Versione PROVVISORIA del contributo presentato al Convegno Annuale

Towards the unchaining of symbolism from knowledge graphs: how symbolic relationships can link cultures.

Bruno Sartini[1], Aldo Gangemi[2]
[1] University of Bologna, Italy – bruno.sartini3(«»)unibo.it
[2] University of Bologna, Italy – aldo.gangemi(«»)unibo.it

ABSTRACT
The aim of the work here proposed is to address the lack of information concerning symbolic meaning in linked open data of the Cultural Heritage domain. Moreover, it is emphasized how this issue limits the interconnections between cultures and cultural heritage items. A review of the current semantic databases and their methods to encode symbolism is presented. Then, an empirical experiment is conducted by describing the symbolism of 15 elements depicted in a CH item using a prototype ontology. The symbolism of those elements has been expanded by including information from a renowned source, and a knowledge base has been created which includes their potential symbolic meaning according to different cultures. This KB has later been matched to a dataset of 3197 paintings, belonging to various genres, extracted from Wikidata. An initial quantitative and qualitative analysis of the results of the matching is presented to demonstrate the potential of a linked data-based semantic representation of symbolism.

KEYWORDS
Symbolism; Knowledge Graphs; Cultural Heritage; Semantic Web

1. INTRODUCTION

The definition of a symbol is a highly debated topic in many contexts such as Philosophy, Linguistics, Anthropology as it is emphasized by Eco (1997). In this work, symbols are considered as elements that convey figurative or implicit meanings. These meanings can change depending on the context. The anchor is a “symbol of security and safety” in the Japanese culture (Otto, 1902, 30). On the other hand, it is used by the German painter Caspar David Friedrich in his painting *The Cross at Rügen* to express “hope of resurrection” (Roberts, 1998, 654).

Semantic web-based knowledge graphs have been applied to cultural heritage due to their dynamic nature able to encode the heterogeneous links that exist between CH objects (Lodi, 2017). But, because of the lack of structured symbolic information in these KGs, few connections can be highlighted on a cultural level, therefore limiting the potential offered by linked data with links based only on standard metadata. Additionally, most of the symbolic information of cultural objects are “chained” to unstructured descriptions that cannot exploit the full potentiality of Semantic Web technologies.

2. STATE OF THE ART

In this section, the following Semantic databases i) ArCo ii) Europeana iii) Wikidata will be analyzed and compared on their extractable information, especially in the context of symbolism and cultural interconnections.

ArCo[1] is the knowledge graph of Italian Cultural Heritage. Its data is encoded with its own data model which introduces new classes and properties and imports external ontologies (Carriero, 2019). Every CH object is described according to cataloging standards converted to RDF. In ArCo the property “subject” along with Dublin Core (i.e. `dc:subject`)[2] is used to link the CH objects with their subjects which are not furtherly linked to any symbolic meaning and neither is the CH object. Nevertheless, in some strings encoded using the Dublin Core property “description”, there are some references to iconographic analyses.[3]

[1] http://wit.istc.cnr.it/arco
[2] https://dublincore.org/specifications/dublin-core/dcmi-terms/
[3] An example of this description can be found in https://dati.beniculturali.it/lodview-arco/resource/HistoricOrArtisticProperty/0800084295.html

Europeana[4] is a semantic database developed for the diffusion, digital transformation, and preservation of cultural heritage. It uses the Europeana Data Model to encode the data[5]. It shares with ArCo the use of `dc:subject` and `dc:description`.

Wikidata[6] is a knowledge base that gathers data from the other branches of Wiki (such as Wikipedia, Wikimedia) converting it to a structured linked data format following its own data model. Differently from ArCo and Europeana, Wikidata has a property named P180 or "depicts" that links a painting to the elements that appear inside it.

The unstructured descriptions provided by ArCo and Europeana limit both the cultural interlinking through symbolism and the information extraction potential. SPARQL queries on those only work by including vague regular expressions (such as "symbol of") which do not guarantee to retrieve all the relevant data (recall). Moreover, their unstructured form makes them incompatible with logical inferences. The combination of these issues tends to obfuscate (or "chain") the potential of knowledge graphs to express symbolic content.

3. EXPERIMENT METHODOLOGY[7]

The painting *Attributes of rDo-rje Kon-btsun De-mo* (fig.1) was chosen as the starting point for the experiment. Its natural language description[8] was manually de-structured (or "unchained") to get the elements depicted in it (crown, gold, mule, white horse, golden earring, peony, mirror, thunderbolt, arrow, flaming sword, lotus, white robe, necklace, veil, phoenix) and their symbolic meanings. Their symbolism was then enriched using knowledge extracted from Olderr's Dictionary of Symbols (2012), which added potential symbolic meanings of those elements according to *Chinese, Buddhist, Tibetan* contexts.

The data was encoded using the turtle serialization of RDF and following a rapidly prototyped ontology, created for the sake of the experiment. In this model, the symbolic relation holding between a symbol and its meaning is called *Simulation* and is "reified" so that we can talk about it. Simulations are associated with a *Cultural Heritage Item, a Cultural Context*. Appropriate relations (*personification of, emblem of*) are used to formally represent simulations so that they fit the expressive power of RDF and OWL[9]. A symbol (called *Simulacrum*) and its meaning (called *Reality Counterpart*) are linked to their Simulation relation.
The code representing the symbolic knowledge of the painting contains 745 triples. A summary of the process that "unchained" the symbolic meaning from the natural language description can be found in the table 1.

[4] https://www.europeana.eu/
[5] https://pro.europeana.eu/page/edm-documentation
[6] https://www.wikidata.org/
[7] All the scripts and results of the experiment are available here: https://github.com/br0ast/symbolisminKGexperiment
[8] Available here: https://wellcomecollection.org/works/dhewzen3
[9] https://www.w3.org/OWL/

Part of Painting Description from Dorje (2003)	Element chosen	Symbolic meaning (from description)	Example of Symbolic meaning extracted (from Olderr's Dictionary)	Turtle encoding of **peony-long life symbolic relationship** using the ontology
The painting shows the attributes of Dorje Kongtsun Demo [...] Below the lotus cushion a row of offerings is arrayed:[…], a vase filled with **peonies symbolizing long life**	Peonies (lemmatized as peony)	long life; Attribute of Dorje Kongtsun Demo	CHINESE CONTEXT: love; affection; wealth; honor.	`symb:peonylonglife a symb:Simulation ; symb:hasContext symb:Tibetan, symb:Chinese, symb:Buddhist ; symb:hasRealityCounterpart symb:longlife ; symb:hasSimulacrum symb:peony .`

Table 1: Process of the unchaining of "peony" symbolism from natural language to turtle structured description

Figure 1: Detail of *Attributes of rDo-rje Kon-btsun De-mo* provided by Wellcome Collection

Then, a knowledge base was created by extracting from the dictionary the symbolic meanings associated with these 15 elements belonging to *general, Greco-Roman, flower language* and *Christian* contexts.

Once the knowledge base was set, 3197 paintings were extracted from Wikidata through its SPARQL portal along with the entities depicted in them (property P180). Resources were chosen from 5 different genres (property P136): Mythological paintings (1000 paintings), Allegory (473), Animal Art (378), Floral Painting (376) and Religious art (1000).

The mapping methodology consisted in string matching between the 15 elements of the knowledge base and the depicted elements of the paintings. For every match, symbolical information was automatically encoded for the painting according to its genre. Table 2 shows how the dictionary contexts were associated to the genres.

	Myth. Paintings	Allegory	Animal Art	Religious Art	Floral Painting
General	X	X	X	X	X
Christian				X	
Greco-Roman	X				
Flower Lang.					X

Table 2: Mapping between painting genres and the dictionary contexts

4. CONCLUSION

In this section, a quantitative analysis on the results of the matching is accompanied by an estimation of possible information that could have been added with a more robust knowledge base. Then, two findings about intercultural linking are presented.

Out of all the 3197 paintings extracted, 288 contain at least one of the 15 elements present in the knowledge base: 72 belong to the Allegory Genre, 53 to Animal Art, 18 to Floral painting, 99 to Mythological Paintings and 50 to Religion. The symbolic relationships matched are 393, described by 7587 triples. By considering the symbolism of all the 2564 unique subjects depicted in the paintings, more than one million triples could be generated.

By comparing the Tibetan painting to the knowledge base, it was found that the Buddhist Goddess Dorje Kongtsun Demo shares the attributes of phoenixes, veil and white robe with the Christian personification of Chastity.
Another discovery regards two paintings: *The Magdalen Reading* by Rogier van der Weyden and *Saint George Slaying the Dragon*[10] by Édouard Debat-Ponsan. Looking at their standard Metadata, the only thing that they share is the Wikidata genre of Religious Art. On the symbolic level, after the matching, both express the concept of lust. The literature on these paintings or the characters depicted inside them (Erhardt, 2012) (Fraser, 2017) confirmed their potential symbolic meaning and corresponding connection.

This work presented an experiment on the inclusion of structured symbolic information into linked data of Cultural Heritage. The initial results of the enrichment of a knowledge database on both the quantitative and qualitative sides are promising.

REFERENCES

[1] Carriero, V. A. et al. ArCo: The Italian cultural heritage knowledge graph. International Semantic Web Conference. Springer, Cham, 2019.

[2] Eco, U. Semiotica e filosofia del linguaggio. Einaudi, 1997.

[3] Dorje, G. A rare series of Tibetan banner. in N. Allan (ed.), Pearls of the orient: Asian treasures of the Wellcome Library. London, 2003.

[4] Erhardt, M. et al. Iconographic Studies from the Middle Ages to the Baroque. Brill Press, 2012.

[5] Fraser, G. Saint George and the Dragon: Saintly and Othered Bodies. The Corvette, 4, 2017: 1-13

[6] Lodi G. et al. Semantic Web for Cultural Heritage Valorisation. In: Hai-Jew S. (eds) Data Analytics in Digital Humanities. Multimedia Systems and Applications. Springer, Cham. 2017.

[7] Olderr, S. Symbolism a Comprehensive Dictionary. McFarland, 2012.

[8] Otto, A.F. et al. Mythological Japan, or The Symbolisms of Mythology in Relation to Japanese Art. Drexel Biddle, 1902.

[9] Roberts, H.E. Encyclopedia of Comparative Iconography. Fitzroy Dearborn Publishers, 1998.

[10]Accessible here: https://www.wikidata.org/wiki/Q3039856, https://www.wikidata.org/wiki/Q3947220