Giving a shape to concepts: insights into ontology modeling through the CHint example

Sofia Baroncini

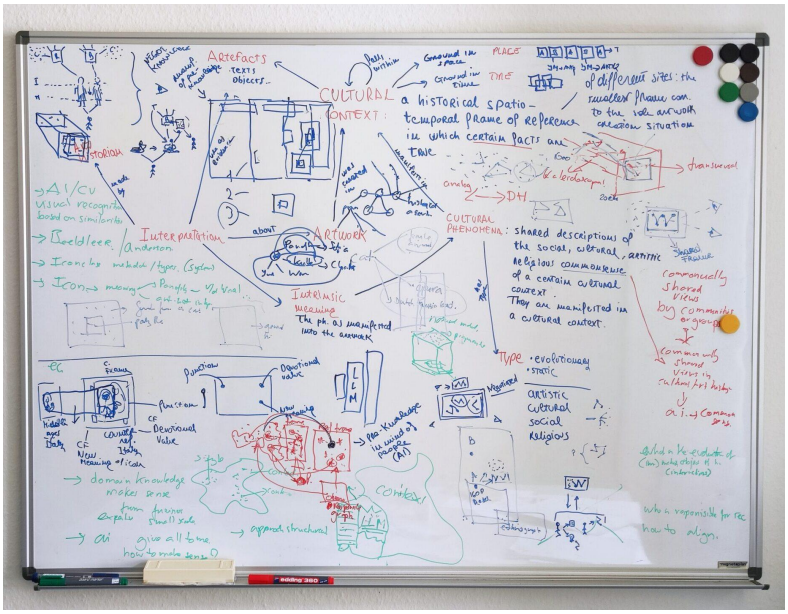
Blog post published September 22, 2026 in *DH Lab*
https://doi.org/10.58079/16u8l

Abstract

Introduction: What is an ontology? Everything we can do on a computer is encoded; namely, a sequence of 0-1 bits is associated with symbolic content. This is the core fundamental principle allowing us to do a range of different operations on a computer, from writing a document in Word to viewing an image.

Keywords

History (Subfield), Semantic Web and Ontologies (Topic), Artificial Intelligence (Subfield), Tutorial, Art History, Ontology Modeling, Semantic Web



Recommended citation

Baroncini, S. (2026, September 22). Giving a shape to concepts: insights into ontology modeling through the CHint example. *DH Lab*. https://doi.org/10.58079/16u8l

Copyright

Copyright © Sofia Baroncini 2026. Distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction: What is an ontology?

Everything we can do on a computer is encoded; namely, a sequence of 0-1 bits is associated with symbolic content. This is the core fundamental principle allowing us to do a range of different operations on a computer, from writing a document in Word to viewing an image.

If everything can be encoded through this kind of association, then **machines can compute meanings as well**. Since the 1950s, the branch of **symbolic AI** has researched the possibility of doing a detailed description of the world, in the form of concepts and relations, and associating them with logical rules. In this way, a computer could do sophisticated operations, such as inferring knowledge not directly stated in the data. For example, if the data states that Sonia is the sister of Juliette and Alice, and the relation is declared as being transitive, then the system could infer that Juliette and Alice are sisters, too.

One application of this approach is **ontologies**, in other words, a **structured way**, processable by computers, of describing a **particular domain of knowledge**. It identifies the **main concepts** in that domain and, most importantly, the **relationships** between them. For example, in a project about works of art, an ontology might define concepts such as Person, Artwork, the Creation Event, Institutions holding the artwork, the Iconography depicted, and so on. In this sense, an ontology does more than provide a list of terms. It provides a shared **conceptual framework** that helps both humans and computers understand **how different pieces of information are connected**.

Furthermore, an ontology can also state **logical rules** allowing the system to infer further information, or to check that data is described correctly. For example, in Figure 1, the ontology states that the concept “Artist” is a subclass of “Person”. If, in the dataset, “Monet” is described as “Artist”, the computer, through a **reasoner** (i.e., a software tool that automatically checks the logical consistency of the ontology and derives new facts from it), can infer that Monet is not only an artist, but also a person. This means that in a dataset in which all people are described according to their profession, and all the professions are stated as a subclass of “Person”, we can easily retrieve all the people in the dataset although none of them is explicitly described as a person.

https://doi.org/10.58079/16u8l

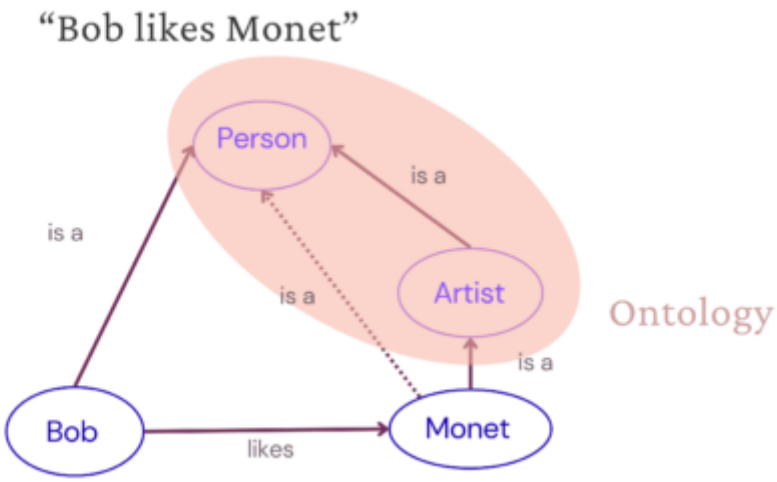


Figure 1. Example of how an ontology can be used to infer further knowledge. If in the graph dataset I insert the information that Monet is an Artist, and the ontology "Artist" is stated as being subclass of "Person", then the computer can infer that Monet is also a "Person", without the need to manually enter that information. Figure by Sofia Baroncini

A key concept is the **distinction between the ontology itself,** defining **classes or concepts**, and data described according to that ontology, representing **individuals or instances** (see Figure 2). Whereas the first describes how the types of entities relate to each other and represents the domain in abstract terms (e.g., "Artist" is a subclass of "Person"), the latter contains the actual content ("Monet"). In other words, it is similar to the difference between an Excel sheet showing only the column names and the same sheet filled with actual data: the columns give us information about how the data is organized ("People"), whereas the rows are the actual individuals ("Monet", "Picasso", etc.). In ontology development, these are also called "**T-BOX**" and "**A-BOX**". The actual content needs to follow the structure defined in the ontology, as shown in Figure 2.

From a technical point of view, ontologies are commonly formalized in a specific language, **OWL 2** (Web Ontology Language 2), which defines the logical relations among entities, and in the **RDF Schema**, a light and less expressive vocabulary to define classes, subclasses, and relations. Ontologies can be easily created with Protégé, an open-source program with a user-friendly interface to create entities and relations and options for saving them in OWL.

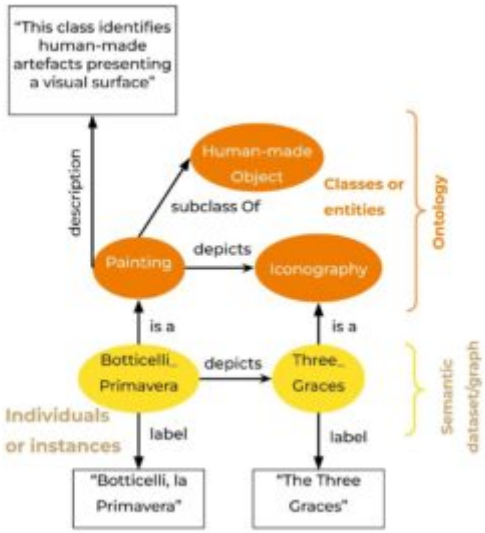


Figure 2. Demonstrative example of an ontology (in orange) describing the classes or entities and their properties or relations in a domain of knowledge, also called "T-Box", and the individuals or instances represented according to the ontology (in yellow), also called "A-Box". As the ontology states that Paintings can depict (relation "depicts") an Iconography, we can reuse this pattern to express that the individual painting "La Primavera" by Botticelli depicts the three graces, another individual having as type the entity Iconography. Relations must be stated in the ontology to be used in data, and must respect the direction of the arrow. For example, when applying this ontology, we can't say that an iconography depicts an artwork – it should always be the contrary. Figure by Sofia Baroncini

Several languages and applications exist. One of the most common areas of application in the DH sector is in the Semantic Web. **Semantic Web** is an extension of the hypertext web in which not only web pages, but also concepts are related to each other. URIs are used not only to identify web pages, but also to identify concepts, entities, and relationships. For example, this URI on Wikidata identifies the Mona Lisa painting, and it also states, through the URI indicating the relationship of "creator", that it was painted by Leonardo da Vinci, associated with another URI.

The benefits of the Semantic Web and ontologies in general are multiple. They help algorithms to clearly identify concepts and differentiate among them, even if they have the same name (e.g., "Venus" as a planet is different from "Venus" indicating the deity of love in Greek mythology). As computers can process this information, computational operations are done with semantic data. An example is protein prediction in the field of biology, done through ontologies and Machine Learning.

In the humanities area, ontologies are mainly used as **standards** to guarantee **interoperability between datasets** maintained by different institutions in the GLAM (Galleries, Libraries, Archives, Museums) sector. If structured data is described according to the same ontology, data can be integrated and queried together, even though separate datasets are created and maintained by

different institutions. In this way, a global, interconnected knowledge graph can be created in a decentralized way. The use of ontologies to describe data is an important component of the creation of **Linked (Open) Data**, i.e., a network of integrated data described using common, open ontologies. The linked data cloud describes the massive, growing amount of LOD datasets and their interconnections (Figure 3).

Following these principles, portals offering access to massive amounts of data were created, such as Europeana, the European portal for Cultural Heritage, or Artresearch.net, a knowledge graph integrating data from the art photographic archives of the PHAROS association.

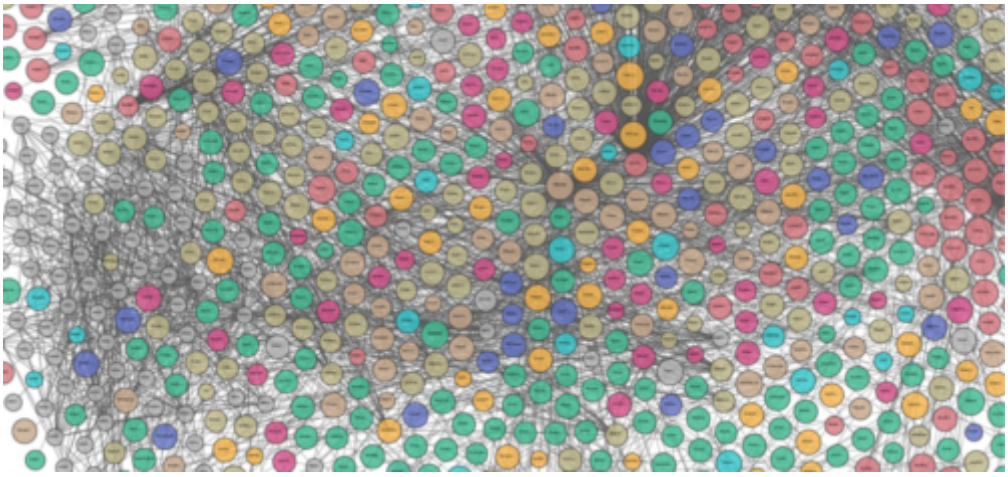


Figure 3. Snapshot of the Linked Data Cloud on 02/09/2026. Source: https://lod-cloud.net/

Ontology Modeling

But how to create an ontology? In this blog post, we will take you through an overview of the steps to create an ontology through the concrete example of the development of the Cultural Heritage Interactions Ontology (CHint) (Baroncini & Heuvel, 2026). The aim of the ontology is to represent various types of interactions between material and immaterial cultural heritage (CH) that occur over time and space. In other words, it relates a CH object with 1) ideas and values that are relevant in the context(s) in which the object is located (i.e., a belief system), 2) the function(s) that such objects acquire in different contexts, 3) the evolution of these features over time, 4) their interaction with contextual ephemeral events and immaterial cultural heritage practices, and 5) different aspects of uncertainty related to the reconstruction of past state-of-affairs of the object.

Before getting into details, it is important to underline that **there's never a single way to model knowledge.** Every model reduces infinite reality by selecting specific features. For example, the floor plan of a building does not render the 3D appearance of the building, nor does a 3D accurate scan record the temperature inside the rooms. Every model follows a specific purpose or research question, and discards a certain amount of information.

The same happens with **data modeling**. Whereas "data" – literally, something "given" from the Latin word *datum* – is perceived as being something neutral, it **always reflects a particular point of view on reality**. This is particularly evident when we deal with **humanities data**. Not only do the

branches of study **reflect a worldview** or way of thinking, but also the information modeled relies on often **uncertain, approximate, and incomplete sources**, which raise a range of subjective interpretations about the past state of affairs.

Several **methodologies** can be used to develop an ontology. Among them, agile and iterative approaches, such as NeOn (Suárez-Figueroa et al., 2015), eXtreme Design (Presutti et al., 2009), and SAMOD, are particularly relevant. Instead of modeling the whole ontology at once, the developer divides the domain into smaller areas and adds them iteratively to the main ontology. During the development of CHint, we adopted the SAMOD method (Peroni, 2016) for modular development, with the aid of the Ontology Requirements Specification Document for the initial definition of the characteristics of the overall ontology (Suárez-Figueroa et al., 2009). For this reason, we will focus on them in the illustration of the development process.

Define the requirements

In the first place, the scope and aim of the project should be defined: what's the aim of the modeling? In which domain? For whom? What questions should the model be able to answer? If a domain expert and a developer are involved, these aspects should be defined together. It is a fundamental phase, as **the decisions that are taken in this step shape all the following process**.

The set of needs that the model should fulfill to be considered finished is called "requirements". Documents such as the Ontology Requirements Specification Document (Suárez-Figueroa et al., 2009) are useful for clarifying these characteristics. In SAMOD, however, these requirements are progressively defined and refined through a series of concrete motivating scenarios. Rather than trying to specify the whole ontology from the beginning, each iteration focuses on a specific part of the domain. Following the SAMOD method, domain experts describe a specific motivating scenario, i.e., a small, concrete description of a problem in the domain that needs to be modeled, accompanied by examples. From this scenario, the developing team formulates informal competency questions, namely, questions that the ontology should eventually be able to answer (for example, "which subjects are represented in Botticelli's *Primavera*"?).

For the development of the Cultural Heritage Interactions Ontology, it was set that the ontology aimed to narrow the gap between records about physical cultural heritage objects (e.g., paintings, altarpieces) and manifestations of immaterial culture (e.g., a certain belief, religious events) that regard them over time and space. Such interaction can indeed deeply change the object, resulting in new functions or iconography – take into account, for example, the religious objects often musealized, appreciated more for their aesthetic value than for their religious function after their decontextualization. The scope of application is the CH sector, and it is intended to be used by digital cultural historians and/or computer scientists interested in developing applications concerning cross-cultural understanding of CH objects. The intended use of the ontology is to "express in a machine-readable format the complex interactions between contemporary culture and immaterial events, and thoughts that eventually influenced CH objects themselves, their function and/or meaning" (Baroncini & Heuvel, 2026, p. 14). The technical requirements were an extensive reuse of available ontologies and the formalization in OWL 2. Whereas the document was initially used to set the aim, scope, and use of the ontology, further sections were filled

during the following development steps. For example, the definition of the glossary of terms, i.e., keywords relevant to the domain, depended on the theories taken into account for the definition of the domain, illustrated in the following section. Another example is the definition of competency questions, defined in the SAMOD iterations during the development. Furthermore, at this initial stage, we identified a set of requirements the ontology should address, developed in specific scenarios during the development steps.

From ideas to sketches: top-down and bottom-up approaches

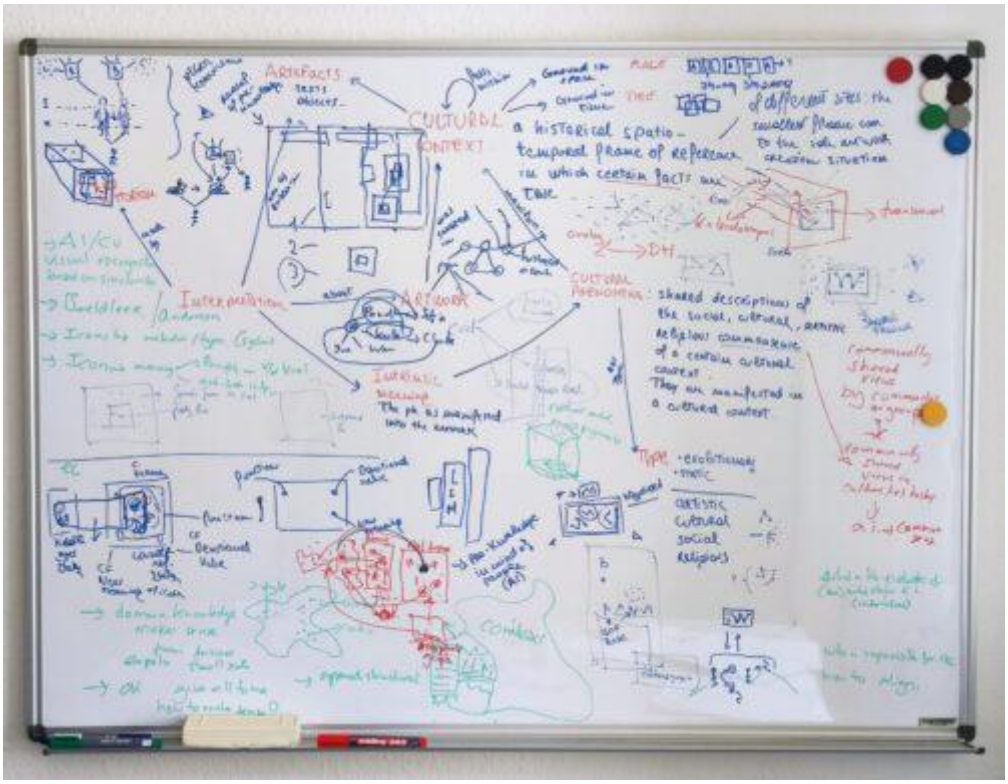


Figure 4. Sketches by Sofia Baroncini and Charles van den Heuvel to grasp the interaction between artworks and immaterial culture over time and space at the beginning of the development of the CHint Ontology. Copyright of the authors.

The first important step to take is to think deeply about the domain. Depending on the cases and theories available, two main approaches (or a mix of them) are possible. The first one, **top-down**, consists of searching for pre-existing domain theories that model the domain and implementing them in the modeling. In this case, the developer adopts a deductive approach, and they move from the theory to the implementation. The second, opposite approach is to induce a general structure from multiple real cases, called **bottom-up**. For example, if there is a dataset about artworks sold in auctions, we can infer how artworks interact with the auction events and buyers from the data itself, and base the ontology on that rich example.

In the DH domain, ontologies often rely on preexisting theories or modeling patterns to represent the domain at hand. Indeed, domain theories are already a conceptualization of the world proposed by historians, philosophers, and humanists who are or were experts in their fields. For

this reason, domain theories can be an invaluable source when formalizing information. Nevertheless, theories may have their limits, as they were not at first created to be a computational model, and they reflect a socioculturally situated worldview, which may be outdated, biased, or limited to a certain culture (e.g., the Western world).

Once the best approach has been identified, according to the available theories or data, it is a good idea to start sketching on a whiteboard or piece of paper how concepts relate to each other. Figure 4 shows the results of a first brainstorming session we had to create the CHint ontology, in which we thought about how the chosen theories could relate to each other. For the development, we have adopted a top-down approach, as we identified relevant theories concerning the evolution of time and space, frames, and the relation of works of art with a belief system.

The importance of standards

It is good practice to reuse existing models if they are in line with the aims of the ontology. As one of the main goals of the Semantic Web and Linked Open Data is the integration of different datasets described with a common structure, it is fundamental to identify which standards and existing ontologies can be reused for the task. Usually, standards for each domain exist, providing a backbone of how the concepts are organized in the domain. For the CH domain, the main standard is CIDOC-CRM, a conceptual framework developed and maintained by the International Council of Museums (ICOM), which serves as a structure to model museum data in different areas, from art to maritime CH. It is a good practice to reuse the standards and existing ontologies as much as possible, as they have already solved core modeling issues, making the work of the ontology developer easier and guaranteeing interoperability.

In the CHint development, this means that existing models such as CIDOC-CRM can provide concepts that are relevant to the domain. Reusing these solutions improves interoperability and makes the resulting ontology easier to connect with other datasets.

Draft. Test. Iterate

Once there is an idea of the domain, scope, data, and theories available, the actual modeling can start. As cited above, many methodologies focus on the development of small portions of the ontology. Each iteration should include at least two types of testing: 1) testing the logical consistency of the ontology through a reasoner, and 2) testing the model against real data (A-BOX). This ensures that there are no logical inconsistencies and that the ontology can actually address the real-world scenario taken into account.

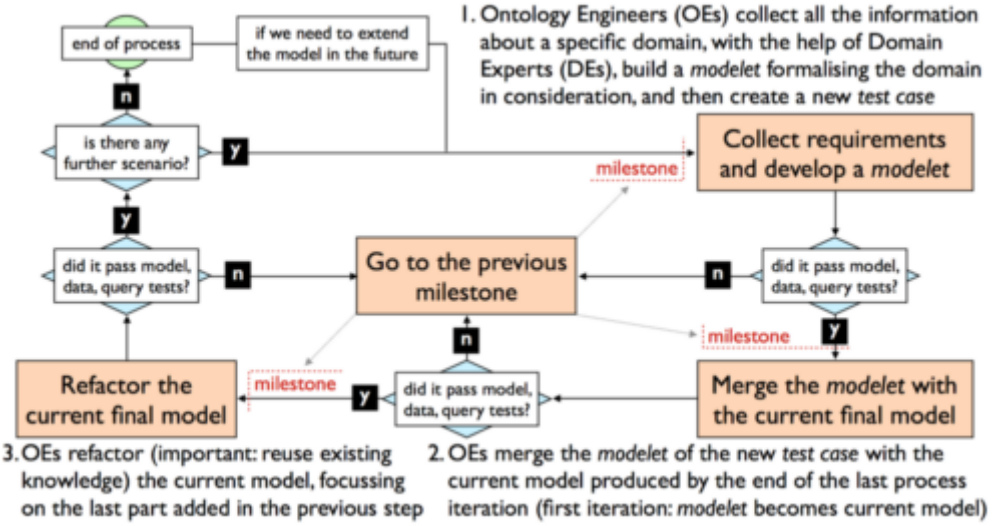


Figure 5. Overview of the ontology development process according to SAMOD. Source: Peroni, S. (2016). https://doi.org/10.6084/M9.FIGSHARE.3189769. Licensed under CC BY 4.0

Draft

As introduced above, the SAMOD method[1] divides the ontology into small motivating scenarios to allow agile development and focus on simpler aspects (Figure 5). Each scenario is a textual description of the area of the domain, and it is accompanied by a glossary of terms describing the core terminology of the scenario.

For the development, we've taken into account a specific case study, chosen for its richness, as a concrete example on which the model should be tested, constituting the A-BOX of the ontology. The case concerns the relics of St. Servatius preserved in Maastricht, the Netherlands, which are regularly brought in procession from the 14th century to the present day. For the richness of documentation, the strong relation with the Catholic community, and the frequency with which it interacts with an immaterial cultural manifestation, it was considered a prime example for the development.

The module *Historical Frame* of the CHint ontology was developed in one iteration and had one motivating scenario.

We quote here, as a way of example, one part of the motivating scenario:

> *A CH object has one or multiple contexts that are related to it. A context expresses the historical, socio-cultural situation in which an artwork exists. By definition, its delimitation is fuzzy, as it corresponds to a certain culture. It includes several traits of various natures, such as religion, social practice, taste, style, etc. Humans, and experts in art history, history, and so on, identify sets of salient traits for each context. In doing so, they frame reality in the sense of Nelson's theory to abstract such traits. In other words, they create observation frames that isolate from the reality contexts that present unifying traits. In doing so, they create a Historical Frame, which*

> *is a portion of the spacetime continuum created to better observe a context that already presents traits of unity. The traits are then collected in Reference Frames, defined according to Minsky's theory, that work as the domain pre-knowledge which is necessary for relating CH objects to their contexts (e.g., placing Raffaello's work in the Italian Renaissance).*
>
> *An object participates in one or more cultural frames, corresponding to 1) the context of artwork creation, and 2) further contexts the object traverses.*

The scenario, indeed, describes the interaction of CH objects with the socio-cultural, immaterial contexts in which they existed – either at the time of their creation, or in following situations. The glossary included key terms such as historical frame, time, event, etc.

For this scenario, we defined Competency Questions (CQs), questions that data described with the ontology should be able to answer. For this reason, the CQs explicitly refer to the case study:

- CQ1: What is the historical frame of the St. Servatius' cup in its original context (i.e., the context in which the object was created)?
- CQ3: What are the St. Servatius' cup historical frames that the object traverses?
- CQ4: What are the beliefs and ideas related to the historical frame "Religious, popular culture of Maastricht from the 14th century to nowadays"?

On the basis of the motivating scenario, glossary, and CQs, the ontology engineer drafts a *modelet* representing how the concepts relate to each other. The design is first a sketch, which is then rendered in an OWL 2 ontology through programs for ontology editing such as Protégé. In this stage, the design should take into account existing standards and available modeling patterns, but without explicitly importing them. Figure 6 illustrates the modelet of the Historical Frame motivating scenario illustrated with the Graffoo representation through the program draw.io. Whereas not explicitly introduced in the modelet to lower the complexity of the task at this step, classes and relations of existing ontologies, mainly standards, were taken into account. For example, the class E92 Spacetime-Volume of CIDOC-CRM was already identified as being in line with the concept of historical frame, as it describes a portion of space and time, or the Cultural Phenomenon class of the ICON ontology, representing deep socio-cultural phenomena recognized in artworks.

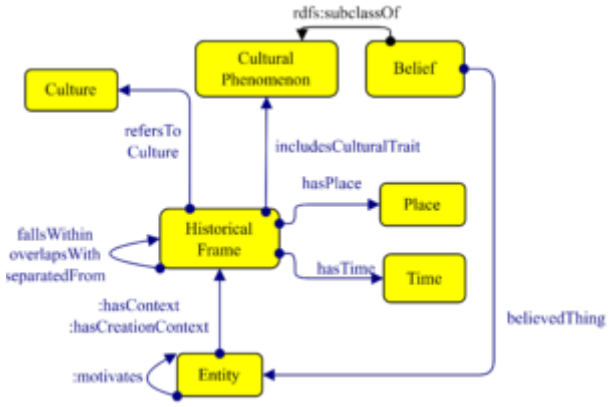


Figure 6. Modelet of the Historical Frame module of the CHint ontology. Created by Sofia Baroncini. Available at Zenodo (DOI: 10.5281/zenodo.20393679). Licensed under CC BY 4.0

Test, iterate, and align

After the creation of the modelet, its logical consistency should be tested through a reasoner, an operation that can be done in Protégé. If the test doesn't raise any error, then a small dataset of real examples described according to the model (A-BOX) needs to be created, and the logical test repeated. If no errors are raised, the Competency Questions should be translated into real SPARQL queries and performed over the dataset. If the queries give the expected result, we can merge the modelet with the main ontology developed in the previous phases.

The last phase of each iteration is the refactoring, namely, the alignment of the current ontology with existing ones, or replacing classes and relations with equivalent or more appropriate terms from existing models. For example, during the modeling of the modelet, we previously identified the class E92 Spacetime-Volume of CIDOC-CRM as suitable to express the concept of Historical Frame, of which it could be a specification. At this stage, we make this alignment explicit, and we declare the class Historical Frame as a subclass of crm:E92. All the tests should be performed again.

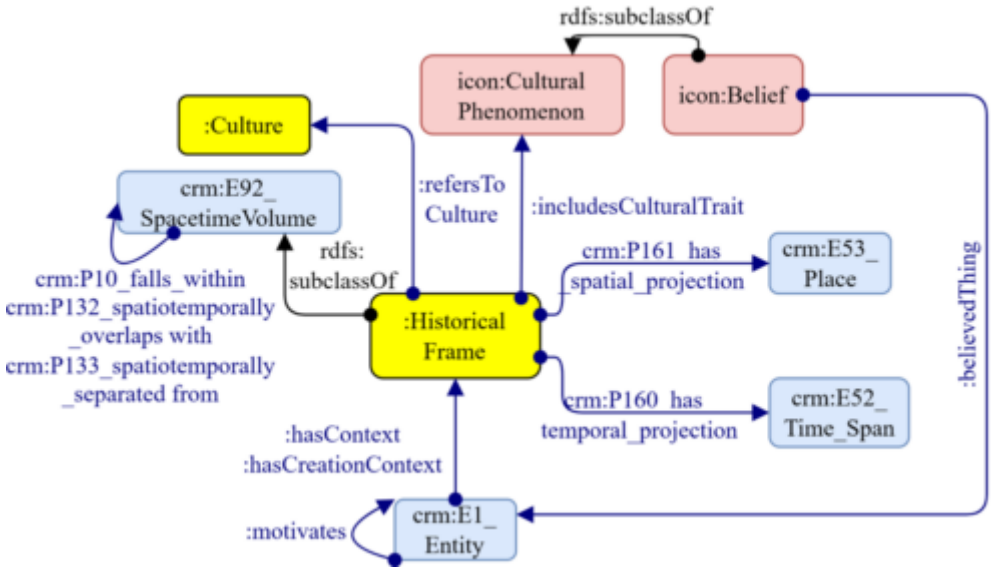


Figure 7. Aligned, final ontological representation of the Historical Frame module. Created by Sofia Baroncini. Available at Zenodo (DOI: 10.5281/zenodo.20393679). Licensed under CC BY 4.0

The process should be iterated on all the scenarios until there are no more scenarios to be modeled. At that point, the ontology is ready to be published.

Publishing and making the ontology FAIR

After the ontology is developed, it is still important to make it FAIR (Findable, Accessible, Interoperable, Reusable). It is good practice to assign it a permanent URI, publish documentation, and include it in the main indexes of LOD vocabularies.

A permanent URI can be assigned through w3id.org, whereas various tools, such as PyLode, can be used to automatically generate the documentation of the ontology from the OWL 2 file.

In the case of CHint, a permanent ID was assigned to the ontology OWL file, the development and documentation, which was created through PyLode.

Finally, it is highly recommended to use tools such as the Ontology Pitfalls Scanner (OOPS!) and FAIR Ontology Pitfalls Scanner (FOOPS!) to detect errors or missing descriptions (e.g., labels, inverse relations) and fix the ontology accordingly before publishing it.

Following these steps, it is possible to create a robust ontology, which can address specific domain aspects and be aligned with standards.

References

Baroncini, S., & van den Heuvel, C. (2026). The Historical Framing Problem: Temporal Modeling of Interactions between Tangible and Intangible Cultural Heritage. *Semantic Web,* 17(4), 1–43. https://doi.org/10.1177/22104968261454485

Peroni, S. (2016). SAMOD: An agile methodology for the development of ontologies. *Proceedings of the 13th OWL: Experiences and Directions Workshop and 5th OWL Reasoner Evaluation Workshop (OWLED-ORE 2016)*. https://doi.org/10.6084/M9.FIGSHARE.3189769

Presutti, V., Daga, E., Gangemi, A., & Blomqvist, E. (2009). eXtreme Design with Content Ontology Design Patterns. *Proceedings of the 2009 International Conference on Ontology Patterns*, 516, 83–97. https://ceur-ws.org/Vol-516/pap21.pdf

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2015). The NeOn Methodology framework: A scenario-based methodology for ontology development. *Applied Ontology*, 10(2), 107–145. https://doi.org/10.3233/AO-150145

Suárez-Figueroa, M., Gomez-Perez, A., & Villazon Terrazas, B. (2009). How to Write and Use the Ontology Requirements Specification Document. On the Move to Meaningful Internet Systems: OTM 2009: Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009, Vilamoura, Portugal, November 1-6, 2009, Proceedings, 2, 966–982. https://doi.org/10.1007/978-3-642-05151-7_16

Featured image: Sketches by Sofia Baroncini and Charles van den Heuvel to grasp the interaction between artworks and immaterial culture over time and space at the beginning of the development of the CHint Ontology. Copyright of the authors.

1. This blog post provides just a general overview of the process to make the methodology more accessible and understandable. For a complete and detailed overview of the modeling process, refer to Peroni (2016). [↩]