture of the Heritage Digital Twin will be guided by rigorous ethical considerations and by a persistent commitment to human oversight. As artificial intelligence systems become more sophisticated in navigating and reconstructing cultural knowledge, their design must remain transparent, verifiable, and accountable. As narrated at the beginning of this paper, the gods fashioned a simulacrum, identical to the real Helen in form, voice, and grace, to deceive mortals and at the same time to preserve and keep her safe from the violence of men. Digital twins too are conceived to preserve and protect, but certainly not to deceive, since by definition their *raison d'être* lies not in illusion but in historical and epistemic fidelity. Yet in pursuing verisimilitude and cognitive depth, the risk is of being misled by the very intelligences we have created. Artificial systems, however powerful, often yield responses that are incomplete, imprecise, or altogether fabricated, i.e., hallucinations mistaken for knowledge. The risk, then, is that we ourselves are deceived by the AI meant to assist us, subtly shifting from curators of memory to unwitting custodians of fictions. The future of Heritage Digital Twins thus resides at the intersection of culture, computation, and semantics [135], transforming them into dynamic epistemological machines capable of simulating worlds, diagnosing faults, predicting futures, and structuring meaning. From this vision emerges a new notion of the Cognitive Digital Twin: a system capable of transcending passive simulation to engage in active reasoning, hypothesis generation, and adaptive control [136].

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app151810061/s1>.

**Author Contributions:** Investigation, A.F., A.H. and M.S. All authors have equally contributed to the present paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

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

## References

1. Graves, R. *The Greek Myths*; Penguin Books: London, UK, 1955; Volume 2, pp. 274–275.
2. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [\[CrossRef\]](#)
3. Semeraro, C.; Lezoche, M.; Panetto, H.; Dassisti, M. Digital twin paradigm: A systematic literature review. *Comput. Ind.* **2021**, *130*, 103469. [\[CrossRef\]](#)
4. Schrotter, G.; Hürzeler, C. The Digital Twin of the City of Zurich for Urban Planning. *PFG—J. Photogramm. Remote Sens. Geoinf. Sci.* **2020**, *88*, 99–112. [\[CrossRef\]](#)
5. Caprari, G.; Castelli, G.; Montuori, M.; Camardelli, M.; Malvezzi, R. Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives. *Sustainability* **2022**, *14*, 6263. [\[CrossRef\]](#)
6. Grieves, M.; Vickers, J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*; Kahlen, F.J., Flumerfelt, S., Alves, A., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 85–113. [\[CrossRef\]](#)
7. Li, L.; Aslam, S.; Wileman, A.; Perinpanayagam, S. Digital Twin in Aerospace Industry: A Gentle Introduction. *IEEE Access* **2022**, *10*, 9543–9562. [\[CrossRef\]](#)



8. Alfaro-Viquez, D.; Zamora-Hernandez, M.; Fernandez-Vega, M.; Garcia-Rodriguez, J.; Azorin-Lopez, J. A Comprehensive Review of AI-Based Digital Twin Applications in Manufacturing: Integration Across Operator, Product, and Process Dimensions. *Electronics* **2025**, *14*, 646. [\[CrossRef\]](#)
9. Bunea, A.A.; Dobre, C.M. Developing an Immersive Virtual Reality Interface for Urban Digital Twins: Integrating CAVE and HMD Systems for Enhanced Urban Analysis. In Proceedings of the 2024 23rd RoEduNet Conference: Networking in Education and Research (RoEduNet), Bucharest, Romania, 19–20 September 2024; pp. 1–5. [\[CrossRef\]](#)
10. European Commission. Destination Earth. Available online: <https://destination-earth.eu> (accessed on 15 July 2025).
11. Nikolinakos, N.T. *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies—The AI Act*, 1st ed.; Number v.53 in Law, Governance and Technology Series; Springer International Publishing AG: Cham, Switzerland, 2023.
12. Niccolucci, F.; Felicetti, A.; Hermon, S. Populating the Data Space for Cultural Heritage with Heritage Digital Twins. *Data* **2022**, *7*, 105. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Niccolucci, F.; Felicetti, A. Digital Twin Sensors in Cultural Heritage Ontology Applications. *Sensors* **2024**, *24*, 3978. [\[CrossRef\]](#)
15. Hedrick, E.; Hatch, D.; Oliver, E. Digital Twins in the Ancient and Historical World. In Proceedings of the 2025 Intermountain Engineering, Technology and Computing (IETC), Orem, UT, USA, 9–10 May 2025; pp. 1–6. [\[CrossRef\]](#)
16. Luther, W.; Baloian, N.; Biella, D.; Sacher, D. Digital Twins and Enabling Technologies in Museums and Cultural Heritage: An Overview. *Sensors* **2023**, *23*, 1583. [\[CrossRef\]](#)
17. Menaguale, O. Digital Twin and Cultural Heritage—The Future of Society Built on History and Art. In *The Digital Twin*; Crespi, N., Drobot, A.T., Minerva, R., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 1081–1111. [\[CrossRef\]](#)
18. Amelio, A.; Zarri, G.P. Cultural heritage digital twin: Modeling and representing the visual narrative in Leonardo Da Vinci's Mona Lisa. *Neural Comput. Appl.* **2024**, *36*, 11859–11876. [\[CrossRef\]](#)
19. Gros, A.; Guillem, A.; De Luca, L.; Baillieul, É.; Duvocelle, B.; Malavergne, O.; Leroux, L.; Zimmer, T. Faceting the post-disaster built heritage reconstruction process within the digital twin framework for Notre-Dame de Paris. *Sci. Rep.* **2023**, *13*, 5981. [\[CrossRef\]](#)
20. Ni, Z.; Eriksson, P.; Liu, Y.; Karlsson, M.; Gong, S. Improving energy efficiency while preserving historic buildings with digital twins and artificial intelligence. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *863*, 012041. [\[CrossRef\]](#)
21. Ni, Z.; Hupkes, J.; Eriksson, P.; Leijonhufvud, G.; Karlsson, M.; Gong, S. Parametric Digital Twins for Preserving Historic Buildings: A Case Study at Löfstad Castle in Östergötland, Sweden. *IEEE Access* **2025**, *13*, 3371–3389. [\[CrossRef\]](#)
22. Zhang, Z.; Dang, A.; Huang, J.; Chen, Y. Advancing Conservation Methods of the Great Wall Cultural Heritage Through Digital Twin. *IEEE Internet Comput.* **2025**, *29*, 48–55. [\[CrossRef\]](#)
23. Brandt, A.K. Beethoven's Ninth and AI's Tenth: A comparison of human and computational creativity. *J. Creat.* **2023**, *33*, 100068. [\[CrossRef\]](#)
24. Rizk, R.; Rizk, D.; Rizk, F.; Kumar, A. A Hybrid Capsule Network-based Deep Learning Framework for Deciphering Ancient Scripts with Scarce Annotations: A Case Study on Phoenician Epigraphy. In Proceedings of the 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 9–11 August 2021; pp. 617–620. [\[CrossRef\]](#)
25. Vats, S.; Mehta, S. Revolutionizing Archaeological Discoveries: The Role of Artificial Intelligence and Machine Learning in Site Analysis. In Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 24–28 June 2024; pp. 1–5. [\[CrossRef\]](#)
26. Cobb, P.J. Large Language Models and Generative AI, Oh My!: Archaeology in the Time of ChatGPT, Midjourney, and Beyond. *Adv. Archaeol. Pract.* **2023**, *11*, 363–369. [\[CrossRef\]](#)
27. Leong, W.Y. Revitalizing Historical Paintings through Artificial Intelligence. In Proceedings of the 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Nan, Thailand, 29 January–1 February 2025; pp. 141–145. [\[CrossRef\]](#)
28. Pisu, F.A.; Porcu, S.; Carboni, R.; Mameli, V.; Cannas, C.; Naitza, S.; Melis, R.T.; Mussi, M.; Chiriu, D. Innovative method for provenance studies in cultural heritage: A new algorithm based on observables from high-resolution Raman spectra of red ochre. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2025**, *329*, 125581. [\[CrossRef\]](#)
29. Mishra, M.; Barman, T.; Ramana, G. Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage. *J. Civ. Struct. Health Monit.* **2024**, *14*, 103–120. [\[CrossRef\]](#)
30. Ararat, M.B.; Yenice, T.K. The use of artificial intelligence in damage assessment of historical buildings. In *Frontiers of Architectural Research*; Elsevier: Amsterdam, The Netherlands, 2025.
31. Gaber, J.A.; Youssef, S.M.; Fathalla, K.M. The role of artificial intelligence and machine learning in preserving cultural heritage and art works via virtual restoration. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *10*, 185–190. [\[CrossRef\]](#)



32. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* **2020**, *8*, 108952–108971. [\[CrossRef\]](#)
33. Mayer, A.; Greif, L.; Häußermann, T.M.; Otto, S.; Kastner, K.; El Bobbou, S.; Chardonnet, J.R.; Reichwald, J.; Fleischer, J.; Ovtcharova, J. Digital Twins, Extended Reality, and Artificial Intelligence in Manufacturing Reconfiguration: A Systematic Literature Review. *Sustainability* **2025**, *17*, 2318. [\[CrossRef\]](#)
34. Charalabidis, Y.; Kontos, G.; Zitianellis, D. Local Digital Twins for Cities and Regions: The Way Forward. In *Decide Better: Open and Interoperable Local Digital Twins*; Raes, L., Ruston McAleer, S., Croket, I., Kogut, P., Brynskov, M., Lefever, S., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2025; pp. 233–259. [\[CrossRef\]](#)
35. Tampère, C.; Ortmann, P.; Jedlička, K.; Lohman, W.; Janssen, S. Future Ready Local Digital Twins and the Use of Predictive Simulations: The Case of Traffic and Traffic Impact Modelling. In *Decide Better: Open and Interoperable Local Digital Twins*; Raes, L., Ruston McAleer, S., Croket, I., Kogut, P., Brynskov, M., Lefever, S., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2025; pp. 203–230. [\[CrossRef\]](#)
36. Arsalan, H.; Heesom, D.; Moore, N. From Heritage Building Information Modelling Towards an ‘Echo-Based’ Heritage Digital Twin. *Heritage* **2025**, *8*, 33. [\[CrossRef\]](#)
37. Fan, C.; Zhang, C.; Yahja, A.; Mostafavi, A. Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. *Int. J. Inf. Manag.* **2021**, *56*, 102049. [\[CrossRef\]](#)
38. Lucchi, E. Digital twins, artificial intelligence and immersive technologies for heritage preservation and cultural tourism in smart cities. In *Digital Twin, Blockchain, and Sensor Networks in the Healthy and Mobile City*; Elsevier: Amsterdam, The Netherlands, 2025; pp. 507–520.
39. Pavlidis, G. Agentic AI for Cultural Heritage: Embedding Risk Memory in Semantic Digital Twins. *Computers* **2025**, *14*, 266. [\[CrossRef\]](#)
40. Singh, M.; Fuenmayor, E.; Hinchy, E.; Qiao, Y.; Murray, N.; Devine, D. Digital Twin: Origin to Future. *Appl. Syst. Innov.* **2021**, *4*, 36. [\[CrossRef\]](#)
41. Leipnitz, M.T.; Petry, R.H.; Rodrigues, F.H.; De Souza Silva, H.R.; Correia, J.B.; Becker, K.; Wickboldt, J.A.; Carbonera, J.L.; Netto, J.C.; Abel, M. Architecting Digital Twins: From Concept to Reality. *Procedia Comput. Sci.* **2025**, *256*, 530–537. [\[CrossRef\]](#)
42. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [\[CrossRef\]](#)
43. Boje, C.; Guerriero, A.; Kubicki, S.; Rezgui, Y. Towards a semantic Construction Digital Twin: Directions for future research. *Autom. Constr.* **2020**, *114*, 103179. [\[CrossRef\]](#)
44. Karabulut, E.; Pileggi, S.F.; Groth, P.; Degeler, V. Ontologies in digital twins: A systematic literature review. *Future Gener. Comput. Syst.* **2024**, *153*, 442–456. [\[CrossRef\]](#)
45. Khani, D.; Moreira, J.L.R.; van Sinderen, M.; Pires, L.F. XMLPO: An Ontology for Explainable Machine Learning Pipeline. In *Frontiers in Artificial Intelligence and Applications*; IOS Press: Amsterdam, The Netherlands, 2024; pp. 93–206. ISSN 0922-6389. [\[CrossRef\]](#)
46. Albers, L.; Große, P.; Wagner, S. Semantic Data-Modeling and Long-Term Interpretability of Cultural Heritage Data—Three Case Studies. In *Digital Cultural Heritage*; Springer International Publishing: Cham, Switzerland, 2020; pp. 239–253. [\[CrossRef\]](#)
47. Felicetti, A.; Meghini, C.; Richards, J.; Theodoridou, M. The AO-Cat Ontology. Technical Report, Zenodo, 2025. Version Number: 1.2. [\[CrossRef\]](#)
48. CIDOC CRM. Available online: <https://cidoc-crm.org> (accessed on 10 July 2025).
49. ISO 21127:2023; Information and documentation—A Reference Ontology for the Interchange of Cultural Heritage Information. ISO: Geneva, Switzerland, 2023.
50. CRMsci—Scientific Observation Model. Available online: <https://cidoc-crm.org/crmsci> (accessed on 11 July 2025).
51. CRMdig—Model for Provenance Metadata. Available online: <https://cidoc-crm.org/crmdig> (accessed on 14 July 2025).
52. Castelli, L.; Felicetti, A.; Proietti, F. Heritage Science and Cultural Heritage: standards and tools for establishing cross-domain data interoperability. *Int. J. Digit. Libr.* **2021**, *22*, 279–287. [\[CrossRef\]](#)
53. Felicetti, A.; Niccolucci, F. Artificial Intelligence and Ontologies for the Management of Heritage Digital Twins Data. *Data* **2024**, *10*, 1. [\[CrossRef\]](#)
54. Alamar, J.; Grootendorst, M. *Hands-On Large Language Models: Language Understanding and Generation*, 1st ed.; O’Reilly Media, Inc.: Sebastopol, CA, USA, 2024.
55. Chua, L.; Ghazi, B.; Huang, Y.; Kamath, P.; Kumar, R.; Manurangsi, P.; Sinha, A.; Xie, C.; Zhang, C. Crosslingual Capabilities and Knowledge Barriers in Multilingual Large Language Models. *arXiv* **2025**, arXiv:2406.16135.
56. Yang, H.; Zhang, Y.; Xu, J.; Lu, H.; Heng, P.A.; Lam, W. Unveiling the Generalization Power of Fine-Tuned Large Language Models. *arXiv* **2024**, arXiv:2403.09162. [\[CrossRef\]](#)
57. Schick, T.; Schütze, H. Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. *arXiv* **2021**, arXiv:2001.07676. [\[CrossRef\]](#)

58. Tudor, C.; Megyesi, B.; Östling, R. Prompting the Past: Exploring Zero-Shot Learning for Named Entity Recognition in Historical Texts Using Prompt-Answering LLMs. In Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025), Albuquerque, NM, USA, 3–4 May 2025; Kazantseva, A., Szpakowicz, S., Degaetano-Ortlieb, S., Bizzoni, Y., Pagel, J., Eds.; pp. 216–226. [\[CrossRef\]](#)
59. Brown, N.B. Enhancing Trust in LLMs: Algorithms for Comparing and Interpreting LLMs. *arXiv* **2024**, arXiv:2406.01943. [\[CrossRef\]](#)
60. Cullhed, E. Instruct-Tuning Pretrained Causal Language Models for Ancient Greek Papyrology and Epigraphy. *arXiv* **2024**, arXiv:2409.13870.
61. Ignatowicz, J.; Kutt, K.; Nalepa, G.J. Position Paper: Metadata Enrichment Model: Integrating Neural Networks and Semantic Knowledge Graphs for Cultural Heritage Applications. *arXiv* **2025**, arXiv:2505.23543. [\[CrossRef\]](#)
62. Meoded, E. Handwritten Text Recognition of Historical Manuscripts Using Transformer-Based Models. *arXiv* **2025**, arXiv:2508.11499. [\[CrossRef\]](#)
63. Zhang, X.; Bos, J. Multi-Modal Semantic Parsing for the Interpretation of Tombstone Inscriptions. *arXiv* **2025**, arXiv:2507.04377. [\[CrossRef\]](#)
64. Schweter, S.; März, L.; Schmid, K.; Çano, E. hmBERT: Historical Multilingual Language Models for Named Entity Recognition. *arXiv* **2022**, arXiv:2205.15575. [\[CrossRef\]](#)
65. ArchaeoBERT. Available online: <https://huggingface.co/alexbrandsen/ArchaeoBERT> (accessed on 20 August 2025).
66. Ehrmann, M.; Hamdi, A.; Pontes, E.L.; Romanello, M.; Doucet, A. Named Entity Recognition and Classification in Historical Documents: A Survey. *ACM Comput. Surv.* **2023**, *56*, 1–47. [\[CrossRef\]](#)
67. Sprugnoli, R.; Tonelli, S. Novel Event Detection and Classification for Historical Texts. *Comput. Linguist.* **2019**, *45*, 229–265. [\[CrossRef\]](#)
68. Zhu, Y.; Wang, X.; Chen, J.; Qiao, S.; Ou, Y.; Yao, Y.; Deng, S.; Chen, H.; Zhang, N. LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities. *arXiv* **2024**, arXiv:2305.13168. [\[CrossRef\]](#)
69. Yang, L.; Chen, H.; Li, Z.; Ding, X.; Wu, X. Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling. *arXiv* **2024**, arXiv:2306.11489. [\[CrossRef\]](#)
70. Tang, N.; Fan, J.; Li, F.; Tu, J.; Du, X.; Li, G.; Madden, S.; Ouzzani, M. RPT: Relational Pre-trained Transformer Is Almost All You Need towards Democratizing Data Preparation. *arXiv* **2021**, arXiv:2012.02469. [\[CrossRef\]](#)
71. Amer-Yahia, S.; Bonifati, A.; Chen, L.; Li, G.; Shim, K.; Xu, J.; Yang, X. From Large Language Models to Databases and Back: A discussion on research and education. *arXiv* **2023**, arXiv:2306.01388. [\[CrossRef\]](#)
72. Choi, Y.; Asif, M.A.; Han, Z.; Willes, J.; Krishnan, R.G. Teaching LLMs How to Learn with Contextual Fine-Tuning. *arXiv* **2025**, arXiv:2503.09032. [\[CrossRef\]](#)
73. Pyrih, V.; Rebmman, A.; van der Aa, H. LLMs that Understand Processes: Instruction-tuning for Semantics-Aware Process Mining. *arXiv* **2025**, arXiv:2508.16270.
74. Jiménez-Ruiz, E.; Grau, B. *LogMap: Logic-Based and Scalable Ontology Matching*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 273–288. [\[CrossRef\]](#)
75. He, Y.; Chen, J.; Antonyrajah, D.; Horrocks, I. BERTMap: A BERT-based Ontology Alignment System. *arXiv* **2022**, arXiv:2112.02682. [\[CrossRef\]](#)
76. Fensel, D.; Şimşek, U.; Angele, K.; Huaman, E.; Kärle, E.; Panasiuk, O.; Toma, I.; Umbrich, J.; Wahler, A. How to Build a Knowledge Graph. In *Knowledge Graphs: Methodology, Tools and Selected Use Cases*; Springer International Publishing: Cham, Switzerland, 2020; pp. 11–68. [\[CrossRef\]](#)
77. Maree, M. Quantifying Relational Exploration in Cultural Heritage Knowledge Graphs with LLMs: A Neuro-Symbolic Approach for Enhanced Knowledge Discovery. *Data* **2025**, *10*, 52. [\[CrossRef\]](#)
78. Bruns, O. The Persistence of Temporality: Representation of Time in Cultural Heritage Knowledge Graphs. In Proceedings of the Doctoral Consortium at ISWC 2023, co-located with 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, 7 November 2023; ISWC 2023, Volume 3678.
79. Faralli, S.; Lenzi, A.; Velardi, P. A Large Interlinked Knowledge Graph of the Italian Cultural Heritage. In Proceedings of the LREC, Marseille, France, 20–25 June 2022; pp. 6280–6289.
80. Carriero, V.A.; Gangemi, A.; Mancinelli, M.L.; Marinucci, L.; Nuzzolese, A.G.; Presutti, V.; Veninata, C. ArCo: The Italian Cultural Heritage Knowledge Graph. *arXiv* **2019**, arXiv:1905.02840. [\[CrossRef\]](#)
81. Barzaghi, S.; Moretti, A.; Heibi, I.; Peroni, S. CHAD-KG: A Knowledge Graph for Representing Cultural Heritage Objects and Digitisation Paradata. *arXiv* **2025**, arXiv:2505.13276. [\[CrossRef\]](#)
82. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805. [\[CrossRef\]](#)
83. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165. [\[CrossRef\]](#)
84. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv:2302.13971. [\[CrossRef\]](#)

85. Gekhman, Z.; Yona, G.; Aharoni, R.; Eyal, M.; Feder, A.; Reichart, R.; Herzig, J. Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations? *arXiv* **2024**, arXiv:2405.05904. [\[CrossRef\]](#)
86. Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 3580–3599. [\[CrossRef\]](#)
87. Agrawal, A.; Suzgun, M.; Mackey, L.; Kalai, A.T. Do Language Models Know When They’re Hallucinating References? *arXiv* **2023**, arXiv:2305.18248. [\[CrossRef\]](#)
88. Sui, Y.; He, Y.; Ding, Z.; Hooi, B. Can Knowledge Graphs Make Large Language Models More Trustworthy? An Empirical Study Over Open-ended Question Answering. *arXiv* **2025**, arXiv:2410.08085. [\[CrossRef\]](#)
89. Sun, J.; Du, Z.; Chen, Y. Knowledge Graph Tuning: Real-time Large Language Model Personalization based on Human Feedback. *arXiv* **2024**, arXiv:2405.19686. [\[CrossRef\]](#)
90. Doerr, M. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Mag.* **2003**, *24*, 75–92.
91. Binding, C.; May, K.; Tudhope, D. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In Proceedings of the Research and Advanced Technology for Digital Libraries: 15th International Conference on Theory and Practice of Digital Libraries (TPDL 2011), Berlin, Germany, 23–26 September 2011; Springer: Berlin, Germany, 2011; pp. 280–291.
92. Doerr, M.; Kritsotaki, A.; Rousakis, Y. CIDOC2VEC: Extracting Information from Atomized CIDOC-CRM Humanities Knowledge Graphs. *Information* **2021**, *12*, 503.
93. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P. Composition-based multi-relational graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 26–30 April 2020.
94. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling relational data with graph convolutional networks. In Proceedings of the European Semantic Web Conference (ESWC), Crete, Greece, 3–7 June 2018; Springer: Berlin, Germany, 2018; pp. 593–607.
95. El Vaigh, C.B.; Garcia, N.; Renoust, B.; Chu, C.; Nakashima, Y.; Nagahara, H. GCNBoost: Artwork Classification by Label Propagation through a Knowledge Graph. *arXiv* **2021**, arXiv:2105.11852. [\[CrossRef\]](#)
96. Meroño Peñuela, A.; Sack, H. Move Cultural Heritage Knowledge Graphs in Everyone’s Pocket. *Semant. Web J.* **2020**, *14*, 323–359.
97. Yuan, H.; Li, Y.; Wang, B.; Liu, K.; Zhang, J. Knowledge Graph-Based Intelligent Question Answering System for Ancient Chinese Costume Heritage. *NPJ Herit. Sci.* **2025**, *13*, 198. [\[CrossRef\]](#)
98. Huang, Y.; Yu, S.; Chu, J.; Fan, H.; Du, B. Using Knowledge Graphs and Deep Learning Algorithms to Enhance Digital Cultural Heritage Management. *Herit. Sci.* **2023**, *11*, 204. [\[CrossRef\]](#)
99. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [\[CrossRef\]](#)
100. Majumder, E.A. Injecting Knowledge Graphs into Large Language Models. *arXiv* **2025**, arXiv:2505.07554. [\[CrossRef\]](#)
101. Zhang, B.; Meroño Peñuela, A.; Simperl, E. Towards Explainable Automatic Knowledge Graph Construction with Human-in-the-loop. *Semant. Web J.* **2023**, 274–289. [\[CrossRef\]](#)
102. Arazzi, M.; Ligari, D.; Nicolazzo, S.; Nocera, A. Augmented Knowledge Graph Querying leveraging LLMs. *arXiv* **2025**, arXiv:2502.01298.
103. Wang, J.; Cui, Z.; Wang, B.; Pan, S.; Gao, J.; Yin, B.; Gao, W. IME: Integrating Multi-curvature Shared and Specific Embedding for Temporal Knowledge Graph Completion. *arXiv* **2024**, arXiv:2403.19881. [\[CrossRef\]](#)
104. Gaši, A.; Ensari, T.; Dağtekin, M. Ontology supported semantic based image retrieval. *Semant. Web.* **2024**, *15*, 2125–2137. [\[CrossRef\]](#)
105. Castellano, G.; Sansaro, G.; Vessio, G. Integrating Contextual Knowledge to Visual Features for Fine Art Classification. *arXiv* **2021**, arXiv:2105.15028. [\[CrossRef\]](#)
106. ARTEMIS. Applying Reactive Twins to Enhance Monument Information Systems. Available online: <https://www.artemis-twin.eu> (accessed on 12 July 2025).
107. PyMuPDF. Available online: <https://pymupdf.readthedocs.io/en/latest/> (accessed on 20 August 2025).
108. SpaCy. Available online: <https://spacy.io/> (accessed on 20 August 2025).
109. All-MiniLM-L6-v2. Available online: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2> (accessed on 20 August 2025).
110. Mistral 7B Instruct. Available online: <https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3> (accessed on 20 August 2025).
111. FAISS. Available online: <https://faiss.ai/index.html> (accessed on 20 August 2025).
112. The AO-Cat Ontology. Available online: <https://zenodo.org/records/7818375> (accessed on 14 July 2025).
113. ARIADNE Research Infrastructure. Available online: <https://www.ariadne-research-infrastructure.eu> (accessed on 12 July 2025).
114. Mazzinghi, A.; Ruberto, C.; Giuntini, L.; Mandò, P.A.; Taccetti, F.; Castelli, L. Mapping with Macro X-ray Fluorescence Scanning of Raffaello’s Portrait of Leo X. *Heritage* **2022**, *5*, 3993–4005. [\[CrossRef\]](#)
115. Leboef, P.; Doerr, M. CIDOC Conceptual Reference Model: A Standard for Cultural Heritage Information. *J. Comput. Cult. Herit.* **2017**, *10*, 1–20.

116. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. *Adv. Neural Inf. Process. Syst.* **2013**, *26*.
117. Xu, D.; Cheng, W.; Luo, D.; Chen, H.; Zhang, X. InfoGCL: Information-Aware Graph Contrastive Learning. *arXiv* **2021**, arXiv:2110.15438.
118. Liang, K.; Liu, Y.; Zhou, S.; Tu, W.; Wen, Y.; Yang, X.; Dong, X.; Liu, X. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *arXiv* **2022**, arXiv:2211.10738. [[CrossRef](#)]
119. Füll, A.; Nissen, V. Design of Explanation User Interfaces for Interactive Machine Learning. *Requir. Eng.* **2025**. [[CrossRef](#)]
120. Barzaghi, S.; Colitti, S.; Moretti, A.; Renda, G. From Metadata to Storytelling: A Framework For 3D Cultural Heritage Visualization on RDF Data. *arXiv* **2025**, arXiv:2505.14328. [[CrossRef](#)]
121. Emonet, V.; Bolleman, J.; Duvaud, S.; de Farias, T.; Sima, A.C. LLM-based SPARQL Query Generation from Natural Language over Federated Knowledge Graphs. *arXiv* **2025**, arXiv:2410.06062.
122. Manzoor, E.; Tong, J.; Vijayaraghavan, S.; Li, R. Expanding Knowledge Graphs with Humans in the Loop. *arXiv* **2023**, arXiv:2212.05189. [[CrossRef](#)]
123. Celino, I. Who is this Explanation for? Human Intelligence and Knowledge Graphs for eXplainable AI. *arXiv* **2020**, arXiv:2005.13275. [[CrossRef](#)]
124. Kamar, E. Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016.
125. Te’eni, D.; Yahav, I.; Zagalsky, A.; Schwartz, D.G.; Silverman, G. Reciprocal Human—Machine Learning: A Theory and an Instantiation for the Case of Message Classification. *Manag. Sci.* **2023**. [[CrossRef](#)]
126. Foka, A.; Griffin, G.; Ortiz Pablo, D.; Rajkowska, P.; Badri, S. Tracing the bias loop: AI, cultural heritage and bias-mitigating in practice. *Ai Soc.* **2025**. [[CrossRef](#)]
127. Tóth, G.M.; Albrecht, R.; Pruski, C. Explainable AI, LLM, and digitized archival cultural heritage: A case study of the Grand Ducal Archive of the Medici. *AI Soc.* **2025**. [[CrossRef](#)]
128. Peters, U.; Carman, M. Cultural Bias in Explainable AI Research: A Systematic Analysis. *J. Artif. Intell. Res.* **2024**. [[CrossRef](#)]
129. Mitchell, M. AI Is Spreading Old Stereotypes to New Languages and Cultures. *Wired* **2025**.
130. Upmann, P. What Role Does Cultural Context Play in Defining Ethical Standards for AI? *AI Glob.* **2024**.
131. Prabhakaran, V.; Qadri, R.; Hutchinson, B. Cultural Incongruencies in Artificial Intelligence. *arXiv* **2022**, arXiv:2211.13069. [[CrossRef](#)]
132. Inuwa-Dutse, I. FATE in AI: Towards Algorithmic Inclusivity and Accessibility. *arXiv* **2023**, arXiv:2301.01590. [[CrossRef](#)]
133. Kukutai, T.; Taylor, J. *Indigenous Data Sovereignty: Towards an Agenda*; ANU Press: Canberra, Australia, 2020.
134. Cherti, M.; Beaumont, R.; Wightman, R.; Wortsman, M.; Ilharco, G.; Gordon, C.; Schuhmann, C.; Schmidt, L.; Jitsev, J. Reproducible Scaling Laws for Contrastive Language-Image Learning. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 2818–2829. [[CrossRef](#)]
135. Tooth, J.M.; Tuptuk, N.; Watson, J.D.M. A Systematic Survey of the Gemini Principles for Digital Twin Ontologies, 2024. Version Number: 1. [[CrossRef](#)]
136. D’Amico, R.D.; Sarkar, A.; Karray, M.H.; Addepalli, S.; Erkoyuncu, J.A. Knowledge transfer in Digital Twins: The methodology to develop Cognitive Digital Twins. *CIRP J. Manuf. Sci. Technol.* **2024**, *52*, 366–385. [[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.

Part of the present work has been supported by **ARTEMIS**, a project funded by the European Union under Grant Agreement n.101188009. However, the views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



**Funded by  
the European Union**