IICONGRAPH: Improved Iconographic and Iconological Statements in Knowledge Graphs

Bruno Sartini[(✉)]

Ludwig-Maximilians-Universität München, Munich, Germany
`b.sartini@lmu.de`

Abstract. Iconography and iconology are fundamental domains when it comes to understanding artifacts of cultural heritage (CH). Iconography deals with the study and interpretation of visual elements depicted in artifacts and their symbolism, while iconology delves deeper, exploring the underlying cultural and historical meanings. Despite the advances in representing CH with Linked Open Data (LOD), recent studies show persistent gaps in the representation of iconographic and iconological statements in current knowledge graphs (KGs). To address them, this paper presents IICONGRAPH, a KG that was created by refining and extending the iconographic and iconological statements of ArCo (the Italian KG of CH) and Wikidata. The development of IICONGRAPH was also driven by a series of requirements emerging from research case studies expressed in competency questions (CQs) that were unattainable in the non-reengineered versions of the KGs. The evaluation results demonstrate that IICONGRAPH not only outperforms ArCo and Wikidata through domain-specific assessments from the literature but also serves as a robust platform for answering the formulated CQs. IICONGRAPH is released and documented in accordance with the FAIR principles to guarantee the resource's reusability. The algorithms used to create it and assess the CQs have also been made available to ensure transparency and reproducibility. While future work focuses on ingesting more data into the KG, and on implementing it as a backbone of LLM-based question answering systems, the current version of IICONGRAPH still emerges as a valuable asset, contributing to the evolving landscape of CH representation within KGs, the Semantic Web, and beyond.

Keywords: Knowledge Graph · Iconography · Iconology · Symbolism · Cultural Heritage · Digital Humanities · Knowledge Graph Generation

1   Introduction

Using Linked Open Data (LOD) in the context of cultural heritage (CH) simplifies the organization, publication, connection, and reuse of knowledge within this domain, and also provides a structure capable of expressing the complex relationships that can emerge between CH artifacts [13]. Over the years, numerous Knowledge Graphs (KGs) have emerged that contain triples on CH, including those referenced in [1,5–7,12,28]. While some serve a general purpose and deal with various domains, others have been specifically crafted to incorporate and represent information about CH. However, recent studies [4] demonstrate that, in the artistic domains of iconography and iconology,[1] current KGs show two main issues: (i) iconographic and iconological statements lack granularity or are dumped[2] in free text descriptions [25], and (ii) cultural symbolism[3] is severely underrepresented.

This paper addresses these gaps by presenting *IICONGRAPH*, a KG developed from the iconographic and iconological statements of Wikidata [28] and ArCo [6], first re-engineered following the ICON ontology [23] structure, and then enriched with LOD on cultural symbolism taken from HyperReal [24]. These two KGs were chosen because they showed the greatest potential in the evaluation work by Baroncini et al. [4], obtaining the highest scores for the correctness of their iconographic and iconological statements, while showing limits when it comes to the level of granularity of these statements.[4] First, the same evaluation is conducted on IICONGRAPH to demonstrate its superior performance compared to the original sources, highlighting its impact through quantitative assessments. Second, the research potential of IICONGRAPH is tested by attempting to address domain-specific competency questions (CQs) that remained unanswered with the original data from Wikidata and ArCo. The rest of the paper is divided as follows. Section 2 gives a background of the work by presenting the resources included in IICONGRAPH, the ontology behind it, and the limitations of the KGs that were chosen as the initial data sources. Section 3 follows by describing the development and release of IICONGRAPH. In Sect. 4, IICONGRAPH is evaluated using the methodology proposed in [4]. The following Sect. 5 describes how the re-engineered KG can now be used to

---

[1] Iconography is the study and interpretation of visual symbols and images, often within the context of art or visual representation. It involves the identification and analysis of symbols, motifs, and elements within images in an artwork. Iconology instead involves the interpretation of images in a broader cultural and historical context, exploring the deeper layers of meaning, cultural ideologies, and socio-political influences associated with visual representations [17].

[2] Considering dumping as the phenomenon in which "important information for which no appropriate field was found, was forced as plain text inside a descriptive field, easy for humans to read but forever lost to any automatic tool" [2].

[3] Intended as the set of symbolic meanings that some CH objects (or the elements depicted in them) convey from specific cultural perspectives.

[4] In the mentioned study, both KGs performed poorly on the "structure" evaluation, which dealt with the possibility to differentiate between iconographic, iconological, and symbolic subjects, among other criteria explained in Sect. 4.

answer domain specific CQs. Section 6 contains a discussion reflecting on the results of the quantitative and research-based evaluations. Then, Sect. 7 mentions related work about the generation of artistic KGs. Finally, Sect. 8 contains final reflections on this work and mentions possible future work.

2   Background, Problem Statement, and Research Requirements

In this section, the resources used to develop and enrich IICONGRAPH are described. In the subsections about ArCo and Wikidata, their current issues in the representation of artistic interpretation domain are highlighted.
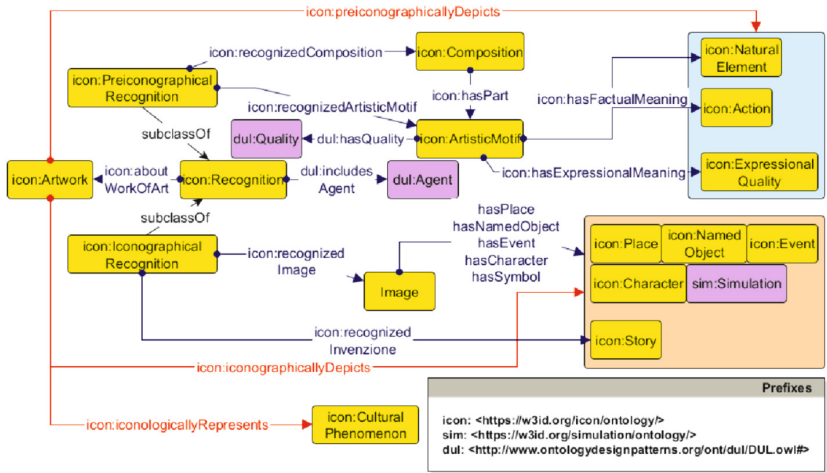
2.1   ICON Ontology 2.0

ICON [23] is an ontology that conceptualizes artistic interpretations by formalizing the methodology of E. Panofsky [17]. According to Panofsky, when performing an artistic interpretation, the interpreter can consider three levels. At a pre-iconographic level, artistic motifs and their factual or expressional meanings are recognized both as single entities and as groups (or compositions). The recognition of a tree, of the action of running or the emotion of crying would be considered pre-iconographical. At an iconographical level, the same motifs are now recognized as what Panofsky calls images, and these images represent characters, symbols, personifications, specific places, events or objects (such as Rome, World War II, and Thor's Hammer). At the same level, the artwork can be seen as depicting a story or an allegory (Panofsky uses "Invenzione" as a general term conveying both stories and allegories). At an iconological level, the artworks are then analyzed in comparison with the cultural context in which they were created, and they become a vessel to convey more in-depth cultural meanings or representing cultural values or cultural phenomena. ICON was updated (in version 2.0) to include three shortcuts that directly link an artwork to the element of pre-iconographic, iconographic or iconologic levels that it depicts or represents [22]. For a more comprehensive overview of ICON, refer to the documentation of IICONGRAPH[5] or to previous publications on the ontology [22,23]. The classes and properties of ICON used in this work are displayed in Fig. 1.

2.2   HyperReal

HyperReal is a KG that contains more than 40,000 instances of symbolism, also called simulations. A simulation is a connection between a symbol (like a cat) and its symbolic meaning (such as divinity) in a particular cultural context (e.g. Egyptian). HyperReal information comes from various sources such as symbols dictionaries [15], and encyclopedias [16] and is structured according to the Simulation Ontology framework. Figure 2 shows the graphical rendering of the *cat-divinity* simulation. The KG is available through its data dump at

---

[5] https://w3id.org/iicongraph/docs/.

Fig. 1. Graphical Rendering of the ICON ontology classes and properties used in IICONGRAPH. The shortcuts of ICON 2.0 are red, imported classes are violet. (Color figure online)

https://w3id.org/simulation/data. HyperReal data are aligned with the corresponding Wordnet [9] and Babelnet [14] synsets to facilitate the process of aligning external data with its symbols and symbolic meanings [21]. HyperReal has been used in the back end of CH-related applications [26], and as a data source for quantitative comparative cultural studies [27]. In the context of IICONGRAPH, it is used to enrich the potential symbolism of artworks. For instance, a painting depicting a cat could be interpreted from an Egyptian point of view as symbolizing divinity. This kind of inference is agnostic from the intention of the creator of the work of art, but contributes to its understanding from a polyvocal and multicultural perspective.
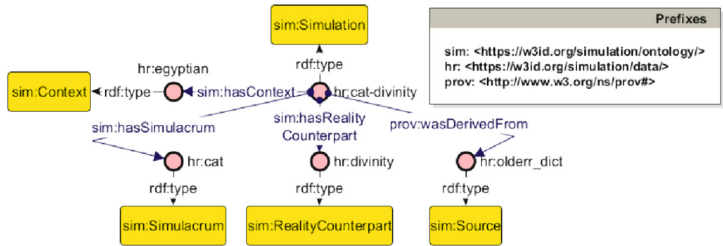
Fig. 2. Graphical Rendering of the *cat-divinity* simulation following the Simulation Ontology schema. The Simulacrum is the specific term used to represent symbols, while Reality Counterpart is the term used to represent symbolic meanings.

2.3   Wikidata

Wikidata [28] is a user-generated, open, comprehensive knowledge base, launched in 2012 by the Wikimedia Foundation, with a wide selection of content, available in various levels of detail and formats. It provides a platform for collaboration, sharing, integration, and a technology system for creating linked data. In the Digital Humanities domain, it is often used to annotate and improve project components, curating metadata to refine the interoperability of authority and local data sets about CH [30]. In the context of this work, the main focus was on the subset of Wikidata's statements regarding artworks and their depictions (the extraction methodology is explained in Sect. 3).

Wikidata Analysis and Problem Statement. When analysing the limitations of Wikidata's iconographic and iconological statements, the main focus is on the property *depicts* (wdt:P180) and its qualifiers. This property links an artwork with an element depicted in it. Its qualifiers, such as *wears* (wdt:P3828), and *expression, gesture, or body pose* (wdt: P6022), give more context to the depicted element. On the one hand, Wikidata contains more than 372,000 *depicts* statements[6] when the subject is a painting (wd:Q3305213), which is a great starting point for digital art history studies. On the other hand, this property is used for all three levels of interpretations, flattening the expressivity of those statements. When it comes to expressing symbolism, there are some exceptions. For, instance *symbolizes* (wdt:P4878) is a qualifier of *depicts* that links a depicted element to what it symbolizes. However, the property is rarely used (only 63 statements related to paintings).[7] **Therefore, the main issue with Wikidata is that when the data are present, the schema is lacking, and when the schema is present, the data are lacking.**

Formulation of Competency Questions for Wikidata. Following the previous statements, in Wikidata it is not possible to retrieve what the most symbolic paintings are and how many serendipitous symbolic connections exist between paintings. Serendipitous connections are defined here as

> "all the new connections that emerged between artworks in Wikidata, caused by the shared symbolic meaning **only**. [...]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 connections. 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" [21].

---

[6] Query last run in December 2023: https://qlever.cs.uni-freiburg.de/wikidata/yKhv77.

[7] Query last run in December 2023: https://w.wiki/8QyF.

By using the current data in Wikidata, zero serendipitous connections emerge.[8] At the same time, it is also currently challenging to distinguish between the pre-iconographical and the iconographical elements depicted in Wikidata's paintings, a task that becomes even more difficult if the objective is distinguishing between the specific types of iconographical subjects (characters, places, attributes, etc.)

Given these premises, the following CQs have been formulated and will be answered in IICONGRAPH.

CQ1. How many serendipitous connections exist among artworks in Wikidata?
CQ2. Which artworks are associated with the most symbolic meanings?
CQ3. How are pre-iconographic and iconographic depictions distributed across Wikidata's *depicts.* statements in paintings?
CQ4. Among iconographical elements, which are the main classes (characters, places, attributes) that emerge as the most frequent?

Regarding CQ1 and CQ2, I hypothesize that after enriching Wikidata with HyperReal, the number of serendipitous connections will substantially increase, and after that it will be possible to rank Wikidata's painting according to their *symbolic temperature*. Addressing CQ3 and CQ4, the hypothesis is that by re-engineering the statements in Wikidata according to the ICON ontology, it will be possible to distinguish and measure the distribution of pre-iconographic and iconographic elements.

2.4   ArCo

ArCo [6] is a KG that describes a wide spectrum of artifacts from the Italian CH, containing items belonging to the architectural, ethnographic, and artistic domains. It follows the structure of the ArCo ontology, spread into different modules to address different levels of description of CH. In the context of this work, only a subset of statements related to artworks (which belong to the class *HistoricalOrArtisticProperty*) will be considered, with more limitations that will be explained in the following subsection.

ArCo Analysis and Problem Statement. ArCo was created by applying Natural Language Processing algorithms to the OCR (Optical Character Recognition) version of printed catalogs. Consequently, even if some of the more technical information was converted into URIs and single nodes in the KG, a great deal of free-text information remains, especially about subjective domains like iconographic readings. Therefore, most of the information regarding iconographical and iconological statements is dumped in a free-text description, not exploiting the full potential of LOD. On the one hand, this puts ArCo in a worse starting position compared to Wikidata, which expresses almost all of the information through URIs and limits the free-text fields. On the other hand, some of the

---

[8] Query last run in December: https://w.wiki/6BZR.

descriptions in ArCo contain detailed interpretations about artworks, even separating pre-iconographic subjects from iconological meanings conveyed by artworks. Additionally, the descriptions have a schematic structure with repeating patterns, especially those related to a series of Italian billboards created in the 20th century. In the current version of ArCo, it is challenging to study the correlations between specific iconographic and pre-iconographic subjects and the cultural event/product they promote (iconological level).

Formulation of Competency Questions for ArCo. Given that the starting point of ArCo is worse compared to Wikidata, only one CQ was formulated, namely:

CQ5. What are the most common iconological meanings associated with Italian Billboards from the 20th century?

The hypothesis is that by transforming the free-text description into structured data following ICON, it will be possible to isolate and then measure the frequency of iconological meanings.

3   IICONGRAPH Development and Release

This section describes how IICONGRAPH was developed and released. Different strategies were adopted for the development according to the issues mentioned in Sect. 2 for the two sources. The main distinction between the two sources is that while Wikidata provides information about the potential relationships between depicted entities (via the qualifiers), requiring a full description using the ICON ontology, ArCo's descriptions are very linear; therefore, only the shortcuts introduced in ICON 2.0 are necessary to describe such information.[9] For both KGs, the generation of the re-engineered LOD was performed in a Python environment via the RDFlib package.[10]

3.1   Wikidata's Conversion

The general pipeline adopted to convert Wikidata was (i) assigning the depicted entities to the classes of ICON, (ii) extracting data about paintings, (iii) aligning them with HyperReal, and then (iv) re-engineering the statements following the ICON ontology. To align Wikidata's depicted entities with ICON classes, we adopted a methodology involving the annotation of the depicted entity types and classes expressed in Wikidata through the properties *instance of* (wdt:P31) and *subclass of* (wdt:P279). Given the impracticality of manually annotating more than 60,000 individual depicted entities, I focused on annotating the top 700 classes and types, ordered by the number of depicted elements assigned to

---

[9] This decision is supported by the work that presents ICON 2.0 [22], in which ArCo's descriptions are a use case example for ICON 2.0.

[10] https://rdflib.readthedocs.io/en/stable/.

them. The top 700 covered more than 85% of the total entities. To ensure objectivity, a no-ambiguity policy guided the single annotator. Each type or class was analyzed on Wikidata using a SPARQL query to verify that all related entities could match the designated ICON class; otherwise, the type or class was discarded. The alignment is not made because of a shared or similar label between the classes of Wikidata and ICON, but rather by choosing the best ICON class to represent the instances of the Wikidata class. For example, the instances of the class *big city* `wd:Q1549591` would be modelled, using ICON, with the `icon:Place` class. Subsequently, all the depictions of big cities will be described in ICON using an `icon:IconographicalRecognition`. At the same time, other classes of Wikidata which have been aligned to `icon:NaturalElement` will make all the elements that belong to classes recognized in ICON via a `icon:PreiconographicalRecognition`. Figure 3 illustrates the distribution of assigned classes and types for pre-iconographical and iconographical elements. After this alignment, the information about the paintings, their depicted elements, the types and classes of the depiction, and their qualifiers were extracted via a SPARQL query. A total of almost 150,000 paintings and their related metadata were extracted. To align Wikidata's entities with HyperReal's symbols for the enhancement, an alignment done in previous work was reused [26]. The conversion of Wikidata yielded more than 29,000,000 triples. More than 3,000,000 symbolic interpretations were inferred, due to the alignment to HyperReal, with an average of around 20 interpretations per painting. For a more detailed description of Wikidata's conversion, refer to the documentation.
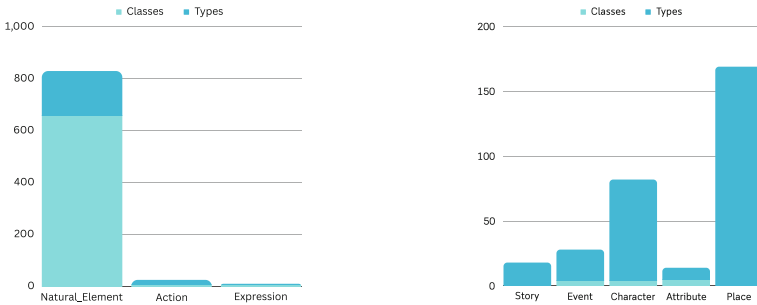
Fig. 3. On the left: manual matching result between Wikidata's types and classes and ICON's classes related to pre-iconographical elements. On the right: manual matching result between Wikidata's types and classes and ICON's classes related to iconographical elements.

3.2   ArCo's Conversion

For the conversion of ArCo, only the shortcut version of ICON was necessary, eliminating the need to assign elements from free-text descriptions to individual ICON classes. Instead, the depicted elements were categorized into the

macrogroups of pre-iconographical, iconographical, and iconological. The process involved extracting ArCo's data using a regex pattern to capture "Iconographic Reading:" (in Italian, "Lettura Iconografica:") in artwork descriptions linked via the Dublin Core description property (`dc:description`). Following the extraction of approximately 23,000 artworks and their descriptions (about 1% of ArCo's total number of artworks), the structure of the descriptions was analyzed to identify other patterns to facilitate the automatization of the conversion. It was noticed that standard descriptions are organized into categories separated by a standard use of punctuation. All descriptions that did not meet this standard were discarded (around 3,000). All iconological meanings, in the context of billboards, were determined to be after the category "Product category/type of event" (in Italian, "Categoria Merceologica/tipo di evento"), where the promotional aspect was described. Ambiguous categories (such as "Names", which included both the people depicted in the billboards but also the CEOs of the companies that were getting promoted) were excluded, and a straightforward approach was employed to distinguish between pre-iconographical and iconographical levels. If an element in the description was written with a capital letter, it was assigned to the iconographic level, otherwise to the pre-iconographical level. Figure 4 visually shows the rationale behind the assignment and parsing of descriptions, exemplified by the artwork available at https://w3id.org/arco/resource/HistoricOrArtisticProperty/0500659063. Before conversion, all descriptions were translated into English using the Google Translate API. Given the simplicity of the texts, this translation did not generate evident errors. Single elements were linked to HyperReal through string matching. In summary, the description from ArCo were translated and then analysed by a simple parser that separates the categories with the punctuation, and then isolates each single element of each description category considering it an iconographical element if written with uppercase or pre-iconographical if lower case. The conversion of ArCo yielded 767,888 triples, which is significantly less than Wikidata because of the difference in number of artworks (150,000 against 20,000), and also because the simplified version of ICON is much less verbose. A total of 457,747 automatic interpretations were generated due to the match with HyperReal.

3.3   IICONGRAPH Release

IICONGRAPH was released according to the FAIR principles [29]. The w3id service was used to obtain persistent URIs for the namespace https://w3id.org/iicongraph/data/, documentation https://w3id.org/iicongraph/docs/ and analysis related to the research case studies https://w3id.org/iicongraph/casestudies/. The same information is accessible via the GitHub repository https://github.com/br0ast/iicongraph/. The prefix *iig*, used in the KG, was registered in http://prefix.cc. The KG is stored in Zenodo, accessible via https://zenodo.org/doi/10.5281/zenodo.10294588. The metadata about the dataset and
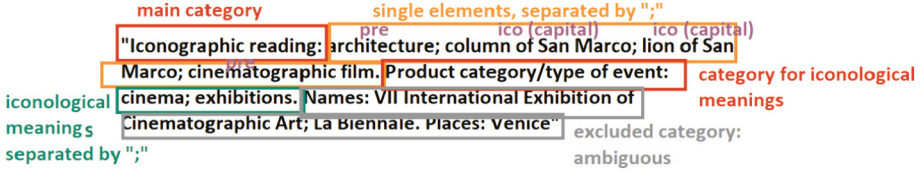
Fig. 4. Visual Example of a standard description in ArCo and the parsing steps applied to it

the provenance of the data is defined in a separate file[11] following the DCAT standard.[12] The ICON ontology used as its schema also respects the FAIR principles, obtaining a score of 90% on the FOOPS tool [11].

4   Quantitative Evaluation

In this section, IICONGRAPH is quantitatively evaluated following the methodology defined by [4]. Three versions of IICONGRAPH will be evaluated, namely IIC-arco, which contains only the re-engineered statements from ArCo, IIC-wikidata, which contains only the re-engineered statements from Wikidata, and IIC-global, containing all the triples. The assessment method considers six criteria, divided into two macro-areas: content and structure. Content considers the evaluation of the correctness of artistic interpretation statements (CR1), and the evaluation on the completeness of artworks interpretations (CR2) (i.e., whether the interpretation mentions, when needed, pre-iconographical, iconographical, and iconological statements). Then, structure addresses the richness of the schema describing the artworks (CR3), the entity linking of artworks with external sources is measured by CR4, CR5 measures how the URIs of the depicted subjects are linked within the same dataset (in technical terms, the outdegree of the subjects' nodes in a graph) and CR6 measures the number of references to external taxonomies of art and culture. All categories are given a weight (CR1,2,3 have a weight of 1, CR4,5 have a weight of 0.6 and CR6 has a weight of 0.8) and the possible scores for each go from 0 to 1. Given the reengineering tasks performed to create IICONGRAPH, this work could only influence the criteria CR2, CR3, CR5 as it does not deal with changing wrong interpretations (CR1), linking artworks between different datasets (CR4), or referring to external taxonomies of art and culture (CR6).

Following the methodology presented in [4], CR2 was calculated by averaging the scores of two annotators that evaluated the description of 100 artworks. The annotators had to decide how many interpretation levels they expected for the artwork. The general guidelines of [4] say that artworks depicting a landscape usually are only interpreted via a pre-iconographical level, most portraits have

---

both pre-iconographical and iconographical meanings, and allegorical, religious, and culturally relevant scenes (depiction of wars, special events for a country or culture) can usually be described using all levels. After averaging the evaluation, IIC-global obtained 0.92, IIC-arco 0.958 and IIC-wikidata 0.97. CR3 was calculated through a comparison of the ICON ontology structure with the gold standard in [4][13]. Given that the schema behind IICONGRAPH is the ICON ontology, developed to describe all the information mentioned in the gold standard, the score of all the versions of the KG in this category was set at 1. CR5 was computed via SPARQL queries on the data, first counting how many subjects in the data were linked to at least more than 1 artwork, and then dividing this number by the total number of subject recognized. The scores obtained are 0.5771 for IIC-arco, 0.4573 for IIC-wikidata and 0.4337 for IIC-global. Since CR1, CR4, and CR6 were not affected by the changes, IIC-wikidata and IIC-arco maintain their scores of [4], while IIC-global receives an average of the two scores. Table 1 shows the scores compared to the other datasets analyzed in [4]. In Sect. 6 the results are analyzed and discussed. All scripts and queries related to the quantitative evaluation are available in the documentation at https://w3id.org/iicongraph/docs/.

Table 1. Overall results of the quantitative evaluation applied to IIC-global, IIC-wikidata, and IIC-arco compared to the results in the state of the art performed in [4]. UF labeled criteria signal that they were not affected by the changes. Ranking signals only the overall top 3 for each category.

	CR1 UF	CR2	CR3	CR4 UF	CR5	CR6 UF	Content	Rank Cont	Structure	Rank Structure	Overall	Rank overall
IIC-global	0.9023	**0.92**	**1**	0.3508	0.4737	0.1404	**0.9111**	2	**0.5357**	2	**0.7234**	2
IIC-arco	0.8278	**0.958**	**1**	0.0026	**0.5771**	0.1238	**0.8929**	3	**0.4823**	3	**0.6876**	3
IIC-wiki	0.9768	**0.97**	**1**	0.699	0.4573	0.157	**0.9734**	1	**0.6065**	1	**0.7899**	1
ArCo	0.8278	0.74	0.3333	0.0026	0.172	0.1238	0.7839		0.1790		0.4815	
Wikidata	0.9768	0.74	0.6667	0.699	0.367	0.157	0.8584		0.4773		0.6678	
Fondazione Zeri	0.9925	0.5117	0.1111	0.0005	0.266	0.5449	0.7521		0.2356		0.4939	
Nomisma	0.9768	0.5	0.2222	0	**0.749**	0.0001	0.7384		0.2239		0.4811	
SARI	0.849	0.3783	0.1111	0.997	0.5	0	0.6136		0.3364		0.475	
Europeana	0.4688	0.236	0.1111	0.0073	**0.6122**	1	0.3524		0.4276		0.39	
ND_Hungary	0.13	0.5392	0.1111	0	0	0	0.3346		0.037		0.1858	
DBpedia	0.655	0.7242	0.2222	0.994	0.41	0	0.6896		0.3549		0.5222	
Yago	0.99	0.4825	0.1111	1	0.1675	0	0.7362		0.2705		0.5034	

5   Research-Based Evaluation

This section shows how IICONGRAPH can be used to answer the CQs formulated in Sect. 2. Regarding CQ1 and CQ2, the methodology consisted of extracting the data on paintings and their symbolic depictions through a SPARQL query performed on IIC-wikidata. Around 79,000 paintings were associated with

---

[13] In summary, a perfect schema would be able to describe actions, preiconographical elements, stories, allegories, iconographical subjects, symbols, iconological subjects, cultural phenomena, and should be able to be used in combination with a taxonomy or controlled vocabulary of art and culture.

a symbolic meaning shared by more than one symbol. Artwork connections were computed using Python, with an iterative process comparing each depicted element between pairs of paintings. The calculation involved determining how many symbolic meanings were shared between the depicted elements of the pairs. At the end of the calculation, **2,481,489,938 serendipitous connections were exposed**. CQ2 was tackled with a SPARQL query, revealing the top 10 most symbolic paintings. **"Entrance into the Ark" (wd:Q209050) by Jan Brueghel The Elder** tops the list with almost 1,500 associated simulations; the rest of the top 10 are detailed in Table 2. In general, paintings with a multitude of animals and plants were associated with most symbolic meanings. Similarly, simple SPARQL queries facilitated the examination of the distribution of pre-iconographic and iconographic representations in Wikidata (CQ3, CQ4), revealing that **64.86% of the depicted elements belong to the pre-iconographical level**. Among iconographic elements, **Characters are the most recognized, with almost 100,000 occurrences**. The results of this analysis are presented in Table 3. CQ5 followed a similar approach. Through a SPARQL query on IIC-arco, the number of paintings associated with each iconological meaning was determined. The top 10 iconological meanings are presented in Table 4. In particular, **the iconological meaning most referred to in 20th century billboards is the promotion of tourism**. In summary, post-reengineering and enrichment, all CQs formulated in Sect. 2 were effectively answered. All scripts and queries developed to address these CQs are provided at https://w3id.org/iicongraph/casestudies to ensure the transparency and reproducibility of the results.

Table 2. Top 10 of the most symbolic paintings in Wikidata, retrieved by a SPARQL query performed on IICONGRAPH. *wd:* is the prefix for https://www.wikidata.org/wiki/entity/

Painting ID	Painting Label	Sim#
wd:Q66107722	**Entrance into the Ark**	**1,488**
wd:Q18809786	**Entry into Noah's Ark**	**998**
wd:Q321303	**The Garden of Earthly Delights**	**851**
wd:Q27980267	Unknown Title	758
wd:Q2510869	Concert in the Egg	747
wd:Q463392	Paradiesgärtlein	723
wd:Q20170089	The Ark	723
wd:Q18917077	The Garden of Eden and the Creation of Eve	721
wd:Q29656879	Earth or The Earthly Paradise	706
wd:Q18573212	The Animals Entering Noah's Ark	662

Table 3. Distribution of Pre-iconographical and Iconographical statements in Wikidata extracted from IIC-wikidata

Level of interpretation	Total	Unique	Specific Element	Total	Unique
Pre-iconographical	**224,981**	**5,131**	**natural elements**	**220,463**	**4,938**
			actions	4,189	147
			expressions	4,667	65
Iconographical	121,893	37,667	**characters**	**98,354**	**27,847**
			events	817	399
			stories	3,436	3,436
			attributes	791	51
			places	17,050	5,438

Table 4. Top 10 iconological meanings associated with the most artworks in ArCo

Iconological Meaning	Artwork #
iig:promotionOfTourism	**4,572**
iig:promotionOfExhibitions	**3,604**
iig:promotionOfTourismPromotionBodies	**3,380**
iig:promotionOfInformationAndCommunication	2,932
iig:promotionOfFoodIndustry	2,814
iig:promotionOfCulturalEvents	2,219
iig:promotionOfTransport	2,211
iig:promotionOfSport	1,928
iig:promotionOfTrade	1,911
iig:promotionOfAgriculture	1,694

6   Discussion of the Results

After a thorough quantitative evaluation, the performance of IIC-global, along with its subsets IIC-wikidata and IIC-arco, outperforms the rest of the KGs examined in [4] in both structure and content scores. The effectiveness of the re-engineering process is evident in the significant improvements observed, particularly in CR5 (subject intralinking potential) for ArCo, where it experienced an impressive increase (more than 300%) from 0.172 to 0.5771. This increase is due to the generation of more subjects expressed in URIs, which increases the number of connections between artworks that share the same subject (now defereanceable compared to the previous text-only version).

The best-performing KG overall is IIC-wikidata, similarly to when the standard version of Wikidata was the top performer before the re-engineering process. Notably, despite the enhancements, ArCo still falls short in the structure criteria, with an overall structure score of less than 0.5. This limitation is attributed

to issues such as references to external taxonomies and the alignment challenges between its artworks and those present in other KGs.

When it comes to the evaluation of research-driven CQs, IICONGRAPH shows great potential for domain-specific analyses, although these results are considered preliminary and show some limitations. In fact, the automatic symbolic interpretations of artwork from a polyvocal point of view (given by Hyper-Real) could be the starting point for more in-depth analysis for art historians, as they only represent potential, creator-agnostic symbolic meanings. Moreover, Table 4 displays elements that could be merged after performing entity disambiguation (i.e., promotion of tourism and promotion of tourism promotion bodies). Despite this, the results underscore the considerable advancement represented by IICONGRAPH and its subgraphs. They not only outperform the state-of-the-art quantitatively but also demonstrate their utility in addressing CQs that were unattainable in the original versions of the KGs. The improvements in both quantitative metrics and research potential underscore the significance of the re-engineering efforts and the enriched representations provided by IICONGRAPH.

7   Related Work

This section provides an overview of the development of artistic KGs or related resources, highlighting differences from IICONGRAPH. Artgraph [7] is a KG developed by combining data from DBpedia and Wikiart, including over 250,000 artworks and associated artists. Its objective is to integrate visual embeddings and graph embeddings from the KG for automated art analysis. However, an examination of Artgraph's properties reveals the same issues found in Wikidata and ArCo, such as the lack of granularity of iconographic and iconological statements due to the absence of interpretative depth. The connection between artworks and subjects relies on a generic "tag" property. Furthermore, the dataset does not incorporate symbolic representation. ICONdataset [3] is a manually annotated KG, containing more than 5,500 art historians' interpretations about more than 400 artworks. It shares with IICONGRAPH the adoption of the ICON ontology as its primary schema. While manual annotation, as employed by ICONdataset, affords complete supervision over the data, ensuring a high degree of accuracy, the inherent drawback lies in its time-consuming nature, evident in the relatively low number of artworks and interpretations. In contrast, IICONGRAPH adopts a semi-automatic approach, resulting in a significant disparity in both the quantity of artworks and interpretations between the two KGs. This distinction emphasizes the scalability and efficiency afforded by a semi-automatic process. Furthermore, IICONGRAPH's incorporation of the HyperReal enrichment introduces an additional layer of symbolic data, augmenting its comprehensiveness, reach, and potentialities in comparison to manually annotated counterparts. MythLOD [18] is an LOD catalog that contains interpretations of more than 4,000 mythological works. It was created by converting a CSV manually populated by domain experts. Its main purpose is to represent in

LOD both the methodology and rationale of the interpretations (iconographic, hermeneutic) and the bibliographic sources which supported the interpretations. However, when it comes to describing the main objects of the interpretations, it relies on the standard Dublin Core[14] subject property (`dc:subject`), which is extremely limited compared to the possibilities offered by the ICON ontology behind IICONGRAPH. Other datasets, such as [1,8,12,19,20] are not mentioned, as they are compared to IICONGRAPH through the evaluation in Sect. 4.

8 Conclusion and Future Work

This paper presented the development and evaluation of IICONGRAPH, a KG created by re-engineering the iconographic and iconological statements of ArCo and Wikidata. IICONGRAPH, IIC-arco and IIC-wikidata outperformed the state-of-the-art of artistic KGs in both structure and content scores. Furthermore, the results of the requirements evaluation based on the CQs demonstrate the suitability of IICONGRAPH to answer domain-specific artistic inquieres. Future work is divided into two main areas. First, the expansion of IICONGRAPH involves ingestion of additional statements from more artistic KGs. Second, in the realm of Large Language Models (LLMs), IICONGRAPH emerges as a valuable resource for developing question answering and chat-based systems focused on CH, addressing a gap identified in the literature regarding symbolic and iconographic knowledge [10]. Additionally, the descriptions in ArCo and the RDF generated to create IICONGRAPH hold promise as fine-tuning arguments for an LLM capable of autonomously generating intricate iconographic LOD from free-text descriptions via prompts. Despite the narrow focus of IICONGRAPH in the field of iconography and iconology, especially in the context of linked open data, it can have an impact to attract interest in digital humanities initiatives that deal with Semantic Web. One key impact of IICONGRAPH is its potential as a valuable resource for applications dealing with cultural heritage data. With structured data on over 170,000 artworks, IICONGRAPH provides a rich source of information for reuse, facilitating nuanced analyses and interpretations. Moreover, the growing field of digital art history stands to benefit from IICONGRAPH's structured data repository. Museums can leverage this resource to enrich their digital collections; if some of their artworks are already included in Wikidata, they can refer also to the IICONGRAPH version, allowing for more robust and iconographically-centered searches, enhancing the discoverability and interpretive potential of their artworks. Furthermore, IICONGRAPH serves as a model for other underrepresented fields within LOD initiatives. By elevating the discourse surrounding iconography and iconology, it can set a precedent for the inclusion and recognition of other specialized domains, such as symbolism in music or Egyptian iconographic data, fostering innovation and inclusivity within the broader scholarly ecosystem. In conclusion, IICONGRAPH stands as a robust and versatile resource that advances the understanding of artistic interpretation within the domains of art history and digital humanities, and

---

[14] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

also presents significant implications for the evolving landscape of LLMs, offering a promising avenue for further exploration and integration into the broader context of CH research.

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ISWC ASWC 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52

2. Barabucci, G., Tomasi, F., Vitali, F.: Supporting complexity and conjectures in cultural heritage descriptions. In: Proceedings of the International Conference Collect and Connect: Archives and Collections in a Digital Age, pp. 104–115. CEUR Workshop (2021). http://ceur-ws.org/Vol-2810/paper9.pdf

3. Baroncini, S., Daquino, M., Tomasi, F.: Exploring iconographical and iconological content in semantic data through art historians' interpretations. In: La memoria digitale. Forme del testo e organizzazione della conoscenza. Atti del XII convegno annuale AIUCD, pp. 9–14. Siena (2023)

4. Baroncini, S., Sartini, B., van Erp, M., Tomasi, F., Gangemi, A.: Is dc:subject enough? A landscape on iconography and iconology statements of knowledge graphs in the semantic web. J. Documentation **79**(7), 115–136 (2023). https://doi.org/10.1108/JD-09-2022-0207

5. Bruno, E., Pasqual, V., Tomasi, F.: Odi and bacodi: a study on destini incrociati by italo calvino with semantic web technologies. In: La memoria digitale. Forme del testo e organizzazione della conoscenza. Atti del XII convegno annuale AIUCD, p. 35. Siena (2023)

6. Carriero, V.A., et al.: ArCo: the Italian cultural heritage knowledge graph. In: Ghidini, C., et al. (eds.) ISWC 2019. LNCS, vol. 11779, pp. 36–52. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30796-7_3

7. Castellano, G., Digeno, V., Sansaro, G., Vessio, G.: Leveraging knowledge graphs and deep learning for automatic art analysis. Knowl.-Based Syst. **248**, 108859 (2022). https://doi.org/10.1016/j.knosys.2022.108859. https://www.sciencedirect.com/science/article/pii/S0950705122004105

8. Daquino, M., Mambelli, F., Peroni, S., Tomasi, F., Vitali, F.: Enhancing semantic expressivity in the cultural heritage domain: exposing the zeri photo archive as linked open data. J. Comput. Cult. Herit. **10**(4) (2017). https://doi.org/10.1145/3051487

9. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. Language, Speech, and Communication, A Bradford Book, Cambridge (1998)

10. Garcia, N., et al.: A dataset and baselines for visual question answering on art. In: Bartoli, A., Fusiello, A. (eds.) ECCV 2020. LNCS, vol. 12536, pp. 92–108. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66096-3_8

11. Garijo, D., Corcho, O., Poveda-Villalón, M.: 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). CEUR-WS, Online (2021). https://foops.linkeddata.es/assets/iswc_2021_demo.pdf

12. Isaac, A., Haslhofer, B.: Europeana linked open data – data.europeana.eu. Semant. Web **4**(3), 291–297 (2013)

13. Lodi, G., et al.: Semantic web for cultural heritage valorisation. In: Hai-Jew, S. (ed.) Data Analytics in Digital Humanities. MMSA, pp. 3–37. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54499-1_1

14. Navigli, R., Ponzetto, S.P.: BabelNet: building a very large multilingual semantic network. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 216–225 (2010)

15. Olderr, S.: Symbolism a Comprehensive Dictionary, 2nd edn. McFarland, London (2012)

16. Otto, A.F.: 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

17. Panofsky, E.: Studies in Iconology: Humanistic Themes in the Art of the Renaissance. Westview Press, Boulder (1972)

18. Pasqual, V., Tomasi, F.: Linked open data per la valorizzazione di collezioni culturali: il dataset mythlod. AIB studi **62**(1), 149–168 (2022). https://doi.org/10.2426/aibstudi-13301. https://aibstudi.aib.it/article/view/13301

19. Rebele, T., Suchanek, F., Hoffart, J., Biega, J., Kuzey, E., Weikum, G.: YAGO: a multilingual knowledge base from Wikipedia, Wordnet, and Geonames. In: Groth, P., et al. (eds.) ISWC 2016. LNCS, vol. 9982, pp. 177–185. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46547-0_19

20. Sabah, R., Sabah, Z.: Modeling nomisma ontology and comparing solutions for uncertainty. Ph.D. thesis, Goethe University Frankfurt (2022)

21. Sartini, B.: Connecting works of art within the semantic web of symbolic meanings. Ph.D. thesis, Alma Mater Studiorium, University of Bologna (2023). https://doi.org/10.48676/unibo/amsdottorato/11020. http://amsdottorato.unibo.it/id/eprint/11020

22. Sartini, B., Baroncini, S.: A comparative study of simple and complex art interpretations in linked open data using icon ontology. In: Proceedings of the International Workshop on Semantic Web and Ontology Design for Cultural Heritage co-located with the International Semantic Web Conference 2023 (ISWC 2023). CEUR Workshop (2023). https://ceur-ws.org/Vol-3540/paper4.pdf

23. Sartini, B., Baroncini, S., van Erp, M., Tomasi, F., Gangemi, A.: Icon: an ontology for comprehensive artistic interpretations. J. Comput. Cult. Herit. **16**(3) (2023). https://doi.org/10.1145/3594724

24. Sartini, B., van Erp, M., Gangemi, A.: 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 2021, pp. 201–208. Association for Computing Machinery, New York (2021). https://doi.org/10.1145/3460210.3493552

25. Sartini, B., Gangemi, A.: 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, pp. 576–580 (2021)

26. Sartini, B., Nesterov, A., Libbi, C., Brate, R., Shoilee, S.B.A., Daniil, S.: Multivocal exhibition: a user-centric application to explore symbolic interpretations of artefacts from different cultural perspectives. In: ExICE - Extended Intelligence for Cultural Engagement Conference, Bologna (2023). https://spice-h2020.eu/conference/papers/posters/ExICE23_paper_7072.pdf

27. Sartini, B., Vogelmann, V., van Erp, M., Gangemi, A.: Comparing symbolism across Asian cultural contexts. In: Digital Humanities 2022 Conference Abstracts,

pp. 358–361. DH2022 Local Organizing Committee (2022). https://dh2022.dhii.asia/abstracts/567

28. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Commun. ACM **57**(10), 78–85 (2014). https://doi.org/10.1145/2629489
29. Wilkinson, M.D., et al.: The fair guiding principles for scientific data management and stewardship. Sci. Data **3**(1) (2016). https://doi.org/10.1038/sdata.2016.18
30. Zhao, F.: A systematic review of Wikidata in Digital Humanities projects. Digital Scholarship in the Humanities (2022). https://doi.org/10.1093/llc/fqac083