MODELING AN ONTOLOGY FOR HERITAGE SCIENCE: CHALLENGES AND KEY STRATEGIES

Erica Scarpa[1], Riccardo Valente[2], Irene Rossi[3]
[1] ISPC - CNR, Italy – erica.scarpa@cnr.it
[2] ISPC - CNR, Italy – riccardo.valente@cnr.it
[3] ISPC - CNR, Italy – irene.rossi@cnr.it

ABSTRACT (ENGLISH)
This paper presents the authors' contributions to the Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) project, in particular the development of an ontology for Heritage Science. Given the inherently multidisciplinary nature of Heritage Science, defining precise boundaries for the domain is a complex task that can hinder effective data organization and limit research potential. After providing a brief overview of the H2IOSC project, the paper delves into the key challenges faced in the ontology development. It further outlines the initial strategies employed to address these challenges, emphasizing the establishment of a permanent interdisciplinary group of Heritage subject-matter experts and the creation of the H-SeTIS online database for surveying existing semantic resources in the Heritage field. The paper also highlights future developments and critical considerations for advancing the project, ensuring its continued relevance and impact and promotes the adoption of good practices for the development of semantic artefacts in the Heritage domain.
Keywords: Semantic Web; Ontologies; Heritage Science; Knowledge Elicitation

ABSTRACT (ITALIANO)
Modellare un'ontologia per l'Heritage Science: sfide e strategie.
Questo articolo presenta il contributo degli autori al progetto Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC), con particolare attenzione alla progettazione di un'ontologia per l'Heritage Science. Data la natura intrinsecamente multidisciplinare di questo dominio, la sua definizione può risultare complessa, con il rischio di compromettere l'organizzazione dei dati e limitare le potenzialità della ricerca. Dopo una breve introduzione al progetto H2IOSC, vengono illustrati i principali problemi nella modellazione di un'ontologia per l'Heritage Science. Vengono inoltre descritti i primi approcci e le strategie adottate per affrontare questa vasta tematica: particolare attenzione è dedicata alla costituzione di un gruppo interdisciplinare permanente di esperti del dominio dell'Heritage e alla creazione di un database online di risorse semantiche (H-SeTIS). Infine, vengono presentati gli sviluppi futuri e gli aspetti critici.
Parole chiave: web semantico; ontologie; Heritage Science; elicitazione della conoscenza

1. INTRODUCTION
This paper explores the ongoing development of an ontology for the Heritage Science (HS) domain: the semantic model will be a key component of the Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) project. Funded by the Next Generation EU plan and the Italian Ministry of University and Research, the H2IOSC project involves twelve institutes of the Italian National Research Council (CNR). Its primary goal is to establish a collaborative, federated cluster connecting the four Italian nodes of European Research Infrastructures (RI) focused on Human Sciences and Cultural Heritage: CLARIN, DARIAH, E-RIHS, and OPERAS.
Specifically, the project aims to create a digital open ecosystem that supports advanced multidisciplinary research while facilitating the creation, access, and (re-)use of open scientific data. Within this framework, the Institute of Heritage Science (ISPC-CNR) plays a crucial role in ensuring resource interoperability for the E-RIHS infrastructure, particularly through the development of an ontology tailored to the needs of HS. This paper presents the methodology and challenges involved in designing this ontology, highlighting the collaborative efforts of subject-matter experts and the innovative approaches used to build a structured, interoperable knowledge framework that advances research and innovation in HS.

2. MODELLING CHALLENGES
HS is generally described as an interdisciplinary research field that integrates the social sciences, natural sciences, and humanities. The concept of HS was first introduced in 2006 by the Science and Technology Select Committee of the British House of Lords (House of Lords Science and Technology Select Committee,

2006). It aims to bridge traditional disciplinary divides by linking Cultural Heritage Conservation with the Social Sciences. However, its formal definition is relatively recent, having been jointly established in 2019 by the E-RIHS RI and the International Centre for the Study of the Preservation and Restoration of Cultural Property (ICCROM)[1].

The recognition of HS as a distinct domain is intended to bring together researchers from diverse disciplines and backgrounds who work on common heritage topics and objects, fostering deeper collaboration and interdisciplinary approaches.

However, the object of study of HS (*i.e.*, Heritage) is a multifaceted, intricate and dynamic concept, with its meaning and perception shifting significantly through space and time. Therefore, before exploring the specific strategy for modeling an ontology for HS, it is essential to understand the challenges involved in defining the scope of the domain itself.

Different approaches to Heritage can coexist, as its interpretation develops in response to local cultures, traditions, and geographical contexts. In some cases, these approaches share common roots, differing only in specific aspects: Heritage is seen as something valuable to preserve and pass down through generations. However, in other instances, theoretical foundations may be fundamentally opposed, as not all societies share the same principles for defining what is considered worthy of transmission. Moreover, the scope of Heritage is not uniform, as it encompasses a wide range of cultural assets—from traditional monuments and works of art to industrial archaeology—each carrying its own unique significance.

Change is also a defining characteristic of Heritage, as modifications in the global concept of Heritage occur rapidly, particularly throughout the second half of the 20th century and into the early 21st century (Harrison, 2020). However, the evolution of Heritage is neither uniform nor consistent. This variability, combined with the relatively recent establishment of HS as a domain, presents significant challenges in developing an HS ontology. Such an ontology must not only accommodate the diverse nature of existing Heritage but also remain adaptable to its future transformations. Ensuring the long-term sustainability of the ontology is crucial for supporting the entire RI and enabling the effective retrieval and integration of complex, evolving data over time.

For the purposes of the H2IOSC project and the development of an HS ontology, it was necessary to define the boundaries of Heritage—an essential step in identifying the domain's characteristics and requirements. It should be emphasized that the boundary definition outlined below is pragmatic, serving as both a reference and modeling strategy for establishing the scope of the ontology. While it is grounded in the most widely recognized and up-to-date theories, it is not necessarily intended to serve as a new formal definition of Heritage.

Among the most significant opposing pairs used for the classification of Heritage entities, Nature-Culture (with Culture here specifically referring to human-made elements) and Tangible-Intangible were identified. For what concerns the first pair, it must be highlighted that what may initially seem like a straightforward classification—a living elephant is different from a wooden statue of an elephant—has more complex nuances. A wooden statue carved by a human is indeed a cultural entity, despite its material being entirely natural. Similarly, a series of carved rock art drawings might appear to fall into a comparable category. However, the immovable nature of their rock support, which entirely belongs to the natural realm, shifts the classification threshold. In such cases, what exactly qualifies as cultural, and what remains natural? Are only the carved drawings considered cultural, while their physical support remains natural? That might seem reasonable, yet carved figures cannot exist independently of the surface that holds them. Does the rock, then, become a cultural entity simply because it bears cultural elements on its surface? If so, would it be correct to classify an entire massive rock as a cultural entity, even if only a small corner contains carvings? Should only the portion with the drawings be considered cultural? Or should the entire rock be viewed as a mixed natural-cultural entity?

These uncertainties deepen when exploring other examples in Heritage. A *bonsai*, for instance, is not entirely a human creation, nor it is purely natural; it is a living organism that humans shaped into a cultural expression. This borderline case illustrates the difficulty of applying a clear-cut distinction between natural and cultural entities within the domain of Heritage.

The contraposition of Nature-Culture was dismissed for the aforementioned reasons, while the pair Tangible-Intangible was kept as more definable and inherently connected to the scope of the E-RIHS RI. The definition of these boundaries was carried out also thanks to the direct involvement of Heritage subject-matter experts (see §5).

[1] https://www.e-rihs.eu/e-rihs-in-a-nutshell/.

3. MODELING AN ONTOLOGY FOR THE HERITAGE SCIENCE DOMAIN

Ontology engineering methodologies begin with knowledge elicitation—the systematic process of gathering, organizing, and refining knowledge to create a structured representation of the concepts, relationships, and semantics characterizing a particular domain (Leenheer, 2009). Subject-matter experts play a crucial role in this process, as their collaboration with knowledge engineers provides the theoretical foundation for an accurate representation (Shadbolt & Smart, 2015). Ontology modeling involves a delicate balance between achieving a shared understanding of key definitions and meanings within a domain and formalizing them into a machine-actionable model. Given the dynamic and evolving nature of the object of study in the HS domain, the negotiation between knowledge engineers and subject-matter experts becomes an even more critical phase, as it influences how data are represented and interpreted. No ontology engineering methodology specifically tailored for the Heritage or HS domain exists, as most methodologies have been developed for business and industry contexts. In recent years, Linked (Open) Data (Heath & Bizer, 2011) and FAIR guidelines (Wilkinson et al., 2016) have had a significant impact on how ontologies have been developed over the past decade. As a result, only the most recent methodologies incorporate these standards. The development of the HS ontology discussed here follows the Linked Open Terms (LOT) methodology (Poveda-Villalón et al., 2022), which incorporates the latest requirements for Linked Data and FAIR principles and promotes the reuse of existing ontologies and vocabularies.

The LOD methodology begins with defining the ontology's requirements specifications, starting with the scope definition. It also involves identifying case studies and real-world scenarios to assess the model's practical applicability. To support these initial steps, two preliminary goals were established: a comprehensive survey of semantic artefacts in the Heritage domain and the establishment of a permanent working group focused on defining the boundaries of HS.

For the first objective, the *Heritage – Semantic Tools and Interoperability Survey* (H-SeTIS, see §4) database served the purpose (Scarpa & Valente, 2024a, 2024b). H-SeTIS is a preliminary survey of all semantic artefacts—"machine-actionable and -readable formalizations of a conceptualization, enabling sharing and reuse by humans and machines" (Hugo et al., 2020, p. 12)—specifically designed for or used within the Heritage domain. Before modeling a complex domain like HS conducting a thorough and detailed review of the state of the art is crucial. This step ensures that the model can address relevant and current questions while meeting the evolving needs of the user community. Additionally, it helps avoid duplicating solutions already developed in existing models.

The second goal is the establishment of a permanent interdisciplinary group of subject-matter experts. This group, gathering around the Heritage Science Ontology RoundTable (‹H/SORT›, see §5), will provide a sustainable framework for collaboration among subject-matter experts from various HS disciplines, including epigraphy, history, archaeology, and conservation science. The group also aims to identify key research questions (KRQs), ensuring that decision-making processes are well-informed, integrating and balancing the different perspectives on Heritage from each discipline.

4. H-SeTIS

The preliminary survey of semantic artefacts within the Heritage field provided a clear state-of-the-art overview for developing these resources. The *Heritage – Semantic Tools and Interoperability Survey* database (https://h-setis.cnr.it/) serves as a robust foundation for ontology development, addressing gaps left by existing aggregators such as BARTOC (Ledl & Voß, 2016) and Linked Open Vocabularies (Vandenbussche et al., 2017), which often fail to capture the complexity of the Heritage domain. H-SeTIS catalogs five core types of semantic artefacts: ontologies, metadata standards, thesauri, application profiles, and software (Scarpa & Valente, 2024a, 2024b). Each resource is documented in detail, including its availability, development status, URI/repository, contributors, language, formats, licensing, keywords, and related bibliographic references. Relationships between resources are mapped to highlight integrations and dependencies. Additionally, editors of all identified semantic artefacts are being contacted to evaluate the responsiveness of each resource and to establish a dedicated contact person.

The H-SeTIS database also includes a curated bibliography of publications referred to the surveyed semantic artefacts and more general resources for semantic web development, managed through Zotero (https://www.zotero.org/). Zotero APIs enable seamless integration of bibliographic data into the H-SeTIS interface, with the shared library publicly accessible (https://www.zotero.org/groups/5434475/). Using Kerko (https://pypi.org/project/Kerko/), the bibliography is displayed within the H-SeTIS front-end.

Although the survey is still ongoing, it has already revealed a rich landscape of semantic artefacts in the Heritage domain. However, it has also exposed major shortcomings in adherence to good practices—one of the key reasons why aggregators like BARTOC and LOD fail to adequately represent semantic artefacts in this field. Many artefacts lack fundamental features such as persistent URIs, or exist solely as conceptual models published in scientific papers. Since its launch in December 2023, the survey has uncovered critical challenges in the Heritage sector, including insufficient documentation, poor adherence to FAIR principles, and limited application of Linked (Open) Data practices. Many resources lack semantic versioning, or active maintenance (Fig. 1). Preliminary findings show that at least half of the surveyed ontologies are inactive or lack status information.
To address these challenges, the ‹H/SORT› working group is developing a collection of best practices for the creation and maintenance of semantic resources in the Heritage field. This initiative, currently in progress, includes defining status descriptions for resources, inspired by the OBO Foundry framework (Smith et al., 2007) but adapted to accommodate the broader range of Heritage-specific artefacts.

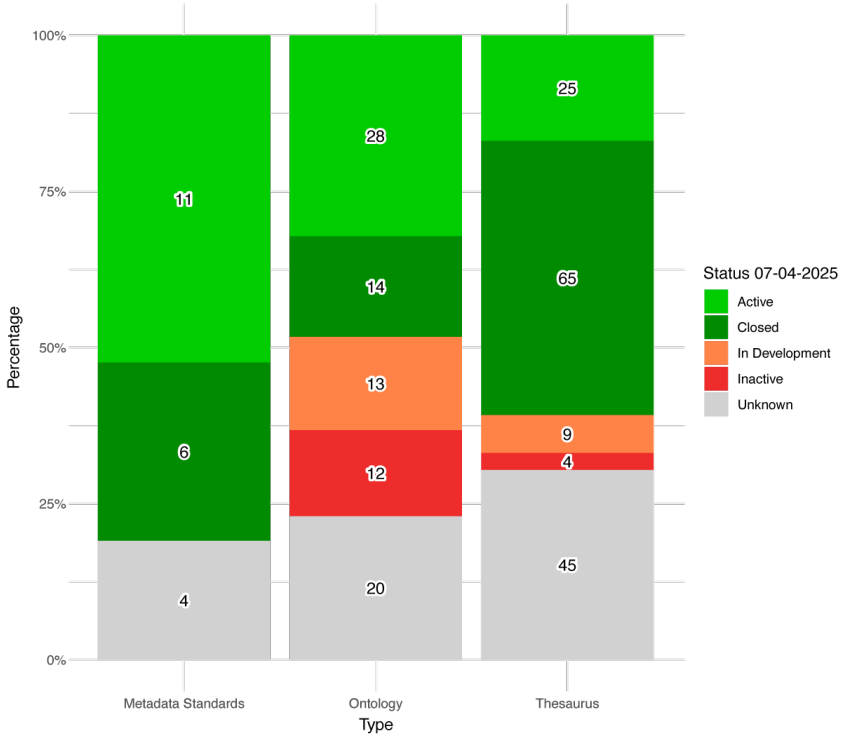

Figure 1 - Availability statuses of three kinds of semantic artefacts recorded in H-SeTIS. Digits in the bar segments display the number of artefacts.

5. ‹H/SORT›: Heritage Science Ontology RoundTable

The LOD methodology for ontology development (Poveda-Villalón et al., 2022) aligns with agile software development, a strategy that promotes close collaboration, face-to-face discussions, and flexible management to prioritize user needs and accelerate development. This approach emphasizes customer satisfaction, iterative improvements, and the creation of essential documentation[2].
Within this operational framework, the ‹H/SORT› working group gathers developers and subject-matter experts to explore the foundations of HS through seamless exchange and discussion, identifying key topics and case studies. The first roundtable of the working group, held in Venice on January 16-17, 2025, introduced an elicitation model focused on collaborative debate among subject-matter experts from specific HS subdomains, identifying diverging opinions on specific topics. The main focus of this first meeting was epigraphy, as the shared case study for all of the four RIs involved in the H2IOSC Work Package devoted to interoperability was the “inscribed object” (*i.e.*, any object carrying an inscription). A second roundtable, held in Milan on March 26, further addressed issues related to knowledge representation for HS, involving a group of experts from STEM disciplines, who brought a different and complementary approach compared to the Humanities.

[2] https://agilemanifesto.org/

The ‹H/SORT› hub (https://hsort-ec1af2.gitlab.io/) serves as a knowledge framework for developing the HS Ontology, where official, lightweight, and focused documentation is continuously updated based on insights from the roundtable discussions and feedback from the ‹H/SORT› working group. The hub also supports ongoing collaboration and tracks developments, ensuring that progress is monitored and results are effectively disseminated.
The ‘Topics’ section of the ‹H/SORT› hub collects a series of short posts that introduce key issues and themes related to Heritage. These posts, designed to be light, concise, and limited in bibliography, are carefully selected to stimulate discussion and debate. One of the key objectives of the ‘Topics’ section is to introduce the main themes and challenges encountered during the development of the HS Ontology to a broader audience. Some topics address core issues in Heritage Studies, such as the characteristics and transformations of Heritage, while others focus on the application of semantic technologies in the Heritage domain. For instance, one post explores the distinction between ‘Provenance’ and ‘Provenience’. The term provenance, commonly used by art historians, refers to the ownership history of an object. Conversely, provenience, typically used in archaeology, refers to the precise physical location where an object was discovered (Scarpa, 2025a). While the difference between these terms—the contrast between a “chain of custody” and a “specific place”—may seem minor or even irrelevant in certain contexts, it is a crucial distinction when dealing with Heritage data and its semantic representation. The varied, and at times contradictory, usage of these terms presents a significant challenge for accurate and consistent modeling.
Another section of ‹H/SORT› is dedicated to case studies, which serve as valuable tools for illustrating real-world applications and scenarios. An effective use of case studies involves selecting examples that differ in their characteristics. Diversity is essential for informing the correct modeling of classes and properties, ensuring that the ontology aligns with actual needs and use cases. For instance, one case study (Scarpa, 2025b) on ‹H/SORT› considers a standard text reproduced as an inscription on several, and quite different, supports (stone vessels, bowls, but even a natural shell). While scholars can easily recognize and interpret such variations, they present modeling challenges that must be carefully addressed to ensure proper representation within a semantic ecosystem.

6. CONCLUSIONS

The development of an ontology for the Heritage Science (HS) domain presents both unique challenges and significant opportunities for advancing research and represent the interdisciplinary nature of HS. By fostering a collaborative framework with subject-matter experts, the ‹H/SORT› working group aims to create a sustainable and interoperable knowledge structure that will support future research in the field. The adoption of current methodologies is central to integrating recent standards, such as Linked Open Data and FAIR principles, into the ontology model. Additionally, the incorporation of updated theoretical foundations seeks to provide an interpretive framework that has been largely unexplored to date. Moreover, the ‹H/SORT› working group is dedicated not only to documenting the ontology itself, with appropriate metadata, but also to thoroughly documenting the development process.

ACKNOWLEDGEMENTS

H2IOSC Project - Humanities and cultural Heritage Italian Open Science Cloud funded by the European Union NextGenerationEU - National Recovery and Resilience Plan (NRRP) - Mission 4 “Education and Research” Component 2 “From research to business” Investment 3.1 “Fund for the realization of an integrated system of research and innovation infrastructures” Action 3.1.1 “Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe” - Project code IR0000029 - CUP B63C22000730005. Implementing Entity CNR.

REFERENCES

Harrison, R. (2020). *Il patrimonio culturale. Un approccio critico* (V. Matera & L. Rimoldi, Eds.).
Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Springer Cham. http://linkeddatabook.com/editions/1.0/
House of Lords Science and Technology Select Committee. (2006). *Science and Heritage. Report with Evidence*. https://publications.parliament.uk/pa/ld200506/ldselect/ldsctech/256/256.pdf
Hugo, W., Le Franc, Y., Coen, G., Parland-von Essen, J., & Bonino, L. (2020). *D2.5 FAIR Semantics Recommendations Second Iteration*. https://zenodo.org/records/5362010
Ledl, A., & Voß, J. (2016). Describing Knowledge Organization Systems in BARTOC and JSKOS. In H. E. Thomsen, A. Pareja-Lora, & B. N. Madsen (Eds.), *Ledl, Andreas; Voß, Jakob (2016). Describing*

Knowledge Organization Systems in BARTOC and JSKOS. In: Thomsen, Hanne Erdman; Pareja-Lora, Antonio; Madsen, Bodil Nistrup (eds.) TKE 2016 the 12th International conference on Terminology and Knowledge Engineering (pp. 168-178). Denmark: Copenhagen Business School (pp. 168–178). Copenhagen Business School. https://research.cbs.dk/en/publications/term-bases-and-linguistic-linked-open-data-proceedings-of-tke-201

Leenheer, P. D. (2009). Ontology Elicitation. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 1966–1972). Springer US. https://doi.org/10.1007/978-0-387-39940-9_1316

Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An Industrial Oriented Ontology Engineering Framework. *Engineering Applications of Artificial Intelligence*, *111*, 104755. https://doi.org/10.1016/j.engappai.2022.104755

Scarpa, E. (2025a, January 20). Provenance and Provenience. *Heritage Science Ontology RoundTable (‹H/SORT›)*. https://doi.org/10.5281/zenodo.15166679

Scarpa, E. (2025b, January 28). Case Study #1: Identical text on multiple objects. *Heritage Science Ontology RoundTable (‹H/SORT›)*. https://doi.org/10.5281/zenodo.15167824

Scarpa, E., & Valente, R. (2024a). A Resource Hub for Interoperability and Data Integration in Heritage Research: The H-SeTIS Database. *Archeologia e Calcolatori*, *35*(1), 543–562. https://doi.org/10.19282/ac.35.1.2024.32

Scarpa, E., & Valente, R. (2024b). *Heritage – Semantic Tools and Interoperability Survey* [Dataset]. https://h-setis.cnr.it/

Shadbolt, N., & Smart, P. R. (2015). Knowledge Elicitation: Methods, Tools and Techniques. In N. Shadbolt, P. R. Smart, & J. R. Wilson (Eds.), *Evaluation of Human Work* (pp. 163–200). CRC Press. https://eprints.soton.ac.uk/359638/

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., & Lewis, S. (2007). The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration. *Nature Biotechnology*, *25*(11), 1251–1255. https://doi.org/10.1038/nbt1346

Vandenbussche, P.-Y., Atemezing, G. A., Poveda-Villalón, M., & Vatant, B. (2017). Linked Open Vocabularies (LOV): A Gateway to Reusable Semantic Vocabularies on the Web. *Semantic Web*, *8*(3), 437–452. https://doi.org/10.3233/SW-160213

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18