Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

CULTURE LETTERARIE E FILOLOGICHE

Ciclo 35

Settore Concorsuale: 01/B1 - INFORMATICA

Settore Scientifico Disciplinare: INF/01 - INFORMATICA

CONNECTING WORKS OF ART WITHIN THE SEMANTIC WEB OF SYMBOLIC MEANINGS

Presentata da: Bruno Sartini

Coordinatore Dottorato

Marco Antonio Bazzocchi

Supervisore

Aldo Gangemi

Co-supervisore

Maria Godefrida Jacoba Van Erp

Esame finale anno 2023

To Léonie and Emma, my symbols of *love*

Abstract

My doctoral research is about the modelling of symbolism in the cultural heritage domain, and on connecting artworks based on their symbolism through knowledge extraction and representation techniques. In particular, I took part in the design of two ontologies: one that models the relationships between a symbol, its symbolic meaning, and the cultural context in which the symbol symbolizes the symbolic meaning; the second models artistic interpretations of a cultural heritage object from an iconographic and iconological (thus also symbolic) perspective. I also converted several sources of unstructured data, a dictionary of symbols and an encyclopaedia of symbolism, and semi-structured data, DBpedia and WordNet, to create HyperReal, the first knowledge graph dedicated to conventional cultural symbolism. By making use of HyperReal's content, I showed how linked open data about cultural symbolism could be utilized to initiate a series of quantitative studies that analyse (i) similarities between cultural contexts based on their symbologies, (ii) broad symbolic associations, (iii) specific case studies of symbolism such as the relationship between symbols, their colours, and the symbolic meanings that they convey. Moreover, I developed a system that can infer symbolic, cultural context-dependent interpretations from artworks according to what they depict, envisioning potential use cases for museum curation. I have then re-engineered the iconographic and iconological statements of Wikidata, a widely used general-domain knowledge base, creating ICONdata: an iconographic and iconological knowledge graph. ICONdata was then enriched with automatic symbolic interpretations. Subsequently, I demonstrated the significance of enhancing artwork information through alignment with linked open data related to symbolism, resulting in the discovery of novel connec-

tions between artworks. Finally, I contributed to the creation of a software application. This application leverages established connections, allowing users to investigate the symbolic expression of a concept across different cultural contexts through the generation of a three-dimensional exhibition of artefacts symbolising the chosen concept.

Abbreviations

Concept	**Abbreviation**
Cultural Heritage	CH
Knowledge Graph(s)	KG(s)
Competency Question(s)	CQ(s)
Iconographical and Iconological (or vice-versa)	*icon*
Resource Description Framework	RDF
Ontology Web Language	OWL
Computer Vision	CV

Contents

List of Figures

List of Tables

Chapter 1

Symbolism and Artistic Interpretations in Cultural Heritage Linked Open Data

1.1 Introduction

In August 2022, the International Council of Museums (ICOM) released a new Museum definition, which updates the previous one of 2007. A museum was formerly defined 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. [2]

The new definition reads as follows:

> A museum is a not-for-profit, permanent institution in the service of society that researches, collects, conserves, **interprets** and exhibits tangible and intangible heritage. Open to the public, accessible and inclusive, museums foster **diversity** and sustainability. They operate and communicate ethically, professionally and with the **participation of**

> **communities**, offering varied experiences for education, enjoyment, reflection, and knowledge sharing. [3]

The notion of interpretation, diversity, participation of communities signals an important shift in how cultural heritage tangible and intangible artefacts are seen, displayed, and understood. Museums thus recognize the importance of all the aspects of cultural artefacts, from physical characteristics to more implicit symbolic meanings that refer to the underlying or hidden meanings conveyed by them. These meanings can be related to the cultural values, beliefs, and practices of the society that produced the artefact, and may not be immediately apparent to someone from a different cultural background. Therefore, people from different backgrounds can interpret the same artefacts in a different, and sometimes opposite way.

The purpose of this thesis is to create a computational system that can interpret artworks from different cultural perspectives, and then highlight connections between artworks based on their symbolic meanings.

Returning to the new ICOM definition, from a digital perspective, the shift mentioned above requires adequate data models that can describe symbolic and artistic aspects of artefacts, which should be granted the same importance as general metadata such as creator, date of creation, physical measurements. A common format to describe, publish, and share cultural heritage (CH) artefacts digitally is Linked Open Data (LOD). One of the main benefits of using LOD to describe cultural heritage artefacts is that it allows for the creation of structured databases of information (also known as Knowledge Graphs)[1] which can be linked together to form an interconnected web of information (the Semantic Web) about these artefacts. This makes it easier to connect the data about these artefacts to other relevant data sources, such as information about the artists, related historical events, or cultural symbolism, providing a more complete and nuanced understanding of

[1] For a conceptual definition of a knowledge graph, I refer to [4]:

> Knowledge Graphs are large semantic nets that integrate various and heterogeneous information sources to represent knowledge about certain domains of discourse.

On a technical level, KGs can be defined as a set of triples (subject, predicate, object) encoded in a serialization of the Resource Description Framework, or RDF [5] the cultural significance of the artefacts, their content, and the context of creation. Additionally, it allows a wide range of stakeholders, such as researchers or interested users, to search for and access information about the artefacts, as well as to understand the relationships between different pieces of information about them. In the context of cultural heritage, LOD can be used to describe a wide range of items, including works of art, historical documents, and archaeological artefacts.

RDF (Resource Description Framework)[2] is a standardised framework for representing and exchanging information of LOD on the Semantic Web.

Ontologies are also an integral part of the Semantic Web. They can be defined as a set of formal, explicit, and well-defined concepts, or terms, and their relationships to each other. Ontologies are expressed in a formal language called OWL (Web Ontology Language), which enables the representation and interpretation of knowledge in a structured and machine-readable form. They provide a means of conceptualizing and structuring information, and are used for a variety of purposes, including natural language processing, knowledge representation, data integration, and artificial intelligence [6]. Ontologies provide a shared vocabulary for a domain, and thus enable machines to better understand and reason about the content.

Given the logical back-end of the ontologies that are used as a schema to describe information within the Semantic Web, reasoning-based inferences can be made on the data to highlight even more connections than the explicit ones declared in knowledge graphs. Currently, the vast majority of these connections emerge from standard metadata, as other aspects of cultural heritage artefacts are often ignored or stored in natural language descriptions that cannot exploit the full potential of the Semantic Web [7].

The idea behind this thesis is that the most interesting connections between works of art lie on a deeper, symbolic level, which is still unexplored and underrepresented in the Semantic Web. This symbolic level comprehends two important aspects: (i) conventional cultural symbolism, meant as the general or cultural specific knowledge of symbols and their meanings that has been studied by scholars

[2]https://www.w3.org/RDF/

and which is generally stored in dictionaries of symbols or symbolic encyclopaedias [8, 9], (ii) symbolic interpretation of artworks, which is the act of associating symbolic meanings to artworks, from both an iconographical perspective, i.e. recognising a symbol in a work of art and therefore associating its symbolic meaning to it, or iconological perspective, such as giving symbolic valence to certain aspects related to the cultural context of the creation of the artwork and its creator [1]. Both aspects are currently underrepresented in Semantic Web, which severely hampers the connections that can emerge from cultural heritage LOD.

The remainder of this chapter is organised as follows. Section 1.2 contains an analysis of knowledge graphs incorporating iconographical and iconological statements, including symbolic representation. Section 1.3 presents the thesis problem statement and associated research questions, as well as the research objectives. Finally, 1.4 summarizes the rest of the thesis.

1.2 Analysis of iconographical and iconological statements of current Knowledge Graphs

For this section, I include a published paper [10] about an assessment on the quality of iconographical and iconological statements in the Semantic Web, in its entirety. This paper was co-written by me and Sofia Baroncini (and our supervisors Aldo Gangemi, Marieke van Erp and Francesca Tomasi), a PhD student in Literal and Philological Culture from the University of Bologna. In particular, I focussed on the Related Work and Evaluation Criteria sections. Both me and Sofia contributed equally to the Introduction, Evaluation of the Results and Conclusion sections. Subsection 1.2.3.1 has been integrated into the paper text in this thesis to give a more comprehensive view of the analysed datasets.

Recent years have witnessed a growing interest in linked open data describing Cultural Heritage [11]. Despite many Cultural Institutions are releasing their data only in a simple tabular form, several Knowledge Graphs (KGs) are addressing the description of artworks in a more structured, logical form. Some of them, e.g.

Wikidata [12], have a general scope and are created collaboratively, while others, (e.g. ArCo [13], Zeri & Lode [14]) are generated by the conversion of authoritative data from cultural institutions.

In this diversified setting, it is important to assess the coverage, accuracy, and reliability of the available data to allow their reuse for domain-specific purposes. While many studies addressed the problem of KG evaluation methods, a survey on Art History information stored in Knowledge Graphs, comprehensive of an assessment of the data quality, is still missing. Therefore, this work aims to evaluate the coverage of the content represented in visual works over existing KGs, with a focus on iconographical and iconological aspects (i.e. artistic subjects and their symbolic and cultural meanings). The phrase "iconographical and iconological" will be referred to as *icon* from now on. We survey KG evaluation methodologies, and adapt some of their metrics to the considered domain of knowledge. Furthermore, theories concerning the *icon* domain are reviewed to assess the extent to which KGs cover information about visual items' subject and content description.

Semantic web technologies offer an opportunity to formally express semantically complex information. For this reason, they are a suitable means to express fields of study as complex as iconography and iconology at the required granularity.

Artworks' content should be analysed both isolated, i.e. by identifying relevant features, and associating them to features of other artworks (e.g. the study of patterns recurring in different subjects [15, 16]). Therefore, the knowledge emerging from an analytic approach is mostly missed when an artwork's content is described just by a general subject term.

The traditional sources of knowledge are natural language descriptions of artworks as found in texts, but texts need knowledge extraction methods to enable further analysis and interlinking, limiting the computational reuse of that knowledge [17].

In addition, since *icon* analysis can potentially involve very different types of cultural objects, often stored by different institutions, the major benefits of storing information about this domain in knowledge graphs include at least:

- The opportunity to answer domain-specific questions through quantitative analysis (e.g.: which attributes and meanings were related to the mythological character of Mercury across the centuries?);

- Accessing and querying interlinked information about worldwide objects that could not otherwise be experienced together (e.g.: all artworks with political implications stored in different museums worldwide);

- Formally expressing the semantic complexity of the topic (e.g.: the levels of meanings of an artwork and its relations to external resources, such as other artworks, texts, etc.).

The main contribution of this work is the assessment of the available data accuracy, reliability, and interoperability in relation to the iconographical and iconological domain of knowledge. Therefore, the major benefit is to provide domain-experts with a clear state of the quality of semantic, domain-specific data available online. Other benefits include improving current data reuse following LOD principles and fostering the creation of a shared semantic description framework for iconology and iconography. With this analysis, we show the reusability potential of the existing KGs based on defined *icon* requirements. Finally, the main findings of this work are shown in a landscape (figure 1.3) in which KGs are positioned according to their performance in the chosen metrics.

From this thesis perspective, this work comprises a first assessment on what can be found in current knowledge graphs and how the information is structured. Key findings of this work will be used to declare the problem statement of section 1.3.

The rest of the section is structured as follows. In subsection 1.2.1 we survey existing methodologies for KGs evaluations, followed by a comparison of theoretical models of artworks interpretation in subsection 1.2.2. subsection 1.2.3 describes the selected graphs, while subsection 1.2.4 illustrates the evaluation method used. Finally, results are presented in subsection 1.2.5 and subsection 1.2.6 contains a reflection on the current state of iconographic and iconological statements in KGs.

The implications of this analysis for the thesis are then described in section 1.3.

1.2.1 Related knowledge graph evaluations

Knowledge graphs differ from traditional relational databases in their structure (graph versus table), the reasoning possibilities that can be applied to them, and facilitated interoperability and interconnections [18]. These differences do require specific methods and metrics to evaluate them. [19] surveys evaluation metrics and methodologies for the tasks of representation learning, knowledge acquisition and completion, with additional analyses over temporal KGs and applications developed from them. [20] provides a series of refinements methods to increase the quality of knowledge graphs. [21] evaluates Cultural Heritage knowledge graphs in terms of their suitability for question answering tasks. [22] proposes a conceptual framework for quantitative and qualitative metrics in the evaluation of knowledge graphs taken from a study of more than 100 scholarly publications. Various general metrics for knowledge graph quality evaluation and applications thereof are provided in [23]. We re-use parts of these metrics, adapting some to focus on the fields of iconography and iconology (see section 1.2.4). [24] presents a similar study, but uses the Goal-question metric paradigm to assess the quality of knowledge graphs. [25] also compares several general domain knowledge graphs in their content coverage. It contains interesting reflections in particular regarding coverage of artistic fields, in which YAGO and DBpedia seem to be the most detailed. [26] uses coverage as well as a metric for evaluation, although this work does not mention cultural heritage related findings. [27] evaluates Wikidata on the basis of schema violations, deprecated entities, looking at its history of updates. [28] also evaluates Wikidata's completeness in the description of data related to cultural heritage. To do so, the information contained in it is compared with the information available on Europeana [29], which is used as a gold standard for completeness. This study does not mention specific aspects related to iconography and iconology. [30] offers a thorough study on the completeness metric when evaluating knowledge graphs. Finally, [31] introduces the concepts of queriability to knowledge graphs, developing a framework for the evaluation of *quality in use*, applying it to DBpedia and YAGO. Queriability is a very intriguing concept when it comes to extracting relatively complex sets of information from knowledge graphs, such complex relationships might be present in knowledge graphs that describe artworks with high granularity. Although to verify the queriability of the *icon* content, a first assessment on what is currently included in a knowledge graph is needed.

In summary, prior work evaluates knowledge graphs suitability for some automatic tasks, or their content, in terms of various metrics that go from completeness to accuracy to quality in use. Some of them focus on specific fields (like cultural heritage). There is no study yet that assesses *icon* aspects in knowledge graphs, which would require a specific evaluation due to the complexity of the information expressed by this domain of knowledge [32] [33]. Therefore, we adapt a selection of the general metrics from the literature to the domain-specific needs, with the addition of a newly created metric. As a result, this section attempts to give a domain-specific overview of the available data quality according to the domain focus of interest and research questions.

1.2.2 Artwork descriptions and interpretations

Nowadays, several approaches for visual images interpretation are available, each considering different aspects [34]. This variety is reflected in interpretation methodologies, which can focus on the objects themselves (formal aspects, content, or materials), on the creator (psychoanalysis) or on the cultural context to which it belongs [35]. Among them, content analysis and understanding are objects of interest in iconography and iconology. Although this field of study was traditionally limited to the interpretation of the artistic subject, the research of Aby Warburg (1889-1929) renewed it [36]. His approach considered the content and forms of the artworks as witnesses of social memory, conducting his analysis in an interdisciplinary way to include religion, culture, and the recurrence of visual patterns through different ages [37, 16]. While iconography can currently be defined as the study of subjects, their attributes and their changes over time, the term *iconology* reflects Warburg's approach, focusing on the socio-cultural interpretation of iconographical and formal variations [32, 38].

Although a methodology for artworks comprehension was considered by Warburg [39], the prevailing theoretical approach in the discipline consists of the subdivision of the artwork's interpretation into 3 or 4 levels, a framework firstly defined by Erwin Panofsky [36]. We refer to [32] for a comparison between the main theories which move from this first formalization attempt. For this study, we adopt Panofsky's theory to evaluate the level of description of artworks in available graphs due to its historical relevance and as it is cited as a reference for subject description by the main cataloguing standards of the field[3]. However, aspects put forward by other art historians will be considered. Here, three layers are identified, namely: pre-iconographical description, iconographical analysis and iconological interpretation. From the first level to the last one, increasing knowledge of conventions, sources and cultural aspects linked to the artwork production are required. When practically applied, the levels constituting the act of interpretation are simultaneous and the interpretation itself is narrowly dependent on subjective intuition [36].

Firstly, objects such as people, actions, emotions, colours, and shapes are recognised (level 1). Then, these objects are interpreted as subjects or iconographies (e.g. Mary) at the second level, which requires the knowledge of the literary sources and visual conventions used in a determined period and context. Then, the reading of iconographies as symptoms of the contemporary society, of the artist's beliefs and personality or as the expression of meanings voluntarily inserted, is the content of the third level.

The levels of this theory are referenced by cataloguing standards for artworks description, such as the Getty's *Categories for the Description of Artworks* (CDWA)[4] and the guide *Cataloguing Cultural Objects* (CCO) [40]. Both of them underline that adopting a simplified description of the approach by Panofsky "can be helpful in indexing subjects for purposes of retrieval"[45]. They define the second

[3] See the *Categories of Description of Works of Art*, available at `https://www.getty.edu/research/publications/electronic_publications/cdwa/18subject.html` and the guide *Cataloguing Cultural Objects* (CCO) [40]

[4] `https://www.getty.edu/research/publications/electronic_publications/cdwa/18subject.html`

[5] For the alignment of the concept of the subject to the main cataloguing standards, we refer to section 16 of *Metadata Standard Crosswalk*, available at `https://www.getty.edu/research/`

and the third level, viz. the identification of themes, narratives, iconographies, and meanings, as the *aboutness* (i.e. what the work is about), whereas the first level corresponds to the *ofness* (viz. what can be seen by a non-expert interpreter [41, pp. 207-208]. If the subject corresponds to the work itself (e.g. the term *architecture* used for describing a cathedral) and does not refer to a subject depicted by the object (e.g. a drawing representing a cathedral), the term *isness* shall be used[4].

Figure 1.1: The Punishment of Tityus, Charcoal drawing, a gift to Tommaso de' Cavalieri, Royal Collection Trust

To illustrate our theory and present an example in which each level of interpretation is covered, we describe Michelangelo's Tityus interpreted in [42]. The drawing (Figure 1.1) shows a laying, naked man whose liver is being devoured by a vulture (level 1, ofness). It represents the story of Tityus (level 2, aboutness), punished by Apollo for having assaulted his mother Leto by chaining him to a rock in Hades while two vultures eternally devour his liver, considered the seat of physical passions (symbol, level 2, aboutness). The story had been commonly interpreted by Michelangelo's contemporaries as an allegory of the tortures caused by immoderate love (allegory, level 2, aboutness). On this basis, Panofsky claims that the artist depicted this story as a symbol of his personal passion for Tommaso Cavalieri

publications/electronic_publications/intrometadata/crosswalks.html (level 3, aboutness), to whom he gifted a corpus of drawings pervaded by Neoplatonic meanings (level 3, aboutness). Table 1.1 shows how this interpretation can be subdivided into levels. For its completeness, this drawing will be considered as an example for artworks' content and meaning evaluation in KG in Section 1.2.4.

Table 1.1: Example of description of an artwork (Tityus, by Michelangelo) interpretation through three levels

Level	Description
1	Nude, laying man, whose liver is devoured by a vulture
2	Tityus; story of Tityius, whose liver is devoured by a vulture; liver as the seat of physical passions; story of Tityus as an allegory of the tortures caused by immoderate love.
3	Agonies of sensual passion, enslaving the soul and debasing it even beneath its normal terrestrial state, according to the Neoplatonic theory; Expression of the agonies of sensual Passion that pervaded Michelangelo after he had met Tommaso Cavalieri, for whom he realized the drawing.

1.2.3 Selection of the knowledge graphs

To collect the most representative RDF data about the description of the artwork, we need to consider which kind of cultural objects can represent a visual subject and can have a cultural meaning. Potentially, every image representing a subject that can be invested with a cultural meaning can be considered by an iconographical-iconological interpretation. To narrow down the research in the art history field, we focus our selection on paintings, sculptures, frescoes, visual subjects on coins (numismatics) and illuminations. Therefore, in this survey, we considered graphs containing data on Cultural Heritage, Museums, Libraries (manuscripts' drawings and decorations) and numismatics. In addition, we included general purpose knowledge graphs likely containing information about artworks such as Wikidata, DBpedia [43] and YAGO [44].

We used the following methodology. We first define our object of interest, namely artworks and information about their subject and meaning. Then, we collect the KGs through i) the analysis of literature concerning a survey or evaluation of CH KGs [45, 21, 46] and ii) direct search on the web, through a manual keyword search on Google Database Index[6] and other main databases search engines.[7][8][9] This led to 56 graphs. These graphs were further pruned according to the criteria of their online availability through a SPARQL endpoint.[10] We considered these criteria fundamental to assessing data that follows the principle of availability and re-usability of the Semantic Web [47], according to its shared standards.[11]

Only 27 out of 56 graphs were active online, 18 of which had a SPARQL endpoint. The KGs for which the SPARQL endpoint was not responsive and the ones having no information about subjects were discarded. Consequently, we obtained 9 graphs. Table 1.2 gives an overview of the number of artworks having a subject in those KGs, distinguishing between URIs[12] and literals.[13][14] This analysis was conducted through SPARQL queries and by consulting the KGs' documentation. The selection process of our analysis highlights how information about cultural heritage is very scarce when considering data that follows Semantic Web principles, as few domain-specific knowledge graphs are available under those conditions. This makes the inclusion of general domain knowledge graphs essential to assess how *icon* aspects are described in the Semantic Web, as the majority of *icon* data in stored in

---

[6] `https://datasetsearch.research.google.com/`

[7] `https://datahub.io/`

[8] `https://triplydb.com/`

[9] `https://www.kaggle.com/`

[10] SPARQL is the query language that is used to retrieve information from RDF data. SPARQL endpoints are online services, linked to specific knowledge graphs, that let users query knowledge graphs through SPARQL queries. For additional information about SPARQL and SPARQL endpoints, we refer to `https://www.w3.org/TR/rdf-sparql-query/`

[11] `https://www.w3.org/standards/semanticweb/query`

[12] A URI (Uniform Resource Identifier), in a Semantic Web context, is the unique identifier for resources. URIs of resources can be semantically linked (using RDF properties) to other resources (and their URIs). For additional information on the concept of URIs, we refer to `https://www.w3.org/TR/webarch/`.

[13] Literals represent basic data types, such as strings, boolean values, integers. They are not assigned a URI, and therefore they can only be referred to as the object of a triple and never as the subject. Literals contain unstructured information (such as natural language descriptions) that might require additional processing before being machine-readable. For additional information about literals, we refer to `https://www.w3.org/TR/rdf11-concepts/`

[14] For an overview of the relations considered to identify subjects and other *icon* information, see section 1.2.4.2

them. From a structural perspective, we would expect the ontological schemas[15] of domain-specific knowledge graphs to describe *icon* information with a higher degree of granularity compared to general ones. This assumption is proved wrong by our results (section 1.2.5), as Wikidata performs better than domain-specific KGs.

One critical aspect we encountered while doing this analysis is the proper identification of what is a work of art. While some graphs use a specific class or property to express it (e.g. *fabio:ArtisticWork* in Zeri & Lode), others do not have a unique way to identify it. In some cases, e.g. Wikidata, many specific classes are used, subclasses of a general "visual work". In others, e.g. SARI's RDS platform, the class "Work" corresponds to many types of cultural objects, specified by a controlled vocabulary. Although this granularity in the artwork description is appreciable, it may generate a few issues when approaching data quantitatively. First, the selection of what is considered an artwork is left to the user, who may be influenced by subjective decisions in this definition. Second, the high number of entities to be included in a SPARQL query can influence the server response.

In the context of this study, we selected which classes could be considered artworks from the analysis of the documentation or from data retrieval. We decided to focus our attention on paintings and sculptures, when available (if the information present in the knowledge graphs made them distinguishable from other artworks), as they are universally considered as artworks with at least a subject. When paintings and sculptures were not available in the studied knowledge graph, we shifted our attention to the most prominent class in the schema that could represent an artwork (as the numismatic items in Nomisma). On the other hand, when the total number of sculptures and paintings was too little for conducting an evaluation, (e.g. in SARI's RDS platform) we included in the analysis broader terms, such as prints,

[15]We consider the ontological schema as the set of ontologies that are used in a knowledge graph as a data model. There exist several general domain schemas, such as Dublin Core `https://www.dublincore.org/specifications/dublin-core/`, Simple Knowledge Organization System (SKOS) `https://www.w3.org/TR/skos-reference/`, or Friend of a Friend (FOAF) `http://xmlns.com/foaf/0.1/`, that are reused in many knowledge graphs. Domain-specific knowledge graph schema might include specifically developed ontologies in their schema, see the ArCo Ontology (`https://w3id.org/arco/ontology/arco`) for the ArCo knowledge graph, or the Nomisma ontology (`https://nomisma.org/ontology`) for the Nomisma knowledge graph

Table 1.2: Overview of the artwork subject presence in the selected graphs*

Short Name	Artwork #	Percentage of artworks having a subject (URI)	Average of subjects (URI) defined per artwork	Percentage of artworks having a subject (literal)	Average of subject (literal) defined per artwork
ArCo	2111726	45.86%	1.01	100	1.22
Fondazione Zeri	20082	99.99%	1.19	0	0
Nomisma	566732	21.16%	1.1	0	0
Wikidata	669857	26.76%	3.37	0	0
SARI	339	72.57%	1.11	0	0
Europeana	13861	9.32%	2.38	33.8	1.64
ND_Hungary	11655	0%	0	55.97	6.04
DBpedia	12250	93.93%	5.697	4.8	1.09
YAGO	29324	12.75%	1.02	0	0

*As of 01/12/2022

illustrations, and graphics. Table 1.3 summarizes classes that define artworks from the selected KGs, along with properties used to link information relevant to iconography and iconology.

1.2.3.1 Description of the selected knowledge graphs

Zeri & Lode [14] is a knowledge graph that was created by converting the Zeri Photo Archive into linked open data. The project involved the reuse of standard ontologies such as CIDOC-crm, and the development of two new ontologies[16] to describe complex objects and mapping the Zeri catalogue data into RDF, resulting in a dataset that demonstrates the complexity of the archive. ArCo [13] is a knowledge graph that describes a diverse range of Italian cultural heritage artefacts, containing items belonging to architectural, ethnographic, artistic domains. Its structure follows the ArCo network of ontologies. Nomisma[17] is a knowledge graph that integrates numismatic data from more than 70 sources. Its structure follows the Nomisma ontology. Wikidata [12] is a heterogenous, collaborative knowledge graph supported by the Wikimedia Foundation. Its structure is based on the Wikidata schema. Europeana [29] is a web portal that aggregates digital cultural heritage collections from various institutions across Europe, using semantic web technologies. The Europeana Data Model (EDM) ontology is used to harmonize the data from the different institutions. RDS Platform is a knowledge graph that contains items from the Swiss Art Research Infrastructure (SARI). It reuses various ontologies (such as GND Ontology [18]) to describe artworks and their content. DBpedia [48] is a heterogeneous knowledge graph generated by converting the structured information of Wikipedia, following the DBpedia ontology. YAGO [44] is a knowledge graph that was created by extracting structured data from various sources, including Wikipedia, WordNet, and GeoNames. It contains over 10 million entities and their relationships, modelled using the YAGO ontology. Finally, the National Digital Data Archive of Hungary [49] is a knowledge graph that describes a diverse range of Hungarian cultural

[16] F Entry Ontology available at `https://essepuntato.it/fentry/current/fentry.html` and OA Entry Ontology available at `https://oaentry-ontology.sourceforge.net/index.html`

[17] https://nomisma.org/

[18] https://d-nb.info/standards/elementset/gnd

heritage items including paintings, photographies, pottery.

Table 1.3: Classes and properties related to the recognition of artworks (sculptures and paintings if available) in selected knowledge graphs

Name (Abbreviation)	**Artwork (paintings and sculptures if possible)**
ArCo	<artwork> a arco:HistoricOrArtisticProperty
Zeri&Lode (Zeri)	<artwork> a fabio:ArtisticWork
Nomisma	<artwork> nmo:hasObverse <something> <artwork> nmo:hasReverse <something>
Wikidata	<artwork> wdt:P31 wd:Q3305213 (Painting) <artwork> wdt:P31 wd:Q860861 (Sculpture)
RDS Platform (SARI)	<artwork> a gndo:Work <artwork> a gndo:formOfWorkAndExpression
Europeana	<artwork> a http://vocab.getty.edu/aat/300033618 <artwork> a http://vocab.getty.edu/aat/300047090
National Digital Data Archive of Hungary (ND_Hungary)	<artwork> a dcmitype:Image
DBpedia	<artwork> a dbo:Artwork
YAGO	<artwork> a schema:Painting <artwork> a schema:Sculpture

1.2.4 Evaluation criteria

Following the approach presented in [50], we define metrics that go beyond accuracy, as we are interested in i) the coverage of the KGs schemas and their data, ii) the references and interlinking with existing taxonomies that identify subjects in art (Iconclass, Getty), iii) alignments, and iv) linking to external knowledge graphs to foster multi-vocality in art interpretations. These general metrics were adapted for the evaluation of the specific domain of knowledge, to obtain a specific quality assessment on domain data. In addition, these metrics acquire a particular relevance for the domain studies, which analyse the relations between cultural objects, their sources and multiple interpretations. Following the theory explained in Section 1.2.2, we are interested in analysing whether the current knowledge graphs distinguish between elements that belong to the first, second, and third level of interpretation. We are therefore looking for clear distinctions when it comes to the description of natural elements depicted in a painting, the recognition of subjects and symbols, and the reflections of the influence of the cultural period in which the artwork was created on the artwork itself and vice versa.

Taking this into consideration, we applied parts of the framework formulated in [23] in the evaluation of the chosen KGs. This study proposes the possibility of a weighting system applied to each metric according to the importance of the task in the context of the evaluation. In our case, we give more weight to the evaluation criteria referring to the elements that were addressed the most in the literature of *icon* studies. Specifically, we assign the maximum weight (1) to those criteria that we consider completely related to iconography and iconology evaluation, 0.8 to those criteria that we consider closely related, and 0.6 to those criteria that we consider partially related. All other criteria are excluded; considering their weight would be 0, they were not computed. Therefore, of all the categories described by [23], we focus only on column completeness, schema completeness, semantic validity, reference to external vocabularies, and interlinking via owl:sameAs.[19] We adapted all metrics cited above to address the specific tasks of evaluation of the *icon* content. As a result of the adaptation, we decided to rename them to address their new specific purpose. Column completeness was changed into Iconographical and Iconological column completeness (IICC), semantic validity became Semantic validity of Iconographical and Iconological triples (SVIIT) schema completeness became Iconographical and Iconological schema granularity (IISG), reference to external vocabularies became references to external taxonomies of art and culture (RETAC) and Interlinking via owl:sameAs became Interlinking of artworks (IA). The differences and specific changes applied to these metrics will be explained in the sub-paragraphs of this section. Finally, we added an entirely new metric to measure intralinking potential for subject comparisons (IPSC).

Table 1.4 summarizes (i) the re-used metrics plus the newly created one, (ii) their adaptation to the *icon* field and (iii) the weight assigned to the metric. We applied these measurements to the knowledge graphs listed in Section 1.2.3. We then

[19]The cited categories will be thoroughly explained in the following part of this section

grouped these metrics in 2 macro-categories, namely i) structure of the knowledge graphs, which includes IISG, IA, RETAC, IPSC, and ii) content of the knowledge graphs, which includes SVIIT and IICC. The results of the analysis and the formulas used to calculate the overall score will be discussed in section 1.2.5.

1.2.4.1 Evaluation Methodology

Of the chosen metrics, three (interlinking of artworks, references to external taxonomies of art and culture, and intralinking potential for subject comparisons) could be processed automatically by analysing the data, one through an analysis of the schemas of the various KGs (iconographical and iconological schema granularity), and two required qualitative evaluations (semantic validity of iconographical and iconological triples and iconographical and iconological column completeness). For all automatic evaluations, a series of SPARQL queries were launched on the analysed graph, some will be listed as examples in the following subsections. For the metrics that required a qualitative evaluation of the content, we extracted random samples of the knowledge graphs[20] and evaluated the graphs manually on those samples through annotations.

Two annotators performed all the annotations. In the annotation process, they could express their inability to evaluate the veracity of some triples if the information contained in the knowledge graph was unreachable (broken links) or too scarce to fully assess its quality. We used Cohen's Kappa (using quadratic weights) [51] to measure the agreement score between the annotators. The triples considered invalid by annotators were mutually excluded when computing these agreement metrics.[21] Given the general agreements of the two annotators for all the different samples annotated, as shown in table 1.5, we decided to average the evaluation scores of the two annotators for both the qualitative categories.

In the following paragraphs, the metrics and our computations to obtain them are described in natural language and their mathematical formulas.

[20] In the case of general-scope knowledge graph, the random sample was extracted from the subgraphs describing sculptures and paintings.

[21] Only 3,3% of total evaluated triples was considered invalid

Table 1.4: Evaluation metrics, the first five criteria are adapted from Färber et al. (2018), the last criterium is newly developed.

Area	Criterium	Adaptation	Weight [0-1]
Content	Semantic Validity	Semantic Validity of Iconographicaland Iconological Triples (SVIIT)	1
Content	Column Completeness	Iconographical and Iconological Col-umn Completeness (IICC)	1
Structure	Schema Completeness	Iconographical and IconologicalSchema Granularity (IISG)	1
Structure	Using External Vocabulary	References to External Taxonomiesof Art and Culture (RETAC)	0.8
Structure	Interlinking via owl:sameAs	Interlinking of Artworks (via variousproperties) (IA)	0.6
Structure	Intralinking Potential for Subject Comparisons (IPSC)		0.6

Table 1.5: Inter-annotator agreement scores as measured by quadratically weighted Cohen's Kappa for semantic validity of iconographical and iconological triples and column completeness per knowledge graph

Knowledge Graph	Semantic Validity	Column Completeness
Yago	1.00	0.65
Nd hungary	0.82	0.62
ArCo	0.77	0.77
Zeri	0.66	0.78
Nomisma	1.00	1.00
Sari	0.78	0.68
Europeana	0.82	0.79
DBpedia	0.89	0.66
Wikidata	0.76	0.90

1.2.4.2 Iconographical and Iconological Schema Granularity

This metric is a re-elaboration of the "Schema completeness" metric in [23]. Schema granularity aims to verify the extent to which the ontologies and vocabularies, and corresponding classes and properties instantiated in the knowledge graphs, cover the domain of interest. In this work, we verify to what extent the schema of the knowledge graph is suited for the complete description of *icon* elements. According to the comparison of theories of art interpretation discussed in section 1.2.2, we formulated the following competency questions [52]:

- What are the pre-iconographical elements that appear in a work of art?
- Which actions are depicted in a work of art?
- What are the subjects of a work of art?
- What are the represented symbols in a work of art?
- What are the represented stories in a work of art?
- What are the represented allegories in a work of art?
- What are the intrinsic meanings associated with a work of art?
- Which cultural phenomena are reflected in a work of art?

- What are the corresponding external taxonomies for the identified iconographical terms?

We then created a gold standard interpretation on the example from Michelangelo's work, able to answer those competency questions, as shown in Figure 1.2. We first aligned the properties used in each KG to our example, and computed schema granularity as the division between the number of properties of the example that have been aligned, and the total number of properties in the example. Given N as the number of properties of the gold standard and N_{akg} as the number of properties of the same gold standard aligned to the properties of the schema of the knowledge graph, we measure the Iconographical and Iconological Schema Granularity (*IISG*) of a knowledge graph as

$$IISG(kg) = \frac{N_{akg}}{N}$$

Table 1.6 shows those properties that were recognized as expressing *icon* content, and were aligned to the gold standard.

We weigh this metric as 1 because a schema that permits to express *icon* statements, respecting the required granularity given by the complexity of their field, is essential to correctly and completely store information on this matter.

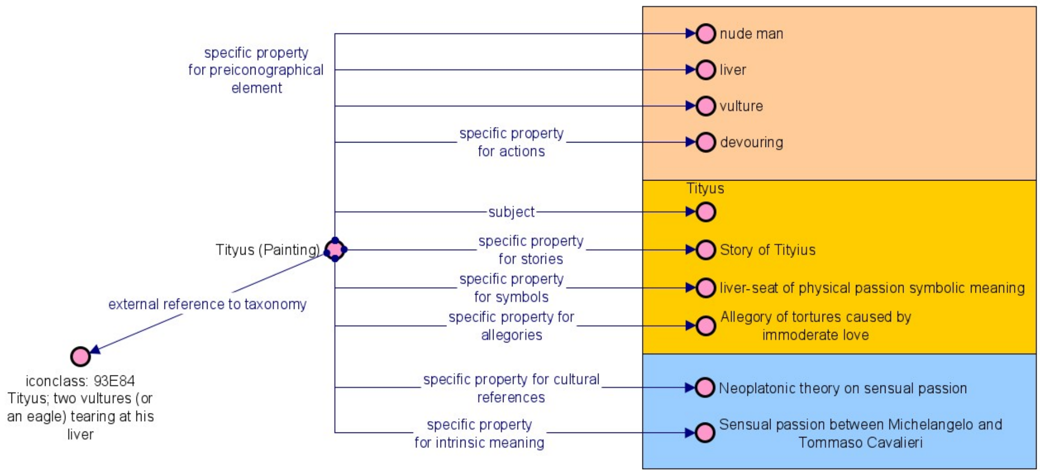

Figure 1.2: The gold standard schema created by applying CQs to the gold example from the literature.

Table 1.6: Properties identifying iconographical and iconological content for each selected knowledge graph

Name	Iconographic and Iconologic Properties
ARCO	arco-cd:hasSubject
	arco-dd:hasIconographicOrDecorativeApparatus
	arco-cd:iconclassCode
	arco-cd:subject
	dc:subject
Zeri	fabio:hasSubjectTerm
Nomisma	nmo:hasPortrait
	nmo:hasIconography
	nmo:hasControlMark
Wikidata	wdt:P180 (depicts)
	wdt:P921 (main subject)
	wdt:P1257 (depicts iconclass notation)
	wdt:P4878 (symbolizes) (qualifier of wdt:P180) wdt:P6022 (expression, gesture or body pose) (qualifier of wdt:P180)
SARI	gndo:topic
	gndo:gndSubjectCategory
Europeana	dc:subject
ND Hungary	dc:subject
DBpedia	dc:subject
	dbp:subject
	dbp:symbol
	dbp:symbols
YAGO	schema:about

1.2.4.3 Semantic Validity of Iconographical and Iconological Triples

This metric was modified from the "Semantic Validity" of [23], in which its purpose is to define whether all the statements of triples in knowledge graphs hold true or not. In our study, we consider the semantic validity of *icon* triples only: we evaluate whether triples that refer to a subject, depicted element, or symbol associated with a painting hold true. To evaluate this, we take a subset of the *icon* statements in each KG. Those statements link the artwork to one of the elements relative to the three layers of interpretation explained in Section 1.2.2, agnostic to the property used. We compute this metric by taking a random sample of 100 iconographical/iconological triples from each knowledge graph, qualitatively evaluating whether the triple is correct (1), partially correct (0.5), or wrong (0). Given S_{ictkg} as the random set of iconographical triples extracted from a knowledge graph, and $S_{evictkg}$ as the evaluation scores set given for each triple $\{sc_1, sc_2 ... sc_x\}$, and x as the sample size[22] to be extracted from the knowledge graph, the Semantic Validity of Iconographical and Iconological Triples ($SVIIT$) is measured as follows

$$SVIIT(kg) = \frac{\sum\limits_{i \in S_{evictkg}} i}{x}$$

This metric offers key insights on the quality of the *icon* content of knowledge graphs, so we give it a weight of 1.

1.2.4.4 Iconographical and Iconological Column Completeness

This metric, in [23], considers the general column completeness of knowledge graphs. In our work, we focus only on the column completeness of *icon* statements. Considering the potentiality expressed in a knowledge graph through the iconographical and iconological schema granularity, we evaluate the column completeness as the schema in use. We extract sub-graphs from the analysed KGs that contain all the *icon* triples associated with 100 randomly selected artworks per KG. This evaluation considers two aspects:

1. the expected number of levels of meaning of an artwork. Generally, a landscape only contains elements belonging to the first level, a portrait contains the first level and then the identification of the subject (second level), and more complex artworks that represent cultural and religious themes can also be analysed at a third, iconological level. Despite the potential for every visual image to have a deeper level of interpretation [38], we decided to expect a third layer only in artworks presenting an explicit cultural subject. This is meant to not affect the artworks' evaluation with the bias of over-interpretation, criticized by some scholars [53]

2. the number of levels covered by the current description in the knowledge

[22] In our case set as 100

graph.

We then divide the covered layers by the expected layers for each artwork in the subset. Having a maximum of three layers, the possible scores for each artwork can be 0 (0 covered layers out of 3 expected, 0/2, 0/1), 0.33 (1/3), 0.5 (1/2), 0.66 (2/3), 1 (1/1, 2/2, 3/3). We do not expect artworks to be described meticulously by indicating every single element of level 1, every single recognizable subject, allegory, symbol of level 2 and every single intrinsic meaning and culturally related meaning of level 3[23]; for this evaluation, having at least one element for every expected level was considered enough. Given A as the set of the randomly sampled artworks in the knowledge graph of size x[24] $\{a_1...a_x\}$, EL as the array of expected layers (a number from one to three) for each artwork

$$EL = \begin{bmatrix} el_1 & el_2 & el_x \end{bmatrix}$$

in A, and CL as the array of covered layers for each artwork

$$CL = \begin{bmatrix} cl_1 & cl_2 & cl_x \end{bmatrix}$$

we create the array SL that contains the divisions between covered and expected layers

$$SL = \begin{bmatrix} \frac{cl_1}{el_1} & \frac{cl_2}{el_2} & \frac{cl_x}{el_3} \end{bmatrix}$$

and then we measure the Iconographical and Iconological Column Completeness (IICC) of a knowledge graph as follows

$$IICC(kg) = \frac{\sum\limits_{i \in SL} i}{x}$$

We consider this metric equally important as having a schema that permits a certain degree of granularity in artwork descriptions; therefore we give full weighting to

[23]Especially considering that in the field of iconography and iconology, there could be potentially endless different interpretations of a painting. It is not possible to list them all

[24]In our case set as 100

this metric (1).

1.2.4.5 Interlinking of artworks

We adapted the metric "Interlinking via owl:sameAs" described by [23] to only apply to artworks. "Interlinking" is considered as the connection between entities belonging to different knowledge graphs. Although less central than the other used metrics (weight = 0.6), we decided to include it because aligning artworks across different knowledge graphs fosters multi-vocality in art interpretation, especially if these knowledge graphs have been manually curated.[25] We measure this metric by dividing the number of artworks in a knowledge graph that are connected to their corresponding versions in external knowledge graphs by the total number of artworks present in a knowledge graph. The main property used to align artwork across different KGs is `owl:sameAs`, but we also looked at other possible alignments from the analysed KGs.[26]

Given KG as the set of triples $\{t_1...t_n\}$ in a knowledge graph (a triple being a sequence of subject, predicate, object $\{s_i, p_j, o_k\}$), A as the set of artworks $\{a_1...a_m\}$ denoted by s_i or o_k, and R_a as the set of relationships $\{r_1...r_z\}$ that are used to align an artwork in a knowledge graph to the same artwork in other knowledge graphs, we consider $A_a = \{a_1...a_w\}$ as a subset of A if

$$\forall a_i \in A_a : a_i \in A \land (\exists p_j \exists o_k : (a_i, p_j, o_k) \in KG \land p_j \in R_a)$$

and we measure Interlinking of Artworks (IA) as

$$IA(kg) = \frac{n(A_a)}{n(A)}$$

Two example queries launched on DBpedia to count the number of artworks

[25]We acknowledge that multi-vocality can be achieved also by giving iconographical and iconological assertions a provenance (even in the same knowledge graph), although for this work we only focus on statements agnostic to the provenance of the interpretation, which would require another specific study

[26]The link to external artworks is expressed 1) in Europeana through the relations `dc:relation` or `edm:relatedTo`, 2) in Wikidata through different `wikibase:identifier`, 3) in ARCO and Zeri&Lode through `rdfs:seeAlso`, beyond `owl:sameAs`

and the number of artworks aligned to different KGs can be seen in listing 1.1 and 1.2 respectively.

Listing 1.1: SPARQL query launched on DBpedia to count the number of artworks

```
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT (COUNT(DISTINCT ?artwork) as ?tot)
WHERE { ?artwork a dbo:Artwork }
```

Listing 1.2: SPARQL query launched on DBpedia to count the number of artworks aligned to external knowledge graphs

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT (COUNT(DISTINCT ?artwork) as ?tot)
WHERE { ?artwork a dbo:Artwork; owl:sameAs ?x }
```

1.2.4.6 References to external taxonomies of Art and Culture

This metric is a re-elaboration of the "Using external vocabulary" metric of [23]. In our work, we focus on the use of vocabulary that belongs to taxonomies of art and culture, which play an important role in artwork descriptions as they provide permanent URIs for specific subjects, scenes, and other *icon* elements represented in artworks. Moreover, they are curated by domain experts, and referring to them gives more authoritativeness to the interpretations. For this analysis, we selected four core taxonomies: Iconclass[27], the Getty Art & Architecture thesaurus[28], the Getty Iconography Authority vocabulary[29], and the Getty Cultural Object Name Authority vocabulary.[30] We measure the references to external taxonomies of art and culture by dividing the number of artworks in a knowledge graph that are associated with at least one of them by the total number of artworks present. Given A as the set of artworks in and KG as the set of triples $\{t_1, t_n\}$ in a knowledge graph (a

[27] http://www.iconclass.org/help/outline

[28] https://www.getty.edu/research/tools/vocabularies/

[29] https://www.getty.edu/research/tools/vocabularies/guidelines/cona_3_6_3_subject_authority.html

[30] https://www.getty.edu/research/tools/vocabularies/cona/

triple being a sequence of subject, predicate, object $\{s_i, p_j, o_k\}$) and T as the set of nodes in a knowledge graph representing a particular subject expressed using a taxonomy of art and culture, we consider an artwork part of the sub set A_t that contains artworks with a taxonomy reference if

$$\forall a_i \in A_t : a_i \in A \wedge (\exists p_j \exists o_k : (a_i, p_j, o_k) \in KG \wedge o_k \in T)$$

and we measure the References to external taxonomies of Art and Culture (RETAC) of a knowledge graph as

$$RETAC(kg) = \frac{n(A_t)}{n(A)}$$

The list of taxonomies of art and cultures used for this analysis contains only those that are referenced at least in one of the analysed knowledge graphs. Increasing the number of taxonomies referenced would not change the methodology of evaluation (and its formula). We welcome potential changes to this list to address *icon* aspects of more specific artworks, such as the reference to the Chinese Iconography Thesaurus[31] for a potential analysis on Chinese *icon* statements in the Semantic Web. References to external taxonomies are strictly related to iconography and iconology, but are not essential to give a complete artwork description. For this reason, we weigh this metric 0.8.

The query shown in listing 1.3 was used to count all the artworks in ArCo referring to a taxonomy of art and culture (Iconclass).

Listing 1.3: SPARQL query launched on ArCo to count the artworks that have a reference to a taxonomy of art and culture (iconclass)

```
PREFIX arco-cd:
<https://w3id.org/arco/ontology/context-description/>
PREFIX arco: <https://w3id.org/arco/ontology/arco/>
SELECT (COUNT (DISTINCT ?s) as ?tot)
WHERE {
?s a arco:HistoricOrArtisticProperty ;
arco-cd:iconclassCode ?res}
```

[31] https://chineseiconography.org/

1.2.4.7 Intralinking potential for subject comparisons

We introduce this metric to highlight the importance of intralinking subjects in the same knowledge graph. We consider "intralinking" as the connection between entities belonging to the same knowledge graph. Having a URI as a subject of an artwork allows grouping artworks per subject and compare them in respect to having a subject as a literal. Moreover, the same subject can then be aligned to other subjects in different knowledge graphs, to foster interlinking in the digital art history LOD field. We measure Intralinking potential for subject comparison by dividing the number of subjects that are linked to more than one artwork by the number of total subjects. Given S as the artistic subjects (expressed as URIs) in a knowledge graph and S_2 as the artistic subjects that are linked to more than one artwork, we measure the Intralinking potential for subject comparison (IPSC) of a knowledge graph as

$$IPSC(kg) = \frac{n(S_2)}{n(S)}$$

As this aspect is relevant but not fundamental for iconographical content representation, we weight it 0.6. Two example queries that count the number of subjects (URIs) in Europeana, and the number of subjects that are linked to more than one artwork can be seen respectively in listing 1.4 and 1.5.

Listing 1.4: SPARQL query launched on Europeana to count all the subjects that are URIs

```
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX ore: <http://www.openarchives.org/ore/terms/>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX aat: <http://vocab.getty.edu/aat/>

SELECT (count(distinct ?sub) as ?tot)
WHERE {?s a skos:Concept ; skos:broader* aat:300033618 .
       ?s2 a skos:Concept ; skos:broader* aat:300047090 .
    {?CHO dc:type ?s ; dc:subject ?sub .
```

```
    FILTER (isURI(?sub))}
   UNION
   {?CHO dc:type ?s2 ; dc:subject ?sub .
    FILTER(isURI(?sub))}
}
```

Listing 1.5: SPARQL query launched on Europeana to count all the subjects that are linked to more than one artwork

```
PREFIX ore: <http://www.openarchives.org/ore/terms/>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX aat: <http://vocab.getty.edu/aat/>
SELECT (COUNT(?sub) as ?tot)
WHERE { FILTER (?tot > 1)
{SELECT ?sub (COUNT(DISTINCT ?CHO) as ?tot)
 WHERE {
  ?s a skos:Concept ; skos:broader* aat:300033618 .
  ?s2 a skos:Concept  ; skos:broader* aat:300047090
  {?CHO dc:type ?s; dc:subject ?sub }
  UNION
  {?CHO dc:type ?s2; dc:subject ?sub }
  }
GROUP BY ?sub }
}
```

1.2.5 Results and discussion

Results obtained from the application of the metrics over the KGs are summarized in table 1.7 and visualized in Figure 1.3. To give a better overview of the results of the metric evaluation, they were then used to place the knowledge graphs inside a two-dimensional landscape. The landscape coordinates are determined by the two macro-aspects, namely content and structure, described in section 1.2.4. We averaged the metrics relative to these two macro-categories to obtain a score for content and structure. These averages are computed taking into consideration the weights of each metric. Given M_s and M_c as the sets of scores of a knowledge graph relative to its structure and content, respectively {IISG, IA, RETAC, IPSC} and {SVIIT, IICC}, and given WM_s and WM_c as the sets of weights given to M_s and M_c, respectively $\{w_{iisg}, w_{ia}, w_{retac}, w_{ipsc}\}$ and $\{w_{sviit}, w_{iicc}\}$, we computed the structure score (SS) of a knowledge graph as follows

$$SS(kg) = \frac{IISG \cdot w_{iisg} + IA \cdot w_{ia} + RETAC \cdot w_{retac} + IPSC \cdot w_{ipsc}}{\sum_{i \in WM_s} i}$$

and the content score (CS) of a knowledge graph as follows

$$CS(kg) = \frac{SVIIT \cdot w_{sviit} + IICC \cdot w_{iicc}}{\sum_{i \in WM_c} i}$$

We divided the graphs in four categories, that represent the four quadrants of the landscape, according to their averaged scores, namely: high in content and in structure (both scores ≥ 0.5), low in content and high in structure (content < 0.5 and structure ≥ 0.5), high in content and low in structure (content ≥ 0.5 and structure < 0.5), low in content and in structure (both scores < 0.5).

Figure 1.3 shows a clear scenario: the content of data is generally correct, but not thoroughly described. In fact, none of the graphs has acceptable results in the structure quadrants, and most of them (7 out of 9) present high scores in content. Nevertheless, this result is given by higher rates in semantic validity (six KGs score more than 0.8) rather than in column completeness (only 3 KGs score more than 0.7). Among them, despite being a general-purpose graph, Wikidata performs the best results. A deciding factor in its performance is that it has the best schema granularity, as several properties can be aligned to the prototype schema of Figure 1.2. In addition, its column completeness scores are higher than some Art History graphs. This is because, in contrast with the approach adopted in the other graphs, the first level of interpretation is often described even when a second or third-level subject is identified (i.e., if “Mary” is included as the depicted entity of the second level, "woman", which refers to the pre-iconographical recognition of Mary, is also included).

The granularity in the levels' description may have an influence on the intralinking metric, since the description of simpler and more generalizable elements of the first level of description can positively affect the capability of comparing artworks that share them. This assumption is evidenced by the fact that graphs such as SARI's platform[32], where the subjects considered are broad concepts (e.g. "persons related to art"), perform better results in intralinking. Although, it is important to underline that the general purpose of the graph and the restricted number of subjects described can affect this evaluation. For example, Nomisma,[33] having as subjects only deities, personifications or Roman emperors, performed the maximum score in this metric.

Other relevant qualitative observations can be made over the results obtained. Firstly, we envision that Art History KGs such as Zeri&Lode, which precisely identifies second-level subjects with an acceptable percentage of interlinking to vocabularies, could foster subject retrieval and semantic computational capabilities by adding information on more levels of interpretation. Additionally, ArCo, created by automatic conversion of cultural heritage catalogues, despite having a high result in column completeness, has low rates in subjects intralinking (0.172) and in relation to external taxonomies (0.123). This may be due to the highly automatic process through which the knowledge graph was created [13]. The automatic creation of URIs for subjects from strings extracted from catalogue data could be improved to avoid duplicates of URIs referring to the same entities, therefore increasing the intralinking potential of the KG. For what concerns references to external taxonomies, Europeana shows the best results. In fact, it is possible to retrieve different types of artworks according to the Getty vocabulary category, allowing feasible reusability and retrieval of information for people knowledgeable about them. Moreover, by defining artwork types in this way, it is also possible to retrieve information without having to know specific classes for types of artworks, shifting from the necessity

[32] https://rds.swissartresearch.net/resource/rdsPages:Start
[33] http://nomisma.org/

to know the specific schema of the knowledge graphs, to the knowledge of general taxonomies applicable to different linked open data datasets. It is interesting to note that, despite having a perfect score in references to taxonomies of art and culture, Europeana does not have any specific property that links an artwork to a taxonomy (it uses dc:subject) which impacted negatively the score obtained in the schema granularity metric. Finally, the National Data Archive of Hungary [49] scores worst in the general categories, given the absence of subjects expressed as URIs, the only use of `dc:subject` to describe *icon* statements, and the complete absence of references to taxonomies.

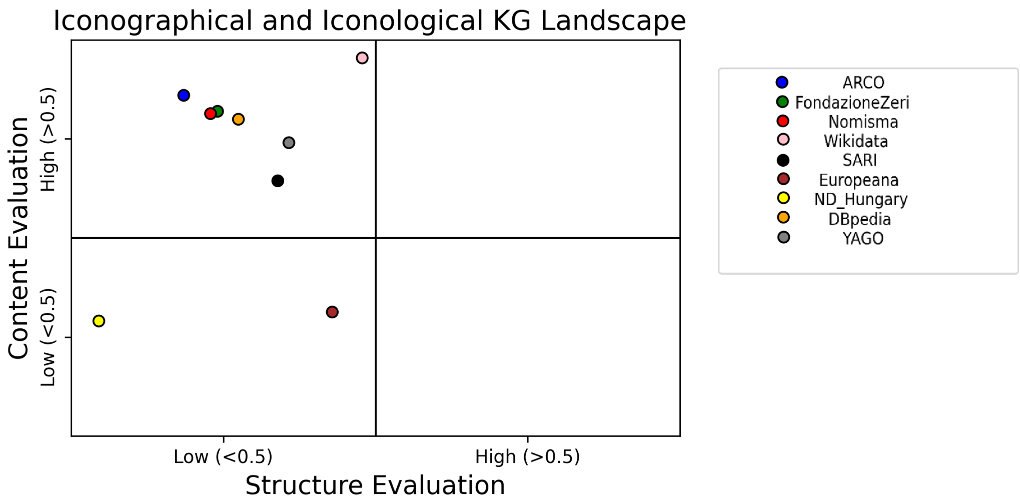

Figure 1.3: The landscape of the knowledge graphs on the quality of their iconographical and iconological statements (content) and the structure of the schemas that describe them (structure).

1.2.6 Conclusions and Future Work

To exploit the capabilities of interlinking, inference and analysis of the semantic technologies applied to *icon* study of artworks, reliable, complete and well-structured data are required. We assess the data quality of current CH KGs that are openly available, providing access to a SPARQL endpoint, and having data on artwork subject descriptions. Our results strongly suggest that only a few KGs mention the artwork's iconography and iconology (Section 1.2.3), although the *icon* descriptions are not sufficiently granular. To assess KGs according to different aspects, we adapt five metrics from prior KG evaluation methodologies (Section 1.2.4), and add

Table 1.7: Results for each metric over the selected Knowledge Graphs*

Short Name	SVIIT (weight 1)	IICC (weight 1)	IISG (weight 1)	IA (weight 0.6)	IPSC (weight 0.6)	RETAC (weight 0.8)
ArCo	0.8278	0.74	0.3333	0.0026	0.172	0.1238
Fondazione Zeri	0.9925	0.5117	0.1111	0.0005	0.266	0.5449
Nomisma	0.995	0.5	0.2222	0	0.749	0.0001
Wikidata	0.9768	0.74	0.6667	0.699	0.367	0.157
SARI	0.849	0.3783	0.1111	0.997	0.5	0
Europeana	0.4688	0.236	0.1111	0.0073	0.6122	1
ND_Hungary	0.13	0.5392	0.1111	0	0	0
DBpedia	0.655	0.7242	0.2222	0.994	0.41	0
Yago	0.99	0.4825	0.1111	1	0.1675	0

*As of 01/12/2022

a new metric. This set of metrics is used to evaluate the content and the structure of sub-graphs describing artworks' *icon* characteristics. We observe that all KGs poorly perform in the *schema structure* as resulting from a combination of metrics, but the major part of them have high or acceptable scores for the *content evaluation* combined metric (Section 1.2.5).

This work gives a critical overview of the complexity involved in the correct and exhaustive creation of domain-specific data. Since the artwork *icon* descriptions are generally correct, the current data can be reliable for data reuse and analysis. Nevertheless, to enhance all the expressivity that may lay in them, a deeper accurate description and a better schema is required. Whereas *icon* descriptions exist, they are not sufficiently interlinked, searchable, and exhaustively described.

1.3 Problem Statement and Research Questions

At the current state, there are several gaps that need to be filled in the Semantic Web and Knowledge Representation domain when it comes to iconographic, iconological and symbolic data.

First, as it will be mentioned in chapter 2, there is a lack of a model that can describe conventional cultural symbolism considering simple and complex symbolic relationships, despite the availability of several sources of unstructured data on symbolism [54, 8, 9, 55], which could be used as a guide for a data-driven modelling of it. At the same time, philosophical theories to describe symbolism exist and could give such models correct terminology and basic structure [56, 57]. As a consequence of this first gap, (i) there is the lack of structured symbolic data, which does not allow quantitative research on symbolism with a certain degree of explainability and provenance, and (ii) it is not possible to infer that an artefact that contains or depicts certain elements can symbolise the symbolic meanings of those elements.

Moreover, artistic interpretations in general are severely lacking on the Semantic Web. As the previous analysis showed, current knowledge graphs do not describe artistic interpretations (including the recognition of symbols and their symbolic meanings) with the proper granularity. Consequently, this lack of structured data about interpretations of artworks denies possible knowledge-driven digital art history studies.

Although some models to describe interpretations exist, presented in chapter 3, they do not consider all the aspects that revolve around an interpretation, as they do not follow the conceptualisation of art interpretation that have been described by art historians throughout the years [58, 59, 38, 16].

As a result, current connections between cultural heritage artefacts within the Semantic Web are mostly made on standard metadata or poorly described artistic interpretations.

With the purpose of filling up the above-mentioned gaps, this thesis aims to answer the following research questions:

- RQ1 To what extent can symbolic relationships in the cultural heritage domain be encoded into computationally ready, structured, semantically linked data?

  - RQ1.1 Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism and symbolic interpretation in the cultural heritage domain?

  - RQ1.2 To what extent and through which means can unstructured data of symbolism be re-engineered into linked data?

- RQ2 To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?

  - RQ2.1 To what extent can it be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?

  - RQ2.2 To what extent can the interconnections of existing knowledge graphs be improved by including such knowledge?

On a more concrete level, to address the mentioned research questions, I formulated the following research objectives:

RO1 Development of an ontology that can describe conventional cultural symbolism.
This objective purpose will cover the current lack of ontologies in the Semantic Web that conceptualise symbolism. The development of the ontology will be essential to answer RQ1 and RQ1.1.

R02 Creation of a knowledge graph that contains instances of conventional cultural symbolism.
This objective will fill up the gap in Semantic Web of linked data that describe conventional symbolism. The process(es) of conversion of the unstructured data that is contained in the sources of the knowledge graph will be used to answer RQ1.2.

RO2.1 Prove the usefulness of such data by performing significant quantitative analysis on it.
By achieving this objective, it will be possible to answer RQ2.

RO3 Development of an ontology, starting from existing models, that can describe artistic interpretations with an adequate degree of granularity defined by the work of renowned art historians.
This objective is complementary to RO1 in its purpose of answering RQ1.1. It will cover the hermeneutic act of interpretation of a work of art and its symbolism. As initially explained in section 1.2, and analysed in chapter 3, several existing models can be improved to reach an adequate granularity in the conceptualisation of artistic interpretations, which allows for the reuse and specialization of some of their existing classes and properties during the development of this ontology.

RO4 Re-engineer iconographic and iconological statements of current knowledge graphs, describing them with the proper granularity using the newly created ontology. Perform inference-based symbolic interpretations on this data to highlight serendipitous connections between artworks based on symbolism. By achieving this objective, it will be possible to answer RQ2.1 and RQ2.2. RQ2.1 will be answered by studying whether the connections created by symbolic meanings can lead to new discoveries, RQ2.2 will be answered by a quantitative evaluation on the growth of a knowledge graph once these new connections are added and information is described with the proper granularity.

1.4 Thesis Summary

This chapter presented the general area of interest of this thesis: symbolism and artistic interpretations on the Semantic Web. Starting from the new ICOM definition of a museum that for the first time mentions the concept of interpretation, it firstly discusses the interest of the cultural heritage domain for this important concept. Then, it highlighted the current gaps of Semantic Web regarding (i) models that describe conventional symbolism and artistic interpretations, (ii) structured data about symbolism, and (iii) structured data about artistic interpretations considering iconographical and iconological recognitions of symbolism. All three gaps need to be filled to achieve the aim of the thesis: connecting artworks through their symbolism within the Semantic Web, effectively creating the "Semantic Web of Symbolic Meanings".

During these three years, some work of this thesis has already been presented under the following peer-reviewed publications:

- Bruno Sartini and Aldo Gangemi. *Towards the unchaining of symbolism from knowledge graphs: how symbolic relationships can link cultures*. In Book of extended abstracts of the 10th national AIUCD conference. AIUCD, Pisa, pages 576–580, 2021.

- Bruno Sartini, Marieke van Erp, and Aldo Gangemi. *Marriage is a peach and a chalice: Modelling cultural symbolism on the Semantic Web*. In Proceedings of the 11th on Knowledge Capture Conference, K-CAP '21, page

201–208, New York, NY, USA, 2021. Association for Computing Machinery, pages 201-208. https://doi.org/10.1145/3460210.3493552

- Bruno Sartini, Valentin Vogelmann, Marieke van Erp, and Aldo Gangemi. 2022. *Comparing Symbolism Across Asian Cultural Contexts Using Graph Similarity Measures*. In Digital Humanities 2022 Conference Abstracts. Tokyo: The University of Tokyo / DH2022 Local Organizing Committee, pages 358-361. https://dh2022.dhii.asia/abstracts/567.

- Sofia Baroncini, Bruno Sartini, Marieke van Erp, Francesca Tomasi, Aldo Gangemi. 2023. *Is dc:subject enough? A landscape on iconography and iconology statements of knowledge graphs in the semantic web*, Journal of Documentation, Vol. 79 No. 7, pages 115-136. https://doi.org/10.1108/JD-09-2022-0207

- Bruno Sartini, Sofia Baroncini, Marieke van Erp, Aldo Gangemi, Francesca Tomasi. 2023. *ICON: an Ontology for comprehensive Artistic Interpretations*, Journal of Computing and Cultural Heritage. Just Accepted. http://dx.doi.org/10.1145/3594724.

- Bruno Sartini, Andrei Nesterov, Claudia Libbi, Ryan Brate, Sarah Binta Alam Shoilee, Savvina Daniil. 2023. *Multivocal Exhibition: a user-centric application to explore symbolic interpretations of artefacts from different cultural perspectives*. In proceedings of ExICE - Extended Intelligence for Cultural Engagement Conference. https://spice-h2020.eu/conference/papers/posters/ExICE23_paper_7072.pdf

Several more papers are currently under review in different venues, and drafts of them are present in the chapters of this thesis.

The rest of this thesis is divided as follows. Chapter 2 describes the modelling of cultural (conventional) symbolism on the Semantic Web, by presenting the development of the Simulation Ontology and the creation of the knowledge graph HyperReal. Chapter 3 follows with the modelling of artistic interpretations on the

Semantic Web by presenting the development of the ICON ontology. Chapter 4 describes a series of analyses that were carried out, and systems that were developed, during the PhD, utilising the newly developed ontologies and knowledge graph. Quantitative symbolic and artistic research using Semantic Web technologies is described. Particular attention is given to the new connections that emerged between works of art once considering their symbolic and artistic characteristics. In summary, this is the chapter that shows the potential of the newly created Semantic Web of Symbolic Meanings. Chapter 5 concludes the thesis with (i) a final discussion on the impact of the work, (ii) a summary of the key findings and key resources developed during the PhD, (iii) a brief discussion of limitations and envisioned plans for future work.

Chapter 2

Modelling Cultural Symbolism on the Semantic Web

This chapter contains excerpts from the paper *Marriage is a peach and a Chalice: Modelling Cultural Symbolism on the Semantic Web* [33] authored by me and my supervisors Aldo Gangemi and Marieke van Erp. I was involved in all the stages of this work, from the initial idea to the writing of all the sections of the paper. This work aims to answer RQ1 *To what extent can symbolic relationships in the cultural heritage domain be encoded into computationally ready, structured, semantically linked data?*, part of RQ1.1[1] *Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism and symbolic interpretation in the cultural heritage domain?*, RQ1.2 *To what extent and through which means can unstructured data of symbolism be re-engineered into linked data?*. Additionally, it satisfies RO1 *Development of an ontology that can describe conventional cultural symbolism* and RO2 *Creation of a knowledge graph that contains instances of conventional cultural symbolism*.

2.1 Introduction

Symbols are strongly related to human expression and communication, for this reason, they can be found in many contexts such as art, literature, music, as well as more recently in movies and commercials. However, the knowledge about symbolism, which encompasses interpretations of cultural objects, along with the informa-

[1] This research question is partially answered by this work and by the work described in chapter 3

tion about canonical symbols from different cultures, is mostly stored in unstructured sources such as encyclopaedias and dictionaries of symbols [9, 54, 55, 60]. Accessing and processing this information quantitatively is therefore currently not possible. This chapter presents a novel approach in modelling cultural symbolic knowledge. We focus on conventional symbolic knowledge as expressed by experts or in a dictionary, rather than symbolic knowledge expressed in a specific work of art. Symbolic knowledge may be found expressed in a specific work of art, or in a conventionalized way, e.g. in expert knowledge, or in a dictionary. This chapter deals with conventional symbolic knowledge independently of the interpretation act of attributing a symbolic meaning to specific cultural heritage objects. The hermeneutic act of interpretation of an object and relative attribution of a particular symbolic meaning to that same object will be covered by chapter 3. Before modelling this type of act, a standard computer-readable way to encode known[2] symbolic knowledge needed to be introduced. Symbols are complex objects: the same symbol might convey different meanings depending on the cultural or artistic context in which it is found [60, 9]. In fact, a rooster in Greco-Roman, Christian, and Chinese cultural contexts is a symbol of vigilance, but in a Gothic context it is a symbol of war [8]. Furthermore, different cultures might convey the same meaning with different symbols. This is for example the case for the concept of marriage, that is expressed in the Chinese culture by the symbol of a peach [55], but in the Celtic context, the concept of marriage is expressed by the symbol of a Chalice [61]. Finally, symbolism can be shared by multiple cultures. Envy was represented as a woman with snakes in her hair in the Greco-Roman culture, the same representation of envy was then adopted by the Christian culture and corresponding works of art [9]. The image of a woman with snakes as hair conveys the same meaning of envy in traditional tales of Japan [62, p. 162]. Linked open data sets in the cultural heritage domain currently lack domain-specific ontologies that can express symbolic relationships in a structured way. [7] explains how the lack

[2] With known it is meant the knowledge that can be found as explicit symbolic information in dictionaries of symbols or that it is stored in open resources because but it has not been labelled as symbolic or encoded in a specific and exhaustive way

of semantic models on the domain of symbolism negatively affects the interlinking of resources through their symbols, and shows promising results in the application of a prototype ontology schema to a small set of data. Given the impact of symbols, not only on the representation, comparison, and evolution of human cultures, but also in common expressions, symbolic knowledge, that has been accumulated throughout centuries, has been brought into a machine-readable and open format to foster quantitative research on this topic. Specifically, the goals of this chapter are to fill the gaps on cultural symbolism that currently exist in linked open data through: (i) the Simulation Ontology, a newly developed ontology that conceptualises symbolic relationships, and (ii) HyperReal, a knowledge graph that contains data from heterogeneous sources converted into the Simulation Ontology schema.

This chapter is divided as follows. Section 2.2 describes all the development steps of the Simulation ontology, section 2.3 deals with the creation of HyperReal, section 2.4 follows with the release details of the ontology and the knowledge graph, and section 2.5 concludes the chapter with a discussion on the answered research questions and the completion of the research objectives.

2.2 Simulation Ontology Design

In this section, I explain the design stages of the Simulation ontology. The ontology was designed following the agile ontology development methodologies of SAMOD [63] and extremeDesign [64]. The design of the ontology was completed in three SAMOD iterations. The first iteration covers the aspects related to the top-down modelling (subsection 2.2.1), the second and third iterations cover aspects related to the data-driven modelling (subsection 2.2.2). For each iteration, the following documentation items were generated: (i) a motivating scenario; (ii) an example of data that covers the scenario in natural language; (iii) a series of competency questions in natural language with the expected results related to the example; (iv) a glossary that contains descriptions of the terms used in the motivating scenario; (v) an ABOX that contains the example encoded using the ontology schema as a turtle[3] se-

[3]https://www.w3.org/TR/turtle/

rialisation, (vi) unit tests that test the competency questions. Some examples of the outputs of the development will be provided in this chapter, the full documentation is available at `https://www.w3id.org/simulation/development`.

2.2.1 Top Down Ontology Development

Eco [56] presents an overview of different theories on symbols and symbolism. Compared to the highly debated topic (according to [56]) of what symbols are and what can be considered a symbol, there is less discussion about the relationship that links the symbol to what it symbolizes. This relationship is partly related to Eco's view on *ratio difficilis*: the meaning of the expression is connected with its content and the "deciphering" or inference of this meaning depends on whether the correlation between the expression and content is based on pre-existing rules or it is something personally created by the issuer of the symbolic linking. Of the theories in Eco's overview, Baudrillard's Simulation and Simulacra [57], mentions the terms simulation, simulacrum and reality counterpart as the three main elements of a symbolic relationship. A *simulation* is intended as the relationship between a symbolic element and its meaning. The meaning expressed in a simulation is different from the literal meaning of the element. Lion as a symbol of courage [8] is a simulation, lion as "a large wild animal of the cat family with yellowish-brown fur that lives in Africa and southern Asia" [65] would not be considered as a simulation. The *simulacrum* is the symbolic element, it is the representation of something else. The *reality counterpart* is the something else represented by the *simulacrum*, not the literal meaning of the simulacrum itself. An olive branch (simulacrum) represents peace (reality counterpart). The simulation *olive branch-peace* is the symbolic relationship that links these two elements. Simulations are not universally valid, some only exist in specific settings or contexts. An owl is the symbol of death in Hindu, Japanese, and Mayan contexts. That means that the simulation *owl-death* exists in those contexts. On the other hand, in a Siberian context, owls are symbols of helpful spirits [8]. From the analysis of Simulation theory, the following competency questions emerged: (CQ1.1) What are the reality counterparts of the simulations that have a specific simulacrum? (CQ1.2) What are the simulations that exist within

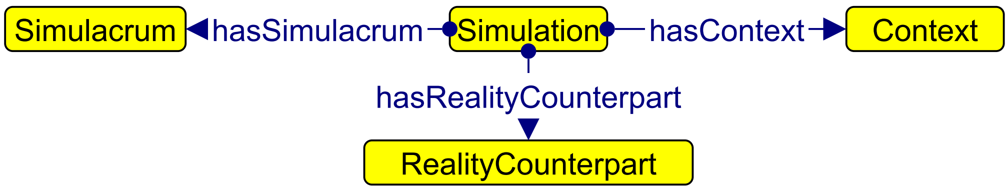

Figure 2.1: Simulation Pattern – Classes and Properties

a certain context? (CQ1.3) What are the simulations in which a certain element participates as either the simulacrum or the reality counterpart? (CQ1.4) What are the simulacra that share the same reality counterpart in their respective simulations, and what is the context in which their simulations exist? A negative competency question[4] was also formulated: (CQ1.5) Are there simulations that have multiple simulacra? As the ontology was not intended as an in-depth description of the philosophical theories of Baudrillard, only the main concepts of his work were reused for the conceptualization of symbolic meanings. To establish a solid foundation in existing good practices, [66] three off-the-shelf ontology patterns were reused to design the OWL2 version of the Simulation ontology: *situation*[5] provides a general structure and vocabulary for n-ary relations; *semiotics*[6] founds Baudrillard's simulations as semiotic acts. An expression (here a simulacrum) denotes a reference (here a reality counterpart), with an interpreted meaning (here muted, but actualisable in the context of a specific interpretation act e.g. in iconology); and *information realization*,[7] which lets us distinguish information objects from their realization or manifestation (here relevant for distinguishing conventional simulations from simulations interpreted e.g. for a specific work of art). In other words, the simulation ontology for conventional symbolic meaning presented here holds between pure information objects (e.g. a lion prototype), and concepts (e.g. force), or stereotyped individuals

[4] Intended as a competency question which should not return any result because it asks for something which does not follow the logical structure of the ontology.

[5] `http://www.ontologydesignpatterns.org/cp/owl/situation.owl`

[6] `http://www.ontologydesignpatterns.org/cp/owl/semiotics.owl`

[7] `http://www.ontologydesignpatterns.org/cp/owl/informationrealization.owl`

(e.g., a generic Persia).[8] The design pattern for simulations is presented in Figure 2.1. Simulacra, reality counterparts, and contexts of symbolic meanings are linked in this pattern through the n-ary relationship (situation) class Simulation. For example, in an Egyptian context, bees signify resurrection. This can be modelled through the `hasSimulacrum`, `hasRealityCounterpart` and `hasContext` properties in our model, as shown in Figure 2.2 To evaluate the initial version of the developed pattern, we qualitatively compare its ontological structure to the structure of a dictionary of symbols. We selected Olderr's [8] dictionary of symbols for this task because it offers more than 40,000 simulations. Moreover, its structure is similar to a dictionary of synonyms and antonyms: for each symbol, a list of potential symbolic meanings is provided along with the context in which those symbolic meanings are valid, often without any additional information. As we will discuss in the conversion Section 2.3 the repetitive patterns in the structure of this dictionary facilitate the automatic conversion of its data into the ontology schema. An example of fitting the pattern to the structure of the dictionary is shown in Figure 2.3. The example shows the "hook" entry of the dictionary. Hook translates in our ontology as the simulacrum. The terms in green (e.g. Christian) represent the contexts in which some hook simulations are valid. The terms in blue represent reality counterparts of simulations that have "hook" as a simulacrum. However, some reality counterparts (in yellow) are introduced by phrases such as "related to", "attribute of" that suggest a specific type of simulation. Furthermore, the terms in red represent some variants of the "hook" entry that have their own symbolic meaning. Both variants and specific simulations were addressed in the design extension of the pattern. The conceptualised classes and properties derived from this qualitative comparison are described in the Subsection 2.2.2.

[8] This is contrasted by using different ontologies that implement simulation occurrences, e.g. a specific interpretation of Persepolis' Achaemenid Persian relief with the Sign of Lion, as a realized simulacrum of Achaemenid Persia power, which will be modelled by the ICON ontology presented in chapter 3

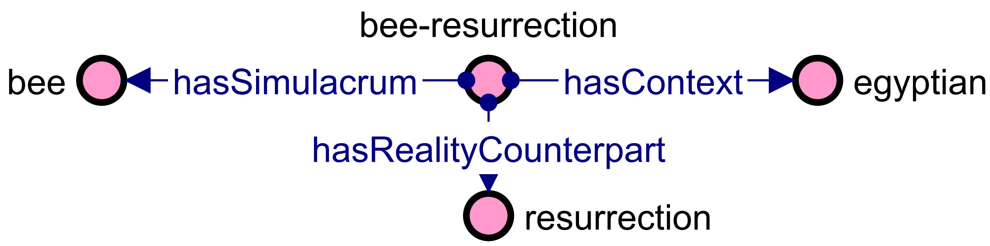

Figure 2.2: Simulation Pattern – Bee-resurrection simulation example

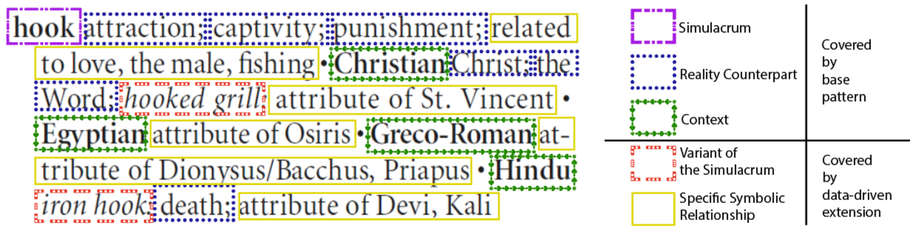

Figure 2.3: Fitting the simulation pattern over Olderr's dictionary "hook" entry

2.2.2 Data-Driven Ontology Enrichment

To cover concepts not dealt with in Baudrillard's Simulation theory, we add 7 new properties and 9 new classes via a bottom-up approach. Simulations are linked to the sources that support their existence, opening up new competency questions: (CQ2.1) What are the simulations and respective reality counterparts that have the same simulacrum but a different source? (CQ2.2) What are the contexts of the simulations listed in a specific source? (CQ2.3) What are the sources of a specific simulation? Another negative competency question was formulated: (CQ2.4) Are there simulations that do not have a source? The source of the simulation was included in the ontology using the class `Source`. PROV-O [67] is the W3C standard ontology to express provenance. The property of PROV-O `wasDerivedFrom`[9] is used to link a simulation to its source. In Olderr's dictionary of symbols [8], simulacra and reality counterparts can have variants that can belong to different simulations compared to the original simulacra or reality counterparts. Variants in this ontology are either represented as (i) a narrower concept than the original simulacrum or reality counterpart, such as *night bird*, which in Olderr's dictionary is a variant

[9] `https://www.w3.org/TR/prov-o/#wasDerivedFrom`

of *bird*, (ii) as a set of things of which the original simulacrum is a part of, such as in the case of *black and white*, which is a variant of *black* in the same dictionary, or (iii) as the simulacrum put in a specific situation such as *bloodstone placed in a glass of water during a drought* as a variant of *bloodstone*. Variants are conceptualized in the ontology with the introduction of the property `hasVariant` that links either a simulacrum or reality counterpart to its variant (which can be either another simulacrum or a reality counterpart). We test this property using additional competency questions to handle the conceptualization of variants: (CQ3.1) What are the variants of a certain element? (CQ3.2) What are the reality counterparts of the simulations with a specific simulacrum or its variants? In Olderr's dictionary, Simulations can be specialized according to the specific symbolic relationship that links a simulacrum with its reality counterpart. If a simulacrum is the emblem of a certain reality counterpart, or an allusion, this specific relationship can be expressed through specialized simulations. Olderr [54] provides some definitions for specific symbolic relationships such as the Allusion, described as

> [...] a reference to an historical person or event, or an artistic or literary work. To have an albatross around one's neck" is an allusion to Coleridge's poem The Rhyme of the ancient Mariner".

or the Association, described as

> [...] something linked in memory or imagination, or by correlation or an analogy with an object, idea, person, or event. The letter "A" is associated with beginning.

Although we suggest potential users of the ontology to follow his definitions, given the highly subjective topic of the ontology, we tried not to limit the use of the classes through logical constraints, as seen in section 2.2.3. In some specialised simulations, the reality counterpart might not be the exact symbolic meaning of a simulacrum. In these cases, the reality counterpart might be something that is prevented, elicited, restored by the simulacrum in a symbolic way. Finally, some simulations might have a reality counterpart that represents the symbolic meaning of the simulacrum, and an additional one that represents something that is prevented, elicited, restored by the simulacrum. For instance, in an Arabian context, an agate is seen as a charm for healthy blood [8]. Therefore, the resulting simulation is `agate-charm-healthyBlood` where agate is the simulacrum, charm is a reality counterpart, healthy blood is an elicited reality counterpart and Arabian is the context. Given that a reality counterpart, such as *healthy blood*, does not change its identity whether it is prevented, elicited, or generally symbolically meant by a Simulacrum, we decided to introduce specific reality counterpart relationships as sub-properties of `hasRealityCounterpart` and not as subclasses of `RealityCounterpart`. To address specific simulations, we formulate the following competency questions: (CQ3.3) What are the simulations in which the simulacra are seen as symbolical protection against reality counterparts? (CQ3.4) What are the simulations that have a specific reality counterpart and other additional reality counterparts, and what specific relationship links those simulations to their reality counterparts? (CQ3.5) What are the simulations and their respective simulacra, contexts, and reality counterparts in which their simulacrum is a symbolical cure for their reality counterpart? The newly added classes and properties are summarised in Table 2.1. Examples of use, and a glossary containing definitions of all classes and properties can be found in the ontology development documentation.

Class	**Specific reality counterpart property**
Association Simulation	no specific property
Correspondence Simulation	no specific property
Manifestation Simulation	no specific property
Relatedness Simulation	no specific property
Attribute Simulation	no specific property
Allusion Simulation	no specific property
Protection Simulation	preventedRealityCounterpart
Emblematic Simulation	no specific property
Healing Simulation	healedRealityCounterpart
no specific simulation class	restoredRealityCounterpart
no specific simulation class	easedRealityCounterpart
no specific simulation class	elicitedRealityCounterpart

Table 2.1: Specific Simulation and related reality counterpart properties

2.2.3 Ontology Axiomatisation

The simulation ontology uses OWL logical axioms to specify its conceptualisation, to enable logical inferences on the symbolic knowledge graph, and to perform automated classification and consistency checking.

Besides the basic classes and properties that have been introduced in the previous sections, and their subsumption and domain/range axioms, we exemplify some OWL axioms that perform restrictions on the main classes, written in Manchester Syntax [68].

- Simulation:

```
hasContext some Context
hasRealityCounterpart some RealityCounterpart
hasSimulacrum exactly 1 Simulacrum
wasDerivedFrom some Source
```

- Healing Simulation:

```
healedRealityCounterpart exactly 1
RealityCounterpart
```

- Protection Simulation:

```
preventedRealityCounterpart exactly 1
RealityCounterpart
```

We also introduce a property chain to provide a direct relation facility between simulacra and reality counterparts:

```
isSimulacrumOf o hasRealityCounterpart
SubPropertyOf symbolicMeaning
```

2.2.4 Ontology Evaluation

The simulation ontology was evaluated in (i) its extraction capabilities with the competency question test, (ii) its compliance to FAIR principles and (iii) its logical and structural consistency.

Simulation	Symbol (Simulacrum)	Symbolic Meaning (Reality Counterpart)	Context
ashTree-odin	Ash Tree	Odin	Norse
ashTree-connection	Ash Tree	Connection	Celtic
ashTree-surrender	Ash Tree	Surrender	Celtic
olive-fertility	Olive	Fertility	General or Unknown
rose-love	Rose	Love	Flower Language
rose-beauty	Rose	Beauty	Flower Language
odin-violence	Odin	Violence	Norse
gazzelle-beauty	Gazelle	Beauty	General or Unknown

Table 2.2: Simulations of the toy dataset created to test the CQs of the first SAMOD iteration.

2.2.4.1 Competency Questions Evaluation

The ontology requirements were expressed through competency questions mentioned in Section 2.2. For each iteration of the ontology, we executed unit tests to verify that the results of the SPARQL query that formalises the competency question matched the expected results of those questions in a toy dataset. All unit tests are available from our GitHub repository in the form of Jupyter notebooks to promote the reproducibility of our results.[10]

The toy datasets used to test the competency questions contain few examples of symbolism from Olderr's dictionary [8] described using the Simulation Ontology, and formalised in turtle.

One example of the competency questions evaluation for each SAMOD iteration will be provided below.

Table 2.2 contains the simulations in the toy dataset for the first iteration (CQ1.1 to Q.1.5). In CQ1.4 "What are the simulacra that share the same reality counterpart in their respective simulations and what is the context in which their simulations exist?" the expected result for the toy dataset would be the simulacra *rose* and *gazelle* which share the reality counterpart *beauty* along with the context of the *rose-beauty* and *gazelle-beauty* which are respectively *flower language* and *general or unknown*. Listing 2.1 shows the formalized query i SPARQL. The retrieved results match the expected ones.

```
PREFIX ex: <https://example.org/>
PREFIX sim: <https://w3id.org/simulation/ontology/>
```

[10]The jupyter notebooks for testing the CQs are available for each SAMOD iteration, respectively in: 1, 2, 3.

```
SELECT DISTINCT ?simulacrum ?context WHERE {
?simulation1 sim:hasRealityCounterpart ?rc .
?simulation2 sim:hasRealityCounterpart ?rc .
?rc sim:isRealityCounterpartOf ?simulation .
?simulation sim:hasSimulacrum ?simulacrum ;
            sim:hasContext ?context
FILTER (?simulation1 != ?simulation2)
}
```

Listing 2.1: CQ1.4 Formalization in SPARQL

Table 2.3 contains the simulations of the toy dataset used for the testing of the competency questions of the second SAMOD iteration. In CQ2.2 “What are the simulations and respective reality counterparts and sources that have the same simulacrum but a different source?” the expected results for the toy dataset are the simulations (i) *olive-fertility* with *fertility* as a reality counterpart and a made up dictionary of symbols 1 (*DS1*) as a source and (ii) *olive-immortality* with *immortality* as a reality counterpart and another made up dictionary of symbols *DS2* as a source. Listing 2.2 shows the formalized query in SPARQL. As for the previous example, the retrieved results from this query match the expected results.

```
PREFIX ex: <https://example.org/>
PREFIX sim: <https://w3id.org/simulation/ontology/>
PREFIX prov:<http://www.w3.org/ns/prov#>

SELECT distinct ?simulation ?rc ?source WHERE {
?simulation1 sim:hasSimulacrum ?simulacrum ;
             prov:wasDerivedFrom ?source1 .
?simulation2 sim:hasSimulacrum ?simulacrum ;
             prov:wasDerivedFrom ?source2 .
filter (?simulation1 != ?simulation2 &&
```

Simulation	Symbol (Simulacrum)	Symbolic Meaning (Reality Counterpart)	Context	Source
ashTree-odin	Ash Tree	Odin	Norse	DS1
ashTree-connection	Ash Tree	Connection	Celtic	DS1
ashTree-surrender	Ash Tree	Surrender	Celtic	DS1
olive-immortality	Olive	immortality	General or Unknown	DS1
olive-fertility	Olive	Fertility	General or Unknown	DS2
rose-love	Rose	Love	Flower Language	DS2
rose-beauty	Rose	Beauty	Flower Language	DS2
damaskRose-freshness	Damask Rose	Freshness	Flower Language	DS2
man-fire	Man	Fire	General or Unknown	LS1
giant-menBeforeTheFall	Giant	Man before the fall	General or Unknown	LS1

Table 2.3: Simulations of the toy dataset created to test the CQs of the second SAMOD iteration.

```
?source1 != ?source2) .
?simulacrum sim:isSimulacrumOf ?simulation .
?simulation sim:hasRealityCounterpart ?rc ;
            prov:wasDerivedFrom ?source .}
```

Listing 2.2: CQ2.2 Formalization in SPARQL

Finally, table 2.4 contains the simulations in the toy dataset used to test the competency questions of the third SAMOD iteration. For the CQ 3.5 "What are the simulations and their respective simulacra, contexts, and reality counterparts in which their simulacrum is a symbolical cure for their reality counterpart?" the expected results are (i) the simulation *amberStone-jaundice* with its simulacrum *amberStone*, reality counterpart *jaundice* and the *islamic* context, and (ii) the simulation *ashLeavesInWine-poison* with its simulacrum *ashLeavesInWine*, reality counterpart *poison* and the *grecoRoman* context. Listing 2.3 shows the formalized query. This result confirms the expected results.

```
PREFIX ex: <https://example.org/>
PREFIX sim: <https://w3id.org/simulation/ontology/>

SELECT DISTINCT ?simulation ?simulacrum
?context ?healedrc
WHERE { ?simulation a sim:HealingSimulation ;
sim:healedRealityCounterpart ?healedrc ;
sim:hasSimulacrum ?simulacrum ;
sim:hasContext ?context .
}
```

Listing 2.3: CQ3.5 Formalization in SPARQL

All the remaining competency questions matched with their expected results.

Simulation	Type of Simulation	Symbol (Simulacrum)	Symbolic Meaning (Reality Counterpart)	Specific Reality Counterpart	Specific Reality Counterpart Relationship	Context	Source
acorn-plague	Protection Simulation	Acorn		Plague	Prevented Reality Counterpart	General Or Unknown	DS1
agate-charm-healthyBlood	Normal Simulation	Agate	Charm	Healthy Blood	Elicited Reality Counterpart	Arabian	DS1
aloe-charm-longevity	Normal Simulation	Aloe	Charm	Longevity	Elicited Reality Counterpart	Egyptian	DS1
amberStone-jaundice	Healing Simulation	Amber (Stone)		Jaundice	Healed Reality Counterpart	Islamic	LS1
ashLeavesInWine-poison	Healing Simulation	Ash Leaves in Wine		Poison		Greco-Roman	LS1
acaciaThorns-neith	Emblematic Simulation	Acacia Thorns	Neith			Egyptian	DS2
bird-theGods	Manifestation Simulation	Bird	The Gods			Hindu	DS2

Table 2.4: Simulations of the toy dataset for the evaluation of the third SAMOD iteration of the Simulation Ontology

2.2.4.2 Automatic Evaluations

Foops! [69] is a web application that evaluates ontologies by verifying that they comply with FAIR principles [70]. According to Foops!, our ontology scores 85%. The tool highlighted the ontology currently lacks versioning information. When a new version of the ontology will be released, this aspect will be covered.

The ontology's syntax was evaluated through the online RDF validator offered by w3.[11] The tool highlighted no pitfalls in syntax.

The ontology was also evaluated by the OOPS tool [71]. All the pitfalls highlighted by this tool except one are about properties or classes that belong to the imported ontology design patterns and therefore are out of our control. The only pitfall regarding the ontology is the suggestion that the Simulation Ontology class *Source* should be equivalent to the Semiotics Ontology Design Pattern class *Reference*. This suggestion was highlighted for the similar meaning between the two words, but the *Reference* class is part of the conceptualisation of the semiotic triangle (or triangle of reference), which is incompatible with the *Source* class which conceptualises the sources that claim the existence of certain simulations.

2.2.5 Related ontologies, taxonomies and structured sources for symbolism

In this section, I analyse the coverage of existing ontologies and knowledge graphs regarding their symbolic content and whether these are modelled as instances of symbolism, and relationships that link a symbol to its symbolic meaning. Furthermore, existing ontologies that partially deal with symbolism but are not yet used in popular knowledge graphs are discussed.

General domain knowledge graphs, such as Wikidata [72] and DBpedia [48] contain properties that link a resource to its symbolic meaning. In Wikidata, the property P4878[12] (*symbolizes*) is a qualifier for statements in which the property P180[13] (*depicts*) is used, and it should represent the symbolic meaning of elements

[11] https://www.w3.org/RDF/Validator/
[12] https://www.wikidata.org/wiki/Property:P4878
[13] https://www.wikidata.org/wiki/Property:P180

depicted in a work of art. Out of more than half a million elements that have been linked to a work of art using the P180 property, only 313 have a *symbolizes* qualifier.[14]

DBpedia uses the property `dbp:symbol`[15] to link a concept to a symbol that represents it. The range of the `dbp:symbol` property is very general, as the property is used both to describe cultural symbols, for instance thunderbolts as the symbol of Zeus[16] but also to express that "metro" is the symbol of the Atocha railway station in Madrid. [17]

Iconclass [73, 74] is a classification system used for attributing of subjects to works of art. Each subject is given a specific code according to the hierachical structure of the system. Specific codes are also given for attributing subjects that appear with symbols, for example, the Iconclass code 11HH(MARY MAGDALENE)[18] is defined as "the penitent harlot Mary Magdalene; possible attributes: book (or scroll), crown, crown of thorns, crucifix, jar of ointment, mirror[. . .]". In this sense, Iconclass contains relevant symbolic information for western art subjects, but its linked open data version consists of a hierarchical SKOS vocabulary. Our work, compared to Iconclass, uses linked open data semantic web technologies to express more specific relationships between symbols and their meaning.

CIDOC CRM [75] is a popular conceptual model for describing museum and cultural heritage artefacts. Its event-centric structure makes it possible to attribute a particular meaning to a specific cultural object. This possibility was extended by the VIR ontology [76], based on CIDOC, using a specific property *symbolize*[19] to link certain elements in a representation of an artwork to their symbolic meaning. The model was enriched to describe complex iconological case studies of interpretation

---

[14] This data was extracted through two queries in the Wikidata SPARQL portal: the query `https://w.wiki/3z56` counts the number of times P4878 is used as a qualifier, the query `https://w.wiki/3z5A` counts the number of elements linked with the property P180. Both queries were run on September 2021, results might be different on new versions of Wikidata.

[15] `http://dbpedia.org/property/symbol`

[16] `https://dbpedia.org/page/Zeus`

[17] `https://dbpedia.org/page/Madrid_Atocha_railway_station`

[18] `http://iconclass.org/rkd/11HH%28MARY%20MAGDALENE%290/`

[19] `https://ncarboni.github.io/vir/#K14_symbolize` URL not working for every browser. Persistent URI of VIR ontology: `http://w3id.org/vir`

[77]. However, the dependency of the interpretation acts makes symbolic relationships encoded using this ontology solely based on the single work of art (i.e. not generalized). In other words, currently CIDOC and VIR permit to describe symbols that are only valid in the context of the work of art in which they are found as a result of an interpretation. The purpose of our work is to first introduce a model that can express general background symbolic knowledge without covering the hermeneutical act of interpretation. Parts of the CIDOC model were used to develop the ICON ontology in chapter 3.

Gartner [78] proposes an ontology to conceptualise iconographical recognition of subjects in artworks according to Panofsky's second level of interpretation.[20] Among the classes that are conceptualised in Gartner's ontology, there is a mention of symbols as recognizing elements for specific works of art's subjects. Unfortunately, the ontology has not been released yet.

Symbolic information can also be found in Wordnet [79], a lexical database of words and relationships between them that has been under development by cognitive linguists since 1985. The dollar sign[21] for example is defined as "a symbol of commercialism or greed", but it does not contain structured information to distinguish between its meaning as denoting a currency and as a symbol of greed. Only in some cases, a meaning is described as figurative, although only in its definition. For instance, for Albatross[22] the definition "(figurative) something that hinders or handicaps" appears in the textual description of the entry. Wordnet aims not to distinguish between symbolic and literal meanings, but the information that can be found in this resource can be extracted and re-engineered to highlight the former from the latter. As explained in Section 2.3, Wordnet's symbolic data was ingested in HyperReal.

[20] Panofsky's second level of iconographic interpretation is about the subject matter of works of art. In this level characters, symbols, places, events, allegories, and stories are associated with the artistic motifs present in the work of art [58].

[21] `http://wordnet-rdf.princeton.edu/id/06834465-n`

[22] `http://wordnet-rdf.princeton.edu/id/05697450-n`

2.3 HyperReal Development

According to Baudrillard, our world is so intertwined with simulations, that we now live in what is called “hyperreality” [57]. Initially defined as a dimension in which the reality is only comprised of simulacra, and all the references to the “real” reality are lost in the simulation, Saussure [80] reprises this concept, and he defines it as a dimension in which the simulacrum and its reality counterpart are blended, and they are almost indistinguishable. Therefore, I decided to give the name *HyperReal* for the knowledge graph that contains triples of symbolism encoded using the Simulation Ontology.

HyperReal was developed by extracting and converting data from different sources: DBpedia, Wordnet, Olderr’s *Symbolism: a comprehensive dictionary* [8] and Alexander Francis Otto’s *Mythological Japan. The Symbolisms of Mythology in Relation to Japanese Art* [55]. The diverse selection of a general domain knowledge graph (DBpedia), a lexical resource (Wordnet), and two unstructured data sources with different characteristics (as Olderr’s Dictionary contains a more repetitive syntax compared to Otto’s Encyclopaedia, which is mostly based on free text) was made with the purpose of tackling the task of converting multiple-heterogenous types of data into Linked Open Data following the Simulation Ontology. For every converted simulation, the respective sources (either DBpedia, Wordnet, Olderr’s dictionary or Otto’s work) have been added using the `prov:wasDerivedFrom` property. Table 2.5 summarises the number of resources that were generated by the conversion.[23] An evaluation of the conversion algorithms can be found in Section 2.3.7. The following subsections describe the specific methodologies applied to convert each of the selected sources.

2.3.1 DBpedia Conversion

We use the DBpedia SPARQL endpoint[24] to retrieve resources related to the property *dbp:symbol* and for resources of type *skos:Concept*[25] that contain the string

[23]The totals are not the exact sum of the previous numbers because some simulacra, reality counterparts, contexts, and simulations are shared by different sources.

[24]`https://dbpedia.org/sparql`

[25]`http://www.w3.org/2004/02/skos/core#Concept`

"symbol" in their label and that have these concepts as their subject (using the property *dct:subject*).[26] For instance, the resource *dbr:Eagle*[27] is the subject of dbc:National_symbols_of_Liechtenstein.[28]

To filter out station signs and chemical codes that are also modelled using the `dbp:symbol` property, but are purely iconic tools, we exclude resources of type Railway Stations and Public Companies. On the one hand, in the triple `?subject dbp:symbol ?object`, ?subject becomes our reality counterpart, and ?object the simulacrum. The *?type* of the same subject, in the triple `?subject rdf:type ?type` was considered as the context of the simulation. Because ?object was either a string or a URI, different approaches were required. In the case of a URI, the simulation was simply created by converting each part of DBpedia into the corresponding elements of the simulation ontology. In case of a string element, if the "," punctuation was found in the string, multiple simulation, with different simulacra and same reality counterpart, were created according to the number of elements resulting from the splitting of the string with the "," character. For instance, the triple `dbr:Zeus dbp:symbol "thunderbolt, eagle"` was converted into two simulations, one with "thunderbolt" as the simulacrum, and the other with "eagle".

On the other hand, in the triple `?subject dct:subject ?object`, `?subject` became the simulacrum and `?object` (cleaned of the initial parts such as "National Symbol of" or "Symbol of") became the reality counterpart. The contexts, in this second case were not extracted for these triples, therefore *General or Unknown* was set as the context for every simulation. For instance, in the triple `dbr:Eagle dct:subject dbc:National_symbols_of_Liechtenstein` , *Eagle* became the simulacrum and *Liechtenstein* became the reality counterpart in a Simulation with a General or Unknown context. For both queries conversion processes, the DBpedia URIs of the resources were kept in the HyperReal by linking the newly created

[26] `http://purl.org/dc/terms/subject`
[27] `http://dbpedia.org/resource/Eagle`
[28] `http://dbpedia.org/resource/Category:National_symbols_of_Liechtenstein`

simulacra and reality counterparts to the corresponding DBpedia resources using the *owl:sameAs* property[29]. Once the filtering phase was done, the data was converted into a turtle file using algorithms written in python with the RDFlib library and following the mapping explained previously. Our DBpedia conversion yielded 3,696 simulations.

2.3.2 Wordnet Conversion

From Wordnet, we extract Symbol and emblem synsets[30] and their hyponyms, as well as synsets that contain phrases such as *symbol of* or *emblem of* in their definition. For each element, we extract its label and definition. The natural language definition of each synset was processed to extract potential contexts and symbolic meanings. For instance, the Penelope synset[31] has "Penelope" as a label and "(Greek mythology) the wife of Odysseus and a symbol of devotion and fidelity[...]" as the definition. The label of the synset was converted into the simulacrum, and we extracted the context and reality counterparts from the definition. In this case, two simulations were created from this synset, *penelope-fidelity* and *penelope-devotion*, both with *greekMythology* as a context. The *penelope* simulacrum was then linked to its original WordNet synset with the property `owl:sameAs`. Our WordNet conversion process yielded 81 simulations.

2.3.3 Olderr's Dictionary Conversion

We used the markup such as boldface in Olderr's dictionary to distinguish between different lemmas, contexts, and variants and converted this to RDF. A simulation URI was formed by joining the simulacrum and reality counterpart labels with a hyphen. The style consistency in the original document made the conversion process almost completely automatic. Some manual corrections had to be done after the automatic conversion in the rare cases of mistakes in the source.[32] A list of bigrams and n-grams were used to detect specific simulations, such as *related to* and

[29] `http://www.w3.org/2002/07/owl#sameAs`
[30] `http://Wordnet-rdf.princeton.edu/id/05773412-n` and `http://wordnet-rdf.princeton.edu/id/06893714-n` respectively
[31] `http://wordnet-rdf.princeton.edu/id/09616318-n`
[32] `https://w3id.org/simulation/code/` lists examples of manual corrections

Source	# of Simulacra	# of Rc	# of Contexts	# of Simulations
Olderr's dictionary	8924	17,990	303	37,763
DBpedia	3006	779	40	3,696
Wordnet	61	76	7	81
Otto's work	106	108	9	185
Total	**11,815**	**18,745**	**321**	**41,623**

Table 2.5: Conversion algorithms evaluation metrics (reporting micro averages)

protection from which resulted in *Relatedness* and *Protection* simulations, respectively. 37,763 simulations were generated from the dictionary.

2.3.4 Otto's work Conversion

All the simulations presented in the symbolism monograph written by Alexander Otto and Theodore Holbrook were firstly manually annotated and then written in a spreadsheet file. Then, the file was converted into RDF, following the simulation ontology schema, with Python scripts that used methods and classes from the RDFlib library. The conversion of *Mythological Japan: The Symbolism of Mythology in Relation to Japanese Art* yielded 185 simulations.

2.3.5 Quality Control

Following the conversion process, we conducted several cycles of error analysis to evaluate the quality of the knowledge graph. We made adjustments to the algorithms and reconverted the knowledge graph as necessary. The most prevalent errors observed were instances where certain simulacra or reality counterparts were not linked to any simulation. Additionally, errors arose during automatic URI generation and the labeling of elements.

In particular, with regards to Olderr's dictionary, some errors stemmed from editing mistakes in the original source. For example, certain lemmas in the dictionary lacked meanings. Furthermore, in instances where multiple meanings were separated by a comma, the first element sometimes consisted solely of two commas (", ,"), indicating a missing meaning due to an editing error in the original source. To rectify these issues, the TXT version of the dictionary underwent manual correction. These errors were detected by some simulacra that were created without a reality counterpart.

All the errors encountered during the DBpedia conversion were attributable to the varying character encodings between the knowledge graph and DBpedia. DBpedia employs an UTF-8 encoding, whereas the algorithm developed for the knowledge graph used ASCII. By converting HyperReal into the UTF-8 encoding, we successfully resolved all the associated errors. These errors were initially detected by the presence of empty URIs due to the failed recognition of some characters.

2.3.6 Reasoning over the knowledge graph

The entire knowledge graph was analysed by the HermiT Reasoner [81] to check inconsistencies. The reasoner showed no inconsistencies. The OWL Profile Checker tool (V1.1.0)[33] positions the knowledge graph in the OWL2-DL profile.[34]

2.3.7 Conversion Evaluation

To evaluate the quality of the conversion of the algorithms, we manually annotated 112 simulations sampled from some lines of Olderr's dictionary and compared it to the automatic conversion of those same lines using precision, recall and F_1 measure of simulacra, reality counterparts, variants, contexts, and simulations and specific types of simulations retrieved.

On average, the conversion achieves a performance of 97% of precision (micro), recall (micro) and F_1 (micro). Table 2.6 summarises the metrics for this evaluation. Errors in the conversion stem from deviations from the dictionary structure, for example when two reality counterparts are provided instead of one.

2.3.8 Entity alignment of HyperReal

All the entities in HyperReal have been processed by Babelfy [82], a tool that performs entity alignment[35] to Babelnet.[36] For each Babelnet entity that was associ-

[33] `https://github.com/stain/profilechecker`

[34] The description logic complexity of the entire knowledge graph is ALCHOIQ (role hierarchies, nominals, inverse properties, qualified cardinality restrictions).

[35] Entity alignment refers to the task of identifying and linking entities across different knowledge bases or ontologies. In the context of the semantic web, entity alignment is used to connect entities from different sources of structured data and enable the integration of data from multiple sources.

[36] Babelnet is both a comprehensive multilingual dictionary, encompassing lexicographic and encyclopedic information on various terms, and an ontology that establishes connections between concepts and named entities through an extensive network of semantic relations. This network, known as Babel synsets, consists of almost 14,000,000 million nodes. More information on Babelnet is

Element	Precision	Recall	F_1
Simulation	0.96	0.97	0.97
Simulacrum	0.97	0.98	0.98
Reality counterpart	0.96	0.97	0.97
Context	0.97	0.98	0.98
Type of Simulation	0.94	0.95	0.94
Variant	0.97	0.98	0.98
Average	0.96	0.97	0.97

Table 2.6: Conversion algorithms evaluation metrics (reporting micro averages)

ated to symbols and symbolic meanings, additional information, when present, was extracted. Namely, broader entities, narrower entities, DBpedia category, DBpedia entity, Wordnet correspondence, label in English (from Wordnet). The DBpedia categories are linked to Babelnet entities through Lemon (Lexicon Model for Ontologies). This is a model that facilitates sharing and linking of terminological and lexicon resources on the Semantic Web, allowing for the representation of linguistic information in conjunction with ontologies [83].

Having links to lexical databases such as Wordnet and Babelnet allows an easier reconciliation of external entities (which might be aligned to their Babelnet or Wordnet correspondence, such as some Wikidata entities) to the symbols and symbolic meanings contained in HyperReal, as it will be explained in some experiments of chapter 4. Furthermore, symbols and symbolic meanings can be grouped together into broader symbolic categories, allowing experiments as it will be presented in section 4.3.

An example of the entity alignment from the simulacrum *butterfly* can be seen in figure 2.4.

The version of HyperReal comprising entity alignments will be released in the future.

2.3.9 Generating a taxonomy of contexts

HyperReal references more than 300 cultural contexts in which simulations exist. These contexts were derived from the literary and digital sources used

available at `https://babelnet.org/about`

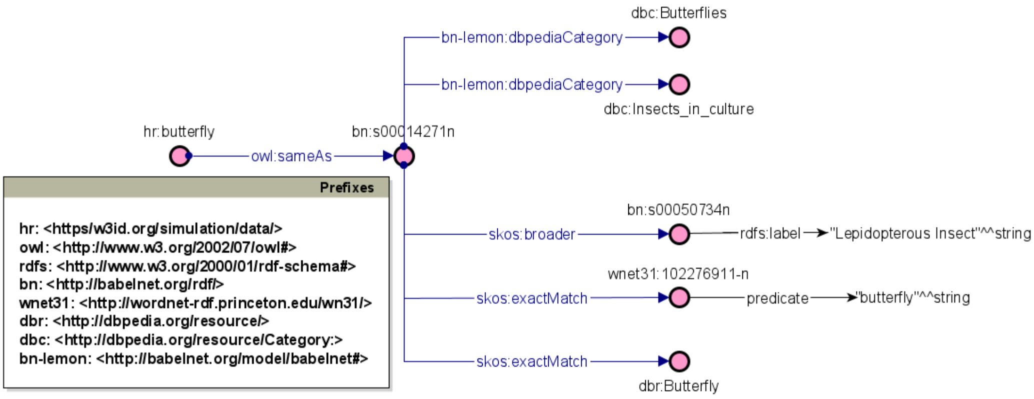

Figure 2.4: Entity alignment for the *butterfly* simulacrum of HyperReal

to build the knowledge graph. Initially, no relationships existed between contexts. As it is mentioned in the future work of the scientific article that describes it [33], these cultural contexts could be connected between each other in terms of relatedness and through hierarchical relationships. By using the SKOS model, I grouped together the cultural contexts in HyperReal in 15 Macro Categories, conceptualised by the class `skos:Collection`, and created a relationship of relatedness (`skos:related`), broadness (`skos:broader`) and narrowness (`skos:narrower`)[37] between the contexts. These new connections and grouping would allow for new possible queries such as “Extract all the simulations that belong to context X and all its broader/narrower contexts” or “Extract all the symbolic meanings that are part of simulations of all the contexts in the Macro-area/Collection Y”. The taxonomy was used in the experiment that compares cultures using their symbolisms presented in section 4.1 of chapter 4. In the future, both the areas and the single contexts will be aligned to relevant cultural taxonomies such as the Getty Art and Architecture thesaurus (AAT).[38]

2.4 Ontology and knowledge graph release

The w3id service was used to obtain persistent URIs for the Simulation Ontology and HyperReal. The current version of the Simulation Ontology is available at

[37]In SKOS, the property `skos:broader` uses the narrower term as the domain and the broader term and the range, and for `skos:narrower` it is the opposite.
[38]`https://www.getty.edu/research/tools/vocabularies/aat/`

https://www.w3id.org/simulation/ontology. Its structure is shown in Figure 2.5. Moreover, the ontology has been inserted into the LOV public registry [84] and it is available there at https://lov.linkeddata.es/dataset/lov/vocabs/simu.

HyperReal is available at https://www.w3id.org/simulation/data/ and includes the import axiom for the ontology schema. The SKOS taxonomy of cultural contexts of HyperReal is released in a turtle serialization at https://w3id.org/simulation/contexts/.

Furthermore, https:www.w3id.org/simulation/development, www.w3id.org/simulation/code, www.w3id.org/simulation/docs/ are the persistent URIs that lead respectively to the ontology development GitHub repository, the scripts used to create HyperReal, and a Widoco [85] documentation of the Ontology.

Finally, the preferred prefix "sim" for the Simulation Ontology was registered on the service https://prefix.cc.

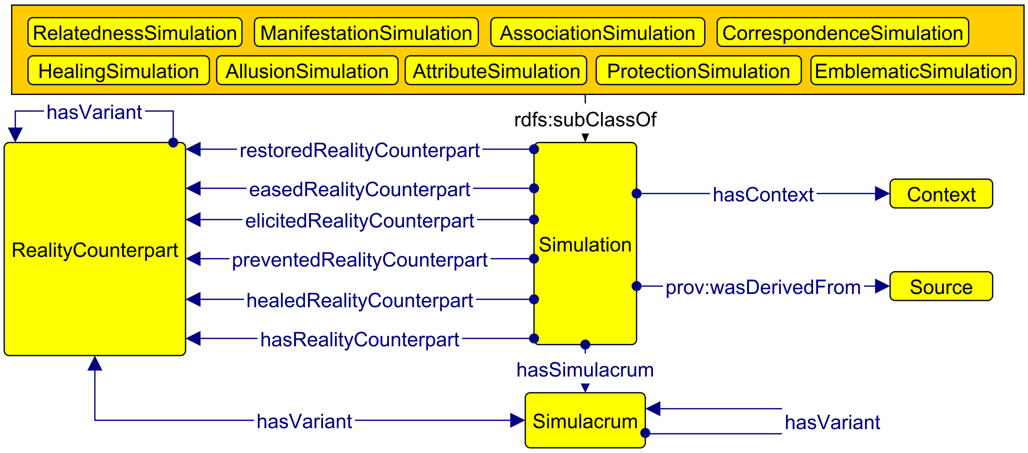

Figure 2.5: Simulation Ontology - Classes and Properties

2.5 Conclusion

I was able to answer RQ1 *To what extent can symbolic relationships in the cultural heritage domain be encoded into computationally ready, structured, semantically linked data?* after testing the competency questions of the Simulation Ontology. They were formulated to cover all the aspects of symbolism that emerged from a top-down perspective on the Baudrillard's elements, and from a bottom-up perspective from the study of symbolic data in the form of the content of the dictionary of symbols. With my conceptualisation of the classes and properties of the Simulation Ontology, I was able to cover all the information required by the competency questions, showing that symbolic relationships described from both a data-perspective and a theoretical perspective on symbolism, can be included in an ontology and in the form of linked open data. Therefore, linked open data is very effective for representing a range of symbolic relationships. Utilising n-ary classes to describe a Simulation as the essential component between a symbol, its symbolic meaning, the cultural context in which the symbol symbolises the symbolic meaning, and the provenance of the statement, satisfies the criteria of Baudrillard's theory. Moreover, the specialisation of symbolic meaning (or reality counterpart) relations was extended to include the representation of symbols that serve as a symbolic remedy for specific illnesses, a symbolic shield against negative effects, or as stimulants to certain emotions or feelings. Specialisation are also applied to the entire simulation, such as the cases of (i) the attribute simulation where the simulacrum serves as the symbolic attribute of the reality counterpart, which is usually a character or personification, or (ii) the association simulation where the symbolic role of the simulacrum and reality counterpart are less rigid and could be interchanged.[39] These specialisations of simulations and symbolic meanings enable the ontology to be well-suited for conceptualising symbolic data that is stored in dictionaries of symbols, satisfying the data-driven requirements.

Investigating the development of the Simulation Ontology only partially addresses RQ1.1 *Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism and symbolic interpretation in the cultural heritage domain?*. In reality, the majority of the requirements from a domain perspective come from semiotics rather than what is stated in RQ1.1, such as the historical, iconographical, and iconological domains. Regarding conventional sym-

[39]One could say that the sea is associated with the Greco-Roman Poseidon/Neptune or vice versa

bolism, the main ideas necessary to conceptualize an ontological model are sourced from the semiotics field. This research question is shared with the study in chapter 3, which concentrates more on historical and art historical domains. The only cultural contribution for this ontology comes from the data of the dictionary employed in the bottom-up approach of the modelling, but it is utilized merely as a template and no formal theory on cultural connections in symbolism was needed.

Furthermore, to answer RQ1.2 *To what extent and through which means can unstructured data of symbolism be re-engineered into linked data?*, the two entirely different methodologies needed to convert unstructured sources of data into the knowledge graph HyperReal witness the difficulty of adopting one general conversion pipeline. On the one hand, the quasi-standardised syntax of the dictionary of symbols made its conversion almost entirely automatic, with some minor corrections applied to the converting algorithms through cycles of evaluations. On the other hand, Otto's encyclopedia lacked a standard structure, and, for the purpose of this thesis, a manual approach was preferred to minimise mistakes that would have spawned from the use of automatic phrase catchers. Nevertheless, both the content of Olderr's dictionary and Otto's Encyclopedia (and to some less extent, the labels that were parsed from DBpedia and Wordnet) could be completely caught and converted into a knowledge graph with a minimum to none loss of information. In fact, I was able to yield 41,623 instances of simulations from the content of all the sources analysed. By performing an evaluation on the conversion that produced most simulations, i.e. the Olderr's dictionary conversion, table 2.6 shows very high levels of precision, recall and F1 (around 0.97 overall) for the detection and conversion of every element that takes part in a simulation. That said, every unstructured source comes with its levels of specificity, so I cannot claim that every possible source can be 100% converted into a knowledge graph without losing information. Estimating on the results achieved so far, unstructured sources can be actively re-engineered to fit on knowledge graph schemas with an acceptable degree of information loss, and with a level of automation that depends on the rigidity of their content in terms of repeating syntax.

Finally, both research objectives 1 *Development of an ontology that can describe conventional cultural symbolism* and 2 *Creation of a knowledge graph that contains instances of conventional cultural symbolism.* are completed with this work. Following the narrative of this thesis, we now have the machine-readable data of conventional symbolism, which can be used to infer symbolic interpretations of artworks given that the list of entities that they depict is provided. What is still missing to create the Semantic Web of Symbolic Meanings is an ontology that can be used to describe the hermeneutic act of artistic interpretation, which can link the conventional symbolism contained in HyperReal to the recognised iconographical elements in an artwork. The next chapter will provide an ontology for this purpose.

Chapter 3

Modelling Artistic Interpretation on the Semantic Web

This chapter is almost completely based on ”ICON: an ontology for comprehensive artistic interpretations” [86], a paper about the development of ICON, an ontology that conceptualises artistic interpretations, with minor adjustments to fit with the narrative of this thesis. This paper was co-written by me and Sofia Baroncini (and our supervisors Aldo Gangemi, Marieke van Erp and Francesca Tomasi), a PhD student in Literal and Philological Culture from the University of Bologna. In particular, I focussed on the subsections 3.2.1, 3.2.3, and 3.2.4 of the state of the art, subsections 3.5.1, 3.5.2, 3.5.3, 3.5.4, 3.5.5, 3.5.5.1, 3.5.5.5, 3.5.5.6, 3.5.6 of the ontology design, subsection 3.6.4 of the evaluation, section 3.7. Both me and Sofia contributed equally to the conclusion (3.8) section.

This work aims to answer RQ 1.1 *Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism and symbolic interpretation in the cultural heritage domain?*, together with what was presented in chapter 2.

3.1 Introduction

Distinguishing between what can be considered an artwork from what is not, and reach a precise definition of art itself can be challenging in a dynamic world in which new forms of art are constantly introduced [87][35]. For this work, we refer to artwork(s) as a “visual object or experience consciously created through an expression of skill or imagination” [88]. Since an artwork, for its nature, usually cannot be thoroughly understood only from its objective characteristics, it is subjective to observers’ interpretations. In this work, we present a new ontology, ICON, which models art interpretations of the artworks’ subject matter and its possible meanings. In the context of art comprehension, an interpretation is intended as “any kind of assignment of meaning or significance to artworks” [87, p. 113]. Through interpretations, art historians can claim different kinds of explanations about the artwork concerning different aspects e.g. artwork’s content, the tendency of the art of a century, or the reception of artworks by the public[89, p. 114]. Among them, the interpretations considering the comprehension of the subject matter of an artwork fall under the domain of iconographic interpretation, which is based only on internal aspects of the artwork (e.g. a child with bows, arrows, and wings depicted in an artwork is recognized as Cupid and as a symbol of love)[89, p. 124]. The interpretation conducted on this basis can be further enriched by other types of interpretations which take as evidence an external source, i.e. the cultural context[89, p. 131]. In this sense, artworks are read as symptoms of the contemporary culture [42]. For example, the fact that, during the Middle Ages, the classical deities were represented deprived of their classical form, can be read as the incapability of the Medieval artists and society of retaining a classical model with its appearance, since it was too far from their taste and from the new Gothic representational conventions. [90].

Therefore, the domain of knowledge of artworks content interpretation is complex and characterized by the subjectivity of the author of each claim. For these characteristics, we believe that Semantic Web technologies, with a focus on ontologies, are a suitable tool to conceptualise this semantic expressivity by means of the high level of granularity and flexibility offered.

Since the domains of iconography and iconology described above concern the description and comprehension of artworks’ content, we model art interpretations according to them.

Among the approaches adopted by the scholars in the context of the artworks' content interpretation, the iconographical-iconological method[1] formalized by Erwin Panofsky had, in the last century, the greatest relevance and influence on contemporary art historians [91]. Panofsky followed the approach firstly adopted by Warburg in his studies [16] which was aimed at understanding the artistic subjects and motifs as witnesses of socio-cultural phenomena. Whereas his studies are nowadays fundamental for the approach itself, the prevailing perspective is the one formulated by Panofsky in a three-layered framework of the artwork's understanding [36]. On that basis, some scholars proposed variations over levels subdivision [92, 76], sometimes including other aspects, such as the artist's psychology [15] or the iconic language of the image [93]. Although the framework is recognized as a valid method or approach, it is commonly accepted to acknowledge that this kind of interpretation is subjective and intuitive when practically applied [36].

For its complete formalization and historical relevance, we will refer to Panofsky's approach as a representative method of the discipline[2], yet considering the enrichment given by other scholars mentioned in section 3.3. In detail, Panofsky subdivides the act of interpretation in three levels, the first two of which fall under the traditional domain of iconography, i.e. the identification of iconographies (level 2) attributes and variants with which they can be represented (level 1), whereas the last level concerns the socio-cultural interpretation of artworks content, closer to Warburg's iconological approach [38].

We consider Panofsky's theoretical approach as suitable for modelling because it is, from our perspective, the most complete attempt in the literature to formalize the discipline in detail. Indeed, it not only defines the mechanisms of interpretation and its components, but also gives precise indications on i) the types of subjects, and ii) how the subjects, their components and meanings are related. Furthermore, it was already used to set the requirements for an adequate description of artistic interpretations, in terms of granularity, in section 1.2 of chapter 1.

[1] Which inspired the name of this ontology

[2] It has to be specified that albeit the relevance of the theory is affirmed, it was mainly developed to be used in Panofsky's area of interest, i.e. the Western Middle Ages and Renaissance Art.

In this work, we consider the concept of interpretation according to the presented theory, i.e. an observer interpreting what is represented by one or more artworks, their possible iconography and meaning. Therefore, other types of interpretations, such as the results of the observation of the physical object (e.g. measurement) and its metadata definitions (e.g. datation, author and title attribution) do not fall within the scope of this work. The concept of subjectivity is limited to the described situation. As the content and meaning interpretation always depend on the viewer perception and background knowledge, multiple, incompatible interpretations may derive from different observations of the same artwork.

The chapter is structured as it follows. In section 3.2 we analyse the state of the art of ontologies and general domain schemas that deal with interpretations, relevant concepts like symbolism, or description of cultural heritage objects. Section 3.3 introduces the theoretical background on art interpretation, on which the ontology is based. Then, section 3.4 explains the requirements that were used to design the ontology, along with potential users and lexical usage. Section 3.5 describes our design process of the ontology, describing in detail all the design iteration that have been undertaken to model different parts of our work, together with the axiomatisation details, the alignments, and reuse of existing ontologies. In section 3.6 the evaluation process of the ontology is explained, with both automatic evaluations provided by relevant tools and quality-based evaluation over the granularity potential of the model. Section 3.7 shortly deals with the release of the ontology and the publication of its documentation. Finally, section 3.8 concludes the chapter with a final discussion about the impact of the ontology, future work and the implications on the research question RQ1.1.

3.2 State of the Art

Interpretations in Semantic Web are a widely discussed topic [94, 92]. In this section, we will analyze i) related work that cover specific aspects of cultural heritage interpretations and possibly iconographical-iconological content in the form of ontologies,[3] ii) how these and related ontologies model the concept of interpretation, iii) iconographic and iconological elements contained in general domain schemas,[4] iv) existing controlled vocabularies and taxonomies designed to classify iconographical and iconological elements, intended as authoritative sources of knowledge that provide permanent URIs for potential subjects, art styles, and other relevant information in the context of art interpretation. Following the re-usability principles of the Semantic Web [95], we reuse parts of the models listed below in our ontology by making alignments between our classes/properties and theirs, or by directly reusing parts of their schema. The alignments are described in section 3.5.5.

3.2.1 Ontologies related to iconography and iconology

In the context of art interpretations, several attempts have been made to create models that cover some specific elements related to interpretations (i.e. symbolic meanings) or the whole act of interpretation of a cultural heritage object. CIDOC-CRM [75] is a widely used ontology in the context of cultural heritage. It has an event-based structure and covers fundamental aspects of the life cycle of a cultural heritage object. Carboni et al. extended it with the Visual Representation Ontology (VIR) [76]. VIR ontology explores the concept of visual representations in artworks, and associates the portion (called *iconographical atom*) of the cultural heritage object to the recognized subject. We use SKOS alignments to refer to parts of CIDOC and VIR in our ontology, and compare the coverage of our ontology and VIR in section 3.6.3. Compared to our ontology, VIR focuses only on subjects of level 2, considering iconographies and their attributes, consequently lacking of a clear distinction between levels. The preliminary study conducted in [92] further extends VIR by the addition of an iconological interpretation class linking to the artwork concepts and external cultural phenomena. It is evaluated over 11 real case studies taken from the literature in iconology, which illustrate a wide vari-

[3] We consider here only those ontologies specifically designed to deal with cultural heritage

[4] By general schemas we intend data models expressed through an ontology that were not designed with the scope of describing only cultural heritage but still contain relevant aspects of our work. The distinction between these and the previous ones lies only in the purpose of the ontologies in question.

ety of aspects included in an iconological analysis. In addition, the work is based on a careful theoretical comparison of the main iconological and iconographical interpretation theories. For its comprehensive overview over iconographical and iconological theories, along with the real-base evaluation, it is used as a source for ontology development here proposed, that has to be seen as its development and refinement. We deepen this study by developing aspects not already considered, such as a more detailed description of level 1 and 2 subjects and the integration of multiple interpretations by different art historians. Gartner [78] proposes an ontology to facilitate and automate the identification of subjects (level 2) in works of arts through logical inferences. No alignments were possible to this ontology because it has not been released. ARCO's ontology [13] was developed to model Italian cultural heritage artefacts by converting information contained in traditional catalogue sheets into linked open data. Among the possible aspects modeled for an artwork, some classes were designed to describe its iconographical apparatus. Apart from this class, the schema does not mention any distinctions between different levels of interpretations. Most of the information about the iconographical and iconological interpretations in ArCo are provided through natural language descriptions with the property `dc:description` or `core:description`[5], not exploiting the full potential of Semantic Web [17]. As we mention in sections 3.4 and 3.5, in the development of ICON we designed specific classes and properties to express this information with the necessary granularity. We reuse some parts of ArCo to refer to the concept of artwork and subject. The Simulation Ontology, presented in chapter 2, models conventional symbolism. Compared to ICON, this ontology does not consider the hermeneutic act of interpretation of associating the symbolic meanings to artworks. Nevertheless, its conceptualization of symbolic meanings using n-ary relationship classes that link a symbol, its symbolic meaning, and the cultural context in which the symbol-symbolic meaning relationship is directly reused to express the symbolic meanings in this work, inserting it in the context of an interpretation of an artwork.

[5]See `https://dati.beniculturali.it/lodview-arco/resource/HistoricOrArtisticProperty/0500653281.html` dc:description value.

3.2.2 Ontological modelling of interpretations

In the context of knowledge organization, several ontologies addressed the concept of interpretation. CIDOC-CRM models assertions with the class `E13_Attribute_Assignment`, which relates the assertion made by one agent to the object considered. Since each assertion reflects the agent's opinion, multiple, contradictory assertions may be represented. The concept of interpretation is applied broadly, including measurements and other types of scientific observations. Similarly, the class `Interpretation` of Arco is intended to describe every piece of information asserted by an agent about an object on the basis of stated sources.[6] The CIDOC-CRM extension CRMinf (CRM inference, also called argumentation model) [96] deepens the concept expressed by `E13_Attribute_Assignment` distinguishing the type of argumentation and if the belief resulting from the argumentation holds true or not. The concept is further explored by the CRMsci (also called scientific observation model) [97], another CIDOC-CRM extension, which integrates CRMinf by formalising the shared scientific process adopted across different domains and the scientific activities involved. In detail, of great interest is the class `crmsci:S4_Observation`, subclass of `crmsci:I1_Argumentation` and of `crm:E13`, expressing the scientific observation of physical events or reality which is done directly or through measurements. It represents the "transition between reality and propositions" [98]. Furthermore, the VIR ontology adds a domain-definition of `crmsci:S4 Observation` by declaring its subclass `vir:IC12 Representation`, which represents an assignment of a solely iconographical status to a physical object.

The same topic is addressed also by the history domain to represent the frequent case of disagreeing historians' interpretations of the same events. As reported by [99], several ontologies afford the theme by modeling different views of the same observed events, such as SEM ontology [100], MIDM (Multiple Interpretation Data Model) [101], the ODP (Ontology design pattern) Event-Model-F[7],

[6] `https://dati.beniculturali.it/lodview-arco-onto/ontology/context-description/Interpretation.html`

[7] `http://ontologydesignpatterns.org/wiki/Ontology:Event_Model_F`

expanding DOLCE+DnS Ultralite (DUL), in which a distinction between facts and interpretations is already stated [102]. HiCO (Historical Context Ontology) [103][8] goes further by adding contextual information to the interpretation, such as interpretation type and criterion.

The concept of interpretation used in our ontology reflects the VIR perspective, narrowing down the field to those interpretations that an observer may do about the visual representation of an artwork, excluding scientific observations about its physical features. Moreover, we further specify the meaning of an interpretation in the context of this work by referring to the definition provided by [101] relatively to the archaeological field, stating that the views produced by scholars are "the result of an interpretive reasoning that includes the subjectivity of the author", due to the uncertainty and incompleteness that often characterizes archaeological data [101].

While modelling an interpretation, it is important to define also what the concept of *meaning* is. Considering semiotics, there are several aspects of meaning that can be modelled [104]. In the context of this study, *meaning* is defined according to the semiotic theory [105] in which a signifier, i.e. an icon, signifies the carried signification, i.e. a meaning [106, pp. 93-94]. Therefore, following Panofsky's modelling, all the subjects identified at each level of interpretation are considered meanings [1]. This definition corresponds to the notion of meaning as a social object introduced in [104].

3.2.3 General schemas containing iconographical and iconological elements

Although iconographical and iconological descriptions are not their main focus, some general domain schemas contain information related to art interpretation. Often these schemas rely on the sole subject property (such as `dc:subject`, `schema:about`) to describe any information about the iconographical and iconological content [43, 44, 29, 14]. One of the outliers in this characteristic is Wikidata and its Wikidata schema [12]. The Wikidata schema contains different properties that link an artwork to its content, such as `wdt:P921` (main subject) or `wdt:P180`

[8] `http://purl.org/emmedi/hico`

(depicts). Moreover, Wikidata allows adding qualifiers to the statements made with the property `wdt:P180` to address specific aspects of elements depicted in the artworks (i.e. their symbolic meaning, qualities).[9] Compared to ICON, Wikidata does not explicitly distinguish between levels of interpretations, linking elements of the first and second level to the artwork with the same property. This limits the possibility of art historian driven research questions to be answered with the Wikidata model. Finally, compared to our work, Wikidata does not include potential intrinsic meaning of artworks provided by the third level of interpretation. A more in-depth analysis of the qualifiers and comparison between Wikidata and ICON is present in Section 3.6.3.

The scope of iconography has some overlaps with the domain of narratology in the description of the plot for what concerns the represented characters and their actions. A good wealth of studies concerns the semantic modelling of the topic[10] [107, 108], among which some focus on the narrative representation in visual images [109, 110]. Even though they do not describe the iconographical subject with the necessary granularity required by an iconographical study, they provide a solution for organising the common archetypical knowledge of stories, events and characters participating in them [109].

3.2.4 Controlled vocabularies for iconography

Controlled vocabularies and taxonomies of art and culture, despite not being ontologies, are essential for standardizing the reference to elements that belong to the first, second, or third level of interpretation. Iconclass [74] is a classification system that mostly deals with iconographical subjects. The Getty art and architecture thesaurus [111] provides permanent identifiers for people, concepts, places that might be contained in artworks. The two aforementioned taxonomies cover a wide amount of information,[11] but other exists that were created ad-hoc for museums or cultural institutions, such as Rijksmuseum's thesaurus [112].

[9] `https://www.wikidata.org/wiki/Property:P4878`

[10] For the modelling of fictional entities from a philosophical perspective, see `https://plato.stanford.edu/entries/fictional-entities/`

[11] Although Iconclass admittedly is based on Eurocentric subjects

These controlled vocabularies and taxonomies can foster the interoperability between different knowledge graphs that use them, but, they do not provide statements regarding the type of element that is depicted in a work of art (whether it belongs to the first, second, or third level of interpretation). For this reason, it is essential for these characteristics to be modelled with a specific ontology.

3.3 Theoretical contribution

In this section, we illustrate Panofsky's theory of art interpretation [90, 1] introducing the theoretical aspects that were fundamental for the ontology design phase. During the act of interpretation of an artwork, the formal aspects, such as forms, colours and compositions, are perceived. When these formal aspects are interpreted as precise objects, the sphere of meanings is considered. According to Panofsky, there are different types of meaning that can be interpreted in an artwork, subdivided in three layers. The depth to which the artwork can be understood depends on the background knowledge of the observer: the more he has knowledge about the artist, stylistic conventions, cultural context of him/her/them, the more the interpretation at each level is correct, including more profound insights on cultural meanings.

The first layer, namely the pre-iconographical description, requires the knowledge of the representational conventions to allow a correct recognition of factual (e.g. objects, people, actions) and emotional meanings, namely primary or natural subjects. In detail, this description is achieved by the recognition of pure forms (i.e. combinations of forms and colours) as carriers of primary subjects. Pure forms such recognized are called ”artistic motifs”, and their combinations are ”compositions”. An enumeration of the recognition of artistic motifs constitutes a pre-iconographical interpretation of the artwork [1, p. 28].

If the observer is familiar with the literary sources known by the artist, then the subjects already identified at level 1, viz. the artistic motifs or compositions, can be recognized at the second level by the combination of them with concepts and themes, obtaining for example characters (e.g. Venus), personifications (e.g.

Table 3.1: Levels of interpretation according to Panofsky [1]

Level	Type	Subject identified	Recognized elements	Example: reading of Leonardo's Ultima Cena	Necessary background
1	Pre-iconographical description	Natural or primary subject, namely factual and expressional meaning	Artistic motifs and their combinations (compositions): pure forms recognized as carriers of primary meanings	13 people, table, food, dishes (all factual meanings) act of talking (expressional meaning)	practical experience
2	Iconographical description	Secondary or conventional subject	Images and their combinations (invenzioni, i.e. stories and allegories): artistic motifs recognized as carriers of a secondary meaning	The last Supper, Jesus, Apostles	literary sources describing themes and conepts familar to the artist
3	Iconological interpretation	Intrinsic meaning or content	Symbolic values: artistic motifs, images, stories and allegories are recognized as manifestations of underlying principles of a cultural context	Manifestation of Leonardo's and Renaissance particular attitude	familiarity with cultural phenomena, tendencies, attitudes

Virtue), or events (e.g. the Battle of Cascina). The artistic motifs such recognized are called images or *Invenzioni*, namely the term used by ancient theorists to identify stories and allegories. Allegories are defined in opposition to stories as "combinations of personifications and/or symbols", although there are many intermediate possibilities between them [1, p. 29, note 1].

Finally, by knowing and understanding the cultural and societal aspects of the artist's time, it is possible to read the artwork and the subjects identified at the previous levels as symptoms of the contemporary society, of the artist's beliefs and personality or as the expression of meanings voluntarily inserted.

The scholar highlights that the first two levels are a description of facts and are under the domain of iconography, whereas the last level is in the domain of iconology, which is a synthetic intuition rather than a description. Table 3.1 resumes the synoptic table in [1, pp. 40-41] integrating it with further explanation of concepts implemented in the ontology modeling and by adding a practical example.

Although some following scholars made some variations of the model,[12] the subdivision of the interpretation in levels is generally accepted. In detail, we highlight that some scholars put the attention on relevant aspects that we considered during the modeling. Van Straten [38] highlights the difference between intentional

[12] We refer to [92] for a further comparison between the major theories

and unintentional meanings by dividing the third level in two layers. In this way, he recognizes that some more profound meanings are voluntarily expressed by the artist (e.g. the concept of ”good wishes” that the artist wants to express in an artwork made for a wedding occasion) and more unconscious, cultural meanings. Another relevant addition is made by Imdahl in [93]. He underlines that the iconic sense of the image should not be ignored, since it is the primary means through which visual arts communicate. For example, the disposition of figures in the space can provide insights on their relationships, actions or in expressional meanings.

Furthermore, the preliminary studies conducted in [92], which considered approximately 50 articles of the major scholars of iconography and iconology, collected in [53, 93, 42, 90, 16, 15, 38], highlight important features that may be involved in an iconographical-iconological interpretation that should not be ignored. Indeed, from the bottom-up analysis emerged that the following aspects may be relevant for the supporting of the third level-meaning, namely i) the direct citation of visual patterns from other artworks ii) the dependency of certain iconographies from specific sources, iii) the role of style, iv) the fact that a cultural meaning generally involves more than one artwork, and v) the fact that scholars often extend claims by other scholars.

3.4 Requirements

Based on the iconographical and iconological literature analysis described in Section 3.3, we produced an ontology requirements specification document (ORSD) following the methodology in [113]. Its terminology was mainly selected from Panofsky’s theory [1]. The output document is described in Tables 3.2 and 3.3.

The purpose of the ICON ontology is to formally represent the domain of knowledge of iconology and iconography with a high granularity level, to allow specific quantitative analysis that can be interesting for domain experts. It is intended to be used by i) cultural institutions willing to publish their data about artwork content in linked data, ii) art historians interested in answering iconographical and iconological research questions in a quantitative way, and iii) developers who plan to use computer vision to associate recognized elements to portions of artworks. Therefore, the ontology aims at being implemented in different contexts, meeting the needs of different types of users. We use the OWL2 format to make the ontology available and reusable.

Therefore, the main non-functional requirement[13] is the reuse and alignment to the standards shared across the community to allow of reusability. Furthermore, the CQs formulated for the functional requirements aim at expressing the various aspects of the iconographical-iconological approach described in section 3.3. We summarise the main themes that can be extracted from the requirements listed in Table 3.3 as follows:

1. The identification of subjects at each level of interpretation needs to be included.

2. The variations of iconographical subjects (e.g. Cupid represented with a bandage and griffon talons, rather than only with traditional attributes, viz. wings and arrows[90]) must be described.

3. The symbolic and cultural meanings attributed to each subject must be included.

In addition, relevant characteristics of the approach are considered, namely:

4. The attribution needs to be subjective.

5. The sources used by the scholar to state its claim need to be present.

6. The clear distinction between the subjects described at a general level (i.e. the background knowledge necessary for iconographical descriptions cited in table 3.1, found in standard vocabularies, describing e.g. Cupid as a ”child with wings and arrows”) and their specific manifestation in a single artwork (e.g. Cupid with griffon talons) needs to be done to allow us to describe variations.

[13] see slot 6a of the ORSD in [113]

7. The ontology must allow the integration of one claim within the agreeing claims quoted by the art historian as a source of shared and accepted knowledge.

8. The ontology must allow to gather sets of agreeing recognitions made in a coherent situation (e.g. a scholar making an interpretation in a specific paper expanding on other scholars' interpretations, therefore including their claims in his own), that may gather the interdependent recognition made at different levels (e.g. a scholar recognizes the level 2 subject ”Cupid”, since he recognized at level 1 the subjects ”child”, ”arrows”, ”wings”).

9. The description of the iconic language of the visual artwork needs to be included, e.g. the relative position of objects and the structure in which they are organized.

10. At least a description of style should be included.

As Panofsky's theory is considered a representative formalization of the iconological approach, we take the majority of the ontology's terms from his theory. Therefore, we decided to populate the pre-glossary of terms (i.e. the relevant terms extracted by the CQs and their answers) contained in Table 3.3, point 7, by extracting the terms which are answering to CQs directly from the definition of his theory. The number following each word indicates its frequency in the selected article[14], in which Panofsky's theory is fully illustrated.

3.5 Ontology Design

The ICON ontology[15] was designed following the SAMOD[63] and eXtreme Design[64] methodologies. SAMOD is an agile methodology that focuses on the application of small iterative steps to model parts of an ontology. Each step is individually documented and combines motivating scenarios that derive from general domain descriptions with data-centric examples of descriptions formalized with the

[14]For this analysis, we referred to the article “Iconography and Iconology: an introduction to the study of Renaissance art” published in [1], since it is the last published revised version.

[15]The ontology is available at `https://w3id.org/icon/ontology/`

Table 3.2: Description of requirements 1-5 according to ORSD methodology

	Ontology Requirements Specification Document (requirements 1-5)
1	Purpose
	The ontology purpose is to formally represent the domain of knowledge of iconology and iconography with a high granularity level, to provide art historians and cultural institutions a way for expressing complex art subjects and meanings, claims about their interpretations and interlinking among them.
2	Scope
	The ontology focuses only on the iconographical and iconological interpretations that can be made about the content and meaning of visual artworks. The ontology has a high level of granularity, to correctly represent i) specific data important for domain experts and ii) the subjectivity of each claim.
3	Implementation language
	The ontology has to be implemented in OWL2 language.
4	Intended End-Users
	User 1. Cultural institutions that have a detailed bibliography about artworks looking for a formal language to express it User 2. Art history scholars with complex research questions only answerable with quantitative methods or wanting to express the data they collected in a formal language User 3. Developers using computer vision to associate recognized elements to portions of artworks
5	Intended Uses
	Use 1. Publish structured data about artworks interpretations online and integrate them with existing data so as to enhance the query potentiality of the cultural institutions' data Use 2. Conduct specific and detailed quantitative analysis to answer research questions in the domain research field Use 3. Provide a semantic structure for knowledge extraction

ontology. We re-use SAMOD methodology for the main part of the design, as we adopt the iteration-like structure and its outputs. In fact, the design process was divided into 4 SAMOD iterations, each dedicated to a particular aspect of the ontology. Each iteration contains a motivating scenario, a glossary with the definition of specific terms, a self-contained ontology prototype that contains only classes and properties relative to the corresponding iteration (with no references to external ontologies), the alignments to external ontologies, the aligned prototype, a series of competency questions formulated both in natural language and SPARQL (referring to the aligned prototype) and a Jupyter notebook that contains unity tests. All the competency questions were tested on real interpretations by Panofsky [42] expressed using the ontology schema. For a more detailed description of the test dataset, see Section 3.6.1. eXtreme Design is another agile methodology that divides the development of an ontology through iterations, but focuses on the re-use of ontology design patterns (ODP). In fact, the methodology tries to solve the "local problems" included in the so-called "local space", or the modelling issues related to the specific ontology that is being developed, with the re-use of modelling patterns that come from the "solution space", such as the ODP. We specifically adopted this

Table 3.3: Description of requirements 6-7 according to ORSD methodology

	Reference Ontology Requirements Specification Document (requirements 6-7)
6	**Ontology Requirements**
6.a	Non-functional Requirements
	NFR1. The ontology must be based on international standards and, when possible, directly reuse them
6.b	Functional Requirements: Groups of Competency Questions
	CQ level 1. CQ 1.1 What level 1 objects are represented in the artwork? CQ 1.2 What objects are natural elements, expressive characteristics or actions? CQ 1.3 What level 1 subjects are formally derived or copied from other artworks level 1 subjects? CQ 1.4 In What compositional structure are the objects organized (e.g. pyramidal arrangement)? CQ level 2. CQ 2.1 What level 2 subjects are identified in each artwork? CQ 2.2 Retrieve respectively all the characters, events, personifications, named objects, and places recognized at level 2. CQ 2.3 In which story or allegory are involved the depicted subjects? CQ 2.4 Do the level 2 subjects have a symbolic meaning? CQ 2.5 which is the object that allows the character recognition at level 2, i.e. the character's attribute? CQ 2.6 What are the representative variations at level 1 of the same level 2 subject in different artworks? CQ 2.7 What are the level 1 variations of the same level 2 subject involved in different stories or allegories? CQ 2.8 What are the level 1 subjects having multiple interpretations at level 2? Which of them are made in the same descriptive situation? CQ level 3. CQ 3.1 What meanings are expressed by the artworks? CQ 3.2 What cultural phenomena are identified? CQ. 3.3 Who identified the cultural phenomena and on which basis? CQ 3.4 What are the artworks involved in the same cultural phenomenon? CQ 3.5 To which specific subjects at level 1 and 2 does the level 3 recognition refers? CQ 3.6 What are the artworks having both a common cultural phenomenon and a common level 2 subject? General CQ. CQ 0.1 What are the sources supporting each subject recognition at each level? CQ 0.2 What is the person responsible for every recognition at each level? CQ 0.3 What are the artworks that are only interpreted on a pre-iconographical level? CQ 0.4 What artworks are interpreted on an iconological level but not on an iconographic one? CQ 0.5 What are the recognitions supporting another one? Of which type are they? CQ 0.6 What artworks or parts of it have a style associated?
7	Pre-Glossary of Terms (Term, Frequency in studied documents)
	Motif(s) 44; Story(ies), 37; Image, 26; Interpretation, 22; Natural, 16; Iconography, 15; Iconographical, 14; Allegory(ies), 11; Intrinsic meaning, 9; Preiconographical Description 9; Iconographical Analysis, 8; Composition, 7; Expressional, 7; Artistic motifs, 5; Factual, 5; Iconological Interpretation, 4; Iconology, 4; Invenzioni, 1

methodology when dealing with the re-use of ontology design patterns that were specialized in the context of our domain. The following paragraphs describe i) each SAMOD iteration, ii) the specialization of ODP to facilitate some modelling issues, and iii) the refactoring of some classes and property through alignment to relevant ontologies.

3.5.1 First design iteration: Recognitions

As explained in Section 3.3, works of art can be analysed through different layers of interpretations that depend on recognitions. A recognition, in the context of this ontology, is an interpretation act made by an agent (or interpreter, which can be a biological or electronic being) that links works of art to something related to their content. From a conceptual perspective, it is a mental entity reflecting the agent's subjective point of view. From a technical viewpoint, it is an n-ary predicate that cannot be modelled using OWL due to expressivity limitations; therefore, it was turned into an n-ary relationship class.[16] Coherent recognitions on the same artwork are collected and documented by interpretation descriptions (requirement 8, section 3.4).[17] In this iteration, we conceptualize the elements that revolve around recognitions. From the n-ary relationship class `icon:Recognition`, several properties were designed (or reused from existing ontologies) to link it to its interpreter(s) (or agents), the artwork that is being interpreted, supporting sources for the recognitions. In particular, the `aboutWorkOfArt` property links the recognition to the artwork (`Artwork` class). Then the `dul:includesAgent` property (from DOLCE[114]) links the recognition to the agent who performed it (requirement 4, section 3.4). The class `InterpretationDescription` is linked to (one or many) `Recogntion` class(es) that comply with it through several properties according to the type of the recognition, namely: `isCompliantWithPreiconogra-`

[16] More observation on the matter can be found in subsection 3.5.5.

[17] The distinction between the mental entity of the recognition and the document entity of the description is necessary not only because a description can contain multiple recognitions, but also as a way of separating through coherent criteria different recognitions made on the same artwork (even by the same interpreter). For example, a cultural institution such as a museum might decide to describe an artwork by collecting only some recognitions made by one interpreter and adding more recognitions made by different interpreters to finalize their description.

phicalRecognition for pre-iconographical recognitions and formal motif recognitions,[18] isCompliantWithIconographicalRecognition for iconographical recognitions, isCompliantWithIconologicalRecognition for iconological recognitions. The CiTO[115] properties cito:citesForInformation and cito:citeAsEvidence can be linked to a Recognition class to provide sources or other information that support a recognition (requirement 5, Section 3.4). Finally, a recognition can also be used to support further recognitions made on the same artwork or another one. For example, Panofsky recognizes that the figure of Chastity sculpted by Giovanni Pisano on the Pulpit of Pisa cathedral is represented with the same appearance of the nude classical iconography of Venus Pudica (formal motif recognition)[19]. This interpretation provides support to the third-level recognition of the characteristics of the Proto-Renaissance movement in the cultural context of the Medieval Tuscany [42, p 157]. To express this using our ontology, the property cito:givesSupportTo can link the supporting recognition to another one (requirement 7, Section 3.4). These elements are also the object of interest of the general competency questions (see Table 3.3, Q0.1 to Q0.5).

Depending on the level of interpretation presented in Table 3.1, four Recognition subclasses have been defined:

- PreiconographicalRecognition (level 1)
- FormalMotifRecognition (level 1)
- IconographicalRecognition (level 2)
- IconologicalRecognition (level 3)

Recognitions at each level of interpretation may be based on the results of the recognition at one of the previous levels. Therefore, they can be linked to-

[18]Both a pre-iconographical recognition and a formal motif recognition are described in the first level of pre-iconographical interpretations of Panofsky, so we use the same property to link them to the InterpretationDescription class

[19]figure available at https://commons.wikimedia.org/wiki/File:Giovanni_pisano,_pulpito_del_duomo_di_pisa,_1302-11,_carit%C3%A0_e_virt%C3%B9_cardinali_03.JPG

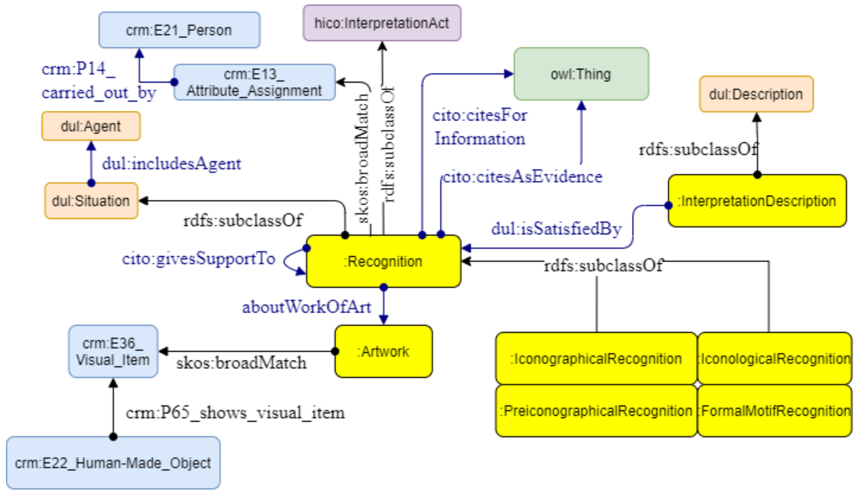

Figure 3.1: ICON ontology classes and properties linked to recognitions

gether but ultimately are modelled as independent of one another. This choice is made since i) the describer may not have available the descriptions of the lower level(s), ii) the corresponding subjects in the other levels may not be relevant for the recognition, iii) it may be possible that a level 3 recognition (i.e. an `IconologicalRecognition`) is linked to level 1 subjects rather than level 2 ones (e.g. iconological interpretations of a landscape painting, which may not have level 2 subjects [1]).

These classes and their specific usage will be further described in the following subparagraphs. Figure 3.1 shows a rendering of the classes and properties of this iteration.

3.5.2 Second design iteration: Pre-iconographical Recognitions (level 1)

In this iteration, we model the recognitions that happen on a Pre-iconographical level. In this level, an interpreter recognizes artistic motifs present in the artwork, and associates to them i) natural objects (a tree, a man, a sword) without identifying specific individuals from those classes (tree of life, Saint Joseph, Excalibur) which are recognized in level 2 (Section 3.5.3), ii) in the form of expressional meanings [20] (emotions of the depicted elements), iii) qualities about these elements (size,

[20] according to Panofsky, the expressional meanings are the subjects that can be interpreted at the first level of recognition through empathy [1, p. 27]

colour, positions), iv) performed actions (see table 3.1 in section 3.3). Assuming that the agent doing the interpretation act might also be a computer, as in the case of the results of object detection through computer vision, we give the possibility to express coordinates of the portion of the image of the artworks where these elements are detected. Furthermore, these coordinates can be expressed using IIIF URIs[116] that point to a specific portion of the work of art. A series of artistic motifs can be grouped together in a composition that can have a compositional structure[21] (e.g. pyramidal). Additionally, an interpreter might recognize similarities between artistic motifs present in a work of art with other artistic motifs of another work of art, recognizing a prototypical artistic motif or composition that is reused in another artwork. For example, the level 1 description of Pisano's figure of Chastity cited above is linked through a formal motif recognition to the level 1 description of Venus Pudica, from which its appearance is derived (i.e. a nude woman covering herself with her arms). Artistic Motifs and compositions are linked to the class `PreiconographicalRecognition` respectively through the properties `recognizedArtisticMotif` and `recognizedComposition`. Only one artistic motif or composition can be linked to a recognition. Compositions are linked to the artistic motifs that take part in them through the `hasPart` property. If the artistic motif refers to a natural object or action with a factual meaning, it is linked to the classes `NaturalElement` or `Action` through the property `hasFactualMeaning`. Otherwise, if what is recognized in the artistic motif is an expressional meaning, the property that links it to expressional meanings is `hasExpressionalMeaning`. If actions, expressional meanings, or natural elements have some specific quality that needs to be highlighted, from the artistic motif the qualities are expressed with the DOLCE `hasQuality` property. When the pre-iconographical recognition is performed by a computer with an object detection algorithm, or when a IIIF URI is provided, it is possible to associate not only the detected objects, but also the coordinates of the image in which they are found. Coordinates of the detected object can be

[21] The compositional structure conceptualization is derived from Imdahl's theory [93]

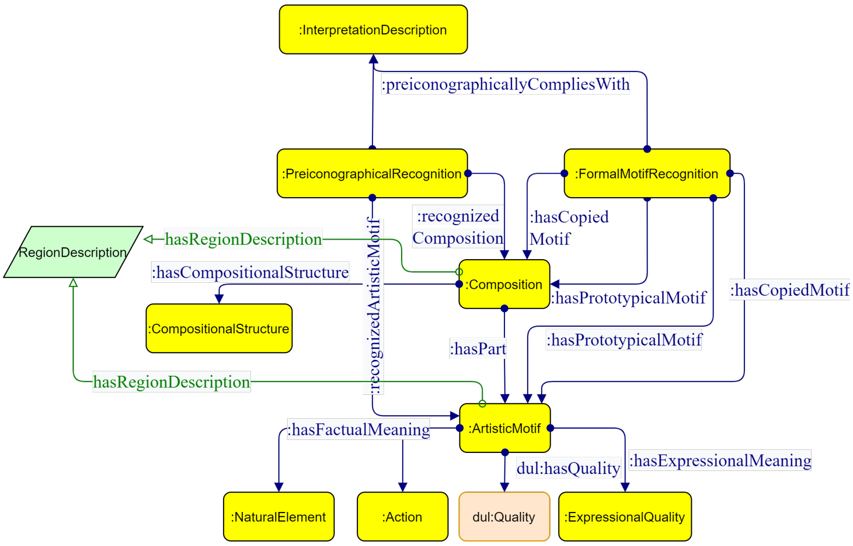

Figure 3.2: ICON ontology classes and properties describing the pre-iconographical level of interpretation (level 1)

expressed through the data property `hasRegionDescription` that has the `ArtisticMotif` or `Composition` classes as the domain. As mentioned above, the use of IIIF URIs for the format of this data property is also welcomed. The `FormalMotifRecognition` class links the prototypical motif to the copied motif, respectively, using the `hasPrototypicalMotif` and `hasCopiedMotif` properties. Finally, all the coherent formal motif recognitions and pre-iconographical recognitions that take part in an interpretation about a work of art, can be linked to an `InterpretationDescription` class, through the property `preiconographicallyCompliesWith`. Figure 3.2 shows a graphical rendering of the classes and properties used in this interpretation level.

3.5.3 Third design iteration: Iconographical Recognitions (level 2)

In this third iteration, we focus on the Panofsky's second level of art interpretation: the Iconographical interpretation. In this level, the interpreter recognises images and invenzioni[22] in an artwork. An image represents the subject depicted as a manifestation in the specific artwork taken into account. It is then linked to second

[22]Invenzione is an Italian word used by Panofsky as an umbrella term for allegories and stories [1, p. 59]

level subjects, which are characters, places, events, named objects[23], symbols, personifications, identifying iconographies from an abstract and general point of view. This distinction between the general subject level (i.e. characters, symbols) and the artwork-specific one (image) is functional to identify the variants of a subject in relation to the specific context (i.e. Thor as represented in a specific painting may differ from its common one). An invenzione, instead, is the subject matter represented by the combination of general subjects linked to the single images recognized.[24]. For example, in an artwork you might recognize three images: the first refers to the general subject of Mary, the second refers to the general subject of Angel Gabriel, and the third refers to the general subject of the Holy Dove. The combination between the general subject of Mary, Angel Gabriel and the Holy Dove is the Annunciation, which, in our ontology terms, would be considered the invenzione. The same invenzione could be present in multiple artworks, but each artwork maintains its uniqueness by having different images. The classes `Story` and `Allegory` are subclasses of the class `Invenzione`. Stories are more likely to contain characters, named objects, places and events, whereas allegories are more likely to contain symbols and personifications. We give the possibility to express symbols as just symbolic meanings recognized, or, for a more thorough description, as Simulations (see section 3.5.5). The classes `Image` and `Invenzione` are linked to the class `IconographicalRecognition` through the respective properties `recognizedImage` and `recognizedInvenzione` (one image or invenzione per recognition). The artistic motif belonging to a pre-iconographical level that refers to the recognition of an image can be linked to it with the property `refersToArtisticMotif` (i.e. the recognition of the image that represents

[23] A named object is a non-living unique element that is often used as an attribute for the recognition of specific characters (Thor's hammer.)

[24] This definition slightly differs from the Panofsky's one: while he describes an invenzione as a form expressing the subject represented by the combination of the single images recognized, we consider it as an individual belonging to the "sphere of secondary or conventional subject matter, viz., the world of specific themes or concepts manifested in images" [1]. This decision is motivated by the fact that, the description of real case studies in the modeling phase, emerged that it would be redundant to general stories and allegories both at the conventional level and their manifestation in the specific artwork. Their variations are already clear, considering which subjects are part of them in each particular case.

Mary Magdalene can be linked to the artistic motif that has the factual meaning of woman). This link is important to ensure that the connection between pre-iconographical elements and the respective iconographical subjects is preserved. If the artistic motif is the principal element that enabled a recognition of an image, then it can be linked to that image through the property `hasRecAttribute` (i.e. the recognition identifying Cupid has recognizing attributes the artistic motifs linked respectively to "wings" and "arrows"). Images are linked to the general subject portrayed through specific properties according to the subject class. The property `hasCharacter` links an image to the class `Character`, likewise: `hasEvent` refers to the class `Event`, `hasPlace` refers to the class `Place`, `hasNamedObject` refers to `NamedObject`, `hasSymbol` refers to `Symbol` and finally, `hasPersonification` refers to `Personification`. The cited ICON classes represent second-level subjects represented in the fictional representational space, therefore including both real and fictional, non-existent subjects (e.g. Medusa, the Greek mythological character appearing in various media), in compliance with the modeling of subjects in narratology [107, 109]. An invenzione is linked to the elements that compose it through the property `composedOf`. Finally, multiple iconographic recognitions that take part in an interpretation of an artwork, are linked to the interpretation using the `iconographicallyCompliesWith` property. Figure 3.3 shows the classes and properties relative to this level of recognition.

3.5.4 Fourth design iteration: Iconological Recognitions (level 3)

Iconological interpretations (third level) focus on the recognitions of intrinsic meanings[25]. An intrinsic meaning links the whole artwork or some parts of it to a cultural phenomenon and a concept that defines it. The `IconologicalRecogniton`

[25] Even if Panofsky's terminology seems to prefer the term *symbolic values* for expressing the interpreted third level aspects of the artwork, we decided to adopt the term "intrinsic meanings" to avoid confusion with the second level symbols.

class is linked to the `IntrinsicMeaning` class [26] through the property `recognizedIntrinsicMeaning`. From there, the n-ary class `Intrinsic Meaning` can be linked to a specific composition, image or artistic motif that can be the focus of the intrinsic meaning through the properties `hasComposition`, `hasImage`, `hasArtisticMotif`. Then, it is linked to the expressed concept through the property `recognizedConcept`. For the range of this property, we reuse the Dolce class
`SocialObject` because there was no need to create an ad-hoc class for this element.[27] Additionally, since an Intrinsic Meaning can also reflect some cultural phenomena, it is linked to the class `CulturalPhenomenon` through the property `recognizedCulturalPhenomenon`. Currently, `CulturalPhenomenon` has 4 subclasses, which specify the type of cultural phenomenon, namely `Attitude`, `Belief`, `CulturalValue`, and `Tendency`. These terms are taken from Panofsky's vocabulary in the description of the third level of artistic interpretation. [28] Finally, all the iconological recognitions that take part in an interpretation made on an artwork are linked to it with the property `iconologicallyCompliesWith`. A graphical rendering of this fourth iteration, representing the third level of the interpretation, can be found in figure 3.4.

3.5.5 Refactoring: reuse and alignment to relevant ontologies and ontology design patterns

To promote ontology interoperability and reusability, we connect to several external ontologies through means of alignments and reuse. We present our alignments

[26] Compared to factual and expressional meanings expressed through a property, an intrinsic meaning needed an n-ary class for representation because of expressivity reasons (owl does not support n-ary predicates).

[27] The concepts, ideas, abstract elements that are linked to intrinsic meanings on an iconological level are very broad [42]. Therefore, we decided to reuse this Dolce class (Social Object) which conceptualises a broad set of possible entities [114]

[28] Although these subclasses could be formally associated with mental entities just as recognitions, they differ in their function. Recognition are modelled on a meta level of the interpretation, as they are used to describe a recognition act made by an interpreter. These subclasses are meant to be the object of the interpretation, as they are associated with the recognition of an intrinsic meaning of the artwork itself. As it will be discussed in the final section, further work will be dedicated to a more thorough description of cultural phenomena and their subclasses.

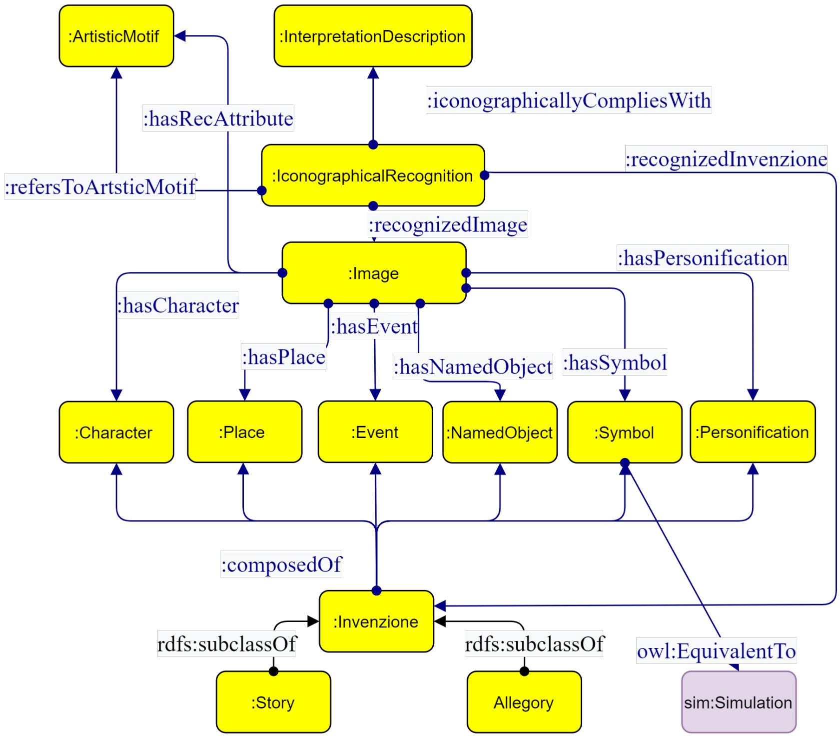

Figure 3.3: ICON ontology classes and properties describing the iconographical level of interpretation (level 2)

and reuse by following guidelines proposed by the state of the art [117, 118]. Our ontology selection for reuse and alignment was guided by different principles: (i) standardization for CIDOC-CRM [75] and FRBRoo[119] because they are considered standard frameworks in the domain, (ii) cognitive and formal analysis for the choice of DOLCE foundational ontology [114][120] in its OWL version (DOLCE Zero), Simulation Ontology [33], VIR [76] HiCO [103] and CiTO[115] as all of them offer design solutions to the competency questions defined from the requirements in section 3.4.

Due to the complexity of the field, the number of ontologies to be reused, and the heterogeneous domains from which they come, we adopted a hybrid reuse approach [117], which, depending on the specific cases explained below, considers either reusing directly the classes and properties of the aforementioned ontologies (either by importing the whole ontology or parts of it), or (indirect reuse) using them as fully extensional ontology patterns, or just as intensional patterns.

Extensional reuse happens when classes or properties of an ontology $O1$

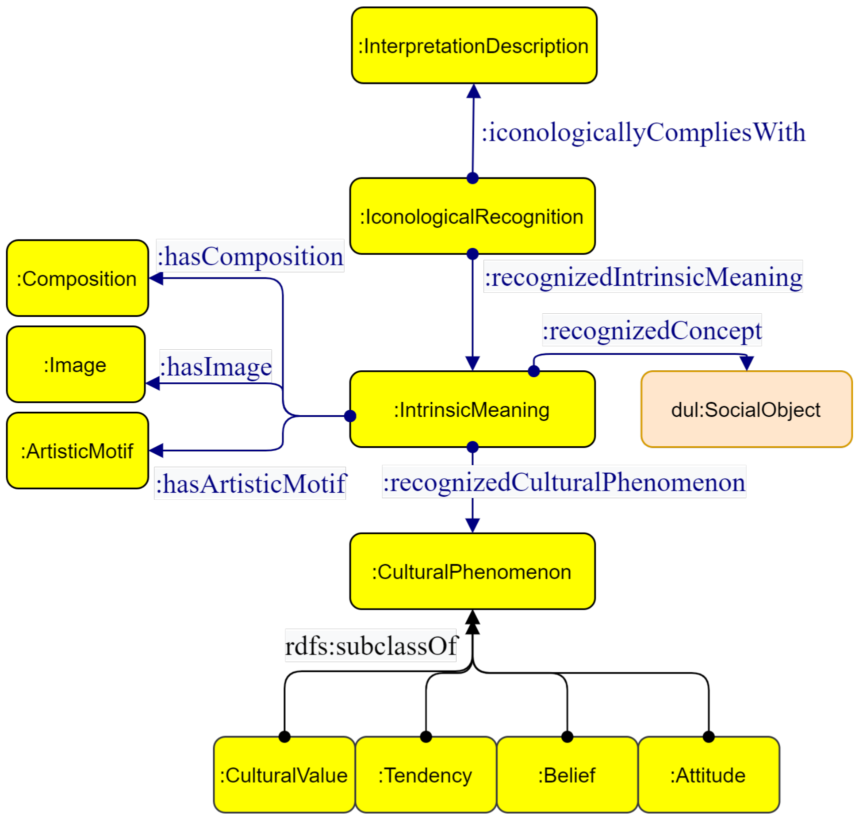

Figure 3.4: ICON ontology classes and properties describing the iconological level of interpretation (level 3)

are logically *aligned* to an external ontology $O2$, which we want to reuse with its full-fledged semantics, because it is compatible, desirable, or necessary. For example, if we extensionally align a $O1$ class `Organisation` to a $O2$ class `dul:SocialObject`, we intend to inherit the semantics of DOLCE's social objects, e.g., that they are not physical.

On the contrary, we use parts of an external ontology $O3$ as purely intensional constructs when we want a limited interoperability, which does not include accepting in $O1$ all the semantics provided in $O3$, because it may be partly incompatible. For example, we may intensionally align a $O1$ class `Image` to a $O3$ class `crm:E36_Visual_Item` because we might not want to inherit the axiom stating that `crm:E36_Visual_Item` is a subclass of `crm:E89_Propositional_Object`.

In order to implement this distinction, indirect reuse is designed using different mapping properties, according to the semantics they provide, and its impact into the resulting reasoning. We have used RDFS (rdfs:subPropertyOf, rdfs:subClassOf) and OWL (owl:equivalentTo) logical properties when we want the alignments to provide first-order extension to ICON schema and data, while we have used SKOS skos:broadMatch, skos:related, and skos:closeMatch for purely intensional mapping, which can be used at query time to integrate data represented with ontologies that may harm the logical integrity of ICON knowledge.

Among the reused ontologies, we have used an intensional (or ”terminological”) mapping for CIDOC, VIR and FRBRoo, because we have noticed potential problems when reasoning is jointly made with both the axioms from ICON, and from those ontologies. For example, a full extensional alignment of the class icon:Image as rdfs:subclassOf crm:E36_Visual_Item would make an automated reasoner infer that icon:Image rdfs:subclassOf crm:E89_PropositionalObject, which is not defendable, since propositional entities typically exclude visual, musical, or other information modalities. In other words, CIDOC contains here a debatable assumption, which should be ignored when reusing data that use CIDOC as their schema. Now, if we use a purely intensional mapping: icon:Image skos:broadMatch crm:E36_Visual_Item, we make a commitment that can be discussed, and the triple can be used to make SPARQL-based data integration, but we will not get the inference that images are propositions.

In this section, we give a thematic overview of the classes and relations reused for satisfying a specific task, and we refer to the documentation (see section 3.7) for further details on the single alignments. Table 3.4 shows the direct reuse of external classes and properties in ICON, and Table 3.5 shows the indirect alignments.

3.5.5.1 Recognitions as situations

According to the guidelines of eXtreme Design [64], we defined our local problem (in our local space) as the expression of recognitions through n-ary relationship classes due to the inability of expressing n-ary predicates in OWL. As explained in the previous paragraphs, our conceptualization of the icon:Recognition class

Table 3.4: Direct Reuse of Classes and Properties in ICON

External Element	Type	Ontology	Usage
Agent	Class	DOLCE	represents interpreter (with dul:includesAgent property)
Quality	Class	DOLCE	represents recognized quality of artistic motifs (linked from icon:ArtisticMotif with the dul:hasQuality property)
SocialObject	Class	DOLCE	used as the symbolic meaning linked to an icon:IntrinsicMeaning class through the property icon:recognizedConcept
includesAgent	Property	DOLCE	links icon:Recognition to the agent (dul:Agent) performing it (also a non-human agent)
givesSupportTo	Property	CiTO	links icon:Recognition to another icon:Recognition that supports it
citesAsEvidence	Property	CiTO	links icon:Recognition to an entity (owl:thing) that is the evidence on which the recognition is based
citesForInformation	Property	CiTO	links icon:Recognition to an entity (owl:thing) that is the source in which the recognition is found (e.g. a bibliographical reference)

Table 3.5: Indirect Reuse of Classes and Properties in ICON: icon element - type - external element

External Element	Type	Ontology	ICON Element	Type of alignment
E5_Event	Class	CIDOC	Action	skos:broadMatch
E36_Visual_Item	Class	CIDOC	Artwork; ArtisticMotif; Composition; Image; IntrisicMeaning	skos:broadMatch
InformationObject	Class	DOLCE	Artwork	rdfs:subClassOf
E13_Attribute_Assignment	Class	CIDOC	Recognition	skos:broadMatch
InterpretationAct	Class	HiCO	Recognition	rdfs:subClassOf
Situation	Class	DOLCE	Recognition	rdfs:subClassOf
Description	Class	DOLCE	InterpretationDescription	rdfs:subClassOf
Simulation	Class	Simulation Ontology	Symbol	owl:equivalentTo
F38	Class	FRBRoo	Character	skos:broadMatch
E1_CRM_Entity	Class	CIDOC	ExpressionalQuality; Invenzione; NaturalElement	skos:broadMatch
E31_Document	Class	CIDOC	InterpretationDescription	skos:broadMatch
E90_Symbolic_Object	Class	CIDOC	Symbol	skos:broadMatch
IC11_Personification	Class	VIR	Personification	skos:closeMatch
Subject	Class	ArCo	Character; Personification; Event; NamedObject; Place; Symbol; Invenzione; Action; NaturalElement; ExpressionalQuality	rdfs:subClassOf
E89_Propositional_Object	Class	CIDOC	Event; NamedObject; Place	skos:related
P138_represents	Property	CIDOC	hasCharacter; hasEvent; hasExpressionalMeaning; hasFactualMeaning; hasNamedObject; hasPersonification; hasPlace; hasSymbol	skos:broadMatch
P140_assigned_attribute_to	Property	CIDOC	associatedForm; refersToArtisticMotif	skos:broadMatch
P141_assiged	Property	CIDOC	recognizedArtisticMotif; recognizedComposition; recognizedImage; recognizedInvenzione	skos:broadMatch
P106_is_composed_of	Property	CIDOC	hasPart	skos:broadMatch
K4_is_visual_prototype_of	Property	VIR	hasCopiedMotif; hasPrototypicalMotif	skos:broadMatch

required a good deal of contextual information (such as the agent performing it, what is recognized in the form of first, second, or third level of interpretation entities, the artwork). We have chosen the situation ontology design pattern[29] as a solution because it was designed to solve modelling issues regarding multiple contextual information connected to the same class in the form of n-ary relationships. The Situation ontology design pattern is reused via the import of DOLCE Ultra-

[29]http://www.ontologydesignpatterns.org/cp/owl/situation.owl

lite.[30] The n-ary relationship ODP is specialized by our `icon:Recognition` class, by making it a subclass of `dul:Situation`.

3.5.5.2 Interpretations as descriptions

The types of recognitions that we have presented are formalised as situations. In the Descriptions and Situations pattern[31] that is also formalized in DOLCE-Ultralite and DOLCE Zero, situations are loosely associated with *descriptions*, i.e., intensional entities that are used criteria for a situation to occur. The pattern is used in most domains: in medicine, a pathological situation depends on the diseases or syndromes that are used to interpret it, and which can have different probabilities to correspond to the actual situation; in Law, different norms may apply to a same legal case; in an everyday situation, an observer may interpret it differently according to her perspective, culture, or intention. In the iconographical and iconological domain, as also applied in the ArCo ontology network [13, 121], all recognitions and high-level interpretations are based on perception criteria, which make a rationale emerge, and eventually motivate a particular interpretation with respect to others. A description is therefore a conceptual entity, constituted by parameters, roles, tasks, etc. [102], which is satisfied by a situation when it involves entities that are classified by one of the parameters, roles, tasks, etc. that constitute a description. For example, the interpretation of a painting (Named *A*) such as "in this painting, there is a lion which symbolizes courage" is compliant with (i) a pre-iconographical recognition (recognizing an artistic motif as a carrier of the factual meaning of a lion), and (ii) an iconographical recognition (recognizing the image of the lion as the simulation of lion-courage). These recognitions would involve the recognizer, a source, the time period, as well as (potentially) additional iconographical aspects. Hence, we formalize this complex relation in terms of compliance: `InterpretationDescriptionPaintingA isCompliantWithPreiconographicalRecognition LionRecogni-`

[30] The aforementioned ODP was reused through the Dolce Ultralite ontology because the Situation class in DOLCE is linked to other classes that are reused in our ontology as well, such as the agent.

[31] `http://www.ontologydesignpatterns.org/cp/owl/descriptionandsituation.owl`

tionInA and isCompliantWithIconographicalRecognition LionCourageRecognitionInA. The property isCompliantWithPre-iconographicalRecognition is made a sub-property of dul:isSatisfiedBy which links a dul:Description (our Interpretation Description is subsumed under description) to one or more dul:Situation (our Recognition is subsumed under situation).

3.5.5.3 Describing artwork content

Since CIDOC CRM offers a way of describing the content of visual elements (crm:E36_Visual_Item, crm:P138_represents, crm:E1_Entity), we modeled the more specific elements recognized in each level of interpretation following this modeling principle as a guideline, and aligning our classes to CIDOC's ones through SKOS relations. As illustrated by figure 3.5, all the classes representing the general subject as represented in the contest of the artwork (i.e. Artistic Motif, Composition, Image, IntrinsicMeaning) are a skos:broadMatch of crm:E36. Furthermore, the recognized subjects at every level are a skos:broadMatch of crm:E1_Entity. In this way, the patterns linking the visual elements recognized in each level and the general subject can be seen as a specification of crm:P138. The identification of the artwork at the abstract level (Artwork, skos:broadMatch of crm:E36_Visual_Item) is intended to make the ontology compliant with the CIDOC-CRM modeling of cultural objects, whereas the alignment of Artwork with dul:InformationObject is motivated by the DOLCE conceptualization of Information Object that fits with our Artwork definition.

3.5.5.4 Interpretation details

The class Recognition has been aligned with classes from HiCO, CIDOC-CRM and DOLCE, as shown in figure 3.1. The class hico:InterpretationAct is intended to represent the context in which a recognition [32] is made, i.e. furnishing more information about the recognition to validate the claim. The recognition

[32] In the context of this study, since we align hico:InterpretationAct to Recognition, we refer to it with the term *recognition* for the clarity sake

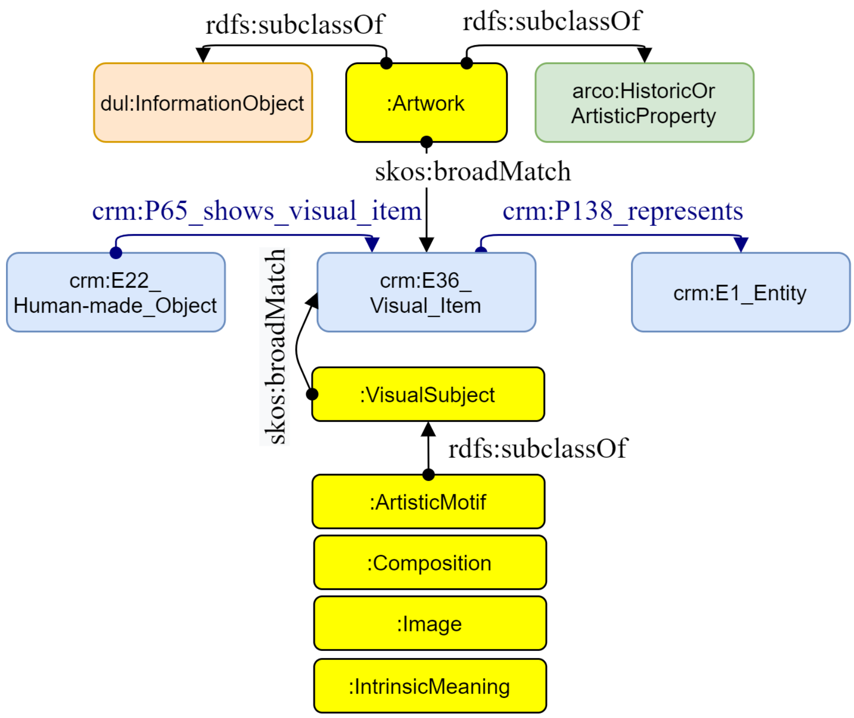

Figure 3.5: Alignment with other ontologies of the artwork description

such represented can be further specified by `hico:interpretationType` and `hico:Interpretation- Criterion`. For its purpose and formal structure, `icon:Recognition` was made a subclass of it. Since also the purpose expressed by `crm:E13_Attribute_Assignment` is of documenting the context in which an assertion about a cultural object was made, it is a `skos:broadMatch` of Recognition, since Recognition is more specific than the more generic concept expressed by `crm:E_13`. Furthermore, `crm:E13` is practically used as an n-ary relationship class linking two individuals through ancillary properties, `crm:P140`, `crm:P141`, identifying respectively the element to which the assignment is made and the assigned one. Therefore, when this logical structure is respected, the respective properties in the subclass of `icon:Recognition` are aligned to `crm:P140` and `crm:P141` through `skos:broadMatch`. Respectively, `RecognizedArtisticMotif` or `RecognizedComposition` at level 1, `RefersToArtisticMotif` and `RecognizedImage` or `RecognizedInvenzione` at level 2.

By the alignment with `hico:InterpretationAct`, and `dul:Situa-`

tion, the ontology not only enhances interoperability but also inherits a variety of means for expressing further detail about each recognition act at each level. For example, the possibility to express an agent using dul:Agent which includes both humans and computers, the time of the recognition using the includesTime property of dolce, the interpretation criterion[33] InterpretationDescription class and type (HiCO) allows the user to fully document the recognition acts, giving a comprehensive representation of the subjectivity of the recognition itself.

The Motif Recognition is developed as a specialization of the VIR property K4i_has_visual_prototype, documenting the use of a visual prototype for an image, enriching the latter by giving the possibility to add further details about the interpretation and to highlight the direct correspondence between the portions of the copying and copied artworks. For example, the derivation of the visual arrangement of the relief *Allegory of salvation* from the Roman relief depicting *Hercules and the Caledonian Boar* described by [90, p. 228; fig. 4-5, p. 231], can be further described by recognizing that the deer in the former is derived from the boar in the latter, and so on. Our property icon:hasPrototypicalMotif was aligned with skos:broadMatch to K4_is_visual_prototype_of.

3.5.5.5 Subjects

As it is the closer definition of artistic subject intended as an object represented by an artwork, we align all the subjects of the ontology to the ArCo's class arco:Subject. Specifically, we indirectly reuse arco:Subject by subsuming icon:Place, icon:NamedObject, icon:Character, icon:Event, icon:Symbol, icon:Personification, icon:Action, icon:NaturalElement, icon:ExpressionalQuality, icon:CulturalPhenomenon, icon:Invenzione, dul:SocialObject to it. In doing so, we also propose a new way of attributing a subject to a work of art compared to ArCo. In fact, while ArCo directly links a subject to the physical representation of the work of art, we link it to an interpretation made on the

[33] in the case of our ontology, interpretation criterion is linked to every single recognition, and not with the interpretations

Table 3.6: Distinction between subjects and visual subjects depending on Panofsky's levels of interpretation

Subject	**Visual Subject**	**Level**
Action	Artistic Motif or Composition	I
Natural Element	Artistic Motif or Composition	I
Expressional Quality	Artistic Motif or Composition	I
Character	Image	II
Event	Image	II
Named Object	Image	II
Place	Image	II
Personification	Image	II
Symbol	Image	II
Invenzione	A series of images	II
Social Object	Intrinsic Meaning	III
Cultural Phenomenon	Intrinsic Meaning	III

visual representation of what is in a physical work of art. By reusing the class `arco:Subject` and not its properties, which consider the physical artwork as the domain, we also avoid possible logical inconsistencies between ArCo's description of physical artefacts and ICON description of visual items. In contrast, the representation of the subjects as manifested in the artwork (i.e., Artistic Motifs, Compositions, Images, and Intrinsic Meanings), are subclass of `icon:VisualSubject`, which is disjoint with `arco:Subject` to underline their different nature and role. Table 3.6 displays the division between subjects and visual subjects according to the different iconographic and iconological levels.

3.5.5.6 Symbols

In an artistic interpretation, an interpreter might recognize a symbol of a specific cultural context in an artwork. For the modeling of symbols, we reuse the entire Simulation Ontology [33] presented in chapter 2. This ontology, designed to conceptualize cultural symbols, uses the n-ary `sim:Simulation` class to link together a symbol, expressed by the class `sim:Simulacrum`, its symbolic meaning, expressed by the `sim:RealityCounterpart` class, the cultural context in which the symbol denotes the symbolic meaning (`sim:Context`) and the source of the claim (`sim:Source`). We aligned our class `icon:Symbol` to the `sim:Simulation` class to allow the expression of symbolic meanings using the Simulation Ontology structure.

3.5.5.7 Expression of Style

The expression of style is an important feature related to iconographical and iconological studies (requirement 10, Section 3.4). Knowing the history of styles is, according to [42], a fundamental requirement for the correct interpretation of level 1 objects. Furthermore, as is evident, among the others, from Warburg's studies on *Pathosformeln* and *Nachleben der Antike*, forms of style are a subject of interest in iconology. Therefore, we reuse CIDOC-CRM to model it according to the solution adopted by linked.art project[34], using the structure

```
crm:E36_Visual_Item crm:P2_has_type <style_type>.
<style_type> crm:P2_has_type <aat:300015646>.
```

where the last object is the Getty AAT vocabulary term defining style. Although the property's domain is `crm:E1_Entity`, it is suggested to use it with `crm:E36_Visual_Item`, in compliance with linked.art directions. Even if we do not express `icon:VisualSubject` and `icon:Artwork` as direct subclasses of `crm:E36`, it is possible to reuse this pattern since ICON's classes are not disjoint with CIDOC's. Therefore, we reause this existing solution to model the requirement 10 of section 3.4 In this way, both the artwork itself and every portion of the image identified at each level can have its own style specification declared.

3.5.5.8 Citations, sources, evidences

As shown in figure 3.1, the CiTO ontology is directly reused to represent the source (`cito:citesForInformation`) from which the Recognition is extracted, the evidence (`cito:citesAs- Evidence`) on which it is based and the supporting (`cito:givesSupportTo`) between acts of recognition. This representation is

[34] `https://linked.art/model/object/aboutness/#style-classification`

fundamental to encourage a documented description of the recognition, its reference and support.

3.5.6 Logical constraints in artistic interpretations

Artistic interpretations are, by definition, subjective. Nevertheless, the characteristics and relationships that surround recognized elements can be subject to logical constraints. Each recognition is made on exactly one artwork, involves exactly one agent, and different recognitions require adequate elements. A pre-iconographical recognition targets as recognized elements, either artistic motifs or compositions. A formal motif recognition instead deals with a prototypical motif and a copied motif, both of them can be either an artistic motif or a composition. Then, an iconographical recognition recognizes exactly one image or one invenzione. Finally, an iconological recognition refers to an intrinsic meaning. Regarding the different recognized elements, the recognition of an artistic motif, in Panofsky terms (and thus in our ontology) implies that the interpreter associates either a natural or expressional meaning to a portion of the artwork. At the same time, the recognition of an image implies the presence of either a character, event, named object, symbol, personification, specific place in the artwork. Furthermore, images, artistic motifs and intrinsic meanings cannot be instantiated without having a recognition that addresses them. To ensure this, we added several restrictions. For example, images must be linked to exactly 1 recognition through the property `isIconographicallyRecognizedBy`. The difference between stories and allegories is that the former generally includes characters, places, events, named objects and the latter is more focused on symbols and personifications. Nevertheless, a story might contain symbols and an allegory may contain characters, so the logical restrictions in these cases are not very strict. An intrinsic meaning can refer to either a cultural phenomenon or a conceptual object.

An exemplification of some restrictions on main classes through OWL axioms follows, formalized in Manchester Syntax:[35]

- InterpretationDescription:

[35] `https://www.w3.org/TR/owl2-manchester-syntax/`

```
isCompliantWithIconographicalRecognition some
IconographicalRecognition

isCompliantWithIconologicalRecognition some
IconologicalRecognition

isCompliantWithPreiconographicalRecognition some
(PreiconographicalRecognition or
FormalMotifRecognition)
```

- Recognition:

```
aboutWorkOfArt exactly 1 Artwork
includesAgent exactly 1 Agent
```

- PreiconographicalRecognition:

```
recognizedArtisticMotif exactly 1 ArtisticMotif
or recognizedComposition exactly 1 Composition
```

- FormalMotifRecognition:

```
hasCopiedMotif exactly 1 ArtisticMotif or
hasCopiedMotif exactly 1 Composition

hasPrototypicalMotif exactly 1 ArtisticMotif or
hasPrototypicalMotif exactly 1 Composition
```

- IconographicalRecognition:

```
recognizedImage exactly 1 Image or
recognizedInvenzione exactly 1 Invenzione
```

- IconologicalRecognition:

```
recognizedIntrinsicMeaning exactly 1
IntrinsicMeaning
```

- IntrinsicMeaning:

```
((recognizedConcept exactly 1 SocialObject)
 and (recognizedCulturalPhenomenon exactly 1
 CulturalPhenomenon))
 or ((recognizedConcept exactly 1 SocialObject)
 or (recognizedCulturalPhenomenon exactly 1
 CulturalPhenomenon))
```

```
isIntrinsicMeaningOf exactly 1
IconologicalRecognition
```

- ArtisticMotif:

```
(hasExpressionalMeaning min 1
ExpressionalQuality) or
(hasFactualMeaning min 1 (Action
or NaturalElement))
```

```
isRecognizedArtisticMotifOf exactly 1
PreiconographicalRecognition
```

- Composition:

```
hasPart min 1 ArtisticMotif
```

```
isRecognizedCompositionOf exactly 1
PreiconographicalRecognition
```

- Story:

```
composedOf some Character or composedOf
some Event
or composedOf some NamedObject or composedOf
some Place
```

- Image:

```
(hasCharacter min 1 Character)
or  (hasEvent min 1 Event)
or (hasNamedObject min 1 NamedObject)
or (hasPersonification min 1 Personification)
or (hasPlace min 1 Place)
or (hasSymbol min 1 (Symbol or Simulation))

isIconographicallyRecognizedBy exactly 1
IconographicalRecognition
```

- Allegory:

```
composedOf some Personification or
composedOf some Symbol
```

This list contains only the directly created axioms. The restrictions inherited by the alignment to external ontologies are available in the documentation.[36]

3.6 Evaluation

The ICON ontology was evaluated in i) its extraction potential through testing the Competency Questions [52] on a real-world interpretations dataset; ii) gauging its granularity potential by comparing data of interpretations written using the model

[36] `https://w3id.org/icon/docs/`

against the same interpretations encoded with other ontologies; iii) logical consistency, FAIRness and syntax using selected tools and services.

The following subsections will deal with the creation of the evaluation dataset and the different evaluation methods.

3.6.1 Creation of evaluation dataset

According to the analysis presented in section 1.2 of chapter 1, existing linked open data hubs do not include detailed descriptions of iconological aspects needed to evaluate the ontology. Therefore, we tested ICON on manually created data extracted from Panofsky's *Studies in Iconology* interpretations [42]. This text was chosen for its historical importance and authoritativeness in the domain at hand. Since iconological interpretations do not have a fixed structure, no automatic recognitions were implemented to extract the knowledge. The entire process is based on the author's qualitative reading and interpretation of the text. Therefore, it has to be considered that the statements obtained depend on the author's subjective comprehension of Panofsky's work. For this reason, the author is indicated as responsible for the graph created, while the Panofsky's text is always cited as the source of each statement. The data creation was conducted as follows. Firstly, we created a tabular data structure reflecting the ontology structure, including information on each level of interpretation. Secondly, we interpreted Panofsky's claims and described the features of interest according to the data structure. The quality of the data was assessed through values validation through controlled lists. Finally, the data were converted into RDF[37] according to the modelling of the ontology described in 3.5 and reusing CIDOC-CRM for the description of artworks' metadata. As a result, the dataset contains a total of 28,864 triples about 152 artworks, 1,980 interpretations, and 928 subjects. Additional statistics about the dataset can be found in Table 3.7.

3.6.2 Competency questions evaluation

The ontology was evaluated on Panofsky's dataset through the competency questions listed in table 3.3. Each SAMOD iteration corresponds to a group of questions

[37] the conversion was done using Python RDFlib library.

Table 3.7: Test dataset overview of number of triples, subjects, and interpretations

	Interpretations	Subjects	Triples	Artworks
Level 1	1,662	491		
Level 2	544	297		
Level 3	274	140		
Total	1,980	928	28,864	152

Table 3.8: Information about artworks cited in the queries

ID	Description
ART1195	Piero di Cosimo, *The Finding of Vulcan*, 1485-1490, Hartford, Wadsworth Atheneum
ART1266	Vatican City, Biblioteca Apostolica Vaticana, *L'Ovide moralisé*, XIV Century, Cod. Reg. 1480, folio 5r. Anomynous, *Saturn*
ART1267	*Saturn*, first third of XV Century, Dresden, Kupferstichkabinett
ART1268	Jacopo Caraglio after Rosso Fiorentino, *Saturn*, Engraving B24, 1526
ART1269	*Saturn and his Children*, in Cim. 10, Middle XV Century, Berlin, Kupferstichkabinett
ART1270	Vatican City, Biblioteca Apostolica Vaticana, Cod. Pal. lat. 1368, folio 1v, XVI Century. Anomynous, *Saturn*
ART1284	Giovanni Rost after Agnolo Bronzino, *Flora*, Florence, Galleria degli Arazzi
ART1285	Albrecht Dürer, *The abduction of Proserpine*, 1516
ART1289	Nicolas Poussin, *Phaethon before Helios*, Berlin, Kaiser Friedrich Museum
ART1346	Michelangelo, *Pen drawing*, Fr. 103, 1504-1505, London, British Museum
ART1534	Rubens, *Saturn devouring a Son*, 1636-1638, Madrid, Prado

at levels 0, 1, 2 or 3, and expected results for each CQ are described. To test the correctness of the single classes, the CQs were further subdivided into more detailed ones closer to the ontology structure. All unit tests that query the test dataset are available through GitHub in the form of Jupyter notebooks.[38] For each level, we describe one or two CQs. Table 3.8 contains an overview of the metadata about the artworks included in the queries.

3.6.2.1 CQ Level 1

The query presented here is part of the CQ 1.2, aimed at retrieving all first level meanings of the artworks considered, distinguishing between *Natural Elements*, Actions and *Expressional Qualities*. The expected results, corresponding to the obtained ones, are shown in Table 3.9.

CQ 1.2.1: *Retrieve all the natural, expressional meanings and actions recognized in the artistic motifs of ART1195.*

```
PREFIX d: <https://w3id.org/icon/data/>
PREFIX icon: <https://w3id.org/icon/ontology/>
```

[38] `https://w3id.org/icon/development`

Table 3.9: CQ 1.2.1 results

natural	expressional	action
man		
	dazed	
woman		
	charitable	
		helping
group of women		
natural landscape		
dog		
		gathering flowers
	surprise	
	amusement	
	pity	
	protectiveness	
	kindliness	
	hospitality	

```
SELECT DISTINCT ?natural ?expressional ?action
WHERE {
?icrec icon:aboutWorkOfArt d:ART1195;
     {?icrec icon:recognizedArtisticMotif ?am}
UNION {?icrec icon:recognizedComposition ?comp.
?comp icon:hasPart ?am }
{?am icon:hasExpressionalMeaning ?expressional}
UNION {?am icon:hasFactualMeaning ?natural.
?natural a icon:NaturalElement}
UNION {?am icon:hasFactualMeaning ?action.
?action a icon:Action}
}
```

The second level 1 query listed below is part of CQ 1.3, aimed at retrieving all level 1 subjects that are formally derived or copied from other artworks level 1 subjects. CQ 1.3.1 applies this question to ART1284 and ART 1285. Results in table 3.10 shows how this structure can allow a detailed and qualitative comparison

Table 3.10: CQ 1.3.1 results

subject	copiedSubject
woman	woman
riding-on	riding-on
unicorn	ram

of the phenomenon of visual motifs copy and migration, since the relation between the single portions interested can be made explicit.

CQ 1.3.1: *What are the level 1 subjects (i.e. copied subjects) copied by ART1284 from ART1285, including the ones identified by a composition? What are the corresponding original subjects in ART1285 (i.e. subjects)?*

```
PREFIX d: <https://w3id.org/icon/data/>
PREFIX icon: <https://w3id.org/icon/ontology/>

SELECT DISTINCT ?subject ?copiedSubject WHERE {
?rec a icon:FormalMotifRecognition;
    icon:aboutWorkOfArt d:ART1284, d:ART1285.
{?rec icon:hasPrototypicalMotif ?am.
?am a icon:ArtisticMotif}
UNION {?rec icon:hasPrototypicalMotif ?comp.
?comp icon:hasPart ?am}
?am icon:hasFactualMeaning |
icon:hasExpressionalQuality ?subject.
{?rec icon:hasCopiedMotif ?copied.
?copied a icon:ArtisticMotif}
UNION {?rec icon:hasCopiedlMotif ?comp.
?comp icon:hasPart ?copied}
?copied icon:hasFactualMeaning |
icon:hasExpressionalQuality ?copiedSubject.
}
```

3.6.2.2 CQ Level 2

The correspondence of level 1 subjects with level 2 ones offers the chance to explore the variations in the subjects' representation, a fundamental research aspect for the domain of iconography and iconology. The query below represents CQ 2.6, aimed at retrieving the representative variations of a level 2 subject, to the Character "blindfold Cupid". This type of research question can be further explored by retrieving the date and place of production of the artwork to obtain a detailed representation of the subject variations over place and time. As a consequence, it can be a useful tool for art historians for integrating qualitative iconographical analysis with a quantitative overview of the phenomenon.

CQ 2.6.1 *What are the variants of the subject "blindfold Cupid"? Retrieve all the level 1 subjects corresponding to this subject along with how many times do they appear.*

```
PREFIX d: <https://w3id.org/icon/data/>
PREFIX icon: <https://w3id.org/icon/ontology/>

SELECT DISTINCT ?lev1 (count(?lev1) as ?tot)
WHERE {
VALUES ?rel {icon:hasFactualMeaning
icon:hasExpressionalMeaning}
?rec icon:recognizedImage ?img;
    icon:aboutWorkOfArt ?art;
    icon:refersToArtisticMotif ?am;
    icon:recognizedImage ?img.
?img icon:hasCharacter d:blindfold-cupid.
{?am a icon:ArtisticMotif; ?rel ?lev1} UNION
{?am icon:hasPart ?a. ?a ?rel ?lev1}

} GROUP BY ?lev1
ORDER BY DESC(?tot)
```

Table 3.11: CQ 2.6.1 results

level 1	tot	level 1	tot	level 1	tot
wings	8	hearts	1	sleeping	1
bandage	7	string-of-hearts	1	natural-landscape	1
bow	6	throne	1	standing-on	1
arrows	6	arrow	1	sphere	1
boy	4	band	1	putto	1
child	4	spear	1	snuggling-in-her-lap	1
griffon-claws	3	standing-on-a-horse	1	talons	1
crown-of-roses	2	horse	1	running	1

3.6.2.3 CQ Level 3

Concerning level 3, we present a query retrieving all the artworks linked to the same cultural phenomenon, focusing on the phenomenon "evolution of the iconography of Saturn". This approach is useful to group all the artworks that are involved in the same cultural phenomenon as a starting point of further analysis, considered fundamental for the researcher. For example, it could be interesting to explore the second level subjects involved in it, their variations at level 1 according to time and space, and the literary sources involved.

CQ 3.4.1 *retrieve the artworks where an intrinsic meaning is associated to the cultural phenomenon CF1087 "Evolution of the iconography of Saturn"*

```
PREFIX d: <https://w3id.org/icon/data/>
PREFIX icon: <https://w3id.org/icon/ontology/>

SELECT DISTINCT ?artwork WHERE {
?rec icon:aboutWorkOfArt ?artwork;
    a icon:IconologicalRecognition;
    icon:recognizedIntrinsicMeaning ?intrinsic.
?intrinsic icon:recognizedCulturalPhenomenon d:CF1087.
}
```

Results: ART1269, ART1270, ART1266, ART1267, ART1268, ART1534, ART1535, ART1289.

3.6.2.4 General CQs

For the general level CQs, we present two competency questions. The first (CQ 0.2.1) is presented to show how different interpretations can be represented. It retrieves all the interpretation descriptions of an artwork and the types of recognition included in it. It is performed over ART1195, which is the object of contrasting interpretations. While describing it [42], the art historian Erwin Panofsky states that his own position diverges from the usual interpretation according to which the artwork depicts the myth Hylas and the Nymphs. He cites the works of other three scholars as references for this general interpretation (A. E. Austin, R. van Marle, L. Venturi). In contrast, he says that it represents the finding of Vulcan by the inhabitants of the island of Lemnos, after he precipitated from Mont Olympus because he was kicked out by his mother. For his interpretation, Panofsky considers various features of the first level (e.g. the general atmosphere of kindness and hospitality, which would be inappropriate to the rape and sexual aggression of the Nymphs to Hylas, described by the myth).

CQ 0.2.1 *What is the person responsible for the recognitions at each level in ART1195? Do they belong to different descriptions?*

```
    PREFIX d: <https://w3id.org/icon/data/>
    PREFIX icon: <https://w3id.org/icon/ontology/>
    PREFIX cito: <http://purl.org/spar/cito/>

    SELECT DISTINCT ?personLabel ?desc ?type  WHERE {
    ?rec icon:aboutWorkOfArt d:ART1195;
        ?rel ?desc;
         crm:P14_carried_out_by ?person;
         a ?type.
        ?desc a icon:InterpretationDescription.
        ?person rdfs:label ?personLabel
} ORDER BY ?desc
```

Table 3.12: CQ 0.2.1 results

Person	Description	RecognitionType
Erwin Panofsky	ART1195-DESC1	PreiconographicalRecognition
Erwin Panofsky	ART1195-DESC1	IconographicalRecognition
Erwin Panofsky	ART1195-DESC1	IconologicalRecognition
A. E. Austin	ART1195-DESC2	IconographicalRecognition
R. van Marle	ART1195-DESC3	IconographicalRecognition
L. Venturi	ART1195-DESC4	IconographicalRecognition

Results are shown in 3.12. The artwork has 4 different Interpretation Descriptions, among which only DESC1 has recognitions at all the levels, whereas the remaining has only level 2 recognitions. To better see the agreement and disagreement of them, it is possible to retrieve the content of such recognitions or to see if the recognitions are already described with CiTO's relations of agreement or disagreement (`cito:agreesWith` `cito:disagreesWith`).

The second competency question (CQ 0.4) retrieves artworks that are described at both levels 1 and 3, but not at level 2. This request shows how the ontology can be used to assess the level of details in which the subjects are described in a dataset according to each level of interpretation.

CQ 0.4 *What artworks are interpreted on an iconological level but not on an iconographical one?*

```
PREFIX d: <https://w3id.org/icon/data/>
PREFIX icon: <https://w3id.org/icon/ontology/>

SELECT DISTINCT ?art WHERE {
?rec icon:aboutWorkOfArt ?art;
     a icon:IconologicalRecognition.
?rec1 icon:aboutWorkOfArt ?art;
   a icon:PreiconographicalRecognition.
MINUS {?rec2 icon:aboutWorkOfArt ?art;
```

```
        a icon:IconographicalRecognition.
    }
    }
```

Result: ART1346.

All the CQs confirmed the expected results in their output. Therefore, it we demonstrate that this ontology allows a meaningful representation of iconological research questions through a quantitative approach.

3.6.3 Comparison with existing ontologies

In this subsection, we propose a qualitative comparison between ICON and existing ontologies. We do so by describing an authoritative example of an interpretation from Panofsky's bibliography, in which there are several intertwined recognitions from iconographical and iconological perspectives. For this purpose, we selected the most complete ontologies for iconographical descriptions, namely: Visual Representation Ontology (VIR) and the Wikidata schema.

We selected the frontispiece of François Perrier's "Segmenta nobilium signorum statuarum..." (1638 Edition) depicting the iconography of Father Time, a subject that emerges from the Renaissance onwards but that, despite it originates from classical sources, was never visually represented in previous times. Panofsky reconstructed its genesis, claiming that it originated by the fusion of a medieval French iconography of Time represented with wings (Temps) and the sinister characteristics of Saturn (e.g. old age, scythe, the act of devouring his children) [42]. Late antique writers had already enriched the figure of the god Saturn with attributes referring to time (i.e. a dragon or snake biting its tail) or through a re-interpretation of the traditional Saturn's attributes, such as the sickle, associated with the times recurring,[39] or the act of devouring the children, reinterpreted as Time devouring "whatever he has created" [42, p. 74]. Panofsky claims that such evolution of the iconography of Time is evidence of the phenomenon of *pseudomorphosis*, accord-

[39]"tempora quae sicut falx in se recurrunt", trad. by the authors: "times which, like the sickle, recur"[122, II, 406]

Table 3.13: ”Time the destroyer” description according to the three levels of interpretation

1	man, nude and old, with a scythe and wings gnawing away a statue snake biting its tail fragments of classical buildings and of statuary
2	Father Time as a Destroyer, symbolically devouring the past by devouring the classical Torso del Belvedere
3	Evolution of the Iconography of Time: 1) Renaissance art ”produced an image of Time the Destroyer by fusing a personification of Temps with the frightening figure of Saturn, and thereby endowed the type of Father Time with a variety of new meanings” [42] 2) Pseudomorphosis: ”certain Renaissance figures became invested with a meaning which, for all their classicising appearance, had not been present in their classical prototypes, though it had frequently been foreshadowed in classical literature” [42]

ing to which, figures with a classical appearance did not exist in the classical visual arts, albeit they were described in classical literature (level 3). In addition to that, the example here described faithfully shows Cesare Ripa’s description of time the destroyer as a demon with iron teeth standing among ruins, a symbol of the fact that time ruins everything without any effort.[40] Table 3.13 resumes the understanding of this artwork at each different level, while figures from 3.6 to 3.9 shows how this example can be modeled with the ontologies considered.

Table 3.14 gives an overview of the comparison results. It is apparent from these schemas that neither VIR nor Wikidata include properties or classes that represent a third level meaning. Nevertheless, they do express important aspects of the domain. On the first level of description, VIR offers only the limited expressivity given by the class Iconographical Atom, which is intended to describe the physical portion of the artwork to which a subject is bound [76]. In contrast, Wikidata offers several level 1 specifications of the subjects through the qualifiers of the property wdt:P180 (e.g., “nudity” and ”old” referred to the subject ”man”, through the qualifier wdt:P1354 ”shown with features”). Nevertheless, none of the ontologies distinguishes the types of objects described, failing to express immaterial items properly, such as actions or emotions (VIR) and the connections existing between these immaterial aspects and the subjects doing them. Therefore, ICON aims at solving this issue by introducing Panofsky’s concept of the Composition of Artistic Motifs. As a

[40]“La Ruina, e la Bocca aperta, e i Denti di ferro mostrano, che il Tempo strugge, guasta, consuma, e manda per terra tutte le cose senza spesa, e senza fatica.” trad. by the authors: ”The Ruina, and the Open Mouth, and the Iron Teeth show, that Time presses, spoils, consumes, and sends all things to the ground without expense, and without any effort”. [123]

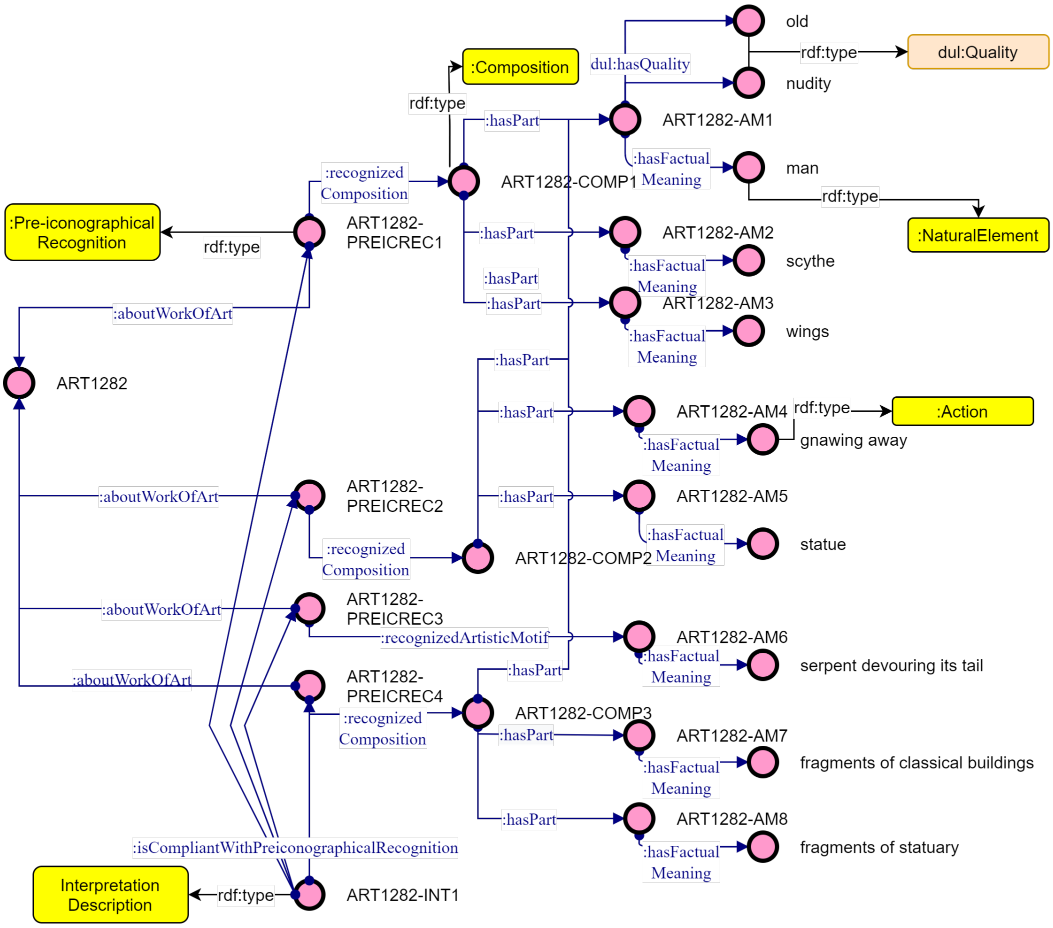

Figure 3.6: Description of the example using ICON (level 1)

result, level 1 objects as depicted in the specific artwork can not only be described in detail, but also gathered in meaningful groups. This structure allows specifying the actors involved in the action, as shown in Figure 3.6 (the subjects “man”, “gnawing away” and “statue” are part of the same Composition ART1282-COMP2), or the actors feeling emotions. At the second level of interpretation, with VIR it is possible to express important characteristics of the representation, such as attributes, personifications, symbols, places and characters (Figure 3.8), but not events. Wikidata, on the contrary, tends not to specify the type of object depicted, even if a symbolical meaning can always be expressed through the qualifier `wdt:P4878` ”symbolizes” (fig. 3.9). The distinction between stories and allegories is not included in both of them, and, for different reasons, the contextual appearance of the subjects cannot be carefully described. This is due to the fact that in VIR it is not possible to properly

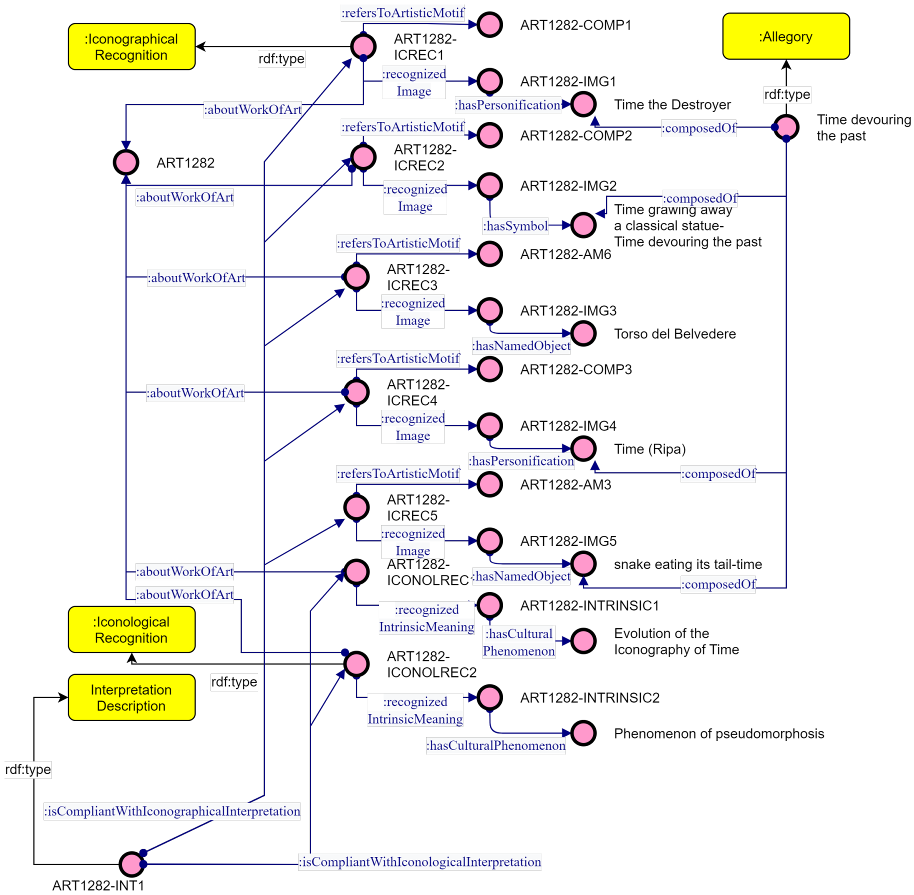

Figure 3.7: Description of the example using ICON (level 2 and 3)

describe the subjects at the first level, whereas in Wikidata, the depicted subjects cannot be related to each other. Therefore, as shown by the example, the act of a person gnawing away at a statue cannot be related to the allegory of Time devouring the past in VIR, and cannot be recognized as an allegory in Wikidata. To solve that, the ICON Iconographical Recognition allows relating the second level subjects to its level 1 representation, and the subjects can be part of Stories or Allegories. In addition, the n-ary class `icon:Image` allows separating the general description of subjects from the contextual one. In this way, subjects can be described carefully including characteristics that would be inappropriate to include in the vocabulary-level description of the subject considered, highlighting variations in their contextual representation. Whereas the VIR and Wikidata features described above allow a description of the first two levels of interpretation, none of them represents the domain of knowledge of iconology by considering the third level of interpretation.

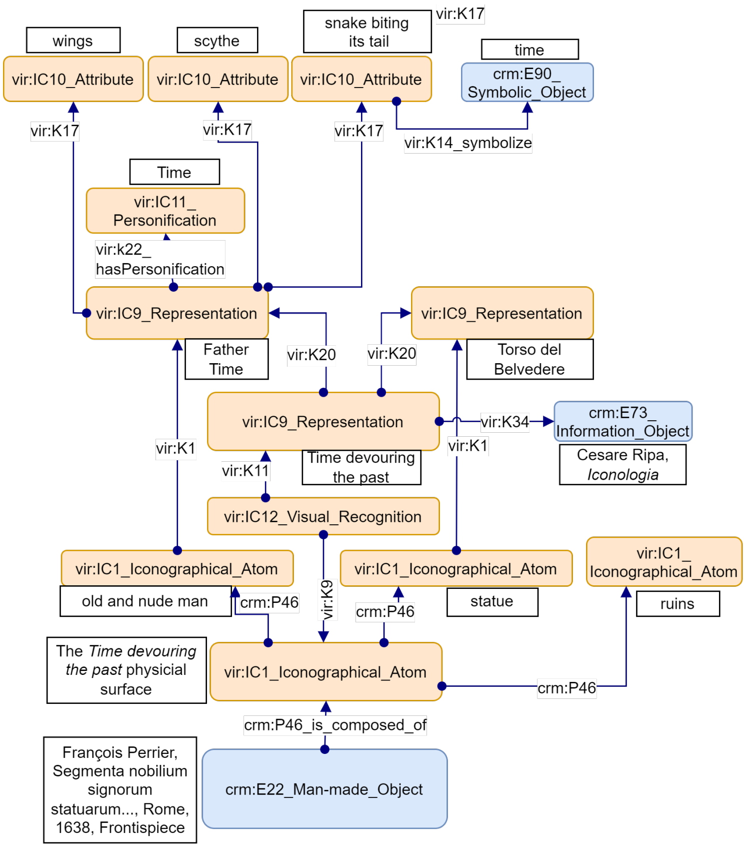

Figure 3.8: Description of the example using VIR

Therefore, ICON introduces `icon:IconologicalRecognition` relating a third level meaning (concept or cultural phenomenon) to the whole artwork or to its specific parts.

Concerning the interpretation's attribution of responsibility, both VIR and Wikidata allow registering the person responsible for the statement in different ways. In VIR, the person responsible and possible sources can be related to the class `IC12 Iconographical Recognition`, on which the whole recognition of the artwork's content without a specification for every subject recognized depends. In Wikidata, every statement can be attributed to a person responsible, but information about sources is not always provided. A consistent difference lies in the fact that whereas VIR expresses the person responsible for the claim, Wikidata considers the person responsible for the data inserted, avoiding a possible interesting comparison of authoritative art historian claims. As a result, ICON introduces an interpretation for each statement, giving subjectivity and authoritativeness for

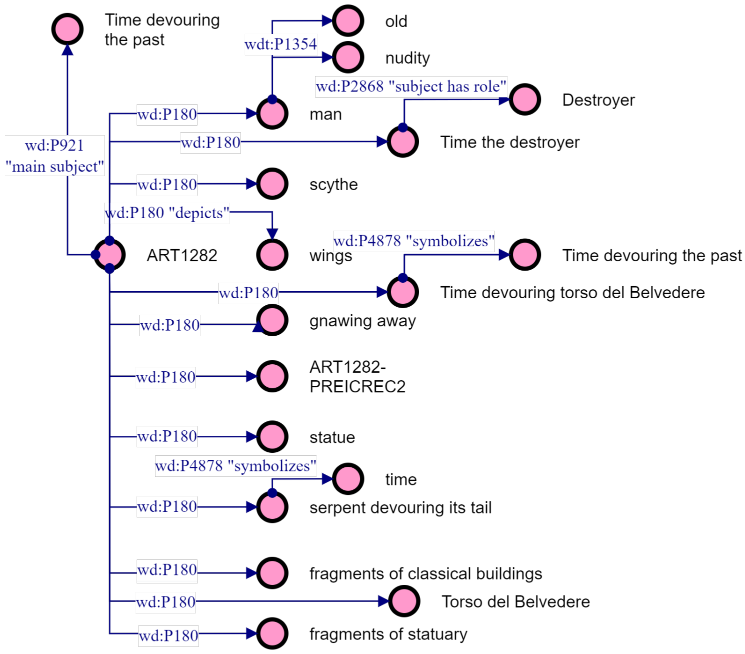

Figure 3.9: Description of the example using Wikidata

Table 3.14: Comparison between Wikidata, VIR and ICON

Ontology	Lev. 1 subj	Lev. 2 subj	Lev. 3 subj	Distinction contextual/general subject description	Interpretation subjectivity	Distinction and relation between levels	Subjects variations description
VIR	attributes only	✓		poor	only for the main interpretation		only attributes and representations
Wikidata	✓	✓		✓	only the data author		only level 1 aspects
ICON	✓	✓	✓	✓	the person responsible for the recognition for each statement	✓	All the levels

each subject recognized. In this way, more agreeing claims can be expressed in the same recognition description, to better represent the realistic case in which an art historian agrees with the claim of others, quoting them and adding further interpretations. In addition, this structure fosters the interoperability of online sources and data integration.

3.6.4 Automatic Evaluation

ICON ontology was evaluated on technical aspects using automatic tools and services. We validated our ontology syntax by the WRC RDF Validation Service.[41] No syntax problems were highlighted by this tool. Then we evaluated the logical consistency of our ontology through the OOPS [71] tool[42] that provides feedback on the ontology in form of highlighted pitfalls of different levels of importance. Most of the issues raised by this tool do not come from our modelling, instead, they are linked to the reused ontologies that might have missing information (for example no ranges and domain in properties or inconsistent labelling). The only highlighted pitfall that was directly linked with classes developed by the ontology is "P30: Equivalent classes not explicitly declared." This issue suggests the possibility that classes such as `icon:Character` and `dul:Quality`, `dul:Role`, and `dul:Reference` should be equivalent. At the same time, it suggests that `icon:Image` should be equivalent to `sim:Simulacrum` and that `icon:Story` should be equivalent to `dul:Narrative`. The first equivalence would be fundamentally wrong because the classes themselves do not represent the same concept. The same can be said for Image and Simulacrum, as in the Simulation Ontology a simulacrum is said to be a general symbol that could have different representations, that is why it is equivalent to our class icon:Symbol. An image is considered instead as a specific representation of a symbol in an artwork; therefore, it is not considered to be the general concept of the symbol itself. Furthermore, `icon:Story` and `dul:Narrative` might have similarities, but the latter class has no description, so we refrained from making ambiguous equivalences. Apart from these three cases, the ontology was evaluated pitfall-free in all the other aspects (considering issues that dealt with originally created classes only). Finally, we analysed our ontology with the FOOPS [69] tool, which evaluates how much an ontology complies with the FAIR (Findability, accessibility, interoperability, reusabil-

[41] `https://www.w3.org/RDF/Validator/`

[42] We suggest replicating this evaluation by pasting the whole RDF/XML file that contains the ontology content into the evaluation website (`https://oops.linkeddata.es/index.jsp`

ity) principles.[43] It is important for ontologies to follow FAIR principles, ensuring that they can be effectively shared, understood, and utilized within the research community and beyond [124]. Our ontology scored 90% in the FOOPS tool. In particular, it received a score of 8.5 out of 9 on reusability, a score of 8 out of 9 on findability, a score of 3 out of 3 on interoperability, and a score of 2 out of 3 in accessibility. The main problem highlighted by this tool is that the ontology is not yet inserted in the linked open data vocabulary (lov).[44] This issue will be addressed in the future. Finally, we verified that the logical axioms of all the external imported classes and properties that were aligned to our ontology did not cause any inconsistency in ICON by running the Hermit Reasoner [125].

3.7 Ontology release and documentation

The current version of ICON ontology contains 28 classes and 58 properties (of which 1 data property). Additionally, it reuses classes and properties taken from 6 ontologies. Its release comes with a documentation that gives a definition for all its classes and properties, along with examples on the use of the ontology. Several permanent URIs were provided by the w3id service:

- `https://w3id.org/icon/docs/`: that links to the documentation of the ontology generated with the widoco tool [85].
- `https://w3id.org/icon/ontology/`: that leads to the owl file of the ontology (RDF/XML serialization)
- `https://w3id.org/icon/development/`: that leads to the GitHub repository where the SAMOD iterations and eXtreme Design unit tests can be found.

[43] The FAIR principles can be summarised as follows. Findability: ensuring that data and metadata are assigned persistent identifiers, richly described, and registered or indexed in a searchable resource. Accessibility: making data and metadata retrievable using standardized communication protocols, even when the data is no longer available. Interoperability: using formal and broadly applicable languages for knowledge representation, employing vocabularies that follow FAIR principles, and including qualified references to other data. Reusability: enriching metadata with accurate attributes, associating detailed provenance, and meeting domain-relevant community standards [124]

[44] `https://lov.linkeddata.es/dataset/lov/`

The preferred prefix of the ontology (icon) was registered on `http://prefix.cc`.

3.8 Conclusion and future work

In this chapter, we presented ICON, an ontology dedicated to the conceptualization of artistic interpretations designed by formalizing the content of several interpretation theories. In line with the principles of reuse and interoperability of the Semantic Web, the ontology reuses (directly and indirectly) several existing ontologies. It is released alongside a documentation that guides potential users in formalizing art interpretations using our model. ICON was evaluated on its extraction potential, syntax, metadata and FAIRness. Moreover, its granularity was highlighted through a comparison to current ontologies on their respective serialization of the same interpretation. The results show how our work elevates the potential of expression of artistic interpretations in the context of Semantic Web by providing a granularity level that was not reached by other ontologies on this topic. Finally, its effectiveness of describing the Iconographical and Iconological complex domain is confirmed by the results of the proposed competency questions, formalized in SPARQL queries, ran on a test dataset containing artistic interpretations.

Although ICON allows users to describe a plethora of concepts related to art interpretation, the ontology presented here could still be improved. Future work will be directed to a more thorough conceptualization of cultural phenomena, personifications and allegories in the same way that cultural symbols were defined by the Simulation Ontology [33]. Furthermore, modeling artistic interpretations required the use of several agglomerated n-ary relationship classes, drawing a long path from the artwork to its meaning. On the one hand, this modeling offers a very high granularity level, as interpretations can be dissected into recognitions representing different levels (pre-iconographic, iconographic, iconological), allowing the potential extraction of very specific information as shown by the testing of the competency questions. On the other hand, potential users might only be interested into separating what is depicted in an artwork into the aforementioned levels, without having to describe the whole process of interpretation. Future work will be devoted to the creation of property chains that can be used to declare that an artwork represents elements of a pre-iconographic, iconographic, or iconological level. Both solutions (series of n-ary classes), and the property chains could be adopted, in a knowledge graph, at the same time, as in ArCo [13], which uses property chains as shortcuts, and n-ary relationship classes to describe the same information with more granularity. As for the alignments, although, as table 5 shows, there are more than 50 alignments to external ontologies, we decided to keep a conservative approach for this version of ICON. Future versions will foster the interoperability of ICON even more by (i) finding other ontologies that cover similar aspects to increase alignments, such as [109], and (ii) by specializing the alignments with the currently aligned ontologies by finding more specific classes and properties. The `icon:Invenzione` class, for instance, is aligned to a very generic `crm:E1_CRM_Entity`. Future versions of CIDOC might introduce new classes that are closely related to ours, leading to a more specific alignment.

On another note, the current debate of interpretation provenance and the difference between "asserting and expressing" [126] in a Semantic Web context using recently introduced technologies of RDF-star[45] and conjectures [126] could use a dataset of contrasting interpretations described with our ontology as a case study.

In conclusion, by combining our model with the external ontologies to which it is aligned, it is now possible to thoroughly describe artworks both in their standard metadata (i.e. creator, date of creation, dimensions, place of creation) and on their content side based on interpretations. The result is that, finally, these two evenly important types of information are now treated equally and can both exploit the potentiality offered by the Semantic Web. This contribution opens up the possibility to link artworks in their content level, allowing content-based research questions in the field of art history to cross into the linked open data realm.

This work is essential to answer RQ1.1 *Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism*

[45] `https://w3c.github.io/rdf-star/cg-spec/2021-12-17.html`

and symbolic interpretation in the cultural heritage domain?. Contrarily to what I wrote about the modelling of conventional cultural symbolism in the previous chapter, for the modelling of artistic interpretation it is necessary to refer to and use the theories formalised by art-historian as a top-down approach. Moreover, as this section has discussed and as a potential future work, more aspects of history and anthropology need to be considered to model concepts such as beliefs, customs and cultural phenomena, which demonstrates the importance of this varied range of subjects in modelling artistic interpretations. All the necessary factors to model iconographic and iconological interpretations are discussed in sections 3.1, 3.4 and tested in subsection 3.6.2.

Finally, the development of this ontology completes research objective 2 *Creation of a knowledge graph that contains instances of conventional cultural symbolism*. Following this thesis' narrative, having a framework to model artistic interpretation makes it possible, for automatically generated symbolic interpretations, to be released in a linked open data format, and be integrated into the creation of the Semantic Web of Symbolic Meanings. Part of the next chapter will be dedicated to a system that can automatically interpret artworks according to different cultural context and release those interpretations in linked open data following the structures of ICON and Simulation Ontology.

Chapter 4

The Semantic Web of Symbolic Meanings

This chapter contains all the analyses that were carried out on the newly created Semantic Web of Symbolic Meanings. Some of them are part of a publication, and others will be submitted to adequate venues after the PhD. If these analyses involved other people, they are acknowledged in the corresponding section. The work presented in sections 4.1, 4.2, 4.3 is used to answer to RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic*, as it proves the usefulness of HyperReal data by showing its potential in fostering research in different domains. Then, section 4.4 presents CYBERNATED, a system that is used to automatically infer symbolic, context-dependent interpretations of artworks based on the symbolism of their depictions. Section 4.5 evaluates CYBERNATED on a dataset annotated by computer vision. The evaluations are both quantitative on the total number of interpretations that can be generated and filtered, and qualitative on the potential of CYBERNATED to also detect the creator's symbolic intention, i.e. detect symbolic meanings that were intentionally referred to by the creator's of the artworks. Then, section 4.6 presents ICONdata, a re-engineered version of the iconographical and iconological statements of Wikidata according to the structure of the ICON ontology (described in chapter 3), and enhanced by the automatic interpretations provided by CYBERNATED. In the same section, the potential of this work is shown by the automatic discovery of serendipitous connections between artworks based on symbolism. Both section 4.5 and 4.6 are used to address RQ 2.1 *To what extent can it (linked open data about symbolism) be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?* as they provide a clear vision (supported by measurable results) on the power of CYBERNATED to generate a significant number of interpretations. Section 4.6 also addresses RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved by including such knowledge?* by analysing how many new connections can be made once linked open data about cultural symbolism from HyperReal is integrated with Wikidata's *icon* statements. Finally, section 4.7 concludes the chapter by presenting an application that was built on the connections between Wikidata, the digital collection of the National Museum of World Cultures (NMVW),[1] and HyperReal. This last work furtherly addresses RQ2 and RQ2.1, and provide a first look at potential applications that could be developed from the Semantic Web of Symbolic Meanings.

4.1 A quantitative comparison of cultural contexts based on their symbolism

This section is an expansion of the analysis proposed in the ADHO2022 abstract "Comparing Symbolism Across Asian Cultural Contexts Using Graph Similarity Measures" [127] written by me and my supervisors (Aldo Gangemi and Marieke van Erp) and Valentin Vogelmann, a researcher for the Dutch National Museum of World Cultures. The original analysis was performed on only an Asian subset of the cultural contexts present in HyperReal. In this version, all the cultural contexts of HyperReal (for which there is enough data) are taken into consideration and also their macro-area(s) derived from the taxonomy described in section 2.3.9. This work is used to address RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?*, as its purpose is (i) to test whether symbolism provides a useful basis for quantitative cultural comparisons (ii) to highlight

[1] More information on this collection is provided in section 4.7

potential connections that can be then qualitatively analysed and verified.

Symbols are pervasive elements in cultures as they convey ideas, values, traditions, and beliefs [128, 129]. Therefore, it is not surprising that symbols are the cornerstone of many comparative cultural studies, such as research on the symbolism of jewellery and ornaments [130], rituals, mottos, icons [131], trees, dragons, and the tree of life [132, 133, 134].

In line with Martinho [135], who suggests a shift towards quantitative methods in cultural studies, and Zepetnek [136], who adapted comparative literature methodologies to investigate similarities between cultures, we recommend a computational approach which makes use of symbols to conduct quantitative comparative cultural analyses. We define a quantitative measure of cultural affinity, which we then put into practice on the data contained in HyperReal.

We measure similarities between pairs of cultural contexts, and between groups of cultural contexts extracted from the cultural context taxonomy.[2] We analyse how symbolism can provide a useful basis for quantitative cultural comparisons, and we aim to establish whether cultural contexts tend to be more similar in terms of their symbols or in terms of the symbolic meanings that their symbols refer to. Moreover, we use the results of quantitative analysis to highlight specific connections between cultures that then we analyse qualitatively.

4.1.1 Measuring similarity: resources and methodology

Being embodied by linguistic expression allows us to measure symbols' and their symbolic meanings' semantic similarity, for which we use the spaCy and Wiki2Vec [137] pre-trained Word2vec models. From spaCy, we use the `en-core-web-lg`[3] pre-trained pipeline. It contains more than 600,000 unique vectors with 300 dimensions, and it is trained on blog articles, news, and comments. It is a good pipeline to measure similarity between common words, such as "tree", "cold", "fire". For Wikipedia2vec, we reuse a model that is pre-trained on all Wikipedia articles from

[2]For the pair of cultural contexts' analysis, we focus only on those cultural contexts that are part of at least 15 simulations. For the grouped cultural contexts, we include all of them

[3]`https://spacy.io/models/en#en_core_web_lg`

a 2018 data dump[4] with vectors of 300 dimensions. We chose to include the Wikipedia2vec model to measure similarity between more context-specific words and entities, such as "Vishnu", which is a Vedic deity. Given the cultural richness included in HyperReal, we have many other examples of specific words which are not included in spaCy. Nevertheless, spaCy performed better than Wikipedia2vec when measuring similarity between common words, so we decided to use the models in combination. We first computed the combination of all the pairs from all the simulacra and reality counterparts in HyperReal which belonged to a specific cultural context (i.e, we excluded the simulations linked to a *general or unknown* context according to the original source). Then, we measured the similarity between each pair using both Wikipedia2vec and spaCy's pipelines. The possible scores for the vector similarities range from 0 to 1. We selected various threshold to consider two words similar or not, specifically: ≥ 0.55, ≥ 0.6, ≥ 0.65, ≥ 0.7, ≥ 0.75 and ≥ 0.8. Then, for each pair of cultural contexts (or pair of groups of cultural contexts) taken into consideration, we extracted the simulacra and reality counterparts that appeared in the simulations of those cultural contexts. We formulate a series of different computations to measure the similarity between each pair of culture:

1. Normal bidirectional similarity: we compute the Jaccard similarity [138] between the sets of symbols and symbolic meanings of the pair of cultural contexts (for example, context A and context B) taken into consideration. The Jaccard similarity is normally calculated according to the division between the lengths of the intersection and the union of two sets; we place in the intersection set all the terms that are considered similar by either the Wikipedia2vec model or the spaCy model. As a result, the similarity measurement is bidirectional; meaning that the similarity between cultural context A and cultural context B is the same as the similarity between cultural context B and cultural context A.

2. Normal unidirectional similarity: in this variant, we only compute the Jaccard similarity with the elements of one cultural context. In fact, we divide the

[4] `https://wikipedia2vec.github.io/wikipedia2vec/pretrained/`

number of elements that one cultural context (again, we call it A) shares with another cultural context (B) by the total number of elements of A. And then we compute it again, dividing the number of elements that B shares with A by the total number of elements of B. This yields a pair of distinct values of similarity that depend on directionality, for instance Similarity(A, B) will be different from Similarity(B, A).

3. Weighted bidirectional similarity: we change the approach of the normal bidirectional similarity, by giving weights to both symbols and symbolic meanings depending on their *symbolic impact* and *symbolic referencing*. We define *symbolic impact* as the number of symbolic meanings associated with a symbol in a specific cultural context, and *symbolic referencing* as the number of times a symbolic meaning is symbolised by a symbol in a particular context. For instance, if a dragon symbol in cultural context A is used in 10 simulations (meaning that it has 10 symbolic meanings), we attribute it a value of 10. Using the same reasoning, if the symbolic meaning of purity in cultural setting A appears in 5 simulations (which implies 5 different symbols in the same cultural context denote it), we assign it a value of 5. The Jaccard similarity hence becomes the total of the weights of the comparable elements between two cultural contexts divided by the sum of all the weights of the elements of the two cultural contexts.

4. Weighted unidirectional similarity: we apply the same weighting methodology explained before to the computation of the unidirectional similarity.

We perform all computation for both symbols and symbolic meanings separately, and then we average the results, obtaining three values of similarity between two cultural contexts according to (i) their symbols, (ii) symbolic meanings, and (iii) a comprehensive score of similarity averaging the previous values. Moreover, all the computations are repeated for all the six similarity thresholds we specified above, meaning that each computation results in 18 values. Then, by applying the four approaches of the computations described in the list above, in total, we get 72 different values that measure the similarity between two cultural contexts (3 multiplied by 6 multiplied by 4). Because we also measure this similarity between continent groups and macro-areas, we get a total of 216 different measurements.

4.1.2 Discussion on the Results

We measure similarities between pairs of cultural contexts that have at least 15 simulations, there are 87 cultural contexts in HyperReal that fit this criterium, as shown in listing 4.1. Then, we also measure similarities between groups of contexts clustered together by the SKOS taxonomy of cultural contexts that was created in section 2.3.9. Macro-area groups of contexts, and geographical (continent-based) groups of contexts are shown respectively in listings 4.2 and 4.3. The 216 results are presented in a series of heatmaps on the "Cultural Context Similarity" section of the website `https://br0ast.github.io/simulationontology/`.[5] For example, the symbol-based weighted bidirectional measurement for the cultural contexts pairs, with a threshold of ≥ 0.8, is available at `https://br0ast.github.io/simulationontology/heatmaps/simplei08simctx.html`.

```
[kb:phrygian, kb:european, kb:folkloric,
kb:indic, kb:greek, kb:buddhist,
kb:christian, kb:flowerLanguage, kb:teutonic,
kb:northAmericanIndian, kb:tungun,kb:siberian,
kb:bambaran, kb:german, kb:islamic,
kb:southAmericanIndian, kb:uralAltaic,
kb:irish, kb:french, kb:hermetic,
kb:witchcraft,kb:minoan,kb:kalmyk,
kb:astrological, kb:gallic,kb:mongolian,
kb:vietnamese, kb:zoroastrian,kb:english,
kb:jain, kb:mexicanIndian, kb:taoist,
kb:yoruban, kb:jewish, kb:phoenician,
kb:mithraic, kb:middleAges, kb:italian,
```

[5] also accessible from the permanent URI `https://w3id.org/simulation/`

```
kb:assyrian, kb:babylonian, kb:gnostic,
kb:amerindian, kb:norse, kb:dogon,
kb:mesopotamian, kb:cartomancy,kb:cambodian,
kb:kabalistic, kb:celtic, kb:arabian,
kb:centralAmericanIndian, kb:germanic,
kb:russian, kb:persian, kb:masonic,
kb:sumerian, kb:japanese, kb:farEastern,
kb:finnish, kb:tarot, kb:pythagorean,
kb:pawnee, kb:oriental, kb:hindu,
kb:roman, kb:african, kb:aztecan,
kb:egyptian, kb:chaldean, kb:oceanic,
kb:slavic, kb:medieval, kb:mayan,
kb:fulani, kb:chinese, kb:american,
kb:australianAboriginal, kb:heraldic,
kb:druidic, kb:alchemic, kb:british,
kb:maori, kb:mediterranean, kb:sumeroSemitic,
kb:grecoRoman, kb:tibetan, kb:zodiacal]
```

Listing 4.1: List of contexts with at least 15 simulations. *kb* is the prefix for `https://w3id.org/simulation/data`

```
[kb:geoSocioCultural, kb:scientific, kb:religious,
kb:culturalesoteric, kb:communication, kb:folkloristic,
kb:social, kb:artistic, kb:philosophical,
kb:cultural]
```

Listing 4.2: List of macroareas. *kb* is the prefix for `https://w3id.org/simulation/data`

```
[kb:asian, kb:northAmerican, kb:european,
kb:southAmerican, kb:african, kb:oceanic]
```

Listing 4.3: List of continent-based contexts. *kb* is the prefix for `https://w3id.org/simulation/data`

From the results, some general remarks can be made on the usefulness of using symbolism to analyse similarities between cultural contexts. First, it is possible to spot strong influences between cultural contexts and macro-areas. For example, looking at the continent-based analyses, it is highlighted how the symbolism of the Asian and African continent influenced the rest of the continents.[6] In fact, they are the top 2 continents when looking at how much other continents have in common with them. The bidirectional measurement shows that the other four continents are similar to the Asian and African more than the other way around. It might also mean that Asian and African cultural contexts' symbolism might embody the rest of the world's. How these analyses can reveal complete embodiments of cultures becomes even more clear when looking at the results of the analysis per pair of cultures.[7] Here, some embodiments confirm logic expectations while others suggest peculiar connection which might need further, specific analyses. As for the expected embodiments, Minoan culture scores 0.8 similarity to the Greco-Roman culture, Vietnamese culture scores 0.76 similarity to Chinese. These connections show how cultures like Greco-Roman clearly embody the Minoan culture, or samewise how the Chinese culture influenced the Vietnamese's.[8] Some more peculiar connections emerge from the similarities between the Pythagorean and Christian cultural contexts, for which the Pythagorean scores a 0.62 (unidirectional) similarity. Another important aspect that emerges from this analysis is how many cultural contexts are systematically connected to the same set of cultures. In fact, cultures such as Buddhist, Christian, Japanese, Chinese, Hindu, Greco-Roman, Japanese, Jewish show the most similarities with all the other cultures in terms of embodiment, as to say that, in a unidirectional analysis, all the other cultures tend to be

[6] For these comments I refer to the weighted bidirectional similarity with a threshold of $\geq$ 0.8 considering the average between symbol similarity and symbolic meaning similarity, available at `https://br0ast.github.io/simulationontology/heatmaps/complexi08avgcontinent.html`

[7] Looking at the weighted bidirectional similarity with a threshold of $\geq$ 0.8 considering the average between symbol similarity and symbolic meaning similarity

[8] With this statement, by no means I intend to undervalue the cultural independence of Vietnam from China, as this analyses only measures similarities between cultures according to their symbologies. It is expected that two cultures of geographically close areas influenced each other, and that symbols got transmitted from one to the other

similar to this set. This might be a result of HyperReal containing more data about this set than the rest of the cultures, and might also be a computational confirmation about the existence of universal symbols and symbolic meanings, highlighting the need to investigate more on the matter. Overall, these findings show how quantitative analyses can be used to point out peculiarities that then need to be studied more carefully as a qualitative matter, highlighting the complementarity between qualitative and quantitative analyses.

As for the differences between symbols and symbolic meanings similarities, the investigated cultural contexts, on average, have slightly higher similarities in terms of their symbols than their symbolic meanings. Table 4.1 provides the averages for all the computations.[9] It indicates that cultural contexts tend to be slightly more similar in symbolic meanings only when considering normal bidirectional similarity. For the rest of the computations, symbols have slightly higher averages. The final averages between the similarities of cultural contexts considering symbols or symbolic meanings again show a very little difference, 0.458 for symbols and 0.442 for symbolic meanings. Nevertheless, this result does not mean that there are not pairs of cultures which tends to be significantly more similar on one of the two measurements. In fact, such is the case for the Kalmyk cultural context[10] which shows high variations when looking at similarities based on symbols or symbolic meanings with other cultural contexts. For example, Kalmyk culture scores 0.47 similarity considering symbols with the Islamic culture, but 0 similarity considering symbolic meaning with the same culture. In fact, the mentioned pair of cultures share the symbol of the crescent moon, but in the Kalmyk culture it symbolises the shapes of faces of people from the "eastern continent"[8], meanwhile for the Islamic culture it symbolises death, divinity, resurrection, and Islam itself. Moreover, a horse head in Kalmyk culture symbolises the West. Horses are also used as a symbol in the Islamic culture, but they symbolise happiness and wealth [8].

[9] The averages comprehend all the thresholds of similarity

[10] The Kalmyk people are a Mongolian-origin ethnic group in Russia with a strong cultural identity and focus on nomadic traditions. They practice Mongolian Buddhism and have a rich musical heritage. They have a history of migration, with the majority of the population now living in the Kalmyk Republic and smaller populations in neighbouring countries [139].

Additionally, Kalmyk culture scores 0 similarity with the Pawnee[11] culture regarding symbols, and 0.52 regarding symbolic meanings.[12] In fact, both Kalmyk and Pawnee cultures share the symbolic meanings of cardinal directions, but they use different symbols to symbolise them. For the Kalmyk culture, an elephant head and the white colour symbolise east, a lion head and yellow colour north, an ox head south, and a horse head and colour west [8]; for the Pawnee culture, north is symbolised by a female brown eagle, and south by a male white eagle [8]. Thus, although cultural contexts, averagely, seem not to be more similar in terms of either symbols or symbolic meanings, these two have complementary effects to explaining cultural similarity.

Table 4.1: Comparison of the symbol and symbolic meaning averages of the similarities between pairs of cultural contexts, considering all the types of similarity. **Bold** indicates the higher value.

Type of similarity	Symbol avg	Symbolic Meaning avg
Normal bidirectional	0.366	**0.385**
Normal unidirectional	**0.503**	0.494
Weighted bidirectional	**0.419**	0.393
Weighted unidirectional	**0.543**	0.497
All	**0.458**	0.442

4.1.3 Future work, limitations, and RQ discussion

After analysing the results of this section, quantitative research on symbolism showed potential to highlight uniqueness and similarities between cultural contexts. Several connections emerged between cultures that need further investigation, such the similarities between Pythagorean and Christian culture. Moreover, the evidence

[11] The Pawnee are a Native American tribe from Kansas, Nebraska, and Oklahoma known for their strong spiritual beliefs, community, hospitality, traditional music and art, and distinct language. They have a developed system of agriculture and base their religion on a reverence for nature. Their culture values reciprocity and generosity, seen in their stories, songs, and art [140].

[12] All the previous remarks about the Kalmyk culture are made considering a threshold of ≥ 0.8 and unidirectional weighted similarity computations. The heatmaps that show these similarities are available at `https://br0ast.github.io/simulationontology/heatmaps/complexi08simctx.html` for the symbols and at `https://br0ast.github.io/simulationontology/heatmaps/complexi08rcctx.html` for symbolic meanings. Because this is a unidirectional computation, the results explained above are found in the heatmaps when "culture 1" is Kalmyk and "culture 2" refers to the other cultures mentioned above.

of the symbolic correlations among the large cultural contexts in HyperReal confirms the presence of primitive symbols and symbolic interpretations that are universally shared among a multitude of contexts. It is necessary to carefully scrutinize these results in future studies, making a comparison with more qualitative investigations by scholars such as Wittkower in [141, 142], attempting to identify any patterns or similarities between the primitive symbols found within HyperReal and those discussed by Wittkower.

With this work, we initiate quantitative methods for investigations into the similarities of cultures based on symbolism. We provide evidence for their usefulness as a complement to established comparative cultural studies, and predict that situating our findings within this field will facilitate new discussions. This result is strongly related to RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?* as it proves the potential of linked data of cultural to symbolism to foster research in the field of anthropology and art history. In summary, having linked open data about cultural symbolism not only permits quantitative analysis on the topic, but can be used to highlight peculiarities that can then be investigated more thoroughly in qualitative studies.

4.2 Do colours influence symbolism? A case study of the white rose

This section, explaining a first experiment on the influence of colours in symbolism, was extracted from the paper "Marriage is a Peach and a Chalice: Modelling Cultural Symbolism on the Semantic Web" [33], written by me and my supervisors Aldo Gangemi and Marieke van Erp. I carried out the analysis described below. This work addresses RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?*.

In this work, we show how HyperReal can be used to investigate the different meanings of a symbol. We investigate whether there is a correlation between the symbolic meaning of a simulacrum and its colour. We chose white roses as the starting symbol, as they occur across many cultures and in different contexts [54].

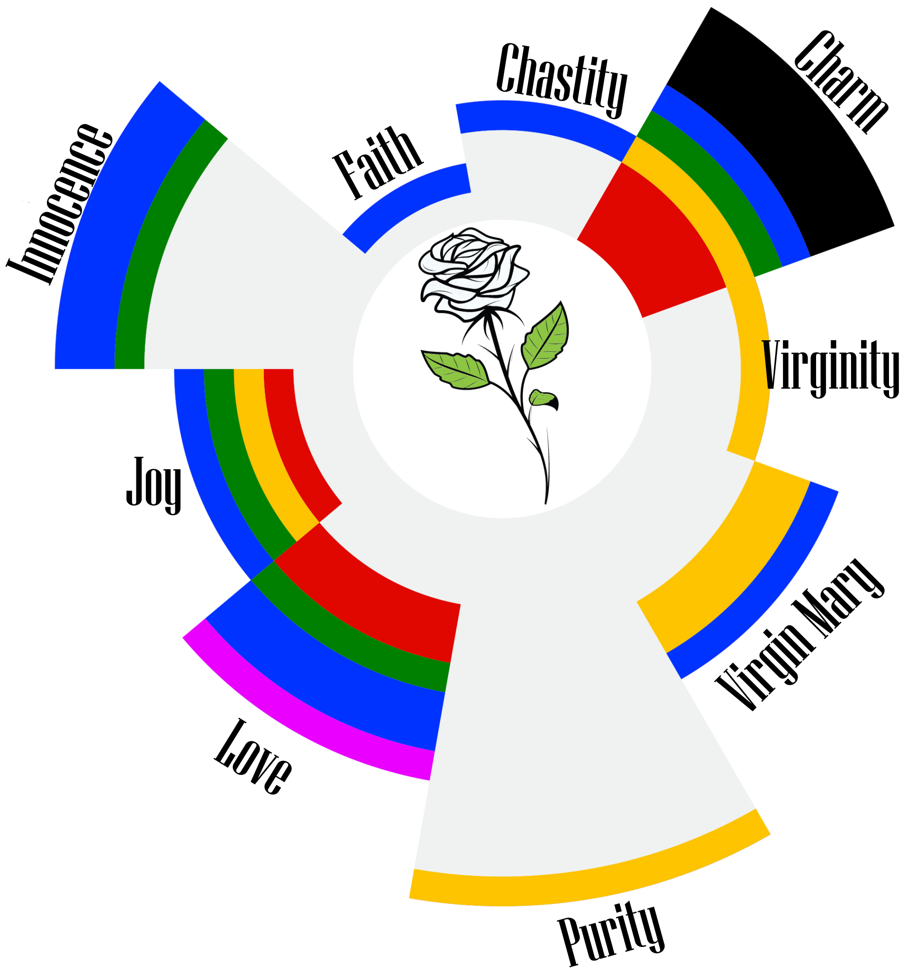

Figure 4.1: Distribution of coloured simulacra across the symbolic meanings of a white rose

4.2.1 Steps of the experiment

We extract the symbolic meanings of a white rose from the knowledge graph through the SPARQL query shown in Listing 4.4. With another series of queries (available on `https://w3id.org/simulation/code`), we extract all simulacra that share a symbolic meaning with a white rose that are associated with the colours red, green, black, white, gold, blue, and purple. We then create nine sets; one for each symbolic meaning of the white rose, and the coloured extracted simulacra were placed in the set they share the symbolic meaning with (in case of more than one matching of symbolic meaning they were placed in more than one set). For instance, one of the symbolic meanings of a white rose is *purity*. *golden hair*, *white knight*, *white swan* share the same symbolic meaning, so they were put in the *purity* set. Only the simulacra's colours were kept, leaving each set with the frequency of colours of simulacra that had the same symbolic meaning.

4.2.2 Discussion on the results

Figure 4.1 shows the distribution of coloured simulacra across the different symbolic meanings of a white rose we could extract from our knowledge graph. Every bar represents a symbolic meaning of the white rose. The length of each colour in a bar is proportional to the number of simulacra of that colour that share the same symbolic meaning. As the figure indicates, the white colour is present in almost all symbolic meanings, with a high peak in *purity* and *innocence*. This suggests that the white colour itself brings these symbolic values to white objects. Moreover, apart from the *charm* symbolic meaning, white has the highest percentage of colour in all meanings of white rose (in *faith* white and blue are split in half). Addressing RQ2, this experiment is an example of how linked data about symbolism could be used for narrow case studies, such as the symbolism of white roses. This study can be used as a template to study the distribution of symbolism across qualities of symbols. In this case, colours were examined, but other elements might be the centre of potential analyses. Another example might be body parts. There are many mentions in HyperReal of symbols that only refer to a specific body part of the symbol, such as the head of a Giant, which is the symbolic attribute of David (king of Judah).

```
PREFIX kb: <https://w3id.org/simulation/data/>
PREFIX sim: <https://w3id.org/simulation/ontology/>

SELECT ?rc WHERE { kb:whiteRose
sim:isSimulacrumOf+/sim:hasRealityCounterpart ?rc .}
```

Listing 4.4: White rose symbolic meaning SPARQL query

4.3 Discovering broad symbolic associations

This section presents a small experiments that was allowed by the entity alignment between HyperReal and Babelnet synsets, WordNet synsets, and DBpedia entities.

In this experiment, I address RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?* by showing the potential of HyperReal to be used in an experiment of broad symbolic associations.

I define broad symbolic associations as the connection between the broader terms of a simulacrum and the broader terms of a reality counterpart. For example, in the simulation rose-love, the broad symbolic association would be flower-emotion. I was able to extract broad symbolic associations thanks to the entity alignments of HyperReal. In fact, as explained in section 2.3.8 aligning the entities of my knowledge graph to corresponding entities and synsets in DBpedia, WordNet and Babelnet, enabled the extraction of their broader terms. Each pair of simulacrum and reality counterpart that are linked to the same simulation can then be replaced by their broader version. Thereafter, I extracted the most frequent pairs of broader symbols and symbolic meaning to study broad symbolic associations. Table 4.3 presents a clear scenario. On one side, the symbols categories are quite diverse, ranging from colours to flora, fauna, people, material, building reptile, on the other side, the most common broad associations are mostly concepts and to a less extent people. `dbr:Concept` is defined by DBpedia as something abstract. Therefore, the data from HyperReal, which reflects to a certain extent conventional symbolic knowledge, suggests that in a symbolic relationship between a symbol and its symbolic meaning, the latter is often abstract, which finds a confirmation in specific studies on the matter [143]. People are the second most common category in broad symbolic associations, which might be linked to the variety of attribute simulations present in HyperReal, that link symbolic attributes to character which might be narrower than `dbr:Person`. Additionally, I developed a function that takes as an input a DBpedia entity that is broader than a set of HyperReal entities, and a role (either simulacrum or reality counterpart), and as an output, it returns the corresponding other elements (i.e, the broader elements of a set of simulacra in case the original input was a reality counterpart, or vice versa) that are more frequently found to pair with the DBpedia entity put as an input. As an example, table 4.2 shows the results of this function using as an input `dbr:Seven_deadly_sins` as the role of reality counterpart. According to HyperReal, when the reality counterpart is narrower than the seven deadly sins (in this case, it means that the reality counterpart is one of the seven deadly sins), then the simulacrum most frequently belongs to broad categories of mammals, adults, reptiles and so on.

At this point, this experiment is still in a preliminary phase, and a more comprehensive version of it will be presented in an academic paper to a suitable venue. Still, regarding RQ2.1, this demonstrates the potential of HyperReal, and more generally, linked open data regarding symbolism, to carry out studies that are intertwined with both anthropology and linguistics.

Table 4.2: Examples of broad simulacra when the reality counterpart is one of the seven deadly sins.

DBpedia Entity for a simulacrum	Freq.
dbr:Mammal	12
dbr:Social_class	8
dbr:Adult	8
dbr:Even-toed_ungulate	7
dbr:Profession	6
dbr:Cleaner	6
dbr:System	5
dbr:Constellation	5
dbr:Reptile	5
dbr:Suidae	5
dbr:Domestic_pig	5
dbr:Pig	5
dbr:Flora	4
dbr:Plant	4
dbr:Person	4

4.4 CYBERNATED: automatiC sYmBolic intErpreter of aRtworks giveN culturAl conTExts and Depictions

This section presents CYBERNATED, a series of algorithms that allows automatic symbolic interpretations of artworks according to the elements that are depicted

Table 4.3: Analysis of the top 25 most frequent pairs of broader symbols and symbolic meanings alignments to DBpedia. "dbr" is a prefix for `http://dbpedia.org/resource/`

Broader Symbol	Broader Symbolic Meaning	Freq.
dbr:Color	dbr:Concept	91
dbr:Symbol	dbr:Concept	85
dbr:Flora	dbr:Concept	72
dbr:Plant	dbr:Concept	60
dbr:Person	dbr:Concept	57
dbr:Symbol	dbr:Person	56
dbr:Tool	dbr:Person	55
dbr:Person	dbr:Person	55
dbr:Card_game	dbr:Concept	53
dbr:Concept	dbr:Concept	52
dbr:Clothing	dbr:Concept	51
dbr:Light	dbr:Concept	50
dbr:Material	dbr:Concept	49
dbr:Flora	dbr:Person	49
dbr:Plant	dbr:Person	49
dbr:Color	dbr:Person	48
dbr:Musical_ensemble	dbr:Concept	48
dbr:Legendary_creature	dbr:Concept	48
dbr:Pack_(canine)	dbr:Concept	47
dbr:Tool	dbr:Concept	47
dbr:Nomenclature	dbr:Concept	47
dbr:Building	dbr:Concept	46
dbr:Reptile	dbr:Concept	46
dbr:Material	dbr:Person	45
dbr:Mammal	dbr:Concept	45

in them. Subsection 4.4.1 describes the matching methodology between external entities and HyperReal's simulacra. Subsection 4.4.2 deals with the inference of the symbolic meanings, and the possible techniques to filter the inferences based on cultural contexts, types of simulations and other specific traits of HyperReal. Subsection 4.4.3 follows, describing the generation of the interpretations from the inferences. Subsection 4.4.4 contains a brief discussion on the potential of CYBERNATED to be used semiautomatically as an aid to organise and filter museums collections through symbolism. Finally, subsection 4.4.5 highlights the weaknesses and strengths of CYBERNATED. For a quantitative evaluation of CYBERNATED and a qualitative evaluation on its potential to detect creator's symbolic intention, refer to section 4.5.

4.4.1 Reconciliation

This step allows me to associate the symbols of HyperReal with the entities depicted in the works of art that are being analysed. I developed a series of algorithms that can aid this reconciliation. First, the labels of the terms can be directly or partially matched to the labels of the symbols of HyperReal. This is generally the approach that provides most matches. However, semantic ambiguity and polysemic words pose significant challenges when using label matching only for reconciling entities. Label matching, which relies solely on comparing the surface form of labels, may lead to inaccurate mappings due to the lack of consideration for underlying semantics and contextual information [144]. Semantic ambiguity refers to situations where a term or phrase can have multiple interpretations or meanings. In the context of entity reconciliation, this means that a label used to represent an entity may be associated with different concepts or entities in different contexts. For example, the term ”Java” can refer to the programming language, an Indonesian island, or a type of coffee. Without considering the semantic context, label matching alone may result in incorrect entity mappings.

Polysemic words further exacerbate the problem.[13] Label matching algorithms typically lack the ability to distinguish between different senses of a polysemic word, leading to potential ambiguities and mismatches in entity reconciliation. A general way to overcome this challenge would be to use Semantic Analysis to detect and exploit the context surrounding the labels that need to be matched [145]. In the case of HyperReal, this solution cannot be adopted due to the synthetic structure of its sources, which, for the most part, lack contextual information. Nevertheless, a more precise reconciliation can be achieved by matching synsets associated to the labels; as HyperReal entities are linked to Babelnet and Wordnet lexical databases, other entities linked to the same synsets can be directly linked to HyperReal. This

[13] Polysemic words are single words that have multiple related meanings. For instance, the word ”bank” can refer to a financial institution or the side of a river.

proved to be useful in the case of the linking to Wikidata, which, for some of its entities, provides alignments to Wordnet and Babelnet. Moreover, having extracted labels from Wordnet, I can also associate the external terms to the Wordnet labels, which are linked to the Wordnet synset, in turn linked to the entities of HyperReal. This methodology can only be adopted if (i) the entities from external sources that undergo reconciliation are as well linked to synsets in Wordnet or Babelnet, or (ii) if it is possible to reconcile them through these lexical databases using the context that surrounds them, and then reconcile them to HyperReal via the matching synsets. Additionally, all these methods can be applied when there is a given list of depicted entities per painting, which is not always the case. In the rest of the work described in this chapter, I will present different scenarios, in which the depicted entities were annotated in the paintings or detected by computer vision algorithms. When such list of depicted entities is not provided, another less precise method can be used. As described in the development of the Multivocal Application described in section 4.7, in case the external sources only contains natural language descriptions of the artworks, a string search can be performed on the descriptions.

4.4.2 Symbolic meaning inferences

Upon connecting the depicted entities of the artworks with their corresponding symbols from HyperReal, it can be inferred automatically that these artworks embody the symbolic meanings of those symbols. Since the simulations in HyperReal are also associated with a cultural context, it is possible to determine in what cultural context an artwork could be interpreted in a particular manner. For example, if painting A contains a fox, it can be inferred, given that a fox-rapacity simulation linked to Jewish context exists, that the painting could be interpreted, from a Jewish perspective, as symbolizing rapacity. It is possible to make CYBERNATED interpret artworks only from certain standpoints, and also restricting the types of simulations that a potential user might be interested in. For example, it could be asked to interpret an artwork only considering potential symbolic attributes of a Christian context, linking it to the Christian characters or figures that are associated to those attributes. Another use case could be the detection of all the protection simulations that are potentially associated to artworks, as to say to interpret only those symbols that are considered a symbolic protection against something. For example, extracting all the artworks that depict a symbolic protection against envy, or inferring only the protection simulations, therefore linking artworks to the symbolic meanings of which they depict a symbolic protection against. These possibilities are expanded even more thanks to the cultural context taxonomy that was created (refer to section 2.3.9 for more information), as it allows users to browse broader and narrower cultural contexts. Consequently, it is possible to perform intricate filtering of the interpretations according to the user's needs.

4.4.3 Generation of the interpretations

Once the inferences are made, CYBERNATED can generate JSON files of interpretations, separated by contexts (if asked), and CSVs. Given that these two formats are quite common as an output, I decided to include another type of output in linked open data. In fact, CYBERNATED can produce as an output of its inferential analysis a LOD version of the interpretations that follows the structure of the ICON ontology. Since we are dealing with a symbolic interpretation, it will produce a series of instances of *iconographical recognitions*, linked to the simulation that is inferred. The interpretation will have as an agent "CYBERNATED", making it possible to distinguish them from other interpretations which might come from another source or are simply manually encoded by the users.

4.4.4 A potential for semi-automation

CYBERNATED could be used as a semi-automatic tool to interpret artworks if used by a domain expert of the art history field, who might want to choose between the proposed interpretations of artworks, which ones are more fitting. At the same time, this series of algorithms could be run on museum collections to cluster the items by their symbolism, allowing museum curators to create exhibitions that follow a symbolic narrative. Considering that museums usually have huge numbers of artefacts between what is exposed and what is stored, being able to automatically analyse their collections and automatically prepare several clusters of symbolism, might accelerate the whole process of curation and exhibition organization. During my PhD, I was unable to test my algorithms on real-case scenarios of museums data, but I have provided some examples of use in section 4.6.

4.4.5 Strengths and Weaknesses

One of the strengths of CYBERNATED lies in the filtering power that allows very precise types of interpretations, thanks to the fine-grained structure of the Simulation Ontology which HyperReal follows. Moreover, the fact that it is completely based on knowledge-driven inference means that it has a high degree of explainability according to the contextual information that are associated with the simulation. By default, all simulations in HyperReal have at least a context (in case of a general context, they are linked to a *General or Unknown* context) and the source of the claim, to ensure the provenance of the information. Potential explanations as to why a certain symbol in a certain culture symbolise a symbolic meaning could be added to the n-ary relationship simulation class if the information is available. One of the weaknesses is that it extremely data dependent, all the interpretation need to exist first in the form of a simulation inside the knowledge graph. Another important aspect that needs to be emphasized is that CYBERNATED does not predict the intention of the creator of the artwork. It just infers symbolic meanings according to the depicted elements, with possible limitations depending on cultural contexts. Nevertheless, an evaluation of its potential to detect the creator's intention is provided in section 4.5, which confirms that as of now, using my methodology, this task is not achievable automatically. In the same section and in section 4.6, CYBERNATED is evaluated quantitatively. All the reflections on CYBERNATED regarding RQ2.1 *To what extent can it [linked open data of cultural symbolism] be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?* will be presented directly in the above-mentioned sections. CYBERNATED is currently being documented and will be released in the future alongside its documentation.

4.5 Quantitative and qualitative evaluation of CYBERNATED on a dataset annotated by computer vision algorithms.

During my PhD, I worked with the researchers of the Saint George on a Bike[14] H2020 project. The purpose of this project is to apply computer vision algorithms to artwork to detect iconographies, depicted entities and potential symbols. In particular, I collaborated with Maria Cristina Marinescu (The principal investigator of Saing George on a Bike) and the scholars Antoine Isaac, Arthem Reshetnikov, Eleftheria Tsoupra, Quim Moré, Cedric Bhihe. We are working on a paper on the enhancement of computer vision detections on artwork with symbolic linked open data to detect symbolism automatically starting from paintings, with the purpose of creating a preliminary visual question answering system that can answer iconographic questions. Their work consists of detecting the depicted entities and their position within the artwork, and my work will provide automatic context dependent inferred interpretations of the artworks according to their detections. The work described in the following section is an evaluation performed on CYBERNATED on a dataset annotated by the computer vision algorithms of the Saint George on Bike researchers. This evaluation is both quantitative, measuring the potential of CYBERNATED to generate and filter interpretations and also qualitative, as the Saint George on a Bike scholars manually evaluated around 1071 interpretations spawned from 11 paintings. We assessed the potential of CYBERNATED to detect the creator's symbolic intention, as to say to detect the symbolic meanings that the creator of the works of art had in mind, by consulting related metadata on the paintings.

4.5.1 Annotated dataset

The annotated data from the Saint George on a Bike project consists of 2,618 paintings spanning between the 14th and the 17th centuries. Their model of object detections is trained to identify 65 classes of potential entities that are typically depicted in artworks. For each painting, they predict which classes are detected in it and

[14] `https://saintgeorgeonabike.eu/`

provide bounding boxes. An example of a prediction is shown listing 4.5.

```
{"00012968.jpg":
        {"classes": ["angel", "nude"],
         "bbxs": [[722.6168599426746,
         1021.8244791030884,
         1227.7735092043877, 1627.5334358215332],
         [637.2079287171364, 427.21840739250183,
         1219.6994082927704, 1072.7770328521729]],
         "scores": [" 96%", " 90%"]
         }
}
```

Listing 4.5: JSON Example of the detection of painting *00012968* ”The Cruxificion” by Jan Provost. In “classes” the detected classes are listed. In “bbxs” the relative positions of the detected classes are listed. In “scores” the confidence scores of the algorithm for each detection are listed.

Table 4.4 shows general information about the annotated dataset.

Table 4.4: Information about the total number of paintings, detected depictions and average number of detected depictions per painting

Dataset	Painting #	Detected Depiction #	Average of detected depictions per painting
Computer vision dataset	2618	5409	2.06

4.5.2 Using CYBERNATED and its filters: a quantitative evaluation

Out of the 65 classes of Saint George on a Bike, 58 could be reconciled to the entities of HyperReal. A first run of CYBERNATED with no filters (meaning that all the possible interpretations are inferred) yielded 120,074 interpretations (1819 unique), for an average of 46.57 interpretations per painting. An example of a JSON output of the interpretation of painting *00012968* can be seen in listing 4.6

```
{"00009378.jpg":
```

```
{"kb:anthonyOfPaduaEspeciallyWithAnUnclothedChrist":
"1",
"kb:johnOfGod": "1",
"kb:ssCatherineOfSienna": "1",
"kb:theCrucifiedChrist": "1",
"kb:theDelphicSibyl": "1",
"kb:theMockingOfJesusChrist": "1",
"kb:ssFrancesXavier": "1",
"kb:theSufferingSavior": "1",
"kb:adversity": "1",
"kb:ignatiusLoyola": "1",
"kb:josephOfArimathea": "1",
"kb:louisIx": "1",
"kb:louisOfFrance": "1",
"kb:martyrs": "1",
"kb:maryMagdalene": "2",
"kb:roseOfLima": "1",
"kb:ssCatherineOfSiena": "1",
"kb:veronica": "1",
"kb:catherineOfAlexandria": "1",
"kb:francisOfAssisi": "1",
"kb:humiliation": "1",
"kb:pain": "1",
"kb:ridicule": "1",
"kb:suffering": "1",
"kb:theAtonement": "1",
"kb:thePassionOfChrist": "1"} }
```


Listing 4.6: JSON Example of the automatic interpretations of painting *00012968* ”The Cruxificion” by Jan Provost. Each automatic interpretation is listed with the number of depicted elements in the painting from which it was inferred

Then, paintings were then associated to the most frequent cultural context(s), *general or unknown* excluded, that was detected in the automatically inferred interpretations. This first filtering stage led to 20,754 interpretation (734 unique). An example of this first filtering applied to painting *00012968* is shown in listing 4.7.


```
{"00009378.jpg": {"suggestedContexts": ["christian"],
"suggestedSimulations": ["kb:crownOfThorns-johnOfGod",
"kb:crownOfThorns-ssCatherineOfSienna",
"kb:crownOfThorns-theCrucifiedChrist",
"kb:crownOfThorns-theMockingOfJesusChrist",
"kb:crownOfThorns-ignatiusLoyola",
"kb:crownOfThorns-josephOfArimathea",
"kb:crownOfThorns-louisIx",
"kb:crownOfThorns-louisOfFrance",
"kb:crownOfThorns-martyrs",
"kb:crownOfThorns-maryMagdalene",
"kb:crucifix-maryMagdalene",
"kb:crownOfThorns-roseOfLima",
"kb:crownOfThorns-ssCatherineOfSiena",
"kb:crownOfThorns-veronica",
"kb:crucifix-anthonyOfPaduaEspecially
WithAnUnclothedChrist",
"kb:crucifix-ssFrancesXavier",
"kb:crucifix-theSufferingSavior",
"kb:crucifix-catherineOfAlexandria",
"kb:crucifix-francisOfAssisi",
"kb:crucifix-humiliation",
"kb:crucifix-pain",
"kb:crucifix-ridicule",
"kb:crucifix-suffering",
"kb:crucifix-theAtonement",
```

```
"kb:crucifix-thePassionOfChrist"]}}
```

Listing 4.7: JSON Example of the automatic interpretations of painting *00012968* ”The Cruxificion” by Jan Provost. Most common context filter is applied.

Automatic interpretations can also be filtered by frequency of symbols that refer to the detected symbolic meaning, under the assumption that a painting that has more than *x* symbols that refer to the same symbolic meaning is more likely to be associated to that symbolic meaning compared to other symbolic meaning which are referred less than *x* times. We ran this filtering with frequency of 2, 3, 4, 5. No interpretations were found with a frequency >4. Table 4.5 shows the results for each frequency value. It is clear from table 4.5 that already with a frequency of at least 2 symbols that refer to the same symbolic meaning, the number of interpretations drops significantly, from 120,074 to 1,795 (-98.5%), together with the number of artworks. In summary, filtering by frequency can significantly reduce the number of automatic interpretations, and, consequently, the number of artworks that are interpreted by CYBERNATED.

Table 4.5: Results for the frequency filtering of the Saint George on a Bike dataset

Data	F >1	F >2	F >3	F >4	F >5
interpretation #	1795	116	12	2	0
painting #	449	72	12	2	0
painting %	17.41	2.8	0.46	0.078	0
filtering %	98.5	99.903	99.99	99.99	100

4.5.3 Qualitative Evaluation on the detection of the creator’s intention

Out of the 120,000 automatic interpretations, 1,074, belonging to 11 paintings, were selected for a manual qualitative evaluation. This evaluation assessed whether CYBERNATED can infer the creator’s symbolic intention. Three annotators evaluated the interpretations, by labelling them as correct (C), incorrect (I) or partially correct (P). The annotators had the possibility of checking descriptions of the paintings and other contextual metadata (creator, date of creation, Iconclass[15] notations) to

[15] `https://iconclass.org/`

aid their evaluation, and could leave the annotation blank in case of uncertain information. After the evaluation stage, Krippendorff's Alpha [146] was computed to measure the agreement between annotators. Although the task of recognizing symbolic meanings and creator's symbolic intention is highly subjective, the annotators showed a considerably high agreement of 0.56.[16] The evaluations were carried out on all the possible interpretations of CYBERNATED. In this way, we could assess a general score of precision for CYBERNATED.[17] Then, the cultural context and frequency filters were applied, which pruned the initial interpretations, and the scores of precision were re-calculated on the new set of filtered interpretations, to measure whether applying filters would increase the possibility to detect a creator's symbolic intention. Precision score was calculated in two ways: (i) the average precision of the annotators was considered, (ii) through a majority vote of the annotations. The initial precision computation led to negative results, as the average precision was 0.13 and the majority vote precision was 0.11. After applying the cultural context filtering, the results slightly improved, as the average precision rose to 0.2 and the majority vote precision increased to 0.17. Then the interpretations were pruned using the frequency filter (F>2). The main issue with this approach is that the filtering gets very strict; for 6 paintings, there are no interpretations that fit this criterium. Additionally, the precision showed no significant increase, as the average precision and majority vote precision slightly rose to 0.18.

4.5.4 General remarks on CYBERNATED evaluation

In respect to the first part of RQ2.1 *To what extent can it be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism?*, from the results in section 4.5.3, it follows that CYBERNATED cannot be used for the task of automatically inferring the creator's symbolic intentions as the scores of accuracy for this task range from 0.12 to 0.2 according to the filtering techniques used. On the other hand, it proves its capability to generate a high number of potential interpretations (with an average of 46.57 interpretations per painting)

[16] Krippendorff's Alpha values go from -1 (complete disagreement) to 1 (complete agreement)

[17] Only precision could be computed, as all the a-priori symbolic meaning of the paintings were not included in the metadata

that can be filtered by minimum frequency or by cultural context. The high number of interpretations offers a good pool of choices for the semi-automatic usage of CYBERNATED, filtering by minimum frequency severely reduces (-98.5% using a frequency of 2) the number of interpretations and allows for a more fine-grained selection of artworks that are linked to certain instances of symbolism. In conclusion, it needs to be mentioned that the potential of CYBERNATED was hampered by the limited number of classes (only 58 that had a match with HyperReal symbols) that could be detected from the computer vision algorithms. In summary, it is possible to generate automatic interpretations utilizing computer vision detections. On a quantitative level, the number of inferred interpretations is related to the number of classes that are detectable by the computer vision system. On a qualitative level, automatic interpretations cannot reliably reflect the creator's will, suggesting the necessity of implementing a semi-automated system for that task.

4.6 ICONdata: highlighting new symbolic connections and artistic serendipitous discoveries in Wikidata

This section is about the re-engineering of iconographical and iconological statements of Wikidata. I first analyse the current limitations of Wikidata in respect to expression of symbolism. Then, I propose an alignment between depicted entities of Wikidata, objects of the predicate wdt:P180 (*depicts*), and classes (they would be an instance of) of the ICON ontology presented in chapter 3. Thereafter, I convert parts of Wikidata into ICONdata, a knowledge graph which contains the original iconographical information of Wikidata described with more granularity. Moreover, I enrich this information by proposing a series of potential symbolic interpretations through knowledge-driven inferences by aligning the depicted entities to HyperReal (described in chapter 2). From the inferred symbolic interpretations, more attention will be given to serendipitous discoveries, defined as all the new connections that emerged between artworks in Wikidata, caused by the shared symbolic meaning

only.[18] ICONdata is still a work in progress, and this preliminary version only contains alignments to pre-iconographical entities (natural elements) and iconographical characters, named objects, places, and stories.[19] Moreover, only entities of type painting (`wd:Q33205213` were re-engineered. The results of this work complete RO4 *Re-engineer iconographic and iconological statements of current knowledge graphs, describing them with the proper granularity using the newly created ontology. Perform inference-based symbolic interpretations on this data to highlight serendipitous connections between artworks based on symbolism.*, and are used to address RQ2.1 *To what extent can it (LOD about cultural symbolism) be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?* and RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved by including such (symbolic) knowledge?*.

4.6.1 Wikidata as a source: current limitations

Wikidata was chosen the source of this re-engineering because of the results of the analysis described in section 1.2 of chapter 1, where it performed best among the knowledge graphs taken into consideration. Nevertheless, it showed various gaps when it comes to the granularity of iconographical and iconological information. In this subsection, I analyse the three different approaches in Wikidata for describing symbolism, assessing how they can currently be used to highlight serendipitous, symbolic connections between artworks.

The property `wdt:P180` (*depicts*) has a central role in all three approaches. It links artworks with their depicted entities, but does not take into consideration the iconographic or iconological level of the representation. The property is used

[18]The last part of the definition is essential because two painting may share a symbolic meaning because they share the symbol that symbolises it. With serendipitous discoveries, I want to emphasize artworks that share the symbolic meaning that is conveyed by different symbols. For example, if Painting *A* and Painting *B* both depict a heart, they will share the potential symbolism of love because they share the same symbol, this would not be a serendipitous discovery. Contrarily, if Painting *A* contains a heart and painting *C* contains a red rose, they share the symbolic meaning of love without sharing the same symbol, which leads to a serendipitous discovery. The concept of serendipity is to be taken in a broader sense, as the "serendipitous" discoveries are hidden at first, and are highlighted only by revealing the symbolic meaning of the depicted entities

[19]These listed elements recall the classes of the ICON ontology presented in chapter 2

336,285 times for paintings (`https://qlever.cs.uni-freiburg.de/wikidata/yKhv77`), considering as paintings all the entities that are linked with the property `wdt:P31` (*instance of*) to the entity `wd:33205213` (*painting*), with 43,888 unique depicted entities (`https://qlever.cs.uni-freiburg.de/wikidata/4n2TkZ`). The first approach regards the qualifier of the *depicts* property, `wdt:P4878` or *symbolize*, which is intended to link the depicted entities to their symbolic meaning. Its use is very limited, and few artworks can be connected through their symbolism. The query in `https://w.wiki/6BZR` shows that out of the 623,728 paintings in Wikidata, only 9 are connected by symbolism for sole the symbol of woman as the Poland nation or as a muse. There are 16 connections, and none of them would be considered serendipitous according to the criterium written above.

As for the second approach, it is possible to extract symbolic connections when looking at the paintings that depict entities which are then linked to what they symbolize through the *symbolize* property, not as a qualifier of the *depicts* property, but as a direct property between the entity and its symbolic meaning, separated from the *depicts* statement (in the explorable web version of Wikidata, this connection appears in the page of the depicted entity, not in the page of the painting). The query available at `https://w.wiki/6BZS` shows that, using this approach, 2287 connections can be extracted between 56 paintings.[20] As query `https://w.wiki/6GXg` shows, none of these connections can be considered as serendipitous.

Finally, the third approach consists of considering depicted elements which are instances of the `wd:Q80071` *symbol* class. As already mentioned, the prop-

[20]Both this and the previous queries were last run on January 2023, results may change following upgrades. Furthermore, there can be multiple connections between the same pair of paintings if they share more than one symbolic meaning. Finally, the numbers retrieved in the queries are doubled from the real results as the same connection from painting A to painting B, is considered distinct from the connection from painting B to painting A. As an example, if painting A depicts an apple which symbolises the original sin, and a dove that symbolises peace, and painting B depicts a snake which symbolises the original sin, and an olive branch symbolising peace, there are two connections. Using the query above, four connections would be shown because both A connected to B because of the original sin, and B connected to A for the same symbolic meaning would be considered, although in reality it is the same connection.

erty `wdt:P31` (*instance of*) is generally used to link elements to the classes they are an instance of.[21] In this case, the symbolism of the depicted entities is expressed using a qualifier of the *instance of* property called *wdt:P642* (*of*). The *of* qualifier represents a specification of the statement it qualifies. In case of the *instance of* property linked to the *symbol* class, through the *of* qualifier it can be expressed what the entity is a symbol of, i.e., its symbolic meaning (as for the previous approach, this information is displayed, in the browser version of Wikidata, in the page of the depicted entity, and not the painting). Using this third approach, as query `https://qlever.cs.uni-freiburg.de/wikidata/QWnqoD` shows, 1412 connections emerge between 77 paintings. Out of those, 30 connections can be considered serendipitous, as the symbols that connect the pairs of paintings to the same symbolic meaning are different. The serendipitous connections emerge from 31 paintings. Figure 4.2 shows an example on how the three different approaches to express symbolism in Wikidata are visualised in the web browser. On the top left, *symbolizes* is used as a qualifier of the *depicts* statement for the untitled painting by Jacek Malczewski, available at `https://www.wikidata.org/entity/Q62027152`, in which a woman symbolizes "the muses". On the top right, *symbolizes* is used as a direct property from the *heart* entity, available at `https://www.wikidata.org/entity/Q826930`, to its symbolic meaning "love". On the bottom left, *of* is used as a qualifier of the *instance of* property that links the *palm branch* entity, available at `https://www.wikidata.org/entity/Q696809`, to its symbolic meaning of "victory", "peace", "eternal life". Regardless of the approach that is used to highlight symbolic connections between artworks, the results on the current state of Wikidata reveal only 30 serendipitous discoveries, as most of the paintings that are linked through symbolism would have been linked regardless by the depicted entities. In other words, as of now, the vast majority of paintings in Wikidata are not linked by shared symbolic meanings, but by shared symbols (and only as a consequence, shared the symbolic meanings). The overall scenario of Wikidata thus presents a twofold issue: on one hand, when the

[21]The term *class* in this case is used in a broader sense, as occasionally the object of this property as the difference in Wikidata between classes and individuals can be inconsistent.

Figure 4.2: Examples of symbolic expressions in Wikidata from its web pages.

data in Wikidata is present, as in the case of the depicted elements (more than 40,000 unique depicted entities), it is not expressed with an adequate granularity (no levels of interpretations); on the other hand, when properties exist to express more complex iconographic information, such as *symbolize*, they are severely underused. The result is that paintings in Wikidata can be queried on a very superficial iconographic and iconological level, and more complex queries that extract symbolism and symbolic connections lead to very scarce results because of data unavailability. Table 4.6 displays the number of connections, serendipitous connections, and unique paintings, symbols, symbolic meanings (that takes part in those connections), that are currently extractable from Wikidata. These numbers will be compared to the results of the serendipitous connection extraction performed in subsection 4.6.6 to address RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved by including such (symbolic) knowledge ?*.

4.6.2 From Wiki to ICON: conversion methodology

This subsection explains how I converted the depicted elements of Wikidata into instances of the `icon:Character`, `icon:Place`, `icon:NaturalElement`, `icon:NamedObject`, and `icon:Story` classes of the ICON ontology. In the first step of this conversion, I extracted the sub-graph that describes artworks, their

Table 4.6: General information on the symbolic connections that emerge from Wikidata using three different approaches.

Method	Artworks #	Connections #	Symbols #	Symbolic Meanings #	Serendipitous Connections #
symbolize as qualifier of *depict* property	9	16	1	2	0
symbolize as a property of a depicted entity	56	2287	3	7	0
of as a qualifier of an *instance of* property linked to a *symbol* class	77	1412	4	7	30
Total	**142**	**3699**	**8**	**16**	**30**

depictions, and provides additional information about those depictions through a query available at `https://qlever.cs.uni-freiburg.de/wikidata/mIehyK`[22] and shown in listing 4.8. I will now break down the query to explain the details of what was extracted. For each painting, all the depicted entities were extracted. If the entities in Wikidata were modelled as individuals, I extracted the class that they are an instance of (using the property `wdt:P31` or *instance of*), else, if the entities were classes themselves[23], I extracted their superclasses (from the property `wdt:P279` or *subclass of*). Moreover, I extracted (if present) all other entities connected to the depictions through the property `wdt:P674`, or *characters*. As explained in sections 4.6.1 and 1.2, Wikidata does not have specific properties to include the recognition of a story in a work of art, as it uses *depicts* for all levels of interpretations. I decided to include the characters linked to the stories depicted in artworks, as they can be described using ICON ontology.[24] Table 4.7 shows total and unique number of entities for each type of extraction. Looking at the numbers of depictions, manually annotating all the unique 43,076 would have been extremely time-consuming. Therefore, I decided to manually annotate only the top 590 most frequent classes that are referred by the depicted entities through the *instance of* property and the top 500 most frequent classes that are referred by the depicted entities through the *subclass of* property. I have annotated whether these classes could contain elements reconcilable to places, named objects, characters, stories, places or natural elements (of pre-iconographical level). With this methodology, I was able to cover almost all the depicted entities, as 88.78% of entities depicted in Wikidata refer to at least one of the annotated classes. Figure 4.3 shows the distributions of the depicted entities into the classes. The histogram clearly shows that the majority of depicted entities can be grouped together into a small set of classes (representing 1 bin of the 8 bin histogram).

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

[22]Query was last run in November 2022, new results might be different

[23]As already mentioned, Wikidata's distinction between classes and individuals is very ambiguous, and the usage of classes and individual is not completely consistent

[24]I also extract other information about qualifiers of the *depicts* property, such as colours, expressions, gestures, worn clothes of the depictions. They will be used for future conversions

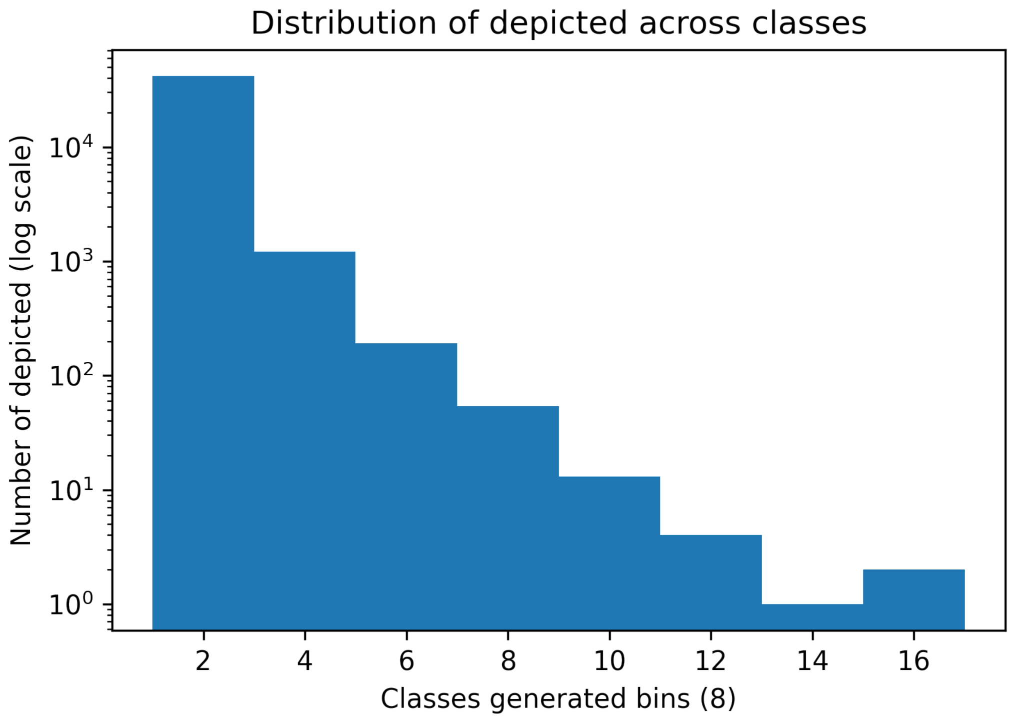

Figure 4.3: Distribution of depicted entities (unique) of Wikidata across their classes, the y-axis has a logarithmic scale

Table 4.7: Summary of the extractions from wikidata

Element	Total #	Unique #
Paintings	130,629	130,629
Depictions	334,407	43,076
Characters (from stories)	4,280	839
Classes (from instance of)	263,959	3,958
Classes (from subclass of)	421,708	5,232

```
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX pq: <http://www.wikidata.org/prop/qualifier/>
PREFIX ps: <http://www.wikidata.org/prop/statement/>
PREFIX p: <http://www.wikidata.org/prop/>
SELECT DISTINCT ?painting ?paintingl ?supclass
?supclassl ?depicted ?depictedl
?wears ?wearsl
```

```
?color ?colorl
?symbol ?symboll
?express ?expressl
?type ?typel
?char ?chart ?charl
?depicted2 ?depicted2l ?depicted2t WHERE {
  ?painting wdt:P31 wd:Q3305213 .
  ?painting p:P180 ?P180node .
  ?P180node ps:P180 ?depicted .
  ?depicted rdfs:label ?depictedl .
  OPTIONAL { ?depicted wdt:P31 ?type .
             ?type rdfs:label ?typel .
             FILTER(lang(?typel) = 'en') }
  OPTIONAL { ?depicted wdt:P279 ?supclass .
             ?supclass rdfs:label ?supclassl .
             FILTER (LANG(?supclassl) = "en") }
  OPTIONAL { ?P180node pq:P3828 ?wears .
             ?wears rdfs:label ?wearsl
             FILTER (lang(?wearsl) = 'en') }
  OPTIONAL { ?P180node pq:P462 ?color .
             ?color rdfs:label ?colorl
             FILTER (lang(?colorl) = 'en') }
  OPTIONAL { ?P180node pq:P4878 ?symbol .
             ?symbol rdfs:label ?symboll .
             FILTER (lang(?symboll) = 'en') }
  OPTIONAL { ?P180node pq:P6022 ?express .
             ?express rdfs:label ?expressl .
             FILTER (lang(?expressl) = 'en') }
  OPTIONAL { ?depicted wdt:P674 ?char .
             ?char rdfs:label ?charl ;
```

```
            wdt:P31 ?chart
            FILTER(lang(?charl) = 'en') }
OPTIONAL { ?depicted wdt:P180 ?depicted2 .
            ?depicted2 rdfs:label ?depicted2l;
            wdt:P31 ?depicted2t
                    FILTER(lang(?depicted2l) = 'en') }
OPTIONAL { ?painting rdfs:label ?paintingl
            FILTER(lang(?paintingl) = 'en') }
FILTER(lang(?depictedl) = 'en')
}
```

Listing 4.8: Query run on Wikidata to extract all the necessary information for the conversion to ICONdata

After the annotation, I started the conversion of Wikidata's *icon* statements to ICONdata by reshaping Wikidata's information into the structure of the ICON ontology. For each depicted entity that was linked to the `icon:NaturalElement` class, the conversion function I developed created an `icon:PreiconographicalRecognition` linked to an `icon:ArtisticMotif`, which was connected to the depicted entity through the property `icon:hasFactualMeaning`. For the depicted entities which were linked to the `icon:Character`, `icon:NamedObject`, `icon:Place` classes, the function generates an `icon:IconographicalRecognition` linked to an `icon:Image`, which is then linked to the depicted entities which can be a character, named object, or a place with the properties `icon:hasCharacter`, `icon:hasNamedObject`, or `icon:hasPlace` depending on the annotation. Finally, for all the depicted elements that were recognized as stories or which were linked in Wikidata with the *characters* property to characters, an `icon:IconographicalRecognition` is generated that is linked to the depicted story through the property `icon:recognizedInvenzione`. Additional characters linked to the depicted elements are included in ICONdata, with the property `icon:composedOf` (the domain of the property is the story, and the range the character).

4.6.3 First conversion results

The initial conversion yielded a total of 216,517 interpretations, out of which, 143,892 are pre-iconographical and 72,625 are iconographical. These numbers are summarised in table 4.8. From this first connection, an important finding already emerges: **in Wikdata, two thirds of the depictions are linked to a pre-iconographic element**.[25] Moreover, as table 4.8 shows, the most prominent iconographic recognition in Wikidata regards characters, the least prominent regards named objects.

4.6.4 Second conversion: automatic symbolic interpretations of Wikidata

During the second conversion phase, I used CYBERNATED to perform entity alignment between the depictions of Wikidata and the symbols in HyperReal. The entity alignment was performed through label matching, and also through the matching of Wordnet and Babelnet synsets, as both Wikidata's entity and my Knowledge Graph symbols, as explained in section 2.3.8, have been aligned to those two linguistic resources. Out of 43,076 depicted entities in Wikidata, 1,063 were aligned to HyperReal's symbols. After the alignment, for each depicted entity in a painting that was aligned to HyperReal, CYBERNATED generated as many `icon:IconographicalRecognition` as the number of simulations that the symbol in HyperReal was linked to, linking them to *image* classes, which were then subsequently linked to the corresponding simulations.

4.6.5 Second conversion results

The second conversion of Wikidata into ICONdata using CYBERNATED yielded 1,979,376 Iconographical interpretations of symbolic meanings. Furthermore, 52,422 paintings (40% of the total) received at least one symbolic interpretation.

[25] This finding will need to be confirmed once all the classes of ICON will be taken into consideration for this conversion

Type of Recognition	Total #	Recognised Characters	Recognised Places	Recognised Named Objects	Recognised Stories
Preiconographical	143,892	X	X	X	X
Iconographical	72,625	59,328	12,039	452	806
total	216,517	X	X	X	X

Table 4.8: Summary of the first conversion of Wikidata into ICONdata

Table 4.9: Number of interpretations generated by each ICON class in ICONdata

ICON Class	Symbolic Iconographic Interpretations #
Character	32387
Natural Element	1944210
Named Object	1831
Place	2556

On average, each painting received around 37 interpretations (average calculated considering only the pool of 52,422 which received at least one interpretation). Table 4.9 shows that most of the symbolic interpretations are generated by natural elements, which again highlights the impact that pre-iconographical recognition have in ICONdata.

4.6.6 Revealing serendipitous connections

I have already defined a symbolic serendipitous connection as a connection between two paintings given by a symbolic meaning that is referred by two different symbols depicted in the pair of paintings. Therefore, a connection that would have not existed only considering depictions. I start this section by explaining the rationale behind the discovering of these connections. Then, I compare the results that I obtain by highlighting serendipitous connections in ICONdata to the connections that are extractable in the current version of Wikidata, explained in section 4.6.1. Finally, I showcase some specific findings of serendipitous connections.

4.6.6.1 A rationale for serendipitous connections: preliminary filtering

There are some preliminary filtering that were adopted to facilitate the detection of serendipitous connections.

Only the set of reality counterparts (viz., symbolic meanings), that I will call `SER-RC`, that appears in more than one simulation can be part of serendipitous symbolic connections; because that means that there are at least two symbols that refer to it.[26] The query to extract these reality counterparts from ICONdata is provided

[26]If this were not the case, a connection based on symbolic meaning would be the same connection based on depicted entities. For example, if pride is part of only one simulation, such as chicken-pride,

in listing 4.9. 1,404 reality counterparts are part of SER-RC

```
prefix sim:<https://w3id.org/simulation/ontology/>
SELECT ?rc (count(distinct ?simu) as ?tot) WHERE
{ ?simu sim:hasRealityCounterpart|
sim:easedRealityCounterpart|
sim:healedRealityCounterpart|
sim:restoredRealityCounterpart|
sim:preventedRealityCounterpart
|sim:elicitedRealityCounterpart ?rc }
GROUP BY ?rc HAVING (?tot > 1)
```

Listing 4.9: Query that extracts the reality counterparts of the SER-RC set

Consequently, only the set of simulations that have a reality counterpart that appears in more than one simulation can be taken into consideration for possible serendipitous discoveries. I call this set SER-SIMT. The query that can extract those simulations from ICONdata is provided in listing 4.10. 6,763 simulations are part of the SER-SIMT set.

```
prefix sim:<https://w3id.org/simulation/ontology/>
prefix sim:<https://w3id.org/simulation/ontology/>
prefix wd:<http://www.wikidata.org/entity/>
prefix wicon:<http://www.example.org/wikicon/>

SELECT DISTINCT ?symbol WHERE {
?ico icon:aboutWorkOfArt ?painting;
     icon:recognizedImage ?img .
        ?img icon:hasSymbol ?symbol .
            ?symbol sim:hasSimulacrum ?simulacra;
            sim:hasRealityCounterpart
            |sim:easedRealityCounterpart
```

that means that no serendipitous connections would emerge from pride because it would be inferred only by the depiction of a chicken.

```
|sim:healedRealityCounterpart
|sim:restoredRealityCounterpart|
sim:preventedRealityCounterpart
|sim:elicitedRealityCounterpart ?rc
. {
SELECT ?rc (COUNT(DISTINCT ?simu) as ?tot)
where { ?simu sim:hasRealityCounterpart
|sim:easedRealityCounterpart|
sim:healedRealityCounterpart
|sim:restoredRealityCounterpart|
sim:preventedRealityCounterpart
|sim:elicitedRealityCounterpart ?rc }
group by ?rc having (?tot > 1)
} }
```

Listing 4.10: Query that extracts the simulations of the SER-SIMT set

Therefore, only the set of simulacra, that I call SER-SIMC, that are part of the simulations of SER-SIMT can be taken into consideration for serendipitous discoveries. The query to extract this set is displayed in listing 4.11. 872 simulacra (82.03% of the originally 1,063 simulacra that were aligned to Wikidata depicted entities) are part of this set.

```
prefix sim:<https://w3id.org/simulation/ontology/>
prefix sim:<https://w3id.org/simulation/ontology/>
prefix wd:<http://www.wikidata.org/entity/>
prefix wicon:<http://www.example.org/wikicon/>

SELECT DISTINCT ?simulacra WHERE {
?ico icon:aboutWorkOfArt ?painting;
     icon:recognizedImage ?img .
        ?img icon:hasSymbol ?symbol .
```

```
            ?symbol sim:hasSimulacrum ?simulacra;
            sim:hasRealityCounterpart
            |sim:easedRealityCounterpart
            |sim:healedRealityCounterpart
            |sim:restoredRealityCounterpart|
            sim:preventedRealityCounterpart
            |sim:elicitedRealityCounterpart ?rc
            . {
            SELECT ?rc (COUNT(DISTINCT ?simu) as ?tot)
            where { ?simu sim:hasRealityCounterpart
            |sim:easedRealityCounterpart|
            sim:healedRealityCounterpart
            |sim:restoredRealityCounterpart|
            sim:preventedRealityCounterpart
            |sim:elicitedRealityCounterpart ?rc }
            group by ?rc having (?tot > 1)
            } }
```

Listing 4.11: Query that extracts the simulacra of the SER-SIMC set

Finally, only the set of paintings, that I call SER-P, that depict at least 1 simulacrum of set SER-SIMT can be part of serendipitous connections. The query to extract this set is displayed in listing 4.12. 48,623 paintings, 92.75% of the paintings that received at least one symbolic iconographic interpretation, are part of this set.

```
prefix sim:<https://w3id.org/simulation/ontology/>
prefix sim:<https://w3id.org/simulation/ontology/>
prefix wd:<http://www.wikidata.org/entity/>
prefix wicon:<http://www.example.org/wikicon/>

SELECT DISTINCT ?painting WHERE {
```

```
?ico icon:aboutWorkOfArt ?painting;
     icon:recognizedImage ?img .
        ?img icon:hasSymbol ?symbol .
            ?symbol sim:hasSimulacrum ?simulacra;
            sim:hasRealityCounterpart
            |sim:easedRealityCounterpart
            |sim:healedRealityCounterpart
            |sim:restoredRealityCounterpart|
            sim:preventedRealityCounterpart
            |sim:elicitedRealityCounterpart ?rc
            . {
            SELECT ?rc (COUNT(DISTINCT ?simu) as ?tot)
            where { ?simu sim:hasRealityCounterpart
            |sim:easedRealityCounterpart|
            sim:healedRealityCounterpart
            |sim:restoredRealityCounterpart|
            sim:preventedRealityCounterpart
            |sim:elicitedRealityCounterpart ?rc }
            group by ?rc having (?tot > 1)
            } }
```

Listing 4.12: Query that extracts the paintings of the `SER-P` set

4.6.6.2 There are more symbolic serendipitous connections in Wikidata than people in Europe

When I first analysed the current content of Wikidata, only 30 serendipitous connections based on symbolism emerged. After converting Wikidata's *icon* statements into ICONdata, and running CYBERNATED on the depicted entities, **1,184,983,553 serendipitous connections emerged, which signifies an increase of 3,949,945,176% from the original data**. This result is extremely impactful for RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved* *by including such knowledge (linked open data about symbolism)?* because it completely broke the "symbolic wall" of Wikidata by exposing more than a billion of serendipitous connections, which remained dormant so far because of the lack of linked open data about cultural symbolism. At the same time, the exploration of this huge number of connections enables (hypothetically) countless new quantitative studies on artworks. Because of this, this result is extremely impactful also for RQ2 *To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic*. Finally, this is significant for RQ2.1 *To what extent can it be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?* as well, because it is a confirmation of the extent of symbolic data to detect serendipitous connections between artworks.

4.6.6.3 Presenting the serendipitous connections of ICONdata

As stated in section 4.6.6.2, the immense number of discovered connections leaves an unparalleled stream of possibilities to explore. In this section, I analyse some of the connections of the most symbolic paintings of Wikidata.

Table 4.10 shows the top 10 of paintings with most interpretations. Some strong correlations emerge from this list. In fact, three paintings, out of which two are the top 2 overall, are about Noah's Ark. Five are about variations of the Earthly Paradise. The remaining two paintings depict animals. Animals are the common factor for the extremely high number of symbolic interpretations. The only painting in which animals are not predominant is "Paradiesgärtlein" by Upper Rhenish master, shown in figure 4.4, which depicts mostly saints, flowers, and plants. This painting is part of 573,623 serendipitous connections, and it is connected to 45,122 paintings out of the 48,623 paintings in `SER-P`. It shares the most connections with the painting that has the most interpretations.

The painting in question is "Entrance into the Ark" by Jan Brueghel the Elder, shown by figure 4.5, (`https://www.wikidata.org/wiki/Q66107722`) with 828 interpretations. This painting alone takes part into 1,183,604 serendipitous connections, and it is connected to 45,584 out of 48,623 possible paintings (con-

Figure 4.4: Paradiesgartlein by Upper Rhenish Master, Public domain, via Wikimedia Commons

tained in SER-P). The painting that shares most connections with (540) is "Hunter and Hounds being Judged" by Paulus Potter (http://www.wikidata.org/entity/Q49842824), which is also part of the top 10.

Looking at the top 10, "The Garden of Earthly Delights" (https://www.wikidata.org/wiki/Q321303) by Hieronymus Bosch deserves a mention for its fame. Shown by figure 4.6, the painting takes part into 728,222 serendipitous connections with 44,962 paintings. It shares 474 connections with "Entrance into the Ark".

These findings indicate that, unsurprisingly, the paintings that received the most interpretations are also the most intertwined in terms of serendipitous connections.

Figure 4.5: Entrance into the Ark by Jan Brueghel the Elder, Public domain, via Wikimedia Commons

Figure 4.6: The Garden of Earthly Delights by Hieronymus Bosch, Public domain, via Wikimedia Commons

Table 4.10: The 10 paintings with most symbolic interpretations in Wikidata accompanied by their number interpretations

Painting	Interpretation #
http://www.wikidata.org/entity/Q66107722	828
http://www.wikidata.org/entity/Q18809786	590
http://www.wikidata.org/entity/Q321303	525
http://www.wikidata.org/entity/Q27980267	443
http://www.wikidata.org/entity/Q463392	418
http://www.wikidata.org/entity/Q29656879	416
http://www.wikidata.org/entity/Q49842824	403
http://www.wikidata.org/entity/Q18917077	399
http://www.wikidata.org/entity/Q20170089	399
http://www.wikidata.org/entity/Q18810005	378

4.6.7 Improving Wikidata as future work

From this analysis, some current gaps of Wikidata have been highlighted. Whereas structural change for Wikidata to distinguish between elements of the first, second, and third level of interpretation would require major changes to its data model, its current gaps on data about conventional symbolic could be filled by extending one of its ways to address symbolism in its data. Of all the approaches presented in section 4.6.1, the second one is the most suitable one to extend Wikidata (while retaining its statement structure) in relation to the Simulation Ontology schema. As a solution, I suggest connecting a symbol to its symbolic meaning using the *symbolize* property and then add the contextual information from the Simulation Ontology (cultural context, provenance, type of simulation, type of reality counterpart relationship) as qualifiers for the *symbolize* property.

4.6.8 Results and discussion

In summary, in this section I presented the conversion of part of Wikidata into ICONdata, by re-engineering its iconographical and iconological statements into the ICON ontology's structure. Then, I explained how CYBERNATED helped to yield around 2,000,000 symbolic interpretations out of the depicted entities in Wikidata. Moreover, I revealed more than 1,000,000,000 serendipitous symbolic connections, defined as connections between pairs of paintings that share a symbolic meaning without depicting the same symbol that symbolises it, which is an increase of almost 4,000,000,000% of the serendipitous connections that can be found currently in Wikidata. These findings give strong answers to RQ2.1 *To what extent can it (linked data about symbolism) be used to detect the potential symbolic meaning of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?* and RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved by including such knowledge (linked data about symbolism)?*. Moreover, with this section I complete the last research objective of the thesis, RO4 *Re-engineer iconographic and iconological statements of current knowledge graphs, describing them with the proper granularity using the newly created ontology. Perform inference-based symbolic interpretations on this data to highlight serendipitous connections between artworks based on symbolism.*. ICONdata is still missing many of the classes of the ICON ontology, such as *Allegory*, *Personification*, *Quality*, *Action*, *Event*. I am currently working on a paper that will describe its complete version. Thereafter, it will be released.

4.7 Building a 3D Multivocal exhibition by connecting artworks to their symbolism

This section describes a project that was developed for the 2022 edition of the HackaLOD hackathon. It was created by me and five other PhD students: Andrei Nesterov, Ryan Brate, Claudia Libbi, Savvina Daniil, and Sarah Binta Sholee. I focussed on the reconciliation between the sources of the data and my knowledge graph HyperReal. We are currently working on a paper detailing this project.

As discussed in chapter 1, LOD is used extensively in the cultural heritage field by providing an unprecedented level of access to information, allowing users to develop applications that can be used to explore and identify relationships between artefacts, places, and people, as well as to trace their origins. By combining data from Wikidata [12], the National Museum of World Cultures digital collection, and HyperReal [33], we developed Multivocal Exhibition, an application that lets users explore how one concept is symbolically expressed by different cultural contexts through a 3D exhibition of artefacts.

4.7.1 Data sources for the app

As Wikidata has already been described thoroughly many times in this thesis, and HyperReal is one of the output of my work, I will briefly describe the remaining source.

The digital collection of the National Museum of World Cultures (NMVW) contains metadata about more than 450,000 items, including ancient artefacts, religious icons, and craftwork from a variety of global cultures. We access this information using a data dump of the collection[27]. It is worth nothing that this the data of NMVW, although being linked data, present the typical issues of current cultural heritage LOD: only the standard metadata of the collection are in linked data, while the information regarding the iconographical content and potential symbolism is chained in unstructured, natural language descriptions.

4.7.2 App development

First, we extracted artefacts' information and what they depict from Wikidata and the NMVW collection. We also extracted information about symbols, their symbolic meanings, and the cultural contexts in which symbols symbolize symbolic meanings from HyperReal. Reconciling the depicted entities of Wikidata with the symbols of HyperReal was not necessary, as they had already been reconciled for the ICONdata re-engineering described in section 4.6. To associate the artefacts in NMVW with the symbols in HyperReal, we translated the latter in Dutch using the Google Translate API[28] and then performed a string search on the whole NMVW collection on artefacts' titles and descriptions. 1,075 HyperReal symbols (translated in Dutch) were found in the NMVW's textual descriptions and titles. CYBERNATED algorithms were run to generate all the potential context-dependent interpretations of the artworks of Wikidata and NMVW. Finally, we developed a graphical user interface (GUI) that let users choose a concept (from the symbolic meanings in HyperReal) and up to three cultural contexts of interest. Thanks to

[27] `https://collectie.wereldculturen.nl/thesaurus/`
[28] https://cloud.google.com/translate/?hl=en

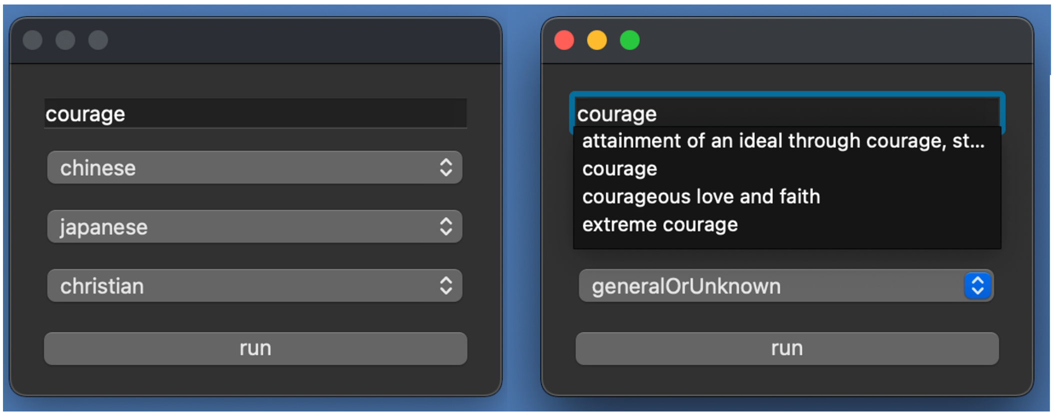

Figure 4.7: GUI of the Multivocal Exhibition App. The concept of courage is being chosen with three cultural contexts of interest

the previous reconciliation, the GUI compiles a JSON file with information about artefacts that depict or contain symbols that symbolize the concept chosen by the users according to the selected cultural contexts. This JSON file then is used as a base for a Unity[29] script that automatically generates a 3-room 3D exhibition of the extracted artefacts, each room dedicated to a different cultural context. In the exhibition, all artefacts come with (i) a description that specifies which symbol(s), in the corresponding culture, refer(s) to the concept chosen by the user, and (ii) a direct link to their metadata pages from Wikidata or NMVW. The purpose of this app is to show how different cultures might interpret artefacts and what they depict according to their symbols, in the hope of fostering an open dialogue about uniqueness and similarities of cultures. Figure 4.7 shows the GUI, and figure 4.8 shows a screenshot of the 3D exhibition working.

4.7.3 Limitation of the App and Future Work

Multivocal exhibition is still a demo. At the current stage, the app allows for the creation of a 9-artefacts, 3-rooms exhibition, explorable in virtual reality by using a laptop.

As future work, we intend to run a user-testing of the app, where small groups of users will experience Multivocal Exhibition together, possibly exploring it with virtual reality visors. We envision collecting their feedback and potential debates

[29] `https://unity.com/`

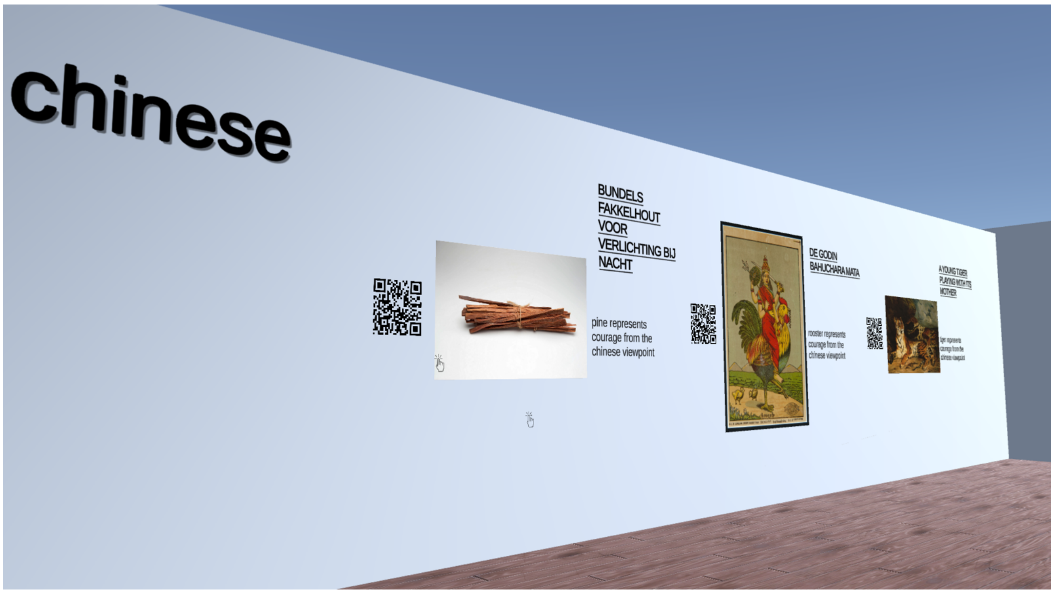

Figure 4.8: 3D Exhibition generated by Multivocal Exhibition App, showcasing artefacts that contain symbols of courage from a Chinese cultural context perspective

that could emerge from the multicultural perspective provided by this app. We hope to stimulate users to share their perspective, from their cultural background, about cultural artefacts and their symbolism.

The potential usage of HyperReal in a museum-like setting was recognised by the jury of the hackathon, who decided to award the Poly-vocality prize to this application. This recognition strengthens the remarks made at the beginning of chapter 1 of this thesis, which emphasized the shift to a more interpretation-oriented definition of museums. Moreover, after the alignments made to HyperReal by CYBERNATED, Wikidata and the NMVW digital collection were connected by symbolism, which relates to RQ2.2 *To what extent can the interconnections of existing knowledge graphs be improved by including this knowledge*. By unchaining the symbols of NMVW data from its natural language descriptions, it was possible to connect it to other KGs through these means. Effectively increasing its interoperability from 0 to an indefinite number, as the serendipitous connections so far were only measured for Wikidata. Nevertheless, considering that 1075 symbols from HyperReal were matched to its dataset, which comprises more than 450,000 artefacts, it can be estimated that the potential number of serendipitous connections that can be exposed in NMWV might be even higher than Wikidata's.

Chapter 5

General Conclusions

In this thesis, I integrated linked open data about cultural conventional symbolism and artistic interpretation in the Semantic Web, creating what I call the "Semantic Web of Symbolic Meanings". I started by assessing the current gaps in the Semantic Web regarding the conceptualisation and expression of symbolism and artistic interpretations. To fill the gaps that I found, I developed two ontologies, the Simulation Ontology and ICON. They model, respectively, conventional cultural symbolism and artistic interpretations. Furthermore, I re-engineered information from heterogeneous sources to create HyperReal, the first knowledge graph completely dedicated to cultural symbolism. These three elements provided the foundations for the Semantic Web of Symbolic Meanings. After filling up the initial gaps, my research shifted to a more empirical direction. I started carrying out quantitative analyses on the data contained in HyperReal, to showcase the potential of linked open data about symbolism in (i) analysing similarities between cultural contexts, (ii) dealing with specific cases of symbolism such as the influence of colours in symbols, and (iii) handling more meta-level topics such as broad symbolic associations. After assessing the potential of the Semantic Web of Symbolic meaning in fostering research in a variety of fields, my research shifted again. In the last part of my thesis, I focus on the development of CYBERNATED, a system that can infer symbolic interpretation of artworks from multiple cultural perspectives by aligning the elements that are depicted in them with the symbols of HyperReal. I evaluate CYBERNATED both quantitatively on the amount of potential interpretation it can generate and its filtering mechanisms, and qualitatively on its ability to detect the artwork's creator's symbolic intention. The evaluation of CYBERNATED culminated with the creation of ICONdata, the re-engineered version of Wikidata's iconographic and iconological statements according to the structure of the ICON ontology. By running CYBERNATED on the depicted entities of artworks in Wikidata, it yielded around 2,000,000 symbolic interpretations, which were ingested in ICONdata. Comparing the serendipitous symbolic connections that existed in Wikidata with ICONdata, I produced an improvement of more than 3,900,000,000% in terms of quantity, exposing more than a billion of undiscovered symbolic connections. The final experiment presented in this thesis regarded the creation of an application that could exploit the newly created knowledge, allowing users to explore how different cultural contexts interpret artworks according to their symbols, and showcasing the potential of the Semantic Web of Symbolic Meanings in a museum-like setting.

5.1 Key resources and key findings

In this section, I revisit my research questions and research objectives by listing a series of key resources and key findings of this thesis. For a better readability, I rewrite the research questions that I presented in section 1.3.

RQ1 To what extent can symbolic relationships in the cultural heritage domain be encoded into computationally ready, structured, semantically linked data?

- RQ1.1 Which artistic, historic, iconographical, and iconological factors need to be considered in the conceptual modelling of symbolism and symbolic interpretation in the cultural heritage domain?
- RQ1.2 To what extent and through which means can unstructured data of symbolism be re-engineered into linked data?

RQ2 To what extent can linked data of cultural symbolism be used to foster quantitative studies on the topic?

- RQ2.1 To what extent can it be used to detect the potential symbolic meanings of artworks and connect them to other works of art through their symbolism? Can the symbolic connections made between artworks lead to serendipitous discoveries?

RQ2.2 To what extent can the interconnections of existing knowledge graphs be improved by including such knowledge?

RO1 Development of an ontology that can describe conventional cultural symbolism.

R02 Creation of a knowledge graph that contains instances of conventional cultural symbolism.

RO2.1 Prove the usefulness of such data by performing significant quantitative analysis on it.

RO3 Development of an ontology, starting from existing models, that can describe artistic interpretations with an adequate degree of granularity defined by the work of renowned art historians.

RO4 Re-engineer iconographic and iconological statements of current knowledge graphs, describing them with the proper granularity using the newly created ontology. Perform inference-based symbolic interpretations on this data to highlight serendipitous connections between artworks based on symbolism.

The key resource and key findings presented below display the code of the corresponding RQ or RO that they address.

5.1.1 Key Resources

RQ1,RQ1.1,RO1 **Simulation Ontology**

Thoroughly explained in chapter 2, this ontology models conventional cultural symbolism by combining a top-down approach from Baudrillard's "Simulacra and Simulation" theory [57], which created the skeleton of the ontology, with a data-driven approach that refined it integrating information from the structure of Olderr's dictionary of symbols [8]. With the combination of these approaches, it was possible to develop an ontology that captures different layers of symbolism. Having 8 specialisations of the main class *Simulation*, and 5 specialisations of the relationship that links a symbolic meaning to a simulation, the ontology was able to answer all the competency questions that tested the extent to which it covered cultural symbolism. Further confirmations come from the fact that its structure could cover ways of expressing symbolism of 4 heterogeneous sources during the development of HyperReal. Contrarily to what was hypothetized in RQ1.1, from a top-down approach, only the field of Semiotics impacted the conceptualisation of the Simulation Ontology. The creation of this resource completed RO1. The Simulation Ontology is available at `https://w3id.org/simulation/ontology`

RQ1.1,RO3 **ICON Ontology**

The ICON ontology conceptualises artistic interpretations. Using a top-down approach focussing on the field of art history, it was modelled after the theory of Panofsky [58, 42, 90, 1]. All the factors that needed to be considered to model this ontology are described in chapter 3. This ontology was essential for the creation of ICONdata, described in 4.6, as the KG reuses the structure of the ontology to describe *icon* information contained in Wikidata with a finer granularity. ICON Ontology is available at `https://w3id.org/icon/ontology`, its development completed RO3.

RQ1.2,RO2 **HyperReal**

HyperReal allowed the foundation of every quantitative analysis of this thesis. It is a knowledge graph that I developed by ingesting symbolic knowledge from Olderr's dictionary of symbols [8], DBpedia, Wordnet, and Otto's encyclopedia of Japanese symbolism [55]. Using the robust Simulation ontology as its data model, it contains more than 40,000 instances of cultural symbolism in the form of *simulations*. A taxonomy created from the cultural contexts of HyperReal is available at `https://w3id.org/simulation/contexts/`. To allow interoperability with other knowledge graph, it was aligned to Worndet, DBpedia, Babelnet.[1] The ingestion of data in HyperReal yielded great results, as the algorithms that automatically converted the major source of its data, the Olderr's dictionary, obtained high scores in precision and recall of the detection and correct encoding of cultural symbolism. Given that HyperReal data has been used in every single experiment of this thesis, and that it spawned results in domains such as linguistics, anthropology, art history, confirms the extent of linked open data about symbolism to foster studies on the topic. Moreover, all the key findings of this thesis confirm this statement. HyperReal is available in a dump at `https://w3id.org/simulation/data`. By creating it, I completed RO2. The alignment-enriched version of HyperReal will be released in the future.

RQ2.1,RO4 **CYBERNATED**

automatiC sYmBolic intErpreter of aRtworks giveN culturAl conTExts and Depictions, or CYBERNATED, is a system that is used to automatically infer symbolic interpretations of artworks from multiple cultural perspectives according to their depicted elements. It has been evaluated quantitatively by assessing its potential to generate interpretation on diverse sets of data. First, it was used to interpret artworks based on the results of computer vision detections. Then, it was also run on the depicted elements that are annotated in Wikidata. Finally, it unchained from natural language descriptions the symbols mentioned in the textual metadata of the Dutch Museum of World Cultures digital collections. For all three datasets, it generated a considerable number of interpretations, averaging 37 per artwork. It was also evaluated qualitatively by making it detect the symbolic intention of artworks' creators. Following the results of a manual evaluation performed in section 4.5, CYBERNATED scores only between 0.12 and 0.2 in precision, which indicates

[1] The alignment mentioned here between HyperReal, DBpedia and Wordnet is not to be confused with the ingestion of their knowledge for its development. As section 2.3 explains, Wordnet and DBpedia's symbolic information were ingested in HyperReal, all the remaining knowledge about Wordnet synsets and DBpedia vocabulary were addressed only during the alignment.

that the system is not yet suited for this type of task. As explained in the section below, CYBERNATED was used extensively to highlight a series of key findings, and to complete RO4.

RQ2.1,RQ2.2,RO4 **ICONdata**

ICONdata is the re-engineering of iconographic and iconological information of Wikidata, described with an adequate granularity. Its development enabled (i) the discovery of many of the key findings that are discussed throughout chapter 4 and in the next section, and (ii) the completion of RO4 alongside the work for the Multivocal Exhibition application described in section 4.7 and summarised below. Its comparisons to Wikidata in section 4.6 extensively address RQ2.2 and RQ2.1. These aspects are also reported as key findings in the next section. ICONdata is still in development, as not all the ICON classes are mapped into Wikidata's depictions. It will be released in the future.

RO4,RQ2.2 **Multivocal Exhibition**

This application combines the power of linked open data about cultural symbolism with the engaging and immersive assets of virtual reality. Its purpose is showing to users how different cultures might interpret artefacts and what they depict according to their symbols, in the hope of fostering an open dialogue about uniqueness and similarities of cultures. The use of CYBERNATED on the Dutch Museum of World Culture digital collection (NMVW) and Wikidata (both used as the sources of artefacts for Multivocal Exhibition), completed, together with the development of ICONdata, RO4. Moreover, addressing RQ2.2, the development of this app showcased the potential to use CYBERNATED to unchain symbolism from natural language description of symbolism, as 1,075 symbols from HyperReal were associated to 450,000 artefacts of the NMVW collection. Multivocal Exhibition was awarded the Polyvocality Award at the 2022 edition of the HackaLOD contest.

5.1.2 Key Findings

RQ2.2 **Semantic Web shows several gaps when it comes to linked data about symbolism**

This finding is explained in sections 1.2, 2.2.5, 3.2. In the first one, it is highlighted that current knowledge graph do not express symbolic and artistic statements with an adequate granularity. In fact, none of the analysed KGs obtains a high score in the structure assessment. Despite not having a positive score in structure, Wikidata performed best in this analysis. For this reason, it was chosen as a candidate for a re-engineering as explained in section 4.6. The other two sections mentioned above discuss the current gaps of Semantic Web in the conceptualization of conventional cultural symbolism and artistic interpretations. Conventional cultural symbolism was almost entirely ignored by the Semantic Web, as there existed only few ontologies that just slightly addressed it. Although they were not conceptualised with an adequate granularity, some ontologies existed that mentioned artistic interpretations. For this reason, ICON ontology reuses many more properties and classes compared to the Simulation Ontology.

RQ2,RO2.1 **There is no significant difference between the similarity of cultural contexts considering symbols or symbolic meanings. Both elements seem complementary in the assessment of potential similarity between cultural contexts** As the results in section 4.1 show, on average, cultures are only slightly more similar on symbols than symbolic meanings, although the difference is only 0.016. Nevertheless, some cultures are significantly more similar in one aspect than the other, such as the Kalmyk-Islamic and Kalmyk-Pawnee pairs.

RQ2,RO2.1 **Similarity between big cultures favours the theory of universal symbolism**

Following the remarks of section 4.1, many cultural contexts share moderate-to-high levels of similarity (0.4-0.8) with the same distinct set of big cultural contexts. This seems to favour the theories of Wittkower about universal symbols [141, 15, 142], although a close comparison between his findings and the findings of my quantitative analyses are needed to confirm this hypothesis.

RQ2,RO2.1 **The *white colour* is extremely influential on the symbolic meanings of the white rose**

Although this is a relatively small finding, 4.2 shows how analyses of linked data about symbolism can reach a fine-grained specificity.

RQ2,RO2.1 **Analyses of broad symbolic associations indicate that while a variety of elements appear included when looking at broader versions of symbols, the same cannot be said for symbolic meanings. In fact, their most frequent broad associations are only abstract concepts and people.**

This finding, explained in 4.3, shows how the entity alignment of HyperReal to Babelnet, DBpedia, and Wordnet enables analyses that explore symbols and symbolic meanings from a linguistics perspective. The experiments on broad associations can be reconciled with more qualitative analyses of [143].

RQ2.1,RO4 **Linked data on cultural symbolism can be used to yield a significant number of interpretations per artwork**

As proven by sections 4.5 and 4.6, the alignment between depicted entities of artworks, including computer vision detections, and linked data on cultural symbolism can produce high numbers of interpretations. An average of around 37 interpretations per artwork were generated across different datasets.

RQ2.1 **Automatic inferences based on linked data on cultural symbolism currently cannot detect the artwork's creator's symbolic intention**

Although the general correctness of interpretations derives from the reliability of the sources, it needs to be highlighted that not all the interpretations of a work of art might be relevant according to the intention of their creator. As the evaluation in section 4.5 showed, it is currently difficult to detect creators' symbolic intentions with the current version of CYBERNATED. The precision for this task ranges between 0.12 and 0.2.

RO4 **Two thirds of Wikidata's depictions belong to a pre-iconographic level of artistic interpretation**
As section 4.6.2 shows, once the *icon* data of a knowledge graph such as Wikidata is described with a proper level of granularity, this enables a series of assessments which also improve its knowledge extraction potential for an art-historical dimension.

RQ2.1,RQ2.2,RO4 **There are more than a billion serendipitous symbolic connections in Wikidata, exposed by the integration of symbolic linked data about cultural symbolism. The new data integration improved the number of serendipitous connection by more than 3,900,000,000%**
This finding showcases what was implied at the beginning of the thesis, that most connections between artworks lie on a deeper, symbolic level which, at the beginning of my work, was not described. By looking at the results of section 4.6.6, not describing the symbolic meanings of works of art in linked open data means ignoring billions of connections.

RQ2.1,RQ2.2,RO4 **Most serendipitous connections are correlated with sceneries of natural life with animals, saints, flowers, and plants**
As the results of section 4.6.6 show, the top 10 of most connected paintings all showed elements belonging to these categories of symbols.

5.2 Current limitations and future work

One of the main limitation of this work is that the inferred interpretations of CYBERNATED are not accompanied by a probability score, currently not computable utilising only knowledge-driven methods. This score would allow for a ranking of most probable interpretations that could be used to enhance CYBERNATED. A possible way to fill this gap would be by training a word2vec similarity model only on texts that deal with cultural symbolism. In this way, it would be possible to measure the similarity between the vectors of simulacra and reality counterparts that belong to the same simulation, and rank the most probable interpretations, using this similarity value. Moreover, the reconciliation system is mostly based on label matching. This, as explained in section 4.4, could pose challenges because of semantic ambiguity of words or polysemic words. Another limitation lies on the absence of chronological connotations in the cultural contexts of HyperReal, which derives from the absence of them from the sources of its data. Some of the cultures that are part of HyperReal existed for thousands of years such as Chinese, Greco-Roman, Japanese; not having temporal connotations limits the exploration of the evolution of symbols from a diachronic perspective within the same cultures, and at the same time makes synchronic types of comparison only possible when considering massive span of centuries, which is why this topic has not been addressed in the thesis. Finally, the top down approach in the development of ICON, the ontology described in chapter 3, was based on authoritative theories of art historians who follow a euro-centric approach to art and icons. Future work could be dedicated to the integration of new types of interpretations from a more diverse range of cultures, considering also new objects of interpretations, such as the calligraphic interpretations in the case of Chinese art [147]. Another improvement to the ICON ontology would be to integrate the art historian-driven structure with a more cognitive-based art interpretation based on aesthetic. The cognitive aesthetic experience of art interpretation has gained significant traction from a linked open data perspective in recent years as user-driven, subjective and emotional aspects of art appreciation are being studied by current EU2020 projects such as SPICE (Social cohesion, Participation, and Inclusion through Cultural Engagement). A network of ontologies was developed in SPICE to model aesthetic and cognitive aspects of art interpretations [148, 149], the subjectivity of these aspects is complementary to the more art history-driven perspective of ICON, derived by the iconographic and iconological theories of Panofksy.

For future work, I intend to add more information about personifications and their attributes. Several scholars have addressed this topic, such as Yassu Okayama, who created the “Ripa Index”[150], a collection of attributes of more than a thousand personifications described by Cesare Ripa in five editions of his "Iconologia"[123].

The Simulation ontology already models the relationships between an entity and its symbolic attributes. Therefore, after pre-processing the information of the Ripa Index, the re-engineering of its content following the Simulation ontology schema can be done automatically. Ingesting knowledge on personifications and their symbolic attributes in HyperReal would open new quantitative investigations and comparisons between symbols and personifications. It would be possible to analyse to what extent symbolic attributes of personifications differ from conventional symbols of the same concept that is personified. Viz., exploring the differences between the attributes of a personification (such as the personification of purity) and the conventional symbols of the same concept (purity).

Additionally, given that HyperReal already describes symbolic attributes of culturally relevant characters (such as deities, mythological and folkloristic characters), parallelisms could be drawn between the iconographic and symbolic representation of personifications and the same representation of other culturally relevant characters, focussing on their symbolic attributes. Several connections already exist and were studied by scholars, such as the attributes of Heracles, which were connected to the attributes of the personification of strength[151]. By quantitatively analysing linked open data on symbolism and personifications, many more serendipitous connections might emerge.

I also intend to integrate references to contemporary symbology in HyperReal, as most of its content describe symbolism that was studied by different scholars who were mostly focused on a historic dimension (which is often not annotated, as discussed in section 5.2). A possible way to include modern and contemporary symbolism in HyperReal would be to survey a heterogeneous group of people (of different age, gender and cultural background) on new symbols and new symbolic meanings that are being referenced in our society. Surveys would be structured to facilitate a conversion to the Simulation Ontology schema, minimizing the pre-processing stage of information as to avoid heavy computations. After the ingestion of the new symbolism into HyperReal, new analyses and parallelism can be built between ancient and modern symbols. Additionally, regarding interpretations, while visual artworks have often been the primary focus of this research, expanding the object of interest to other artefacts can open up new avenues of research. For tangible artefacts, architectural heritage, and cultural objects can be included, for intangible artefacts, music, traditions, and rituals can be included. As an example, HyperReal could be integrated with current knowledge graphs dedicated to intangible heritage, such as olfactory [152] heritage and music heritage [153].

A whole part of the future work will be dedicated to the deployment of applications. As it was done for the Multivocal Exhibition App, the creation of a GUI for CYBERNATED, would allow non-expert users to generate symbolic interpretations of artworks. Moreover, ICONdata could be explored employing APIs that allow the extraction of iconographical and iconological statements from it, and also a GUI that gives the possibility to explore the serendipitous symbolic connections that were highlighted by my analysis. Users might be able to start from a Wikidata artwork of their choice, and then explore how that artworks is connected to others through symbolism or other iconographical and iconological filters, inheriting all the functions that were developed for CYBERNATED.

One final note that I would like to specify, is that at the beginning of my research I started with 30 connections of Wikidata and the extremely overused `dc:subject`. I soon realised that to start quantitative research about symbolism in cultural heritage using linked open data, which was the original starting point for this PhD, I had to fill up countless gaps that (three years ago) prohibited such types of analysis. Eventually, my doctoral research shaped into a series of resources and tools that can enable the type of work that I had in mind at the start of the PhD. Using figurative language, my thesis was developed more horizontally than vertically, in the sense that it created the base to foster quantitative research on the topic. Still, I am delighted that I managed to create chapter 4. I am confident that there is a lot more that can be discovered from my analyses, and countless more experiments could be made. In fact, I believe that I barely scratched the surface of this topic, which gives me exciting ideas for the future. On a conclusive note, hopefully now it will be possible to write a thesis about “Exploring the connections of Works of Art within the Semantic Web of Symbolic Meanings”. As this thesis ends, let the endless stream of research begin.

Acknowledgements

This PhD has been funded by the Emilia Romagna Region (grant agreement no. 462 25/03/2019). I am deeply grateful to my supervisors Marieke van Erp and Aldo Gangemi, who mentored me in this roller coaster of an adventure. I also want to express my gratitude to the reviewers of the papers I published, and the external reviewers of this thesis. I would like to thank the University of Bologna for providing me with tools and spaces for my research, and to the Digital Humanities Laboratory (DHLab) situated within the KNAW Humanities cluster institutes, for having me as a guest researcher for six months.

Appendix A

Toy Datasets

```
@prefix ex: <https://example.org/> .
@prefix sim: <https://w3id.org/simulation/ontology/> .

### Simulations

ex:ashTree-odin a sim:Simulation ;
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:odin ;
    sim:hasContext ex:norse .

ex:ashTree-connection a sim:Simulation ;
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:connection ;
    sim:hasContext ex:celtic .

ex:ashTree-surrender a sim:Simulation ;
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:surrender ;
    sim:hasContext ex:celtic .

ex:olive-fertility a sim:Simulation;
```

```
    sim:hasSimulacrum ex:olive ;
    sim:hasRealityCounterpart ex:fertility ;
    sim:hasContext ex:generalOrUnknown .

ex:rose-love a sim:Simulation;
    sim:hasSimulacrum ex:rose ;
    sim:hasRealityCounterpart ex:love ;
    sim:hasContext ex:flowerLanguage .

ex:rose-beauty a sim:Simulation;
    sim:hasSimulacrum ex:rose ;
    sim:hasRealityCounterpart ex:beauty ;
    sim:hasContext ex:flowerLanguage .



ex:odin-violence a sim:Simulation;
    sim:hasSimulacrum ex:odin ;
    sim:hasRealityCounterpart ex:violence ;
    sim:hasContext ex:norse .

ex:gazelle-beauty a sim:Simulation;
    sim:hasSimulacrum ex:gazelle ;
    sim:hasRealityCounterpart ex:beauty;
    sim:hasContext ex:generalOrUnknown .

### Simulacra

ex:rose a sim:Simulacrum ;
        sim:isSimulacrumOf ex:rose-beauty ,
        ex:rose-love .
```

```
ex:ashTree a sim:Simulacrum ;
        sim:isSimulacrumOf ex:ashTree-odin ,
                           ex:ashTree-connection ,
                           ex:ashTree-surrender .

ex:olive a sim:Simulacrum ;
        sim:isSimulacrumOf ex:olive-fertility .

ex:gazelle a sim:Simulacrum;
        sim:isSimulacrumOf ex:gazelle-beauty .

### Simulacra and Reality Counterparts

ex:odin a sim:Simulacrum, sim:RealityCounterpart ;
        sim:isSimulacrumOf ex:odin-violence ;
        sim:isRealityCounterpartOf ex:ashTree-odin .

### Reality Counterparts only

ex:connection a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf ex:ashTree-connection .

ex:surrender a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf ex:ashTree-surrender .

ex:beauty a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:rose-beauty ,
                                   ex:gazelle-beauty .
```

```
ex:love a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:rose-love .

ex:fertility a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:olive-fertility .

ex:violence a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:odin-violence .



### Contexts

ex:norse a sim:Context ;
        sim:isContextOf ex:ashTree-odin ,
                        ex:odin-violence .

ex:celtic a sim:Context ;
        sim:isContextOf ex:ashTree-connection,
                        ex:ashTree-surrender .

ex:generalOrUnknown a sim:Context ;
        sim:isContextOf ex:olive-fertility ,
                        ex:gazelle-beauty .

ex:flowerLanguage a sim:Context ;
        sim:isContextOf ex:rose-love ,
                        ex:rose-beauty .
```

Listing A.1: Toy Dataset for the evaluation of the CQs of the first SAMOD iteration of the Simulation Ontology

```
@prefix ex: <https://example.org/> .
@prefix sim: <https://w3id.org/simulation/ontology/> .

### Sources

ex:dictionaryOfSymbols1 a sim:Source ;
    sim:isSourceOf ex:ashTree-odin ,
                   ex:ashTree-connection ,
                   ex:ashTree-surrender ,
                   ex:olive-immortality .

ex:dictionaryOfSymbols2 a sim:Source ;
    sim:isSourceOf ex:olive-fertility,
                   ex:rose-love,
                   ex:rose-beauty,
                   ex:damaskRose-freshness .

ex:literarysource1 a sim:Source ;
    sim:isSourceOf ex:man-fire ,
                   ex:giant-manBeforeTheFall .

### Simulations

ex:ashTree-odin a sim:Simulation ;
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:odin ;
    sim:hasSource ex:dictionaryOfSymbols1 ;
    sim:hasContext ex:norse .

ex:ashTree-connection a sim:Simulation ;
```

```
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:connection ;
    sim:hasSource ex:dictionaryOfSymbols1 ;
    sim:hasContext ex:celtic .

ex:ashTree-surrender a sim:Simulation ;
    sim:hasSimulacrum ex:ashTree ;
    sim:hasRealityCounterpart ex:surrender ;
    sim:hasSource ex:dictionaryOfSymbols1 ;
    sim:hasContext ex:celtic .

ex:olive-immortality a sim:Simulation;
    sim:hasSimulacrum ex:olive ;
    sim:hasRealityCounterpart ex:immortality ;
    sim:hasSource ex:dictionaryOfSymbols1 ;
    sim:hasContext ex:generalOrUnknown .

ex:olive-fertility a sim:Simulation;
    sim:hasSimulacrum ex:olive ;
    sim:hasRealityCounterpart ex:fertility ;
    sim:hasSource ex:dictionaryOfSymbols2 ;
    sim:hasContext ex:generalOrUnknown .

ex:rose-love a sim:Simulation;
    sim:hasSimulacrum ex:rose ;
    sim:hasRealityCounterpart ex:love ;
    sim:hasSource ex:dictionaryOfSymbols2 ;
    sim:hasContext ex:flowerLanguage .

ex:rose-beauty a sim:Simulation;
```

```
    sim:hasSimulacrum ex:rose ;
    sim:hasRealityCounterpart ex:beauty ;
    sim:hasSource ex:dictionaryOfSymbols2 ;
    sim:hasContext ex:flowerLanguage .

ex:damaskRose-freshness a sim:Simulation;
    sim:hasSimulacrum ex:damaskRose ;
    sim:hasRealityCounterpart ex:freshness ;
    sim:hasSource ex:dictionaryOfSymbols2 ;
    sim:hasContext ex:flowerLanguage .

ex:giant-manBeforeTheFall a sim:Simulation;
    sim:hasSimulacrum ex:giant ;
    sim:hasRealityCounterpart ex:manBeforeTheFall ;
    sim:hasSource ex:literarySource1 ;
    sim:hasContext ex:generalOrUnknown .

ex:man-fire a sim:Simulation;
    sim:hasSimulacrum ex:man ;
    sim:hasRealityCounterpart ex:fire ;
    sim:hasSource ex:literarySource1 ;
    sim:hasContext ex:generalOrUnknown .

### Simulacra

ex:rose a sim:Simulacrum ;
        sim:isSimulacrumOf ex:rose-beauty ,
                           ex:rose-love ;
        sim:hasVariant ex:damaskRose .
```

```
ex:damaskRose a sim:Simulacrum ;
        sim:isSimulacrumOf ex:damaskRose-freshness ;
        sim:isVariantOf ex:rose .

ex:ashTree a sim:Simulacrum ;
        sim:isSimulacrumOf ex:ashTree-odin ,
                           ex:ashTree-connection ,
                           ex:ashTree-surrender .

ex:olive a sim:Simulacrum ;
        sim:isSimulacrumOf ex:olive-fertility ,
                           ex:olive-immortality .

ex:man a sim:Simulacrum ;
        sim:isSimulacrumOf ex:man-fire ;
        sim:hasVariant ex:manBeforeTheFall .

ex:giant a sim:Simulacrum ;
        sim:isSimulacrumOf ex:giant-manBeforeTheFall .

### Reality Counterparts

ex:odin a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:ashTree-odin .

ex:connection a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf ex:ashTree-connection .

ex:surrender a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf ex:ashTree-surrender .
```

```
ex:beauty a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf ex:rose-beauty .

ex:love a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:rose-love .

ex:fertility a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:olive-fertility .

ex:immortality a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf
    ex:olive-immortality .

ex:freshness a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf
    ex:damaskRose-freshness .

ex:fire a sim:RealityCounterpart ;
        sim:isRealityCounterpartOf ex:man-fire .

ex:manBeforeTheFall a sim:RealityCounterpart ;
    sim:isRealityCounterpartOf
    ex:giant-manBeforeTheFall ;
    sim:isVariantOf ex:man .


### Contexts

ex:norse a sim:Context ;
```

```
        sim:isContextOf ex:ashTree-odin .

ex:celtic a sim:Context ;
        sim:isContextOf ex:ashTree-connection,
                        ex:ashTree-surrender .

ex:generalOrUnknown a sim:Context ;
        sim:isContextOf ex:olive-fertility ,
                        ex:olive-immortality ,
                        ex:man-fire ,
                        ex:giant-manBeforeTheFall .

ex:flowerLanguage a sim:Context ;
        sim:isContextOf ex:rose-love ,
                        ex:rose-beauty ,
                        ex:damaskRose-freshness .
```

Listing A.2: Toy Dataset for the evaluation of the CQs of the second SAMOD iteration of the Simulation Ontology

```
@prefix ex: <https://example.org/> .
@prefix sim: <https://w3id.org/simulation/ontology/> .

### Sources

ex:dictionaryOfSymbols1 a sim:Source ;
    sim:isSourceOf ex:acorn-plague,
                   ex:agate-charm-healthyBlood,
                   ex:aloe-charm-longevity .

ex:dictionaryOfSymbols2 a sim:Source ;
```

```
    sim:isSourceOf ex:acaciaThorns-neith,
                   ex:bird-theGods .

ex:literarySource1 a sim:Source ;
    sim:isSourceOf ex:amberStone-jaundice,
                   ex:ashLeavesInWine-poison .

### Simulations

ex:acorn-plague a sim:ProtectionSimulation ;
    sim:hasSimulacrum ex:acorn ;
    sim:preventedRealityCounterpart ex:plague ;
    sim:hasContext ex:generalOrUnknown ;
    sim:hasSource ex:dictionaryOfSymbols1 .

ex:agate-charm-healthyBlood a sim:Simulation ;
    sim:hasSimulacrum ex:agate ;
    sim:hasRealityCounterpart ex:charm ;
    sim:elicitedRealityCounterpart ex:healthyBlood ;
    sim:hasContext ex:arabian ;
    sim:hasSource ex:dictionaryOfSymbols1 .

ex:aloe-charm-longevity a sim:Simulation ;
    sim:hasSimulacrum ex:aloe ;
    sim:hasRealityCounterpart ex:charm ;
    sim:elicitedRealityCounterpart ex:longevity ;
    sim:hasContext ex:egyptian ;
    sim:hasSource ex:dictionaryOfSymbols1 .

ex:amberStone-jaundice a sim:HealingSimulation ;
```

```
    sim:hasSimulacrum ex:amberStone ;
    sim:healedRealityCounterpart ex:jaundice ;
    sim:hasContext ex:islamic ;
    sim:hasSource ex:literarySource1 .

ex:ashLeavesInWine-poison a sim:HealingSimulation ;
    sim:hasSimulacrum ex:ashLeavesInWine ;
    sim:healedRealityCounterpart ex:poison ;
    sim:hasContext ex:grecoRoman ;
    sim:hasSource ex:literarySource1 .

ex:acaciaThorns-neith a sim:EmblematicSimulation ;
    sim:hasSimulacrum ex:acaciaThorns ;
    sim:hasRealityCounterpart ex:neith ;
    sim:hasContext ex:egyptian ;
    sim:hasSource ex:dictionaryOfSymbols2 .

ex:bird-theGods a sim:ManifestationSimulation ;
    sim:hasSimulacrum ex:bird ;
    sim:hasRealityCounterpart ex:theGods ;
    sim:hasContext ex:hindu ;
    sim:hasSource ex:dictionaryOfSymbols2 .

### Simulacra

ex:bird a sim:Simulacrum;
        sim:isSimulacrumOf ex:bird-theGods .

ex:acaciaThorns a sim:Simulacrum;
        sim:isSimulacrumOf ex:acaciaThorns-neith .
```

```
ex:ashLeavesInWine a sim:Simulacrum;
        sim:isSimulacrumOf ex:ashLeavesInWine-poison .

ex:amberStone a sim:Simulacrum;
        sim:isSimulacrumOf ex:amberStone-jaundice .

ex:aloe a sim:Simulacrum;
        sim:isSimulacrumOf ex:aloe-charm-longevity .

ex:agate a sim:Simulacrum;
    sim:isSimulacrumOf ex:agate-charm-healthyBlood .

ex:acorn a sim:Simulacrum;
        sim:isSimulacrumOf ex:acorn-plague .

### Reality Counterparts

ex:plague a sim:RealityCounterpart;
    sim:isPreventedRealityCounterpartOf
    ex:acorn-plague .

ex:healthyBlood a sim:RealityCounterpart;
    sim:isElicitedRealityCounterpartOf
    ex:agate-charm-healthyBlood .

ex:charm a sim:RealityCounterpart;
        sim:isRealityCounterpartOf
        ex:agate-charm-healthyBlood,
        ex:aloe-charm-longevity .
```

```
ex:longevity a sim:RealityCounterpart;
        sim:isElicitedRealityCounterpartOf
        ex:aloe-charm-longevity .

ex:jaundice a sim:RealityCounterpart;
        sim:isHealedRealityCounterpartOf
        ex:amberStone-jaundice .

ex:poison a sim:RealityCounterpart;
        sim:isHealedRealityCounterpartOf
        ex:ashLeavesInWine-poison .

ex:neith a sim:RealityCounterpart;
    sim:isRealityCounterpartOf ex:acaciaThorns-neith .

ex:theGods a sim:RealityCounterpart;
    sim:isRealityCounterpartOf ex:bird-theGods .
```

Listing A.3: Toy Dataset for the evaluation of the CQs of the third SAMOD iteration of the Simulation Ontology

Bibliography

[1] Erwin Panofsky. *Meaning in the visual art : papers in and on art history*. Doubleday Anchor book, 59. Anchor Books, New York, 1955.

[2] ICOM. ICOM International Statuten | ICOM. http://icom-oesterreich.at/page/icom-international-statuten, August 2007. Vienna.

[3] ICOM. Museum Definition. https://icom.museum/en/resources/standards-guidelines/museum-definition/, August 2022.

[4] Dieter Fensel, Umutcan Şimşek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler. *Introduction: What Is a Knowledge Graph?*, pages 1–10. Springer International Publishing, Cham, 2020.

[5] Brian McBride. *The Resource Description Framework (RDF) and its Vocabulary Description Language RDFS*, pages 51–65. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[6] Michael Uschold and Michael Grüninger. Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11, 01 1996.

[7] Bruno Sartini and Aldo Gangemi. Towards the unchaining of symbolism from knowledge graphs: how symbolic relationships can link cultures. In *Book of extended abstracts of the 10th national AIUCD conference.*, pages 576–580, Pisa, 2021. AIUCD.

[8] Steven Olderr. *Symbolism a Comprehensive Dictionary*. McFarland, London, 2 edition, 2012.

[9] Helene Roberts. *Encyclopedia of Comparative Iconography : Themes Depicted in Works of Art*. Routledge, Miton Park, United Kingdom, September 2013.

[10] Sofia Baroncini, Bruno Sartini, Marieke van Erp, Francesca Tomasi, and Aldo Gangemi. Is dc:subject enough? a landscape on iconography and iconology statements of knowledge graphs in the semantic web. *Journal of Documentation*, 79(7):115–136, 2023.

[11] Edie Davis and Bahareh Heravi. Linked data and cultural heritage: A systematic review of participation, collaboration, and motivation. *J. Comput. Cult. Herit.*, 14(2), may 2021.

[12] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledge base. *Commun. ACM*, 57(10):78–85, sep 2014.

[13] Valentina Anita Carriero, Aldo Gangemi, Maria Letizia Mancinelli, Ludovica Marinucci, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. Arco: The italian cultural heritage knowledge graph. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web – ISWC 2019*, pages 36–52, Cham, 2019. Springer International Publishing.

[14] Marilena Daquino, Francesca Mambelli, Silvio Peroni, Francesca Tomasi, and Fabio Vitali. Enhancing semantic expressivity in the cultural heritage domain: Exposing the zeri photo archive as linked open data. *J. Comput. Cult. Herit.*, 10(4), jul 2017.

[15] Rudolf Wittkower. *Allegory and the migration of symbols*. Thames & Hudson, London, England, February 1987.

[16] Aby M Warburg. *The renewal of pagan antiquity - contributions to the cultural history of the European renaissance*. Texts & documents. Getty Publications, Los Angeles, CA, August 1999.

[17] Bruno Sartini and Aldo Gangemi. Towards the unchaining of symbolism from knowledge graphs: how symbolic relationships can link cultures. In *Book of extended abstracts of the 10th national AIUCD conference. AIUCD, Pisa*, pages 576–580, 2021.

[18] Valentina Janev, Damien Graux, Hajira Jabeen, and Emanuel Sallinger. *Knowledge graphs and big data processing*. Springer Nature, 2020.

[19] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514, 2022.

[20] Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8:489–508, 2017.

[21] Maria Angela Pellegrino, Vittorio Scarano, and Carmine Spagnuolo. Move cultural heritage knowledge graphs in everyone's pocket. *Semantic Web*, 14(2):323–359, 2023.

[22] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Soeren Auer. Quality assessment for linked data: A survey. *Semantic Web*, 7(1):63–93, 2016.

[23] Michael Färber, Frederic Bartscherer, Carsten Menne, and Achim Rettinger. Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web*, 9(1):77–129, 2018.

[24] Behshid Behkamal, Mohsen Kahani, Ebrahim Bagheri, and Zoran Jeremic. A metrics-driven approach for quality assessment of linked open data. *Journal of theoretical and applied electronic commerce research*, 9(2):64–79, 2014.

[25] Daniel Ringler and Heiko Paulheim. One knowledge graph to rule them all? analyzing the differences between dbpedia, yago, wikidata & co. In

Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz), pages 366–372. Springer, 2017.

[26] Nicolas Heist, Sven Hertling, Daniel Ringler, and Heiko Paulheim. Knowledge graphs on the web - an overview. In *Knowledge Graphs for eXplainable Artificial Intelligence*, 2020.

[27] Kartik Shenoy, Filip Ilievski, Daniel Garijo, Daniel Schwabe, and Pedro Szekely. A study of the quality of wikidata. *Journal of Web Semantics*, 72:100679, 2022.

[28] Nuno Freire and Antoine Isaac. Wikidata's linked data for cultural heritage digital resources: An evaluation based on the europeana data model. In *International Conference on Dublin Core and Metadata Applications*, pages 59–68, 2020.

[29] Antoine Isaac and Bernhard Haslhofer. Europeana linked open data – data.europeana.eu. *Semant. Web*, 4(3):291–297, jul 2013.

[30] Subhi Issa, Onaopepo Adekunle, Fayçal Hamdi, Samira Si-Said Cherfi, Michel Dumontier, and Amrapali Zaveri. Knowledge graph completeness: A systematic literature review. *IEEE Access*, 9:31322–31339, 2021.

[31] Tong Ruan, Yang Li, Haofen Wang, and Liang Zhao. From queriability to informativity, assessing "quality in use" of dbpedia and yago. In *European Semantic Web Conference*, pages 52–68. Springer, 2016.

[32] S. Baroncini, M. Daquino, and F. Tomasi. Modelling art interpretation and meaning. a data model for describing iconology and iconography, 2021.

[33] Bruno Sartini, Marieke van Erp, and Aldo Gangemi. Marriage is a peach and a chalice: Modelling cultural symbolism on the semantic web. In *Proceedings of the 11th on Knowledge Capture Conference*, K-CAP '21, page 201–208, New York, NY, USA, 2021. Association for Computing Machinery.

[34] Gillian Rose. *Visual methodologies*. SAGE Publications, Thousand Oaks, CA, January 2001.

[35] Laurie 1941 Adams. *The methodologies of art: an introduction*. Westview Press, 2nd edition, 2010.

[36] Marion G Müller. Iconography and iconology as a visual method and approach. In *The SAGE Handbook of Visual Research Methods*, pages 283–297. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, 2014.

[37] Orietta Rossi Pinelli. *La storia delle storie dell'arte*. Einaudi, Turin, Italy, 3 edition, 2019.

[38] Roelof van Straten. *An introduction to iconography*. Routledge, London, England, 2 edition, October 2012.

[39] Matthew Rampley. From symbol to allegory: Aby warburg's theory of art. *Art Bulletin*, 79(1):41–55, March 1997.

[40] Patricia Harpring, Elisa Lanzi, Linda McRae, et al. *Cataloging cultural objects: A guide to describing cultural works and their images*. American Library Association, 2006.

[41] Maja Žumer, Marcia Lei Zeng, and Athena Salaba. *FRSAD: conceptual modeling of aboutness*. ABC-CLIO, 2012.

[42] Erwin Panofsky. *Studies in iconology: humanistic themes in the art of the Renaissance*. Westview Press, Boulder, Colo., 1972.

[43] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

[44] Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. Yago: A multilingual knowledge base from

wikipedia, wordnet, and geonames. In *International semantic web conference*, pages 177–185. Springer, 2016.

[45] Antonis Bikakis, Eero Hyvönen, Stéphane Jean, Béatrice Markhoff, and Alessandro Mosca. Editorial: Special issue on semantic web for cultural heritage. *Semant. Web*, 12(2):163–167, January 2021.

[46] Iztok Savnik, Kiyoshi Nitta, Riste Skrekovski, and Nikolaus Augsten. Statistics of knowledge graphs based on the conceptual schema. *CoRR*, abs/2109.09391, 2021.

[47] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, March 2016. Number: 1 Publisher: Nature Publishing Group.

[48] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In Karl Aberer, Key-Sun Choi, Natasha Noy, Dean Allemang, Kyung-Il Lee, Lyndon Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux, editors, *The Semantic Web*, pages 722–735, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[49] Csaba Fülöp, Gergő Kiss, László Kovács, and András Micsik. Using a metadata schema registry in the national digital data archive of hungary. In *Research and Advanced Technology for Digital Libraries*, Lecture notes in computer science, pages 314–322. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[50] Richard Y Wang and Diane M Strong. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33, 1996.

[51] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[52] Michael Uschold and Michael Grüninger. Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11, 01 1996.

[53] E H Gombrich. Icones symbolicae: The visual image in neo-platonic thought. *J. Warburg Courtauld Inst.*, 11(1):163–192, January 1948.

[54] Steven Olderr. *Reverse symbolism dictionary: symbols listed by subject*. McFarland, Jefferson, N.C., 2005. OCLC: 156738525.

[55] Alexander Francis Otto. *Mythological Japan: The Symbolisms of Mythology in Relation to Japanese Art, with Illustrations Drawn in Japan, by Native Artists - Primar*. Bibliolife DBA of Bibilio Bazaar II LLC, Charleston, SC, USA, 2014. Google-Books-ID: 0Ki8oAEACAAJ.

[56] Umberto Eco. *Semiotica e filosofia del linguaggio*. Einaudi, Turin, Italy, 1996. Google-Books-ID: pYFvAAAACAAJ.

[57] Jean Baudrillard. *Simulacra and Simulation*. University of Michigan Press, Ann Arbor, MI, USA, 1981.

[58] Erwin Panofsky. *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*. Harper & Row, New York, NY, USA, 1962.

[59] Erwin Panofsky. *Studies in iconology: humanistic themes in the art of the Renaissance*. Westview Press, Boulder, Colo., 1972. OCLC: 1151325340.

[60] J. G. Frazer. *The Golden Bough: A Study in Magic and Religion*. Palgrave Macmillan UK, London, United Kingdom, 1922.

[61] Shawn R. Tucker, editor. *The Virtues and Vices in the Arts: A Sourcebook*. The Lutterworth Press, 1 edition, 2015.

[62] Henri L. Joly. *Legend in Japanese art; a description of historical episodes, legendary characters, folk-lore myths, religious symbolism*. London ; New York : John Lane, The Bodley Head, 1908.

[63] Silvio Peroni. A simplified agile methodology for ontology development. In *Proceedings of the 13th OWL: Experiences and Directions Workshop and 5th OWL reasoner evaluation workshop (OWLED-ORE 2016)*, pages 55–69, Cham, Switzerland, 01 2016. Springer.

[64] Valentina Presutti, Enrico Daga, Aldo Gangemi, and Eva Blomqvist. Extreme design with content ontology design patterns. In *Proceedings of the 2009 International Conference on Ontology Patterns - Volume 516*, WOP'09, pages 83–97, Aachen, DEU, 2009. CEUR-WS.org.

[65] Cambridge University Press. lion. `https://dictionary.cambridge.org/dictionary/english/lion`, 2021.

[66] P. Hitzler, A. Gangemi, K. Janowicz, A. A. Krisnadhi, and V. Presutti. *Ontology Engineering with Ontology Design Patterns: Foundations and Applications*. IOS Press, Amsterdam, the Netherlands, 2016.

[67] Timothy Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. Prov-o: The prov ontology, 2013.

[68] Matthew Horridge, Nick Drummond, John Goodwin, Alan Rector, Robert Stevens, and Hai Wang. The manchester owl syntax. In *Proceedings of the*

*OWLED*06 Workshop on OWL: Experiences and Directions*, Athens, GA, USA, 10–11 November 2006. CEUR-WS.

[69] Daniel Garijo, Oscar Corcho, and María Poveda-Villalón. FOOPS!: An ontology pitfall scanner for the fair principles. In *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021)*, Online, 11 2021. CEUR-WS.

[70] Daniel Garijo and María Poveda-Villalón. Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web. *Studies on the Semantic Web*, 49:39–54, 2020. Publisher: IOS Press.

[71] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 10(2):7–34, 2014.

[72] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, September 2014.

[73] L. D. Couprie. Iconclass: an iconographic classification system. *Art libraries journal*, 8:32–49, 1983.

[74] H. van de Waal, Rudi H. Fuchs, L. D. Couprie, and E. Tholen. *Iconclass : an iconographic classification system / H. van de Waal ; completed and edited by L. D. Couprie with R. H. Fuchs, E. Tholen*. North-Holland Pub. Co Amsterdam, 1973.

[75] C. Bekiari, G. Bruseker, M. Doerr, C.-E. Ore, S. Stead, and A. Velios. Definition of the cidoc conceptual reference model v7.1.1, 2021.

[76] Nicola Carboni and Livio de Luca. An Ontological Approach to the Description of Visual and Iconographical Representations. *Heritage*, 2(2):1191–1210, 2019.

[77] S. Baroncini, M. Daquino, and F. Tomasi. Modelling art interpretation and meaning. a data model for describing iconology and iconography, 2021.

[78] Richard Gartner. Towards an ontology-based iconography. *Digital Scholarship in the Humanities*, 35(1):43–53, April 2020.

[79] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. A Bradford Book, Cambridge, MA, USA, May 1998.

[80] Ferdinand de Saussure. *Course in General Linguistics*. Columbia University Press, June 2011. Pages: 336 Pages.

[81] Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, and Zhe Wang. Hermit: An owl 2 reasoner. *Journal of Automated Reasoning*, 53:245–269, 10 2014.

[82] Andrea Moro, Alessandro Raganato, and Roberto Navigli. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244, 2014.

[83] John P. McCrae, Dennis Spohr, and Philipp Cimiano. Linking lexical resources and ontologies on the semantic web with lemon. In *Extended Semantic Web Conference*, 2011.

[84] Pierre-Yves Vandenbussche, Ghislain Atemezing, María Poveda-Villalón, and Bernard Vatant. Linked open vocabularies (lov): A gateway to reusable semantic vocabularies on the web. *Semantic Web*, 8:437–452, 01 2017.

[85] Daniel Garijo. Widoco: a wizard for documenting ontologies. In *International Semantic Web Conference*, pages 94–102. Springer, Cham, 2017.

[86] Bruno Sartini, Sofia Baroncini, Marieke van Erp, Francesca Tomasi, and Aldo Gangemi. Icon: An ontology for comprehensive artistic interpretations. *J. Comput. Cult. Herit.*, apr 2023. Just Accepted.

[87] Stecker Robert. *Artworks : Meaning, Definition, Value.* Penn State University Press, 1996.

[88] T. Editors of Encyclopaedia Britannica. Art. `https://www.britannica.com/art/visual-arts`, 2020.

[89] Kerr Houston. *An introduction to art criticism: histories, strategies, voices*. Pearson, Boston, 2013. OCLC: 900357124.

[90] Erwin Panofsky and Fritz Saxl. Classical mythology in mediaeval art. *Metrop. Mus. Stud.*, 4(2):228, March 1933.

[91] Donald Preziosi. *The art of art history : a critical anthology*. Oxford history of art. Oxford University Press, Oxford, 2009.

[92] S. Baroncini, M. Daquino, and F. Tomasi. Modelling art interpretation and meaning. a data model for describing iconology and iconography, 2021.

[93] Max Imdahl. Iconica. l'intuizione delle immagini. *Aisthesis. Pratiche, linguaggi e saperi dell'estetico*, 2012.

[94] Marilena Daquino, Valentina Pasqual, and Francesca Tomasi. Knowledge representation of digital hermeneutics of archival and literary sources. *JLIS. it, Italian Journal of Library and Information Science*, 11(3):59–77, 2020.

[95] Elena Simperl. Reusing ontologies on the semantic web: A feasibility study. *Data & Knowledge Engineering*, 68(10):905–925, 2009.

[96] S. Stead, M. Doerr, C.E. Ore, and A. Kritsotaki. Crminf: The argumentation model. an extension of cidoc-crm to support argumentation (0.10.1). Technical report, ICOM/CIDOC Documentation Standards Group, 2019.

[97] M. Doerr, A. Kritsotaki, Y. Rousakis, G. Hiebel, and M. Theodoridou. Cdefinition of the crmsci: An extension of cidoc-crm to support scientific observation (1.3). Technical report, ICOM/CIDOC Documentation Standards Group, 2021.

[98] Martin Doerr, Chryssoula Bekiari, Athina Kritsotaki, Gerald Hiebel, and Maria Theodoridou. Modelling scientific activities: proposal for a global schema for integrating metadata about scientific observation. In *Access and understanding–networking in the digital era: The 6th annual conference of CIDOC, the International Committee for Documentation of ICOM, Dresden, Germany*, 2014.

[99] Ingo Frank. Multi-perspectival representation of historical reality. In *JOWO*, 2019.

[100] Willem Robert Van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. Design and use of the simple event model (sem). *Journal of Web Semantics*, 9(2):128–136, 2011.

[101] M Van Ruymbeke, P Hallot, and R Billen. Enhancing CIDOC-CRM and compatible models with the concept of multiple interpretation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, IV-2/W2:287–294, August 2017.

[102] Aldo Gangemi and Peter Mika. Understanding the semantic web through descriptions and situations. In Robert Meersman, Zahir Tari, and Douglas C. Schmidt, editors, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 689–706, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[103] Marilena Daquino and Francesca Tomasi. Historical context ontology (hico): a conceptual model for describing context information of cultural heritage objects. In *Research Conference on Metadata and Semantics Research*, pages 424–436. Springer, 2015.

[104] Davide Picca, Alfio Massimiliano Gliozzo, and Aldo Gangemi. Lmm: an owl-dl metamodel to represent heterogeneous lexical knowledge. In *LREC*, 2008.

[105] Umberto Eco. *Trattato di semiotica generale*. La Nave di Teseo Editore spa, 2016.

[106] Luc Steels. *Foundations for Meaning and Understanding in Human-centric AI*. Venice International University, 2022.

[107] Fabio Ciotti. Toward a formal ontology for narrative. *MATLIT: Materialidades da Literatura*, 4(1):29–44, 2016.

[108] Mischa M Tuffield, Dave E Millard, and Nigel R Shadbolt. Ontological approaches to modelling narrative. In *Proceedings of 2nd AKT DTA Symposium*, 2006.

[109] Rossana Damiano and Antonio Lieto. Ontological Representations of Narratives: a Case Study on Stories and Actions. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 76–93, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[110] Lei Xu, Albert Merono-Penuela, Zhisheng Huang, and Frank Van Harmelen. An ontology model for narrative image annotation in the field of cultural heritage. In *Proceedings of the 2nd Workshop on Humanities in the Semantic web (WHiSe)*, pages 15–26, 2017.

[111] Patricia Harpring. Development of the getty vocabularies: Aat, tgn, ulan, and cona. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1):67–72, 2010.

[112] Chris Dijkshoorn, Lizzy Jongma, Lora Aroyo, Jacco Van Ossenbruggen, Guus Schreiber, Wesley Ter Weele, and Jan Wielemaker. The rijksmuseum collection as linked data. *Semantic Web*, 9(2):221–230, 2018.

[113] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. How to write and use the ontology requirements specification document. In *On the Move to Meaningful Internet Systems: OTM 2009*, Lecture notes in computer science, pages 966–982. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[114] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with dolce. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 166–181. Springer, 2002.

[115] David Shotton. Cito, the citation typing ontology. *Journal of biomedical semantics*, 1(1):1–18, 2010.

[116] Stuart Snydman, Robert Sanderson, and Tom Cramer. The international image interoperability framework (iiif): A community & technology approach for web-based images. In *Archiving Conference*, pages 16–21. Society for Imaging Science and Technology, 2015.

[117] Valentina Carriero, Marilena Daquino, Aldo Gangemi, Andrea Nuzzolese, Silvio Peroni, Valentina Presutti, and Francesca Tomasi. The landscape of ontology reuse approaches. *Appl. Practices Ontol. Des., Extraction, Reason*, 49:21, 2020.

[118] Inès Osman, Sadok Ben Yahia, and Gayo Diallo. Ontology integration: approaches and challenging issues. *Information Fusion*, 71:38–63, 2021.

[119] Pat Riva and Maja Zumer. Frbroo, the ifla library reference model, and now lrmoo : a circle of development. In *Proceedings of the IFLA WLIC 2018 Conference*, 2017.

[120] Stefano Borgo, Roberta Ferrario, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Daniele Porello, Emilio M. Sanfilippo, and Laure: Vieu. Dolce: A descriptive ontology for linguistic and cognitive engineering. *Applied Ontology*, 17(1):45–69, 2022.

[121] Valentina Anita Carriero, Aldo Gangemi, Maria Letizia Mancinelli, Andrea Giovanni Nuzzolese, Valentina Presutti, and Chiara Veninata. Pattern-based design applied to cultural heritage knowledge graphs. *Semantic Web Journal*, 12(2):313–357, 2021.

[122] Remigio Sabbadini, Georgius Thilo, and ERMANNUS HAGEN. Servii grammatici qui feruntur in vergilii carmina commentarii. *Rivista di Filologia e di Istruzione Classica*, 16:322, 1888.

[123] Cesare Ripa. *Iconologia del cavaliere Cesare Ripa, perugino*, volume 3. nella stamperia di Piergiovanni Costantini, 1764.

[124] María Poveda-Villalón, Paola Espinoza-Arias, Daniel Garijo, and Oscar Corcho. Coming to terms with fair ontologies. In C. Maria Keet and Michel Dumontier, editors, *Knowledge Engineering and Knowledge Management*, pages 255–270, Cham, 2020. Springer International Publishing.

[125] Birte Glimm, Ian Horrocks, Boris Motik, Giorgos Stoilos, and Zhe Wang. Hermit: An owl 2 reasoner. *Journal of Automated Reasoning*, 53(3):245–269, Oct 2014.

[126] Marilena Daquino, Valentina Pasqual, Francesca Tomasi, and Fabio Vitali. Expressing without asserting in the arts. In *Proceedings of the Italian Research Conference on Digital Libraries*, Padova, Italy, 2021.

[127] Bruno Sartini, Valentin Vogelmann, Marieke van Erp, and Aldo Gangemi. Comparing symbolism across asian cultural contexts. In *Digital Humanities 2022 Conference Abstracts*, pages 358–361. DH2022 Local Organizing Committee, 2022.

[128] A. L. Kroeber and Clyde Kluckhohn. *Culture: A Critical Review of Concepts and Definitions*. Peabody Museum Press, Cambridge, Massachusetts, 1952.

[129] Richard W. Brislin. Comparative research methodology: Cross-cultural studies. *International Journal of Psychology*, 11(3):215–229, 1976.

[130] Mobinā Zavvāri and Amirhossein Chitsāziyān. A comparative study on the symbolism in turkmen and baluch ornaments in iran. *Journal of Iranian Handicrafts Studies*, 3(2):121–130, 2021.

[131] Ian Manners. Symbolism in European integration. *Comparative European Politics*, 9(3):243–268, July 2011.

[132] Laura Rival. *The social life of trees: anthropological perspectives on tree symbolism*. Routledge, 2021.

[133] Lina Yuan and Yunling Sun. A comparative study between chinese and western dragon culture in cross-cultural communication. In *Proceedings of the 6th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2020)*, pages 74–77. Atlantis Press, 2021.

[134] Stephen J. Reno. Religious symbolism: A plea for a comparative approach. *Folklore*, 88(1):76–85, 1977.

[135] Teresa Duarte Martinho. Researching culture through big data: Computational engineering and the human and social sciences. *Social Sciences*, 7(12), 2018.

[136] Steven Tötösy de Zepetnek. From comparative literature today toward comparative cultural studies. *CLCWeb: Comparative Literature and Culture*, 1(3):2, 1999.

[137] Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto. Wikipedia2Vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from Wikipedia. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 23–30, Online, October 2020. Association for Computational Linguistics.

[138] Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579, 1901.

[139] Elza-Bair Mataskovna Gouchinova. *The Kalmyks*. Routledge, 2013.

[140] Gene Weltfish. *The lost universe: Pawnee life and culture*, volume 653. U of Nebraska Press, 1977.

[141] Rudolf. Wittkower. *Allegory and the migration of symbols / Rudolf Wittkower*. The collected essays of Rudolf Wittkower ; 3. Thames and Hudson, London, 1987.

[142] Rudolf Wittkower. Eagle and serpent. a study in the migration of symbols. *Journal of the Warburg Institute*, 2(4):293–325, 1939.

[143] Gary Lupyan and Bodo Winter. Language is more abstract than you think, or, why aren't languages more iconic? *Philos Trans R Soc Lond B Biol Sci*, 373(1752), August 2018.

[144] Hanna Köpcke and Erhard Rahm. Frameworks for entity matching: A comparison. *Data & Knowledge Engineering*, 69(2):197–210, 2010.

[145] Jeffrey Parsons and Yair Wand. *Attribute-Based Semantic Reconciliation of Multiple Data Sources*, pages 21–47. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[146] Klaus Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[147] Valentina Pasqual, Marta R. Bisceglia, Adriana Iezzi, Martina Merenda, and Francesca Tomasi. Modeling chinese contemporary calligraphy: the write dataset. In *Digital Humanities 2022 Conference Abstracts*. DH2022 Local Organizing Committee, 2022.

[148] Enrico Daga, Luigi Asprino, Rossana Damiano, Marilena Daquino, Belen Diaz Agudo, Aldo Gangemi, Tsvi Kuflik, Antonio Lieto, Mark Maguire, Anna Maria Marras, Delfina Martinez Pandiani, Paul Mulholland, Silvio Peroni, Sofia Pescarin, and Alan Wecker. Integrating citizen experiences in cultural heritage archives: Requirements, state of the art, and challenges. *J. Comput. Cult. Herit.*, 15(1), jan 2022.

[149] Jason Carvalho, Enrico Daga, Luigi Asprino, Aldo Gangemi, Mark Maguire, Paul Mulholland, Adam Stoneman, and Chukwudi Uwasomba. Spice linked

data hub - data infrastructure and tools to support citizen curation. In *ExICE - Extended Intelligence for Cultural Engagement*, March 2023.

[150] Y. Okayama. *The Ripa Index: Personifications and Their Attributes in Five Editions of the Iconologia*. Davaco Publishers, 1992.

[151] Amy C. Smith. Eurymedon and the evolution of political personifications in the early classical period. *The Journal of Hellenic Studies*, 119:128–141, 1999.

[152] Pasquale Lisena, Daniel Schwabe, Marieke van Erp, Raphaël Troncy, William Tullett, Inger Leemans, Lizzie Marx, and Sofia Colette Ehrich. Capturing the semantics of smell: The odeuropa data model for olfactory heritage information. In Paul Groth, Maria-Esther Vidal, Fabian Suchanek, Pedro Szekley, Pavan Kapanipathi, Catia Pesquita, Hala Skaf-Molli, and Minna Tamper, editors, *The Semantic Web*, pages 387–405, Cham, 2022. Springer International Publishing.

[153] Jacopo de Berardinis, Albert Meroño Peñuela, Andrea Poltronieri, and Valentina Presutti. The harmonic memory: A knowledge graph of harmonic patterns as a trustworthy framework for computational creativity. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 3873–3882, New York, NY, USA, 2023. Association for Computing Machinery.