CHAPTER 8

Evaluation

This chapter illustrates how the resources created in the doctoral project, i.e., the Iconology Dataset and the ICON Ontology, were evaluated.

The dataset is evaluated with respect to its data quality, its ability to answer domain research questions, and its compliance with the FAIR principles. The ontology is tested through 1) competency questions performed on a subset of the Iconology dataset, 2) a comparison with other ontologies, and 3) an automatic evaluation of its syntax correctness, logical consistency, and compliance with FAIR principles.

8.1 Dataset evaluation

In the vision of the semantic web, to guarantee the correct interoperability of data, a crucial aspect is the quality assessment of published datasets. This led to consistent data validation and verification techniques available in the literature (Zaveri et al., 2016), as illustrated in Section 3.6.1. For the quality assessment of the Iconology Dataset, we performed two types of evaluation. First, we evaluated the dataset regarding the intrinsic, contextual, representational, and accessibility aspects by performing a selection and adaptation of the metrics proposed by Färber et al. (2018). This ensures that the dataset is acceptably complete for what concerns the content, error-free from a formal point of view, consistent, accessible, and interlinked. In particular, the evaluation of data consistency against the constraints of the used ontologies was a crucial point, for which we decided to perform a more thorough evaluation through SHACL shapes. The second step of the evaluation involves assessing the dataset's adherence to the FAIR principles through online self-assessment surveys.

8.1.1 Metrics

For the quality assessment on the Iconology Dataset, we adopt the applicable metrics of the study by Färber et al. (2018), as, compared with the state of the art presented in Section 3.6.1, it provides a clear and complete methodology for the evaluation of data quality based on the available literature. They propose a total of 34 criteria, divided into 11 dimensions, grouped into four categories. 23 of the 34 criteria were adopted, as

Published in: Sofia Baroncini, Diving into artworks interpretations through the lenses of semantic data. An Application on Panofsky's Iconological Studies, Heidelberg: arthistoricum.net 2026, https://doi.org/10.11588/arthistoricum.1462

they were compatible with the dataset. All criteria score 0 to 1, where 1 corresponds to a high-quality grade concerning the selected metric. We sometimes specialized the metrics to suit the dataset-specific features better. All re-used metrics are listed in Table 8.1. For each, the adopted metric and the results are provided.[1] In the following sections, we better describe the metrics for each category and their eventual refinement, providing motivations for each score. The overall result of all the adopted metrics is 0.84, showing that the Iconology dataset has a good quality grade in several aspects.

8.1.1.1 Intrinsic category

Intrinsic data quality is the first category taken into account. It denotes the inherent quality of the data itself, namely its context-independent characteristics. The intrinsic category includes three dimensions: accuracy, trustworthiness, and consistency.

Accuracy aims at assessing the formal correctness of the data, considering the aspects of syntax and semantics. Indeed, it verifies that the data respect RDF syntactic rules and that the literals adhere to standards. In our case, we verified whether the dates, expressed as literals, comply with the ISO 8601 standard. The semantic validity metric aims to assess whether the content expressed by literals is semantically sound. As the information provided in the Iconology dataset is extracted from Panofsky's bibliographic references, we adapted the metric to the specific case. In detail, we extracted the information concerning a sample of five randomly selected artworks, and we verified if the content of triples, including literals, holds true by comparing them with the information provided in the bibliographic reference. We focussed on 1) triples that include the artwork label to verify the correctness of artwork metadata and 2) triples describing the subjects recognized in the artworks to verify if the actual content was correct. All the metrics of this dimension performed well, with an average of 1.

The second dimension, namely **trustworthiness**, aims at assessing if the information provided in the dataset is correct, true, real, and credible. Färber et al. (2018) formalize this dimension into two aspects: trustworthiness on the Knowledge Graph level and the statement level. Although a further metric dedicated to the representation of unknown and empty values is provided, we do not apply it to the current dataset evaluation, as this type of data is not included in it. The first metric focuses on the adopted data curation and insertion methods. Specifically, it considers whether 1) data were manually inserted by experts, community contributors, or automatically extracted, and 2) whether the operation took place in an open or closed system. As the Iconology Dataset was manually created and curated in a closed system, the score 1 is assigned. The second metric is then used to assess the presence of the provenance assertion. As iconographic and iconological statements are documented through the ICON ontology relations, this metric gets 1 as a score.

1 A notebook with the analysis performed is available in the project repository

The **consistency** dimension evaluates whether data in a dataset conflict with each other. As introduced in Chapter 4, due to this dimension's relevance in the literature, we dedicate a specific evaluation of it through SHACL to verify if instances are compliant with the logical constraints established at the ontology level. The SHACL shapes were extracted from the ontologies used (e.g., ICON, CIDOC-CRM, RDFS, DOLCE, CiTO, PROV-O, etc.) through the free online tool provided by Astrea[2](Cimmino et al., 2020). Among reused ontologies, FOAF, XML Schema, and DCAT-AP_IT were excluded from SHACL shapes generation, as their prefix was not regularly resolved by the tool, that is, no RDF was retrieved when the ontology prefix was passed.[3]

The dataset was validated against the obtained definition through the tool ELI Validator,[4] created in the context of the European Legislation Identifier project. Validation was performed iteratively on the data, which were consequently fixed according to the results.

In the last iteration, some errors remained unsolved. An error is that instances of Allegory and Story do not have the type of superclass `Invenzione` declared, a specification which, according to the author, would make data less readable. Second, the ICON ontology requires that the range of the property `icon:recognizedConcept` is a `dul:SocialObject`. As, in the context of the current modeling framework, we extensively reuse CIDOC-CRM for the description of fundamental aspects of the included objects (i.e., artworks, books, people), we decided to use the closest entity of CIDOC, namely crm:E28_Conceptual_Object, rather than the DOLCE's one. Another error is that the dataset does not declare the inverse property of relation `pro:hasRoleInTime`. To avoid lower performance in data retrieval, we included the inverse properties only of ICON, which is the ontology that describes the core aspects of iconography and iconology.

The overall score performed by the dataset for the intrinsic category is 1.

8.1.1.2 Contextual category

Contextual data quality underscores the need to evaluate data quality within the specific context in which data are used. According to Färber et al. (2018), it includes the dimensions of relevance, completeness, and timeliness. For the evaluation of the dataset, we adopt the only dimension of completeness. As data mainly reflect the art interpretations expressed by a single art historian in a determined range of time, the dimensions of relevancy, which evaluates the presence of a ranking system of divergent assertions, and timeliness, which evaluates if the data are updated for a successful usage for the task at hand, are not applied.

2 https://astrea.linkeddata.es/

3 the generated SHACL shapes are available in the project's repository

4 https://webgate.ec.europa.eu/eli-validator/validate

Completeness expresses the degree to which the data is comprehensive enough to achieve the desired goal required by the specific context in which they are used. Färber et al. (2018, p. 8) identifies three metrics for this dimension, namely schema completeness, column completeness, and population completeness. As the metric of population completeness verifies whether the considered KG covers the basic population, we consider it not applicable to this case, in which a selection of Panofsky's claims was chosen for the dataset population as a case study.

Schema completeness determines whether the adopted schema can fully describe the domain of interest by comparing its classes and attributes to a golden standard. As the dataset focuses on a detailed description of iconographical and iconological interpretations, we adopt as a gold standard the golden example schema formulated in Section 3.6.3.2 (see Figure 3.1), which is already used to evaluate the completeness of the schema of currently available KGs in relation to the *icon* statements. Similarly, we adopt the same criterion, consisting of 1) the alignment of the properties used in the KG to the gold schema and 2) the schema completeness as the division between the number of properties of the example that have been aligned and the total number of properties in the example. As all the properties were aligned, the score obtained is 1.

Column completeness determines whether the attributes of classes defined at the ontology level are present in the instance level of the Knowledge Graph. We specified this metric for the core classes of the dataset, namely the Human-made Objects (E22), Artworks, and the four types of Recognition. As the relations describing them intend to provide means for a more thorough description, when necessary, we narrow the scope solely to the relations considered necessary or highly recommended. The final score is calculated by dividing the number of instances of each specific class that occur with the property considered by the total number of instances of that class. The dataset performed 0.85 as a score for this metric.

The overall score performed by the dataset for the intrinsic category is 0.83.

8.1.1.3 Representational category

Representational data quality encompasses elements tied to data format and its meaning. It comprises two key dimensions, viz. the ease of understanding by a human interpreter and interoperability, focusing on machine-readability.

Ease of Understanding indicates how easily data can be understood by a human data consumer without any confusion or ambiguity. To this end, the criteria of 1) description of resources, 2) labels in multiple languages, 3) provision of an understandable RDF serialization and 4) the presence of self-describing URIs are used to determine whether the data comply with this dimension (Färber et al., 2018, p. 10).

The first criterion evaluates whether the described resources present a textual description specified through the relations `rdfs:comment`, `rdfs:label` or `schema:description`, and it is calculated by dividing the number of resources that have such description by the total number of resources.

The second criterion aims at verifying if textual descriptions are provided in multiple languages, an aspect that would improve the ease of understanding of the dataset. As we did not provide labels in other languages, this metric performs 0.

According to Färber et al. (2018), some RDF serializations are more comprehensible to a human than the RDF/XML standard. Therefore, the third metric assesses whether data are provided in other serialization formats. Since the dataset is also released in Turtle format, this metric gives the maximum score.

The last metric of the ease of understanding dimension concerns the presence of self-describing URIs. In fact, their use radically increases the understandability of data. Although extensive use of self-describing URIs was made in the creation of the Iconology Dataset (see Section 6.3.4.1), the characteristics and needs of some resources prevented their use in all cases. The unique identification of Cultural Objects and artwork was a fundamental aspect, which would not have been guaranteed by the use of self-describing URIs as artworks may have the same descriptive metadata (e.g., title, artist). Second, the length of the description provided for Cultural Phenomena prevented the creation of self-describing URIs based on their label, making necessary the creation of an alphanumeric string to identify them. For these aspects, the result for this metric is 0.5.

The overall score obtained by this dimension is 0.58.

Interoperability focuses on evaluating the representation of data from a technical perspective. Therefore, it considers whether the data are concise and whether they can be easily integrated with other data. These aspects are evaluated through 1) the presence of blank nodes or reification classes, 2) the provisioning of several serialization formats, 3) the use of external vocabulary, and 4) the interoperability of proprietary vocabulary. As the internal vocabulary is not defined, we consider the last criterion to be not applicable in our case.

Blank nodes and reification classes are necessary when further information is provided about the assertion. However, they may prevent straightforward data operations, such as data integration and data processing on the client side. As both CIDOC-CRM and ICON extensively use reification to provide provenance information about the assertion, but the use of blank nodes is avoided, this metric scores 0.5.

Secondly, since we followed best practices for data publishing, we provided data in two formats, Turtle and RDF/XML (score: 1).
Since no classes and relations were internally defined in the dataset, but all the vocabulary used belongs to standards or newly-created ontologies that are externally defined and aligned with standards, all the terms used for resource description are taken from external vocabularies. Therefore, the external vocabulary usage metric performs 1 as a score.

The interoperability dimension scored an overall score of 0.83.

8.1.1.4 Accessibility category

The final element of the assessment is accessibility, which looks at how data can be obtained. It is divided into the dimensions of accessibility, licensing, and interlinking.

The dimension of **accessibility** measures the degree to which data can be accessed or obtained quickly and easily. In the following paragraphs, we elaborate more on the adopted criteria.

The dereferencing possibility of resources examines whether the URIs of RDF resources are accessible and return RDF information. To this end, we extracted a sample of 20 URIs, equally divided into internally created and external to the dataset, and checked whether RDF information was returned. As a result, all the URIs of the sample were deferenceable. The availability of the KG verifies the Graph availability over time. Since it has just been released, we can not evaluate this criterion by performing the metric proposed by Färber et al. (2018). Instead, we consider whether the dataset is hosted on a server that guarantees its availability over time. As the DH.arc University research center hosts the dataset, we consider this metric satisfied. The criterion *provisioning of public SPARQL endpoints* checks whether a public SPARQL endpoint is provided. Since we provide a SPARQL endpoint,[5] this metric scored 1. Similarly, since the dataset is provided in two formats, the criterion *provisioning of an RDF export* also obtains the maximum score. The *support of content negotiation* metric verifies whether the endpoint supports different data formats during the content negotiation request and whether the server provides the requested content type. The ICON endpoint supports several data formats (e.g., JSON, CSV, XML/RDF, TURTLE), but it does not return results accordingly, so this metric scores 0.5. Finally, as structured metadata are provided,[6] the criterion *provisioning of metadata about a KG* obtained the maximum score. The overall accessibility score is 0.9.

The **licensing** dimension aims at assessing whether an appropriate license for reuse is provided. Since we published it under a Creative Common license, specified in the data through the relation `dct:license`, this metric obtained 1 as a score.

The **interlinking** dimension aims to assess whether entities expressing the same concept are linked to external sources to allow the exploration of further knowledge. It includes two criteria, viz. 1) the interlinking through `owl:sameAs` and 2) the validity of external URIs. The first checks the amount of resources linked to external ones through the specified property. To calculate it, we selected only actual resources (i.e., we filtered out the reification classes). We then calculated the total of resources that have at least an `owl:sameAs` property and divided it by the total of resources. As we performed a semi-automatic alignment and every entity was manually checked

5 As the SPARQL endpoint is hosted on a university server, it is considered stable. Since the dataset has small dimensions, no timeout issues have been experienced.

6 https://w3id.org/icon/data

before being added to the dataset, we do not perform further metrics of precision and recall of the alignment algorithm. Although resource alignment is a central part of the development of the dataset, few resources were aligned with this property (score: 0.13). The main reason for this is the lack of respective sources in other KGs because of the dataset's high granularity and domain specificity.

Linking to external resources is a pillar of the semantic web. However, such links can become obsolete and not be available anymore. Hence, the last criterion in the accessibility section verifies whether the external links included in the dataset are active. To this end, we selected a sample of 25 URLs representative of the aligned main source. In fact, the sample consisted of 10 URLs of pictures (`foaf:picture`), 5 from Iconclass, and 5 from Getty Vocabularies. This criterion performed well (score: 1).

Table 8.1: Performed evaluation metrics

Name	Metrics	Result
Intrinsic category		
Accuracy	Data accuracy is defined as the degree to which the dataset is exact, reliable and free of errors (Färber et al., 2018, p. 4)	
Syntactic Validity of RDF Documents	Validation through TurtleValidator[7]	1
Syntactic Validity of Literals	Check if dates comply with the ISO 8601 standard through regular expressions	1
Semantic Validity of Triples	Adapted metric: select triples of a sample of 5 artworks that have literals and verify if they are true by comparing them with the textual source from which the information was extracted or with the artwork picture	0.99
Accuracy average		**1**
Trustworthiness	The degree to which the information is accepted as correct, true, real, and credible. Trustworthy applications need to be able to verify the origin of data.	
Trustworthiness on KG level	Metrics: 1: manual data curation, manual data insertion in a closed system, 0.75: manual data curation and insertion, both by a community, 0.5: manual data curation, data insertion by user or automated knowledge extraction from structured data sources, 0.25: automated data curation, data insertion by automated knowledge extraction from structured data sources, 0: automated data curation, data insertion by automated knowledge extraction from unstructured data sources	1
Trustworthiness on statement level	Metrics: 1: provenance (through the use of properties) on statement level is used, 0.5: provenance on resource level is used, 0: otherwise	1
Trustworthiness average		**1**

7 https://github.com/IDLabResearch/TurtleValidator

Table 8.1 *(continued)*

Contextual category		
Completeness	The extent to which data are of sufficient breadth, depth, and scope for the task at hand	
Schema Completeness	Definition of a gold standard for iconographical and iconological description from a golden example. The properties used in the KG are aligned to the gold standard. The schema completeness metric is obtained by dividing the number of properties of the example that have been aligned on the total number of properties in the example.	1
Column Completeness	Ratio of the number of instances that have both the class k and a value for the relation r, to the total number of instances that have class k. We apply this metric to the core classes, namely: Human-Made Objects, Artworks, Recognitions, and to their necessary or highly recommended relations	0.85
Completeness average		**0.92**
Representational category		
Ease of Understanding	Dimension defining the grade of understandability of a data source by a human data consumer	
Description of resources	The metric is computed as the number of described resources divided per the number of all considered resources. We apply the computation only actual instances, excluding reification classes.	0.94
Labels in multiple languages	Presence of labels in other languages than English	0
Understandable RDF serialization	Data are provided in RDF in formats more human-understandable than RDF/XML (e.g., N3, N-Triple, and Turtle). **Metric:** Verify if the serialization formats provided include other formats that RDF/XML.	1
Self-describing URIs	Metrics: 1 self-describing URIs always used 0.5 self-describing URIs partly used 0 otherwise	0.5
Ease of understanding average		**0.61**
Interoperability	Interoperability focuses on the representation from a technical perspective	
Avoiding blank nodes and RDF reification	Metrics: 1: no blank nodes and no reification; 0.5: either no blank nodes or no reification; 0: otherwise	0.5
Provisioning of several serialization formats	This criterion verifies the presence of multiple serialization formats	1
Using external vocabulary (ontologies)	Presence of external vocabulary (i.e., classes and relations) for resource description.	1
Interoperability average		**0.83**

Table 8.1 *(continued)*

Accessibility category		
Accessibility	The extent to which data are available or easily and quickly retrievable	
The dereferencing possibility of resources	Over a sample of 20 URIs, we check if the URIs are accessible via HTTP requests and if they return useful RDF data by analyzing the HTTP status codes.	1
Availability of the KG	**Adapted metric:** whether access to the dataset is provided by a server that guarantees its continuity	1
Provision of public SPARQL endpoints	Whether an official SPARQL endpoint is publicly available.	1
Provisioning of an RDF export		1
Support of content negotiation	Metrics: 1 content negotiation is supported, correct content type is returned; 0.5: content negotiation is supported, but results are not in the desired content type; 0: otherwise	0.5
Provisioning of metadata about a KG	Presence of structured metadata on KG	1
Accessibility average		**0.92**
License	Provision of machine-readable licensing information	1
Interlinking	The degree to which entities representing the same concept are connected, within or between multiple data sources, is referred to as interlinking.	
Interlinking via owl:sameAs	Proportion of resources that have a connection to an external knowledge graph via `owl:sameAs`. We exclude reification classes from this metric	0.13
Validity of external URIs	We assess the accuracy of external URIs by analyzing a sample set of URIs to determine if there is a timeout, a client error (HTTP response 4xx), or a server error (HTTP response 5xx). It is tested over a random sample of 25 URIS, composed of 5 external URLs of pictures, 5 from Iconclass, 5 from Getty Vocabularies	1
Interlinking score		**0.55**
Overall score		**0.84**

8.2 Evaluation: domain-specific research questions

To evaluate to which extent traditional iconological issues can be quantitatively addressed through the modeling and data created, we assessed which of the 20 sub-questions presented in Chapter 7 (see Table 7.1) could be answered. All the sub-questions were translated into SPARQL queries and performed against our dataset in our analysis.

As shown in Table 8.2, about 85% of the queries could be answered, while 15% were not fully answered due to missing or limited data that hamper a reliable analysis of the phenomenon. For example, symbols rarely appear multiple times (i.e., the dataset includes very diverse subjects), hindering the observation of their evolution over time.

Table 8.2: Percentage of addressed and partially addressed sub-questions expressing the RQs of the domain

Theme	Fully addressed		Partially addressed	
	Questions	Percentage	Questions	Percentage
Cultural phenomena	SQ1, SQ2, SQ3, SQ5, SQ6	83,34%	SQ4	16,67%
Symbols	SQ7, SQ8, SQ10, SQ11	80%	SQ9	20%
Iconographies	SQ12, SQ13, SQ15, SQ16, SQ17	83,33%	SQ14	16,67%
Evidence	SQ18	100%		0
Visual citations	SQ19, SQ20	100%		0
Total		85%		15%

8.2.1 FAIR evaluation

As the FAIRness of the data is a crucial aspect of Linked Data (Wilkinson et al., 2016), we conducted a self-evaluation using the DANS self-assessment questionnaire.[8]. The dataset obtained an overall score of 76% for compliance with the FAIR principles, divided into Findable (77%), Accessible (75%), Interoperable (92%) and Reusable (61%).

8.3 Ontology evaluation

The ICON ontology was evaluated in i) its extraction potential by testing the Competency Questions listed in Table 5.6 of Section 5.2 (Uschold & Grüninger, 1996) on real-world interpretations, ii) its granularity potential by comparing the data of the interpretations written using the model against the same interpretations encoded with other ontologies, iii) logical consistency, FAIRness and syntax using selected tools and services.[9]. The following subsections deal with the creation of the evaluation dataset and the different evaluation methods.

8.3.1 Evaluation dataset

We tested ICON on the section of the Iconology Dataset containing interpretations extracted from Panofsky's *Studies in Iconology* (Panofsky, 1972). This text was chosen for its historical importance and authoritativeness in the domain at hand. As a result, the dataset contains a total of 28,864 triples about 152 artworks, 1,980 interpretations, and 928 subjects. Additional statistics on the dataset can be found in Table 8.3.

8 Available at https://satifyd.dans.knaw.nl/ A copy of the Iconology Dataset Evaluation results is available in the project repository

9 This section was published in Sartini et al. (2023) In detail, Sections 8.3.1, 8.3.2, and 8.3.3 correspond to Sections 6.1, 6.2, 6.3 of the article, for which S. Baroncini is responsible. Section 8.3.4 corresponds to Section 6.4, for which B. Sartini is responsible.

Table 8.3: Quantitative overview of the test dataset (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

8.3.2 Competency questions evaluation

The ontology was evaluated in Panofsky's dataset using the competency questions listed in Table 5.6. Each iteration of the SAMOD methodology corresponds to a group of questions at levels 0, 1, 2, or 3, and the 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 on GitHub in the form of Jupyter notebooks.[10] For each level, we describe one or two CQs. Table 8.4 contains an overview of the metadata about the artworks included in the queries.

Table 8.4: 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. Anonymous, *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. Anonymous, *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

8.3.2.1 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 ones obtained, are shown in Table 8.5.

10 https://w3id.org/icon/development

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

Listing 8.1: SPARQL query of CQ1.2.1.

```
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 an icon:Action}
}
```

Table 8.5: 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	

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 ART1285. Results in Table 8.6 show how this structure can allow a detailed and qualitative comparison 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)?*

Listing 8.2: SPARQL query of CQ1.3.1.

```
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.
}
```

Table 8.6: CQ 1.3.1 results

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

8.3.2.2 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 artwork's date and place of production to obtain a detailed representation of the subject variations over place and time. Consequently, it can be useful for art historians to integrate qualitative iconographic 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 they appear.*

Listing 8.3: SPARQL query of CQ2.6.1.

```
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 8.7: 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

8.3.2.3 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 with the cultural phenomenon CF1087 'Evolution of the iconography of Saturn'*

Listing 8.4: SPARQL query of CQ3.4.1.

```
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.

8.3.2.4 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 (Panofsky, 1972), the art historian Erwin Panofsky states that his position diverges from the usual interpretation according to which the artwork depicts the myth Hylas and the Nymphs. He cites the works of three other 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 Mount 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?*

Listing 8.5: SPARQL query of CQ0.2.1.

```
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 8.8: 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 Table 8.8. The artwork has 4 different Interpretation Descriptions, among which only DESC1 has recognitions at all the levels, whereas the remaining has only second-level recognitions. To better see their agreement and disagreement, 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 detail 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?*

Listing 8.6: SPARQL query of CQ0.4

```
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 is possible to state that this ontology allows a meaningful representation of iconological research questions through a quantitative approach.

8.3.3 Comparison with existing ontologies

To comply with RQ2, we propose a qualitative comparison between ICON and existing ontologies. We do so by describing an authoritative example from Panofsky's bibliography, in which there are iconographical and iconological interpretations. For this purpose, we selected the most complete ontologies for iconographical descriptions, namely: Visual Representation Ontology and Wikidata.

We chose the frontispiece of François Perrier's 'Segmenta nobilium signorum statuarum' (1638 Edition)[11] depicting the iconography of Father Time, a subject that emerged from the Renaissance onward but that, despite its origins 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) (Panofsky, 1972). 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,[12] or the act of devouring the children, reinterpreted as Time devouring 'whatever he has created' (Panofsky, 1972, p. 74). Panofsky claims that such an evolution of the iconography of time is evidence of the phenomenon of *pseudomorphosis*, according to which figures with a classical appearance did not exist in the classical visual arts, although they were described in the 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

11 Figure available at https://www.britishmuseum.org/collection/object/P_1895-1031-28

12 'tempora quae sicut falx in se recurrunt', trad. by the authors: 'times which, like the sickle, recur'(Sabbadini et al., 1888, pp. II, 406)

with iron teeth standing among ruins, a symbol of the fact that time ruins everything without any effort.[13] Table 8.9 resumes the understanding of this artwork at each different level, while Figures from 8.1 to 8.4 show how this example can be modeled with the ontologies considered.

Table 8.9: Description of the artwork 'Time the destroyer' according to the three levels of interpretation

Lev.	Description
1	man, nude and old, with a scythe and wings gnawing away a statue snake biting its tail fragments of classical buildings and 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' (Panofsky, 1972) 2) Pseudomorphosis: 'certain Renaissance figures became invested with a meaning which, for all their classicizing appearance, had not been present in their classical prototypes, though it had frequently been foreshadowed in classical literature' (Panofsky, 1972)

Table 8.10 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 (Carboni & de Luca, 2019). 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 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 8.1 (the subjects 'man', 'gnawing away' and 'statue' are part of the same Composition ART1282-COMP2), or the actors feeling emotions.

13 '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 any effort'. (Ripa, 1764)

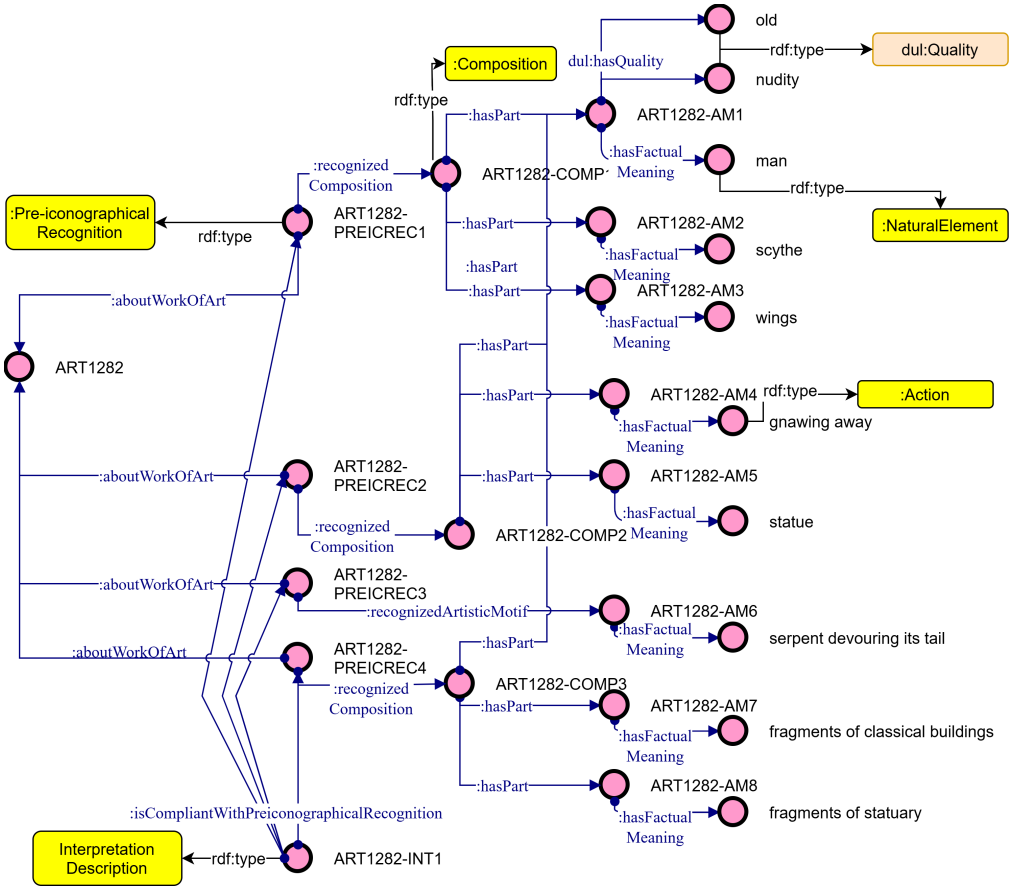

Figure 8.1: Description of the example using ICON (level 1)

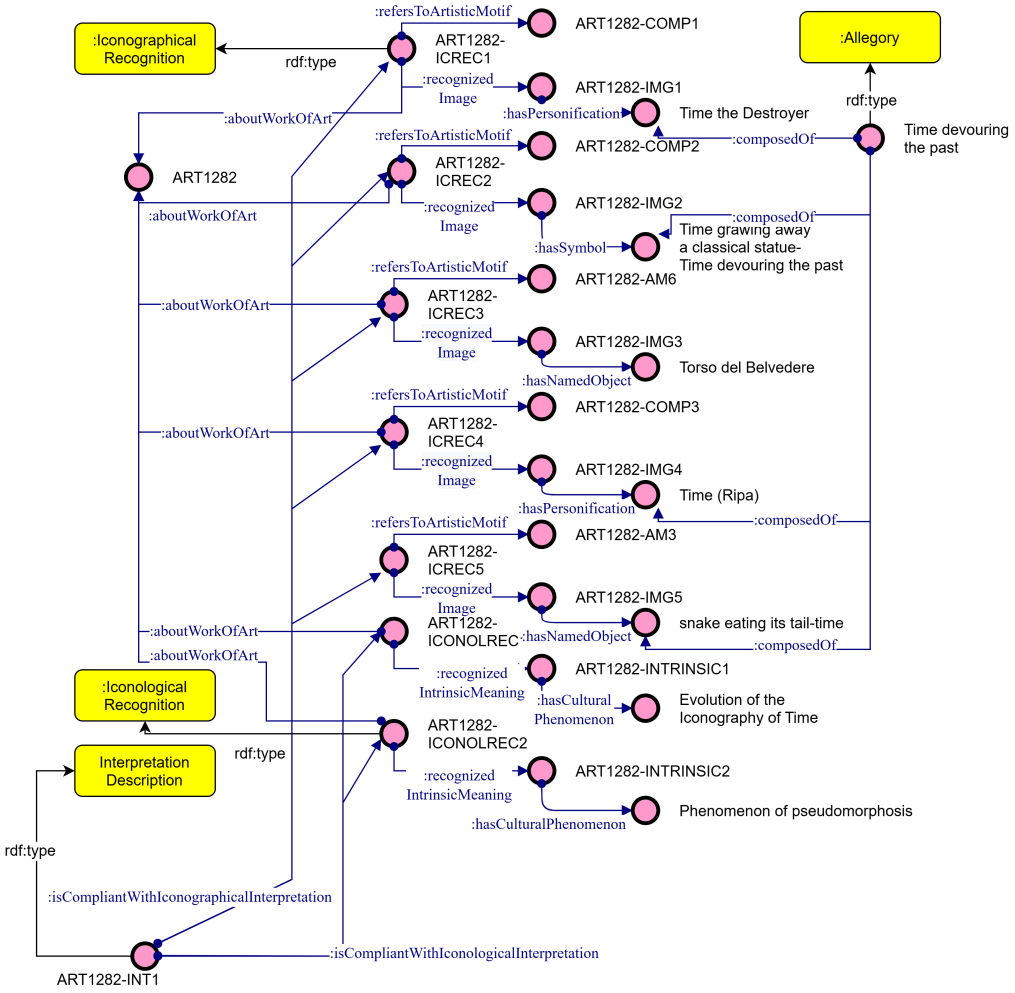

Figure 8.2: Description of the example using ICON (level 2 and 3)

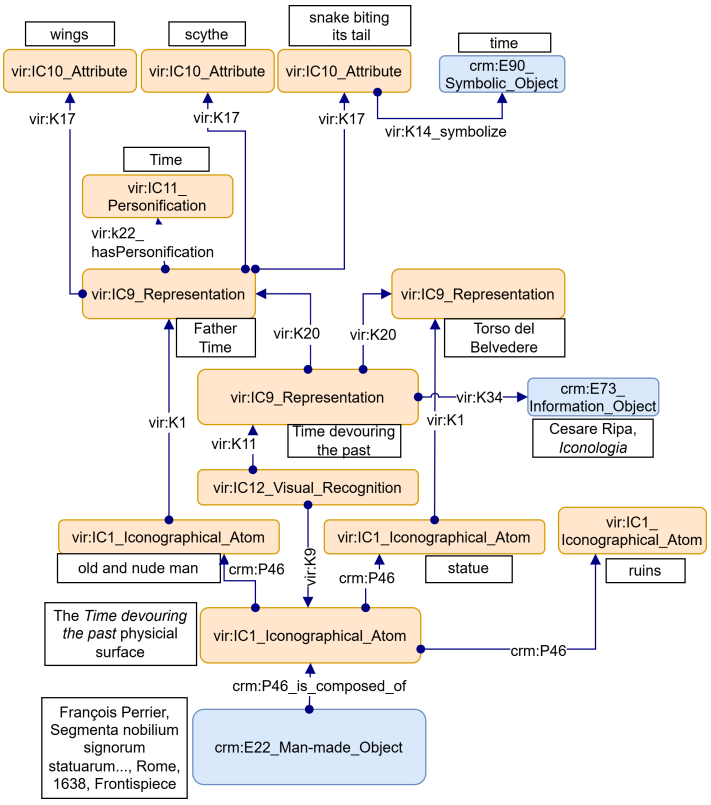

Figure 8.3: Description of the example using VIR

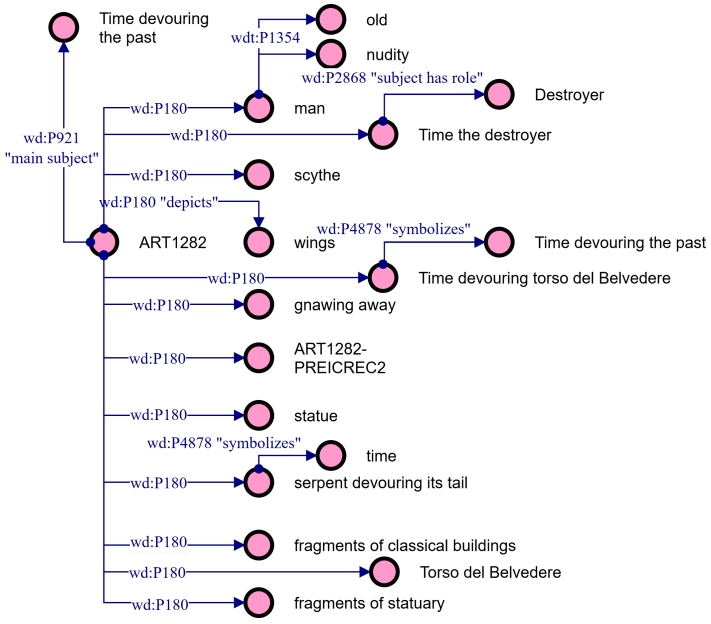

Figure 8.4: Description of the example using Wikidata

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 8.3), 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' (Figure 8.4). 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 because in VIR it is not possible to properly 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. Therefore, ICON introduces `icon:IconologicalRecognition` relating a third-level meaning (concept or cultural phenomenon) to the whole artwork or its specific parts.

Regarding the attribution of responsibility to the interpretation, both VIR and Wikidata allow one to register 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, specified through statements properties, 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 each recognized subject. In this way, more agreeing claims can be expressed in the same recognition description to represent better the realistic case in which an art historian agrees with the claim of others, quoting them and adding further interpretations. Additionally, this structure fosters the interoperability of online sources and data integration.

Table 8.10: 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

8.3.4 Automatic Evaluation

Apart from the evaluations made to verify the explicit goals declared in our research questions, we also evaluate some more technical aspects of our ontology using automatic tools and services. We validated our ontology syntax with the W3C RDF Validation Service.[14] No syntax problems were highlighted by this tool. Then we evaluated the logical consistency of our ontology through the OOPS (Poveda-Villalón et al., 2014) tool[15] 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 modeling; instead, they are linked to the reused ontologies that might have missing information (for example, no ranges and domain in properties or inconsistent labeling). The only highlighted flaw that was directly related to the 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, which is why it is equivalent to our class icon:Symbol. An image is instead considered 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,

14 https://www.w3.org/RDF/Validator/

15 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

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 created classes only). Finally, we analyzed our ontology with the FOOPS (Garijo et al., 2021) tool, which evaluates how much an ontology complies with the FAIR principles. Our ontology scored 90%. In particular, it received a score of 8.5 out of 9 on reusability, 8 out of 9 on findability, 3 out of 3 on interoperability, and 2 out of 3 on accessibility. The main problem highlighted by this tool is that the ontology is not yet inserted in the linked open data vocabulary (LOV).[16]. Finally, we verified that the logical axioms of all the external imported classes and properties that were aligned with our ontology did not cause any inconsistency in ICON by running the HermiT Reasoner[17] (Glimm et al., 2014).

16 The ontology was suggested to the vocabulary dataset, available at https://lov.linkeddata.es/dataset/lov/

17 Version 1.4.3.456