Archeologia e Calcolatori
36.1, 2025, 481-492
doi 10.19282/ac.36.1.2025.27

DEMYSTIFYING THE CIDOC CRM: A LIGHTWEIGHT INTRODUCTION

1. Introduction

The CIDOC Conceptual Reference Model (hereafter, CIDOC CRM) is a formal ontology primarily designed for museums and cultural institutions to describe and organize their data. An ontology is a model representing some subject matter (Uschold 2022, 3) which, in the case of the CIDOC CRM, is factual knowledge about the human past. Widely recognized in the Heritage field, the CIDOC CRM is since 2006 the ISO standard for the representation of museum and cultural heritage knowledge. However, it is also well-known that mastering the model requires a long learning process, and its improper use, without adequate training and understanding, can lead to unintended or suboptimal results. Over its lifetime, the CIDOC CRM has undergone several significant revisions, which are thoroughly documented in an extensive, albeit jargon-heavy, body of literature. In the past ten years, a suite of CIDOC-compliant extensions has been developed to enhance archaeological data integration (CRMarchaeo, CRMba, CRMsci, CRMgeo; Meghini *et al.* 2017, 22). Further extensions – both official and unofficial – have broadened the model's potential user base well beyond its original museum-focused domain. While the CIDOC CRM initially concentrated on representing knowledge within memory organizations, it has since evolved to represent factual information about human history, the things that humans have done, created, and experienced throughout time.

The Heritage-Semantic Tools and Interoperability Survey (Scarpa, Valente 2024a), conducted as part of the H2IOSC project (www.h2iosc.cnr.it), highlighted the widespread adoption of the CIDOC CRM. However, the survey, undertaken to prepare for the development of an ontology for the Heritage Science domain within the same project, also revealed a fragmented scenario, marked by limited reusability standards, poor data integration, and an overall project-led approach to semantic interoperability in the Heritage field (Scarpa, Valente 2024b). This paper offers a concise and accessible overview of the CIDOC CRM, outlining its core principles, advantages, challenges, and implications for use to facilitate better integration and reuse. At the time of writing, the most recent official ISO release is version 7.1.3 (Bekiari *et al.* 2024), while versions 7.2.4 and 7.3 remain in draft.

2. The CIDOC CRM in a nutshell

2.1 *Origin and recent developments*

The CIDOC CRM was originally conceived as an entity-relationship data model. However, following the ICOM meeting in Stavanger, Norway, in 1995, the newly formed CIDOC Documentation Standards Working Group (DSWG) – created by merging the Data Model WG and the Data Terminology WG – decided to transition to an object-oriented modeling approach (Crofts, Reed 1996). This decision was taken to extend the original entity-relationship model and make it more flexible and is part of a shift to object-oriented modeling in museum knowledge that took place in the mid-1990s (Bearman 2008, 43). As a consequence, the CIDOC CRM has been reengineered since 1996 by the DSWG into an ontology aimed at addressing the challenges of semantic interoperability across various types of museum data and their connections to archival and library materials. The so designed object-oriented model was submitted in 1999 to become an ISO standard, achievement accomplished in 2006 (ISO 2006), with two major updates in 2014 (ISO 2014) and 2023 (ISO 2023).

2.2 *The domain*

The domain that the CIDOC CRM aims to represent – *i.e.*, the specific area of knowledge the model focuses on – centers on cultural heritage, history, and related knowledge. In its latest release, the CIDOC CRM is defined as a model «intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information and similar information from other domains» (Bekiari *et al.* 2024, 9). Within this framework, cultural heritage encompasses «the things preserved by the memory institutions, *i.e.* museums, sites and monuments records ('SMR'), archives and libraries» (Doerr 2009, 464). Doerr (2003, 75) argued that the challenge in modeling cultural heritage lies not in defining it but in accounting for its inherent diversity and incompleteness. To address this complexity, CIDOC CRM structures knowledge through the historical events associated with objects, rather than focusing solely on the objects themselves (Bekiari *et al.* 2024, 33-34). The CIDOC CRM emphasizes understanding cultural heritage by examining the events that have shaped and influenced objects over time, rather than merely cataloging surviving artifacts. Without knowledge of the events surrounding an object's creation and use, its true meaning cannot be fully understood, since an object is defined by the events that characterize its existence.

Before exploring the implications of CIDOC CRM's modeling approach, it is important to note a significant shift in focus with the introduction of the RDF Schema (RDFS) implementation since v.7.1.1 (see *infra*). Initially, the CIDOC CRM was designed as a museum-centered model. In version

4.0, its scope was «summarised in simple terms as the curated knowledge of museums» (Crofts *et al.* 2004, ii). However, by version 7.1 and later, its focus expanded to encompass «the curated, factual knowledge about the past at a human scale» (Bekiari *et al.* 2021, ii), reflecting a broader historical perspective. As with any conceptualization of a subject matter, the CIDOC CRM is deeply rooted in its domain of knowledge and, therefore, reflects the specific needs, structures, and practices of the museum-oriented cultural heritage field. Understanding these aspects is essential to fully leveraging the model's potential.

2.3 *The basic tenets*

The domain representation provided by the CIDOC CRM is grounded in the philosophical distinction between two high-level concepts: persistent entities that maintain their identity over time (endurants) and entities that unfold in time (perdurants). This distinction is not unique to the CIDOC CRM but reflects a broader modeling strategy adopted by other upper-level ontologies, such as BFO (Otte *et al.* 2022), UFO (Guizzardi *et al.* 2022), and DOLCE (Gangemi *et al.* 2002, 168). For example, the CIDOC CRM class E10 Transfer of Custody is a perdurant and, consequently, a subclass of E2 Temporal Entity (for a description of CIDOC CRM basic classes see Bruseker *et al.* 2017, 111-114). An instance – a specific occurrence of a class – of E10 Transfer of Custody is 'the return of Picasso's Guernica to Madrid's Prado in 1981' (Bekiari *et al.* 2024, 66), an event that unfolds over time rather than persisting unchanged. This event describes both the artist Picasso and his painting, emphasizing the relational context. In contrast, the Rosetta Stone, classified as an E22 Human-Made Object (a subclass of E77 Persistent Item), retains its identity over time and is therefore an endurant (Bekiari *et al.* 2024, 74).

As Bearman (2008, 56) pointed out, this event-based modeling approach defines objects by their contexts, connecting them through a network of events. It emphasizes relationships between entities as actions, such as discovery and creation, highlighting the unity of diverse objects through these interactions. This emphasis on events enables the CIDOC CRM to move beyond simply cataloging objects, instead capturing the rich context of their existence within the flow of history.

Identity is another crucial concept in ontology modeling that has relevance to correctly apply the CIDOC CRM. To accurately determine the instances each class describes, it is essential to establish identity principles that distinguish entities and assess to which extent they remain the same over time despite undergoing changes. In the CIDOC CRM, real-world entities fall under E77 Persistent Item and are defined as entities that remain recognizable over time (*i.e.*, endurants) by some structural characteristics. In the CIDOC

CRM ecosystem, an item, such as a monument, artwork, or artifact, may experience transformations, but as long as certain essential characteristics persist, it is still recognized as the same entity. The CIDOC CRM focuses on documenting these changes while preserving the entity's identity through time. A clear example is the Great Sphinx of Giza (Bekiari *et al.* 2024, 34): although it lost its nose at some point in history, its defining characteristics remain intact – no one would dispute that the Sphinx before and after this change refers to the same entity. Even if the Sphinx were to lose its front paws, it would still be recognized as the Great Sphinx of Giza, as its overall form and essence remain.

The CIDOC CRM does not explicitly define a threshold at which identity is lost, because the concept of identity is inherently subjective and its interpretation can vary depending on the context. While the Sphinx retains its identity despite the loss of its nose, or even its paws, there comes a point, such as complete fragmentation, where its recognition as the same entity becomes questionable (Suchánek 2022, 7). At what exact point does the entity stop being the Sphinx? In this regard, E77 Persistent Item does not define a rigid boundary – the Sphinx identity depends on how it is recognized in documentation and historical interpretation. The description of E77 Persistent Item suggests that identity is not based on an immutable set of characteristics but rather on continuity and recognition across time. The question becomes: Does the identity of an entity disappear when it physically disintegrates, or does it persist in some conceptual form (e.g., through documentation, historical continuity, or reconstruction)? In the CIDOC CRM, even if a physical object ceases to exist, its identity can persist as an E89 Propositional Object, an immaterial entity documented through records, historical accounts, or reconstructions.

3. The CIDOC CRM in practice

3.1 *The CIDOC CRM and the Semantic Web*

Like many concepts, the term Semantic Web refers both to the object itself and its study. The concept of the Semantic Web was officially introduced in 2001 (Berners-Lee *et al.* 2001) as an extension of the World Wide Web enriched with meaning, although its foundational ideas were established in the late 1990s. As a research field, the Semantic Web began to take shape around the same time and has undergone at least two significant shifts: from ontologies to linked data, and from linked data to knowledge graphs (Hitzler 2021). Tim Berners-Lee's proposal of linked data principles for publishing structured data (Berners-Lee 2006) promotes the identification of resources with URIs to link the resources themselves, rather than merely linking the pages that describe them. Adopting the linked data approach enhances discoverability at both deeper and broader levels, facilitating greater data integration and reuse. The Semantic Web evolution has culminated (though it is far from finished) in what is now often referred to as the 'Semantic Web stack', a set of interconnected technologies and standards designed to support the representation, exchange, and querying of structured data on the web. By providing computers with structured, linked data and rule-based instructions, the Semantic Web enables automated reasoning, which is central to its 'semantic' aspect, deriving meaningful insights and new knowledge from pre-existing data. More recently, the introduction of the FAIR principles for research data established a new set of best practices (Wilkinson *et al.* 2016), which have also been increasingly applied to and shaped ontology development.

As the brief 'biography' of the CIDOC model outlined above suggests, the CIDOC CRM was not originally designed for seamless integration into the Semantic Web, as both were evolving during the same period, each striving to define its shape and purpose. While the earliest relational model underlying the CRM was not oriented toward Semantic Web technologies, which did not yet exist, the development of the CIDOC CRM as an ontology from 1996 onward strongly aligned with what would become the core vision of the Semantic Web. In addition, the CIDOC CRM, like many other semantic artefacts within and beyond the Heritage field, was not initially developed following FAIR-compliant methodologies, linked data principles, or other Semantic Web technologies.

Tibaut and Guerra De Oliveira (2022) measured the appropriateness of the CIDOC CRM 2014 ISO release for reusability. Their in-depth evaluation of the CIDOC CRM's metrics and value scores with reference to specific parameters, provides valuable insight into the CIDOC CRM's structure. Their approach involved developing a new ontology evaluation framework, building upon the OQuaRE (Ontology Quality Requirements and Evaluation) framework (Wilson *et al.* 2022). They applied an improved version to evaluate the CIDOC CRM across several parameters, including the number of external ontologies used, composability (*i.e.*, how well it incorporates external ontologies through a name-spacing mechanism), and aggregability (the contribution of external ontologies within the CIDOC CRM). While Tibaut and Guerra de Oliveira did not directly evaluate version 7.1.3, their analysis focused on fundamental structural and design characteristics of the CIDOC CRM model itself, many of which are inherent to its conceptual approach and persist across version updates.

The CIDOC CRM ISO 21127:2014 obtained the lowest scores for the Reliability and Functional Adequacy parameters. Reliability measures how well the ontology keeps working over time without issues. It is based on several factors, including how deep the structure is (from general to specific concepts). The CIDOC CRM's score for this parameter is low because it has a very detailed, 10-level structure that includes both broad and very specific concepts, which makes it harder to maintain consistently. This is the case of concepts such as E34 Inscription, a class defined as comprising 'recognisable texts that can be attached to instances of E24 Physical Human-Made Thing'. E34 Inscription is a subclass of (in order): < E37 Mark < E36 Visual Item < E73 Information Object < E90 Symbolic Object < E28 Conceptual Object < E71 Human Made Thing < E70 Thing < E77 Persistent Item < E1 CRM Entity (Alexiev *et al.* 2013, 7). This is CIDOC's deepest branch; note however, that multiple inheritance is also present, as E34 Linguistic Object is also a superclass of E34 Inscription. This detailed structure affects its overall reliability score.

It must also be stressed that, at present, the CIDOC CRM does not natively reuse identical concepts from other ontologies. It defines its own unique classes and properties specifically crafted to meet the needs of the domain it describes, *i.e.*, it has a high ontological commitment. Unlike other ontologies that may reuse concepts from shared vocabularies (such as FOAF or Schema.org) by importing them or directly linking to their URIs, CIDOC CRM originally established a distinct conceptual model. This means that even when its concepts are semantically similar to those found in other ontologies, the CIDOC CRM defines its own versions rather than reusing or aligning with existing concepts from external sources by default. The structure of the v7.1.3 definition document indicates that it continues to define its own core concepts rather than strictly building upon or formally mapping to these external ontologies within the core model document itself. Therefore, Tibaut and Guerra De Oliveira's (2022, 17-18) recommendation to integrate with or reuse such ontologies, or to use linking mechanisms like SKOS to enhance interoperability, remains highly relevant for v7.1.3.

3.2 *RDFS and OWL implementations*

The CIDOC CRM's formal definition uses semantic data modeling principles and is expressed through first-order logical axioms, but is also designed to be convertible to and implementable in formats like RDF, RDFS, and OWL. Up to version 7.0, the CIDOC CRM was formalized as an XML schema. Since version 7.1.1, it has been formalized in the RDF data model and its RDFS extension for representing concepts, relationships, and hierarchies. Additionally, an OWL formalization based on version 7.1.3 has been developed. These formats, however, have expressivity limitations that prevent their use for the full formal specification of the CRM's intended semantics (Meghini, Doerr 2018).

While RDF provides a foundational structure for representing data as triples (subject-predicate-object), RDFS builds upon RDF by enabling basic ontology modeling. However, RDFS remains limited in expressiveness: it does not support advanced features such as property quantification or complex class expressions, though it does enable basic inference. In RDFS, the properties rdfs:domain (defining the class to which the subject of a triple necessarily belongs) and rdfs:range (defining the class of the object) are essential for specifying relationships between properties and classes. However, the strict type of inferences they trigger can complicate the representation of broader or more nuanced data structures (Uschold 2022, 84-85). If the property ex:writtenBy has the domain ex:book and the range ex:author, then for any triple using this property, the subject is inferred to be a book and the object is inferred to be an author. Therefore, the triple ex:NaturalHistory ex:writtenBy ex:PlinyTheElder is valid because Natural History is correctly classified as an instance of the class ex:book, and Pliny the Elder is correctly classified as an instance of the class ex:author. However, it is important to note that rdfs:domain and rdfs:range do not merely state that 'books are written by authors'; rather, they assert that anything that appears as the subject of ex:writtenBy is inferred to be a ex:book, and anything that appears as the object is inferred to be an ex:author. This means that if a triple uses ex:writtenBy, the subject will automatically be treated as a book and the object as an author, regardless of how they were originally classified. In the case of a property such as P108 has produced, the subject of a triple must be an instance E12 Production (domain) and the object an instance of E24 Physical Human-Made Thing (range) (Bekiari *et al.* 2024, 171). The triple ex:CreationOfTheMonaLisa crm:P108_has_produced ex:MonaLisa is valid because subject and object belong to the correct classes. But if the same property is used for a triple such as ex:BattleOfWaterloo crm:P108_has_produced ex:StrategyDocument the model would incorrectly infer that the Battle of Waterloo is an instance of E12 Production, which is a category for production activities designed to create new items (like making a physical object or artefact). Note that according to the CIDOC CRM, the Battle of Waterloo is an instance of E7 Event.

Compared to RDFS, OWL provides more sophisticated means of representing knowledge. An OWL implementation of the CIDOC CRM was developed by researchers at Erlangen University. This version, known as CIDOC Erlangen (https://erlangen-crm.org/), uses OWL-DL 1.0 to formalize existential and cardinality restrictions. The lack of semantic enforcement for cardinality and other restrictions in the original CIDOC CRM has significant implications for data modeling and interoperability. Without formalized constraints, there is a risk of inconsistent interpretations among different systems and implementations, leading to potential inaccuracies in data representation. For instance, while the documentation states that the property P98 brought into life has a cardinality of 'one to many', this does not prevent users from misapplying the property in ways that violate these guidelines. For example, the property P11 had participant, which connects the class E5 Event with E39 Actor, is defined with a cardinality of 'many to many (0,n:0,n)' (Bekiari *et al.* 2024, 121-122; Sanfilippo *et al.* 2020, 108-109). This allows events to exist without any participants and permits actors to engage in zero or multiple events. However, this flexibility can result in divergent formalizations: while some users may adhere to the original cardinality, others might impose restrictions on E39 Actor, requiring at least one actor for each event. While models based on the latter interpretation can encompass those of the former, the reverse is not true, leading to potential mismatches in data representation.

Meghini and Doerr (2018) argued that RDFS and OWL, while practical for system implementation, are not ideally suited as abstract formal models. In particular, the use of International Resource Identifiers (IRIs), which are essential in RDF and OWL to uniquely identify resources, is considered unnecessarily complex and ill-suited for expressing the formal structure of CIDOC CRM as they are over-complicated for human understanding and unnecessary when trying to describe or model concepts. Consequently, the CIDOC CRM is formalized using first-order logic, a system widely used in fields such as mathematics and computer science to describe relationships and properties with simplicity and clarity.

3.3 *CIDOC CRM knowledge graphs*

As outlined above, the CIDOC CRM, which is a highly expressive ontology, struggles with the limited expressiveness of ontology languages such as RDFS and OWL: shortcuts have been designed, which serve the purpose of representing relationships at different levels of granularity (Bekiari *et al.* 2024, 16). For example, the property P2 has type is a shortcut linking an E1 CRM Entity to an E55 Type (E1 CRM Entity → P41i was classified by → E17 Type Assignment → P42 assigned → E55 Type). As an event-based model, the CIDOC CRM defines the assignment of types as an event or action (E17 Type Assignment). So, instead of directly stating 'this object is of type sculpture', the CIDOC CRM records that someone assigned the type 'sculpture' to the object. The first-order logical definition of the shortcut P2 has type ($P2(x,y) \Leftarrow (\exists z)\,[E17(z) \wedge P41i(x,z) \wedge P42(z,y)]$) implies that the statement $P2(x,y)$('x has type y') holds whenever there exists some z that is an instance of E17 Type Assignment, such that x is the object assigned by z (via P41i), and z assigns the type y (via P42). This allows simply stating that an object 'has type' 'sculpture', bypassing the explicit modeling of the E17 Type Assignment event where that classification occurred.

This event-based modeling nature of the CIDOC CRM makes it difficult for non-experts to query Knowledge Graphs (KGs) that adopt it. A user might be interested in information like where a specific object, e.g. a clay tablet, was found and by whom. In a CIDOC-based KG, to retrieve e.g. information about a clay tablet found by sir Leonhard Woolley during the excavations at Ur, the user must navigate long ontology path patterns, *i.e.* sequences of relationships or classes that the user must follow within an ontology to navigate from one concept to another. In this case, the user would need to trace relationships like the tablet being part of the E18 Physical Thing class, then connect it to the finding event (E5 Event) through P12i was present at, and then link the finding event to E39 Actor via P14i carried out by, ultimately reaching the actor, sir Leonard Woolley (Tzompanaki, Doerr 2012, 51-52; Mountantonakis, Tzitzikas 2025). The path radius, which refers to the distance in terms of ontology relationships the users are willing to traverse from the starting point in the query, can be high. The further one needs to go in the ontology to connect the starting object (the clay tablet) to relevant contextual information (like the excavation event or the actor involved), the more complex and challenging it becomes for non-experts to perform effective queries. Several attempts have been made to simplify SPARQL queries in CIDOC CRM KGs (Alexiev 2012; Alexiev *et al.* 2013; Mountantonakis, Tzitzikas 2025).

One of the major challenges that CIDOC CRM-based knowledge graphs face is the integration of preexisting data. Heritage information is highly diverse, often stored in disparate, heterogeneous systems with varying schemata, documentation practices, languages, and levels of detail. The collaborations with Ontotext (Alexiev 2012; Alexiev *et al.* 2013), conducted as part of the ResearchSpace project (https://researchspace.org/) using a vast pool of preexisting data such as the British Museum database, highlighted the significant challenge of querying the complex graphs that result from applying CIDOC CRM to large datasets (Alexiev 2012; Alexiev *et al.* 2013). By employing a Fundamental Relations approach, Ontotext was able to simplify CIDOC CRM's potentially complex queries, resulting in enhanced semantic querying of the knowledge graph of the British Museum comprising 2,051,797 museum objects, 415,509 thesaurus entries, 195,208,156 explicit statements, and a 42 GB repository.

4. Final remarks

The CIDOC CRM is a complex ontology, a fact reflected in its deep class and property hierarchies, extensive use of multiple inheritance, and large number of properties. Its complexity also stems from its philosophical underpinnings, such as the distinction between endurants and perdurants, and the focus on identity, as well as the rich, detailed definitions provided for its components. The CIDOC CRM is highly expressive, capable of capturing the complexity and nuances of cultural heritage information with fine granularity and allowing for intricate interconnections. It supports modeling relationships at different levels of detail, often offering both a fully elaborated path through intermediate entities and a shortcut property that bypasses them. However, RDFS does not support property quantification, and strong shortcut semantics are not expressible in OWL. As a result, neither language fully meets the expressiveness needs of the CIDOC CRM.

Erica Scarpa
Istituto di Scienze del Patrimonio Culturale - CNR
erica.scarpa@cnr.it

Acknowledgements

This research was supported by the 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

Alexiev V. 2012, *Implementing CIDOC CRM search based on fundamental relations and OWLIM rules*, in *Semantic Digital Archives: Proceedings of the 2nd International Workshop on Semantic Digital Archives (Paphos 2012)*, CEUR Workshop Proceedings 912 Paphos, 95-105 (https://ceur-ws.org/Vol-912/paper8.pdf).

Alexiev V., Manov D., Parvanova J., Petrov S. 2013, *Large-scale reasoning with a complex cultural heritage ontology (CIDOC CRM)*, in V. Alexiev, V. Ivanov, M. Grinberg (eds.), *Workshop Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013)*, CEUR Workshop Proceedings La Valetta, Malta, 80-93 (http://ceur-ws.org/Vol-1117/paper8.pdf).

Bearman D. 2008, *Representing museum knowledge*, in P.F. Marty, K. Jones (eds.), *Museum Informatics*, New York, Routledge, 35-57 (https://doi.org/10.4324/9780203939147).

Bekiari C., Bruseker G., Canning E., Doerr M., Michon P., Ore C.-E., Stead S., Velios A. 2024, *Definition of the CIDOC Conceptual Reference Model v.7.1.3*. (https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.3.pdf).

Bekiari C., Bruseker G., Doerr M., Ore C.-E., Stead S., Velios A. 2021, *Definition of the CIDOC Conceptual Reference Model v.7.1*. (https://cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.1%20%5B8%20March%202021%5D.pdf).

Berners-Lee T. 2006, *Linked Data* (https://www.w3.org/DesignIssues/LinkedData.html).

Berners-Lee T., Hendler J., Lassila O. 2001, *The semantic web*, «Scientific American», 29-37.

Bruseker G., Carboni N., Guillem A. 2017, *Cultural heritage data management: The role of formal ontology and CIDOC CRM*, in M.L. Vincent, V.M. López-Menchero Bendicho, M. Ioannides, T.E. Levy (eds.), *Heritage and Archaeology in the Digital Age*, Quantitative Methods in the Humanities and Social Sciences, Cham, Springer International Publishing, 93-131 (https://doi.org/10.1007/978-3-319-65370-9_6).

Crofts N., Doerr M., Gill T., Stead S., Stiff M. 2004, *Definition of the CIDOC Conceptual Reference Model v.4.0*. (https://cidoc-crm.org/sites/default/files/cidoc_crm_version_4.0.pdf).

Crofts N., Reed P. 1996, *CIDOC data modeling working group*, «Archives and Museum Informatics», 10, 1, 89-91 (https://doi.org/10.1007/BF02802421).

Doerr M. 2003, *The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata*, «AI Magazine», 24, 3, 75-75 (https://doi.org/10.1609/aimag.v24i3.1720).

Doerr M. 2009, *Ontologies for cultural heritage*, in S. Staab, R. Studer (eds.), *Handbook on Ontologies*, Berlin-Heidelberg, Springer, 463-486 (https://doi.org/10.1007/978-3-540-92673-3_21).

Gangemi A., Guarino N., Masolo C., Oltramari A., Schneider L. 2002, *Sweetening ontologies with DOLCE*, in A. Gómez-Pérez, V.R. Benjamins (eds.), *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Berlin-Heidelberg, Springer, 166-181 (https://doi.org/10.1007/3-540-45810-7_18).

Guizzardi G., Benevides A.B., Fonseca C.M., Almeida J.P.A., Sales T.P., Porello D. 2022, *UFO: Unified Foundational Ontology*, «Applied Ontology», 1, 17, 167-210 (https://doi.org/10.3233/ao-210256).

Hitzler P. 2021, *A review of the semantic web field*, «Communications of the ACM», 64, 2, 76-83 (https://doi.org/10.1145/3397512).

ISO 2006, *Information and Documentation. A Reference Ontology for the Interchange of Cultural Heritage Information (ISO Standard No. 21127:2006)* (https://www.iso.org/standard/34424.html).

ISO 2014, *Information and Documentation. A Reference Ontology for the Interchange of Cultural Heritage Information (ISO Standard No. 21127:2014)* (https://www.iso.org/standard/57832.html).

ISO 2023, *Information and Documentation. A Reference Ontology for the Interchange of Cultural Heritage Information (ISO Standard No. 21127:2023)* (https://www.iso.org/standard/85100.html).

Meghini C., Doerr M. 2018, *A first-order logic expression of the CIDOC Conceptual Reference Model*, «International Journal of Metadata, Semantics and Ontologies», 13, 2, 131 (https://doi.org/10.1504/ijmso.2018.098393).

Meghini C., Scopigno R., Richards J., Wright H., Geser G., Cuy S., Fihn J., Fanini B., Hollander H., Niccolucci F., Felicetti A., Ronzino P., Nurra F., Papatheodorou C., Gavrilis D., Theodoridou M., Doerr M., Tudhope D., Binding C., Vlachidis A. 2017, *ARIADNE: A research infrastructure for archaeology*, «Journal on Computing and Cultural Heritage», 10, 3, 1-27 (https://doi.org/10.1145/3064527).

Mountantonakis M., Tzitzikas Y. 2025, *Generating SPARQL queries over CIDOC-CRM using a two-stage ontology path patterns method in LLM prompts*, «Journal on Computing and Cultural Heritage», 18, 1, 1-20 (https://doi.org/10.1145/3708326).

Otte J.N., Beverley J., Ruttenberg A. 2022, *BFO: Basic Formal Ontology*, «Applied Ontology», 17, 1, 17-43 (https://doi.org/10.3233/AO-220262).

Sanfilippo E.M., Markhoff B., Pittet P. 2020, *Ontological Analysis and Modularization of CIDOC-CRM*, in B. Brodaric, F. Neuhaus (eds.), *Frontiers in Artificial Intelligence and Applications*, IOS Press, 107-121 (https://doi.org/10.3233/FAIA200664).

Scarpa E., Valente R. 2024a, *Heritage. Semantic Tools and Interoperability Survey* (https://doi.org/10.71795/8t2z-hh65).

Scarpa E., Valente R. 2024b, *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).

Suchánek M. 2022, *OntoUML Specification Documentation* (https://ontouml.readthedocs.io/en/latest/index.html).

Tibaut A., Guerra De Oliveira S. 2022, *A framework for the evaluation of the cultural heritage information ontology*, «Applied Sciences», 12, 2, 795 (https://doi.org/10.3390/app12020795).

Tzompanaki K., Doerr M. 2012, *Fundamental Categories and Relationships for Intuitive Querying CIDOC-CRM Based Repositories* (https://publications.ics.forth.gr/tech-reports/2012/2012.TR429_Intuitive_querying_CIDOC-CRM.pdf).

Uschold M. 2022, *Demystifying OWL for the Enterprise*, Synthesis Lectures on Data, Semantics, and Knowledge, Cham, Springer International Publishing (https://doi.org/10.1007/978-3-031-79482-7).

Wilkinson M.D., Dumontier M., Aalbersberg Ij.J. *et al.*, *The FAIR guiding principles for scientific data management and stewardship*, «Scientific Data», 3, 1, 1-9 (https://doi.org/10.1038/sdata.2016.18).

Wilson R.S.I., Goonetillake J.S., Ginige A., Indika W.A. 2022, *Ontology quality evaluation methodology*, in O. Gervasi, B. Murgante, E.M.T. Hendrix, D. Taniar, B.O. Apduhan (eds.), *Computational Science and Its Applications. ICCSA 2022*, Cham, Springer International Publishing, 509-528 (https://doi.org/10.1007/978-3-031-10522-7_35).

ABSTRACT

This paper provides a concise overview of the CIDOC Conceptual Reference Model (CIDOC CRM). The CIDOC CRM is a formal ontology initially developed for museums and cultural institutions to describe and organize their data. It serves as the ISO standard for representing museum and cultural heritage knowledge. The paper outlines the model's core principles, advantages, challenges, and implications for its use, touching upon the CIDOC CRM's relationship with the Semantic Web and challenges in implementing it with technologies like RDFS and OWL. It also highlights challenges in querying CIDOC-based knowledge graphs and integrating pre-existing heterogeneous data.