ARTO: An Artwork Object Ontology for Descriptive and Contextual Captioning

Can Yang[1], Bernardo Pereira Nunes[1], Sergio Rodríguez Méndez[1], Yige Chen[1], Rubén Manrique[2], and Marco Antonio Casanova[3]

[1] Australian National University, Canberra, ACT 2600, Australia
[2] Universidad de los Andes, Bogotá, 111711, Colombia
[3] PUC-Rio, Rio de Janeiro, 22451-900, Brazil

Abstract. Existing artwork ontologies focus mainly on artwork management (e.g., basic metadata) lacking detailed content and contextual representation. To fill this gap, we propose the ARTwork object Ontology (ARTO), designed to provide a comprehensive representation of artwork from both descriptive and contextual perspectives. The Artwork Descriptive Model represents the artistic expression of artwork, including visual elements, scenes and emotions, while the Artwork Contextual Model captures background information such as historical framework and related events. The ontology is designed using a data-driven approach, building on existing available data and incorporating concepts from established cultural heritage ontologies such as CIDOC-CRM and EDM. We evaluated ARTO using the OntOlogy Pitfall Scanner and interviewed 10 art experts to validate and refine our model.

Keywords: Contextual Object Model, Descriptive Object Model, Ontology, Captioning, Artworks

1 Introduction

With the development of digital technologies, cultural heritage institutions have evolved how they catalogue and manage artworks [17,33], which relies heavily on accurate and consistent metadata. However, institutions still face the challenge of finding a more efficient approach when it comes to captioning [8]. Traditionally, creating artwork captions is a manual and time-consuming process requiring significant knowledge and expertise [10], resulting in limited coverage and overall inconsistency [21]. Despite the advancements in artwork management, artwork captioning remains a challenge. Our proposed ontology aims to capture and model the complex and comprehensive knowledge in the art domain, providing a solid foundation for generating captions in a broad scope.

The existing artwork-related ontologies have certain limitations in supporting and covering the different dimensions required by the captioning task. As defined by Anderson [1] and Barrett [2], artwork captioning can be categorised into three types: descriptive, contextual and interpretive. Descriptive caption endeavours to create a superficial visual representation of the artwork so that the viewers

can reconstruct the image of the object right in their own minds. Contextual caption aims to provide the historical, cultural and social background to which the artwork pertains, thus helping viewers to better appreciate and understand the artwork with the underlying context. Interpretive caption offers subjective interpretations, fostering critical thinking and prompting viewers to develop their personal comprehension of the artwork.

It is worth noting that extant ontologies often fail to capture multiple perspectives and focus primarily on basic information and broad contexts of artworks, such as title, format, size and time of creation [12]. The lack of detailed and insightful content representations of artwork ontologies also affects the accuracy and completeness of artwork descriptions, as they may not objectively, comprehensively and completely reflect the hidden meanings of the artwork, thus limiting the viewer's understanding of the artwork and the story behind it. Thus, it is needed to develop ontologies that incorporate multiple, integrative perspectives. Besides, established captioning approaches are mainly for realistic images. Although there have been some attempts at image captioning for artworks, such as [8,19,18,35,28,16,20], most of them directly apply existing methods used for recognition tasks to train machine learning models using art-related data often overlooking the significant differences between artworks and realistic images. Captions for artworks highlight more cultural and historical information and interpretation of images, rather than simply listing their contents. Thus, this paper introduces a novel artwork caption ontology that aims to provide comprehensive and holistic information about artworks with a focus on paintings. Given the complexity and challenge of generating captions for paintings in the artwork captioning domain, demonstrating the applicability of the ontology can provide a solid theoretical foundation for its broader utility.

The proposed ontology is structured in two parts:

(i) *Artwork Descriptive Model*, emphasising artistically specialised content in artworks, including visual elements, objects, scenes and emotions;

(ii) *Artwork Contextual Model*, providing a comprehensive collection of artwork-related information, including historical context and related events.

Although we followed a data-driven approach to build ARTwork Object Ontology (ARTO), we validated it through interviews using the Think-aloud Protocol (TAP) [4]. The interviews helped to validate the structure and content of our model from a practical perspective and to ensure that it meets the requirements and vision of the captioning task in the artwork field. The logical side of ARTO was also evaluated using the OntOlogy Pitfall Scanner! (OOPS!) [24]. The main contribution of our study is the development of an ontology for artwork captioning, including the *Artwork Descriptive Model* and the *Artwork Contextual Model* validated by a well-designed protocol with the potential to be generalised to other subfields of arts and domains.

2 Related Work

Existing ontologies and vocabularies have made significant contributions to the organisation and representation of knowledge in the art domain.

Ontologies related to Artwork: Several artwork ontologies have been proposed for different scenarios. CIDOC-CRM [3] is a general model for describing concepts and relationships in the cultural heritage domain. The Visual Representation (VIR) ontology [6] extends CIDOC-CRM to describe the representation and relationships of visual features. However, it lacks a detailed description of visual elements, such as colour and composition. Based on CIDOC-CRM, the Functional Requirements for Bibliographic Records (FRBR) [13] model extends it to better represent bibliographic information and its relationship with art and cultural objects. Similarly, ArCo [7] is also built upon CIDOC-CRM, specifically focusing on Italian cultural heritage. Europeana[4], an institution that aggregates millions of digital items from European cultural collections, defined the Europeana Data Model (EDM) [12] to facilitate interoperability and data integration among different sources. The Visual Resources Association Core (VRA CORE) [32] is designed specifically for managing data on visual culture, such as paintings, sculpture, books and performance art. It integrates artwork information such as title, artistic style, material types, etc. Another relevant work is the ICON ontology [27] that focuses on representing the knowledge of iconology and iconography, covering aspects such as background information, artwork interpretation and reference sources.

Vocabularies for Artwork: Getty Research Institute has developed the Art & Architecture Thesaurus (AAT) [14] and the Union List of Artist Names (ULAN) [15] vocabularies which provide a foundation for the knowledge in the art domain. The AAT is a structured vocabulary covering concepts in art and related domains. The ULAN focuses on standardising artist names and has been employed to construct visualisations of artist social networks [30].

Although existing ontologies have been widely used in the art domain, they have limitations for the task of artwork captioning, particularly in their modelling of content and events. Many existing models attempt to include an extensive range of properties and classes to construct a complex ontology covering all art types. However, they primarily focus on organising and managing basic metadata and contextual information of artworks, lacking in the modelling and representation of the artwork content itself. Therefore, it is necessary to propose a new ontology specialised for artwork captioning, which captures the rich visual details, relationships and contextual information that are necessary for accurate and informative artwork captions. By focusing specifically on paintings and simplifying the scenarios, we can create a comprehensive and effective ontology that captures the detailed elements and relationships within this type of artwork.

[4] https://www.europeana.eu/en

3 Artwork Ontology Model

To overcome the current limitations of ontology in comprehensively describing artworks, we have designed ARTO that aims to capture the multi-perspective characteristics of artworks. This ontology is represented by two models: the Artwork Descriptive and the Artwork Contextual models. The Artwork Descriptive Model focuses on the expressive content of the artwork itself, such as visual elements, scenes and emotions, providing a framework for the audience to perceive and imagine the artwork directly. The Artwork Contextual Model, on the other hand, details the context information of the artwork, for example, metadata, creation background, provenance and artists' experience, thereby offering a systematic approach to capturing and presenting a richer understanding of perspectives. By combining the content and context of the artwork, our model can provide comprehensive information about artworks and promote viewers' deeper comprehension and appreciation.

3.1 Methodology

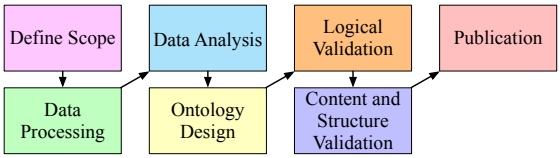

Fig. 1. Ontology Design Workflow

The construction of the artwork ontology follows the NeOn methodology [29], which emphasises reusing and refactoring of existing ontologies. Figure 1 presents the workflow of creating the artwork captioning ontology. We initiated the design process by collecting and analysing data from online galleries and museums to build the conceptual structure of the ontology. Then, we identified the core concepts, such as artists, events, style and creation techniques, and defined the relationships between them, such as "belongs to", "has event" and "created by", forming a semantic network. Throughout this process, we consider reusing existing ontologies or vocabularies, such as the CIDOC-CRM [3], EDM [12] and the Event ontology [26], to promote interoperability and leverage established conceptualisations.

As for the evaluation and validation of the ontology design, we employed OOPS! [24] to assess the logical consistency and completeness of the ontology. We also invited art experts to review the ontology's structure and content, incorporating their feedback to refine and improve the ontology. Finally, we have published our Artwork Ontology[5] where we provide human-readable and machine-readable documentation, describing its structure and content. To

[5] ARTO is available at the following persisting identifier: https://w3id.org/arto.

accommodate different users and systems, the ontology is available in multiple serialisation formats, such as RDF/XML and Turtle. We use the MIT license agreement to specify the usage rights and permissions. We have established version control and update mechanisms to incorporate new knowledge and address any identified issues, ensuring the ontology remains up-to-date and relevant.

3.2 Ontology Design Rationale

Table 1. Data Sources Statistics

Data Source	**Number of Artworks**	**Has Schema?**	**Number of Attributes**	**Copyright**
Europeana	2,594,696	Yes	17	Fully Public
Wikiart	172,397	No	17	Partially Public
Artic	93,836	No	9	Partially Public
American art	39,388	No	13	Partially Public
Louvre	512,613	No	9	Public Metadata
Nation Museum	82,463	No	6	Partially Public
NGA.gov	139,723	No	26	Fully Public
MET museum	218,930	No	12	Partially Public
Brooklyn Museum	67,529	No	13	Partially Public
Getty	86,848	Yes	10	Partially Public
British Museum	2,123,229	No	16	Partially Public
Art bank	10,727	No	9	For research and study

Designing an ontology requires both specialised expertise in the field of art and a deep understanding of the existing data intended to populate the ontology. The ontology not only needs to capture the broad categories and relationships inherent in the art but also be fine-tuned to the specific characteristics, techniques and contexts that are present in the collection. To develop ARTO, we adopted a data-driven approach. We began by collecting artworks from well-known art websites listed in Table 1, such as Wikiart, Art Institute of Chicago and the National Gallery of Art. These websites typically provide two types of information: basic artwork details (e.g., title, artist, location, year) and their contextual descriptions. We first extracted and organised the basic artwork information according to their meanings, as different websites often use varying terminologies. The other part is about the overall description of the artwork, which varies greatly. Popular artworks usually have a lot of descriptions, including specific visual elements of the artwork and interpretations, the artist's creative background and even the artist's biography. Most lesser-known artworks, however, have little to no description. We further analysed the content of these descriptions and classified them. Using the collected data, we constructed a preliminary ontology. After scraping and analysing multiple websites and data sources, we recognised the need for a more detailed representation of artworks. ARTO defines the structure of the artwork-related information (see Figure 2),

aimed at providing a systematic and comprehensive framework for context-rich image captioning. This ontology covers:

- **Basic Attributes:** Includes basic attributes that every artwork would have, such as title, date of creation, genre, medium and style.
- **Visual Components:** It focuses on the physical and visual aspects of the artwork, including colour, texture, composition, line, shape and dimensions. It also covers the objects and scenes depicted in the artwork and the techniques and materials the artist used to create these visual components.
- **Content:** The content of the artwork includes the themes, symbolic elements and emotions presented in the artwork.
- **Background Events:** The background information, including the artist's background, the artwork's provenance and exhibition history.

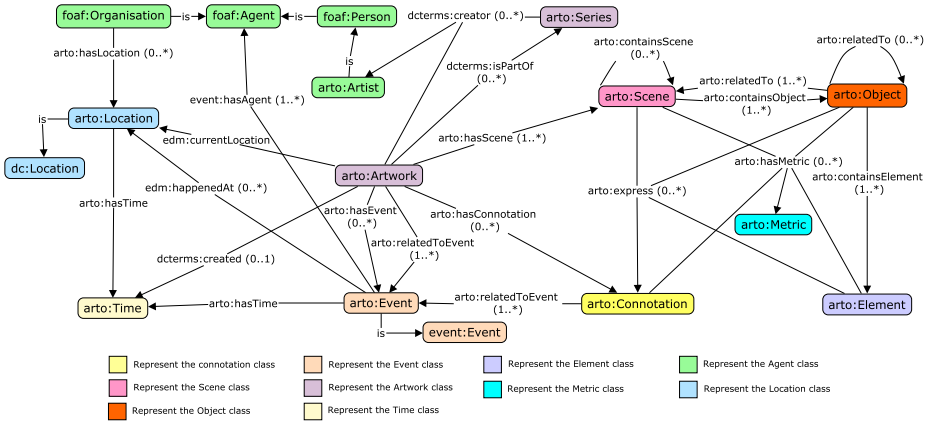

Fig. 2. Core Model of the Artwork Object Ontology. *The colours are used consistently in this paper to represent different classes and their instances.

3.3 Artwork Descriptive Model

Looking at the existing ontologies related to artworks, most of them do not attempt to model the content of artworks, leaving a significant gap. The content is an essential part of an artwork, as artists convey specific themes and emotions through elements such as colour, texture and composition, all of which are key components of the artwork. Due to the complexity of artwork content, traditional classification methods struggle to capture their rich connotations. Similarly to the Document Object Model (DOM) [34], which neutrally captures the inner structure of documents, we propose the Artwork Descriptive Model aiming at describing the content of artworks. We note that this model currently targets artworks that contain distinguishable real-world objects, such as "landscape paintings" or "portrait paintings". Figure 3 presents the concept map of the Artwork Descriptive Model, which decomposes artwork content into four levels: basic element, object, scene and connotation:

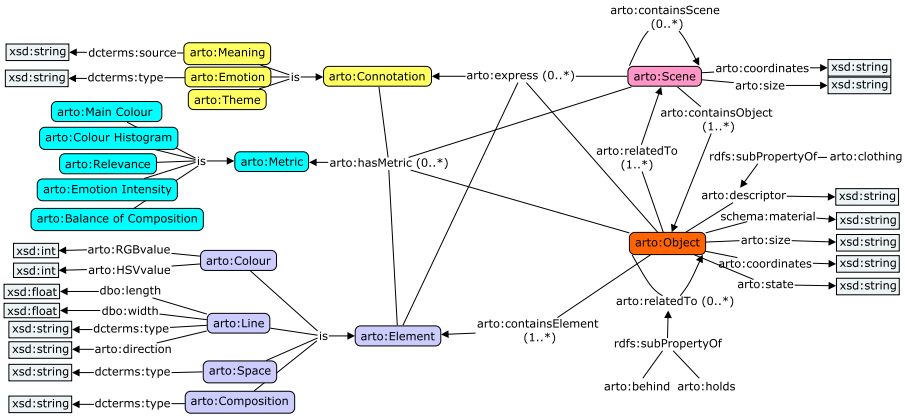

Fig. 3. Artwork Descriptive Model

- **Basic Elements**: Basic elements contain colour, line, shape, texture, space and composition. Of these, colour and line form the foundation, as their combination can create shape and texture. Space pertains to the representation of objects and their three-dimensional relations within the artwork and composition involves the arrangement and layout of these objects.
- **Object**: Objects are the main components of the artwork. Various objects interrelate and collectively form the entire artwork. Each object has its unique state and attributes that reflect the details and depth of the artwork.
- **Scene**: A scene can be viewed as a portion of the artwork composed of a group of objects that convey a specific meaning or sentiment. A complex artwork is typically not constituted by only one scene but rather by multiple meaningful scenes that jointly articulate the overall theme of the artwork. Each scene plays an indispensable role in the whole artwork.
- **Connotation**: From the high-level perspective to analyse the meaning, theme and emotion of the artwork. While the theme of most artworks can be discerned directly from their content, some works, especially those that are implicit or abstract, cannot be recognised directly. For such artworks, understanding the background of its creation, and the artist's experiences becomes the key to truly understanding the essence of the artwork.

To map the Artwork Descriptive Model to the ontology, each layer of the model is represented as a class. To facilitate quantitative analysis, the `Metric` class represents measurable aspects of the content as we explain in what follows.

The `Element` class represents the fundamental visual components of an artwork, with `Colour`, `Line`, `Space` and `Composition` as its key subclasses for paintings. Each subclass of `Element` has its own unique properties. For instance, the `Colour` class includes properties of `RGBvalue` and `HSVvalue`, while the `Line` class has properties like type, direction, length and width. These properties enable a detailed description of the basic elements within an artwork.

The `Object` class represents the identifiable and meaningful entities depicted in an artwork, which are composed of various `Elements`. Objects can be fur-

ther described by their specific properties, such as `size`, `material` and `state`, which indicates the condition or appearance of the object. The properties of an `Object` vary depending on its type, therefore, the `descriptor` property allows for extendable subproperties to describe any aspects of different objects. Figure 7 illustrates the extensibility through the "clothing" subproperty.

The `Scene` class represents a meaningful aggregation of objects. The `relatedTo` relationship links an object to the scene it appears. And an object can be assigned to one or more scenes.

The `Connotation` class is at the top of the hierarchy, capturing the high-level semantics and interpretations of an artwork. It has three subclasses: `Meaning`, `Emotion` and `Theme`. The `Meaning` represents the underlying message or symbolism of the artwork. The `Emotion` describes the emotional expression of the artwork. The `Theme` subclass encapsulates the central idea or subject of the artwork. In artwork, the `Scene`, `Object` and `Element` are all relevant to the expression of `Connotation`.

For structural clarity and better analysis of the content of the artwork, we designed a `Metric` class to represent the various possible metrics, including some statistical information about the basic elements such as colour statistics, emotion intensity and metrics of other visual elements. These classes jointly build a multi-level ontology for representing the content of artworks.

3.4 Artwork Contextual Model

While the Artwork Descriptive Model focuses on the intrinsic elements and content of the artwork, the Artwork Contextual Model provides the necessary contextual information to situate the artwork within its historical, cultural and artistic context. This model: (i) characterises artworks by capturing the essential properties, such as the genre, medium and style; (ii) brings out underlying events or facts that are represented by the characters or objects in the artwork; (iii) captures events related to the artwork, such as exhibitions, provenance, or events related to the artist, such as where they studied, and travelled, as well as identify the time and location for each event; and, (iv) establishes connections between artworks and various agents involved in their life cycle—from creation to presentation.

The Artwork Contextual Model builds upon and extends existing ontologies in the cultural heritage domain, such as CIDOC-CRM [3] and EDM [12]. These ontologies provide a solid foundation for representing cultural heritage information. Our model aims to narrow down the context of artworks and simplify the representation of concepts. We borrowed concepts representing classes and properties from CIDOC-CRM and EDM, such as `E5_Event`, `E53_Place` and `E52_Time-Span`, to build our Artwork Context Ontology.

In addition, our model is also inspired by the Provenance (PROV-O)[6] and Event Ontologies [26]. PROV-O offers a way to represent the provenance information of things, which is crucial for tracing the history and ownership of

[6] https://www.w3.org/TR/prov-o/

artworks. The Event Ontology provides a framework for describing events and their relationships to entities, which fits well with our goal of capturing events related to artworks and artists. By leveraging the modelling structure and approach of these ontologies, we built our event modelling, which can effectively represent historical context, events behind creation, provenance and others.

Furthermore, our model incorporates elements from other art-related ontologies, such as the Art and Architecture Thesaurus [14] for representing artistic styles, techniques and materials, and the Union List of Artist Names [15] for identifying and disambiguating artists. By reusing and extending existing ontologies, the Artwork Context Model (Figure 4) provides a more detailed representation of artworks while maintaining interoperability with other data sources. The Artwork Context Model covers properties and the context of artworks. The ontology comprises five main classes: `Artwork`, `Event`, `Location`, `Time` and `Agent`, emphasising extensibility and offering flexible design for different classes.

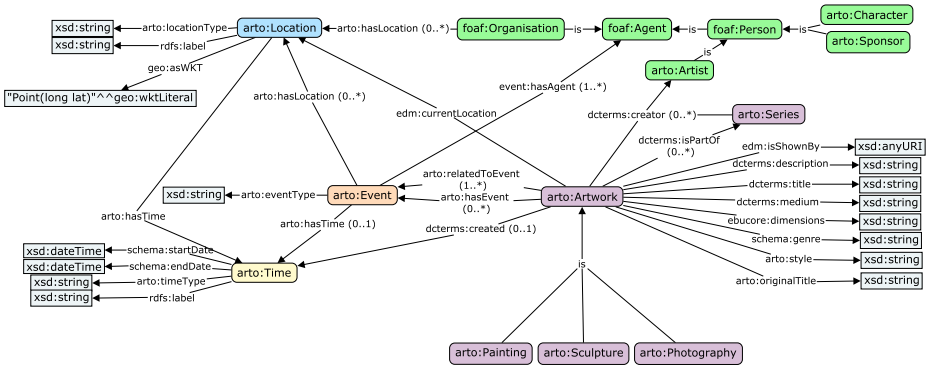

Fig. 4. Artwork Contextual Model

The `Artwork` class mainly refers to visual art, covering a range of categories such as Painting, Sculpture, Photography, Printmaking and Digital Art. For this paper, the emphasis is on `Painting` as the subclass of `Artwork`. While our illustration in Figure 4 only includes `Painting`, `Sculpture` and `Photography,` we encourage and welcome the integration of diverse and varied types of artworks. We proposed that the categories under `Artwork` are open-ended and flexible, to embrace the richness of diversity in the art world. Additionally, to maintain relevance and applicability, the `Series` class represents a collection of artworks. The ontology includes the fundamental properties of an artwork, including title, theme, description, elements, style, medium, format and image resource. It also allows for the representation of more specific properties relevant to different types of artworks.

The `Time` class represents one of the most relevant aspects of an artwork. Time potentially encodes events that influenced its creation. The temporal data provides the chronology of artworks and also implies the historical, cultural or personal contexts. Representing time is not a straightforward task, especially

for historical data. The time data might be annotated with precise dates or be associated with broader epochs, seasonal timelines, or even vague indications like 'late Renaissance'. These diverse temporal expressions mandate a flexible representation model. Therefore, we have devised a `timeType` categorisation to accommodate specific years, months, days, seasons and broader, more qualitative expressions such as 'early/late 19th century.' All instances of time are treated as intervals, which means every time instance has a `startDate` and an `endDate`. This interval-based methodology provides a structured and consistent means of representing time. Therefore, the `Time` class is adaptable and extensible. We note that, for our purposes, borrowing concepts directly from the Time Ontology[7] would unnecessarily add more complexity to our model.

Regarding the `Location` class, we have implemented a strategy similar to `timeType`, considering the frequent correlation between location and time in historical contexts. The discernment of specific locations within historical records poses considerable challenges due to the dynamism and variability inherent in geographical delineations over time. To streamline this intricate process, we have instituted a `locationType`, which encompasses a wide array of geographical entities such as countries, states, cities, districts, streets, and more, allowing for the representation of diverse spatial granularities. Moreover, we have integrated geographical coordinates to serve as precise locational indicators, facilitating more accurate spatial analysis. Therefore, our ontology adopts the `geo:wktLiteral` datatype from GeoSPARQL [22] standard to represent the specific coordinates of the location. In situations where historically existent nations or regions lack unequivocal coordinates or where boundary definitions have evolved, we endeavour to align them with contemporary geographical demarcations to deduce an approximate scope. This alignment process is meticulous, ensuring that the approximated locations maintain as much historical accuracy as feasible, respecting the historical integrity of the regions involved. To connect and represent historical location names, we set an `existedIn` property that links the location with a time limitation. Additionally, our approach maintains flexibility to accommodate updates and refinements in location data. As new information becomes available or as historical understandings evolve, adjustments can be made to the `Location` class to reflect more accurate or nuanced geographical understandings, ensuring the continual relevance and accuracy of the spatial representations within our framework. This multifaceted approach to location allows for a comprehensive and adaptable representation of geographical entities and their interrelations over time.

In the given context, an `Agent` emerges as the main participant in any event, primarily comprising organisations and persons. `Organisation` can represent a myriad of institutional forms, such as museums, galleries and art schools. The `Person` class can be broken down into more specific classifications, such as artists, sponsors and characters. Moreover, this classification is dynamic and inclusive, accommodating the incorporation of additional roles as various events are incorporated. It is open-ended, allowing for the integration of new participant

[7]

types as they are identified, ensuring the `Agent` class remains comprehensive and reflective of the diversity and complexity of participation in events.

The structuring of the `Event` class is fundamentally anchored in three cardinal elements: time, location and agents. The concepts modelling is inspired by PROV-O, Event ontology [26] and DOLCE [5] which provides a comprehensive framework for capturing and representing real-world events. Our approach ensures the capture of various background information pivotal to understanding the artworks. For instance, it encompasses events occurring during the creation of the artwork, such as the artist's experiences, interactions, and the prevailing historical and cultural environment, all of which intricately weave into the fabric of the resultant artwork. These elements offer insights into the motivations, influences and conditions that shaped the artwork. It is also designed to integrate events post-creation, encompassing aspects like provenance history and exhibition history. These elements collectively contribute to the evolving narrative of the artwork, marking its journey through time and space and its interactions with various entities. To maintain an optimal balance between flexibility and comprehensiveness, we introduced an 'eventType' attribute. This captures the evolving and diverse events without compromising on specificity. We employed RDF-star[8] to represent the `Event` class for its enhanced capability to represent more complex relationships and events compared to traditional RDF, which is inherently constrained to binary relationships. When we use RDF to describe events, we need an additional event entity to serve as an intermediate entity to connect all the entities related to the event, such as people, time and location. However, with RDF-star, different entities are directly associated and it can better handle complex events and model their direct relationships. RDF-star not only extends our ability to encapsulate intricate events associated with artworks but also provides a more concise and easy-to-understand representation.

3.5 Artwork Object Ontology Example

Fig. 5. "The Shepherdess"[10], also known as "The Little Shepherdess", is a painting by William-Adolphe Bouguereau completed in 1889. The title is taken from the Southern French dialect. The painting depicts an idyllic, pastoral scene of a lone young woman in peasant attire posed for the artist, balancing a stick (likely her crook) across her shoulders, standing barefooted in the foreground. In the background are oxen grazing in a field.

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

Figure 5 illustrates a painting titled "The Shepherdess" by William-Adolphe Bouguereau. From the background information, we create a context ontology example of this painting using RDF-star format (Figure 6).

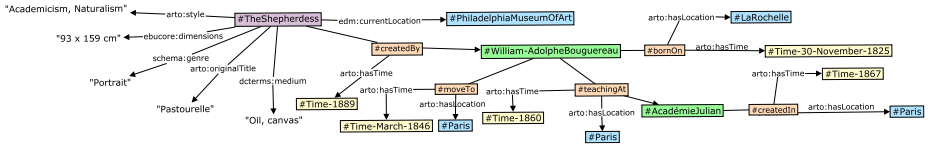

Fig. 6. ARTO Contextual Model of "The Shepherdess"

Additionally, Figure 7 provides information about the content of the artwork. It deconstructs the whole painting into different levels.

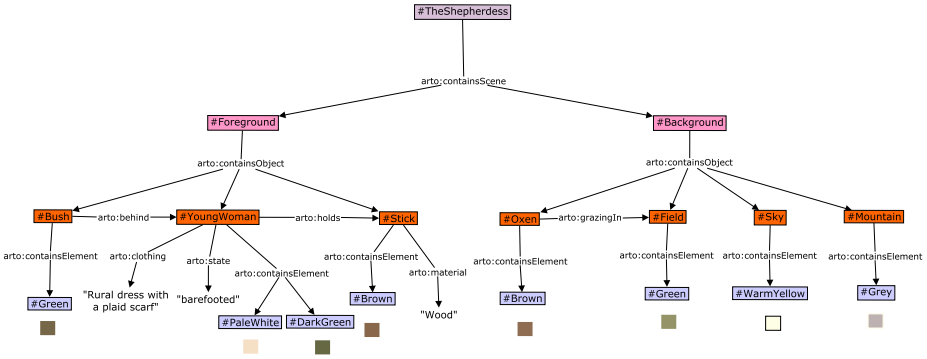

Fig. 7. ARTO Descriptive Model of "The Shepherdess"

4 Evaluation

In this section, we detail the evaluation process of the ontology, involving OOPS! [24] assessment and Think-aloud Protocol with 10 art experts.

4.1 OOPS! Assessment

To validate the quality and completeness of the proposed ontology, we utilised OOPS! [24], which is an online tool that helps to detect and troubleshoot potential errors and pitfalls during ontology creation. It serves as a valuable resource for ontology developers during validation processes and can detect 40 common ontology design flaws, such as loops between defined classes, lack of annotations, and use of the wrong data types. By uploading the ontology to OOPS! we

[10] https://en.wikipedia.org/wiki/The_Shepherdess

were able to systematically identify and address a range of potential problems in the ontology and resolve them. Analysis of the structure and content of the ontology by OOPS! revealed that the initially designed ontology carried risks of inappropriate class relationships, inconsistent use of attributes, and logical inconsistencies, all of which were modified. Notably, the only other hidden risk left was the inverse relationship not explicitly declared, which is considered irrelevant in the context of this ontology. This evaluation not only helps us to develop logical consistency in the ontology but also ensures its quality.

4.2 Think-aloud Protocol (TAP)

TAP is a widely used method that can be considered a form of synchronised verbal reporting [25,4,31,9]. This protocol was applied[11] to gather valuable input and guidance from art experts, with a focus on identifying the key features of artwork caption, validating the feasibility of the proposed ontology and providing feedback for improvement. In the TAP, subjects are asked to verbalise their immediate thoughts, feelings and cognitive processes while performing a specific task [11]. Through this method, we were able to identify and record the complete thought process of the artwork captioning task.

Protocol Design: The general procedure of the TAP[12] is: firstly, the art experts were expected to select two of their most familiar and two least familiar paintings from the given ten examples. Subsequently, they were required to verbally describe the four works and then label the captions they generated according to the categorisation of captions (descriptive, contextual and interpretive) [1,2]. Afterwards, they listed and ranked the aspects they considered most important when describing the artwork. The experts were asked to share their approach to describing the artwork, and their insights on describing subjective aspects of the artwork, such as emotion and subject matter. Next, they outlined the elements that they found most challenging to describe and their thoughts on how to verify the accuracy of the description. Finally, after presenting the experts with the initial design of the ontology, the experts needed to apply it to make captions for four artwork examples and evaluated whether the generated content was comprehensive and structured well. It is worth noting that at the beginning of the interview, the participants had no prior information about the ontology, thus reducing potential cognitive biases and enabling them to provide insights solely based on their extensive expertise.

Participants: We recruited 10 art experts who have advanced proficiency in painting or the visual arts, including professors at art colleges or practising artists. Participation was voluntary and participants' information was kept confidential throughout the study. Before participating in the study, all subjects signed an informed consent form agreeing to be part of the study and to be recorded. Each interview lasted approximately one hour.

Results: By analysing the process, content and challenges of the interviewees when generating artwork captions, we verified whether the ontology design

[11] The ANU Human Research Ethics Committee approved this research (2023/1399).
[12] Interview details and related information can be found at `https://w3id.org/arto`.

aligned with practical needs and covered sufficiently comprehensive information. At the same time, the suggestions for improvement and concerns raised by the art experts also guided the ontology's refinement. Below, we discuss several key findings revealed by the interview results and explain their impact on validating and improving the ontology's design.

In terms of the aspects involved in captions, 90% of the participants emphasised the importance of understanding the background of the artist and the creation environment, 60% focused on the content aspects of the artwork, such as composition, colour and characters, and 40% mentioned the meaning and interpretation of the artwork. In addition, we found that there are some differences in their focus when describing artworks they are familiar and unfamiliar with. For familiar artworks, they were more inclined to provide background information about the artwork, while for unfamiliar artworks, background details were usually not mentioned, and only one person mentioned that they might infer the background of the artwork. These findings indicate that although they have different focuses, the ontology covers key information for artwork captioning. This is further validated by their evaluation of the artwork caption types (descriptive, contextual and interpretive) and ontology examples. Regarding the type of captions, all participants agreed that the categorisation method was good, which was further verified by performing Ground Theory[23] analysis on the caption examples they created. Their feedback on the provided ontology examples was also positive. Most participants agreed on the comprehensiveness of the ontology.

Art experts expressed different views on the emotional analysis of artworks, with most emphasising the importance of understanding the content of the artwork and the personal experience of the artist in interpreting emotions. However, one participant questioned the need to objectively analyse the emotional content in an artwork, arguing that it can only be understood through the artist's experience. Additionally, they indicated the most challenging aspects of creating captions. Half of the participants stated that it was difficult to describe it from a professional perspective and 30% thought it was difficult to know the creator's intent. Another 30% found the contextual information difficult to verify, which ARTO can assist with. Our ontology contains information on content that is analysed from a professional perspective, and we also include contextual information that can better help people understand the background and support individual interpretation.

Apart from the direct evaluation of the ontology design, some aspects of the interview results provided insights for future works. Experts presented different opinions on the importance of different aspects. All mentioned the importance of content, but only around 60% mentioned the background and interpretation of the artwork. This provides support for the proportion of captions and emphasises the importance of the artwork content. To further investigate the importance of different aspects, we conducted a detailed survey where participants ranked 11 aspects from most relevant (1) to least relevant (11). The results showed that the "Artist" had the highest average importance rank, followed by "Content", "Composition", and "Background about the Artist". On the other hand, "Ob-

jects", "Emotion", and "Visual Elements" were considered less important on average. Experts emphasised the importance of incorporating background events in artwork captions. They highlighted the significance of the artist's personal experiences on their technique and style, as well as the relevance of the social and historical context during the artwork's creation, which provides insights into its thematic expressions. Moreover, the motivation behind the creation can provide meaning and emotion to the artwork. To capture these contexts, our ARTO-based knowledge graph will include a comprehensive range of events, and the caption generation model will emphasise these background details to enrich the viewer's understanding and appreciation of the artwork. Regarding the interpretation of the captions, half of the participants suggested maintaining objectivity and avoiding personal opinions, while the other half suggested collecting comprehensive information and using publicly recognised understandings or interpretations. This will help us to provide objective interpretations when automatically generating captions, presenting viewers with widely accepted interpretations while allowing them to form their own interpretations of the artwork. The insights from art experts provide a professional perspective for our ontology's subsequent task of artwork image captioning, helping us generate high-quality and comprehensive artwork captions. Overall, the results proved the adequacy and comprehensiveness of ARTO considering different types of captions, which met the needs of artwork captioning.

5 Conclusion and Future Work

This paper proposes two high-level models designed primarily for artwork captioning tasks: (i) the Artwork Descriptive Model, which represents the content of the artworks, such as visual elements, scenes and emotions; and, (ii) the Artwork Contextual Model, which captures background information, such as cultural and historical events. By integrating these two models, our proposed ontology enables a deeper understanding and representation of artworks for automate captioning.

The ontology was validated using two different methods. OOPS! was used to verify technical aspects, and the Think-aloud Protocol to gather expert knowledge and model art-related concepts besides understanding the artwork captioning process through the perspective of art experts.

Future work includes the integration of the ontology as part of a comprehensive pipeline[13] to automatically generate captions and to be applied to various artwork-related tasks beyond image captioning.

References

1. Anderson, T.: Defining and structuring art criticism for education. Studies in Art Education **34**(4), 199–208 (1993), http://www.jstor.org/stable/1320404
2. Barrett, T.M.: Criticizing art : understanding the contemporary / Terry Barrett. Mayfield Pub. Co Mountain View, Calif (1994)

[13] More detailed information can be found at https://w3id.org/arto.

3. Bekiari, C., Bruseker, G., Doerr, M., Ore, C.E., Stead, S., Velios, A.: Definition of the cidoc conceptual reference model v7.1.1. The CIDOC Conceptual Reference Model Special Interest Group (2021). https://doi.org/10.26225/FDZH-X261
4. Bernardini, S.: Think-aloud protocols in translation research: Achievements, limits, future prospects. Target. International Journal of Translation Studies **13**(2), 241–263 (2001)
5. Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E.M., Vieu, L.: Dolce: A descriptive ontology for linguistic and cognitive engineering. Applied Ontology **17**, 45–69 (2022). https://doi.org/10.3233/AO-210259, https://doi.org/10.3233/AO-210259, 1
6. Carboni, N., de Luca, L.: An ontological approach to the description of visual and iconographical representations. Heritage **2**(2), 1191–1210 (2019). https://doi.org/10.3390/heritage2020078, https://www.mdpi.com/2571-9408/2/2/78
7. Carriero, V.A., Gangemi, A., Mancinelli, M.L., Marinucci, L., Nuzzolese, A.G., Presutti, V., Veninata, C.: Arco: The italian cultural heritage knowledge graph. In: Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I., Hogan, A., Song, J., Lefrançois, M., Gandon, F. (eds.) The Semantic Web – ISWC 2019. pp. 36–52. Springer International Publishing, Cham (2019)
8. Cetinic, E.: Towards generating and evaluating iconographic image captions of artworks. Journal of Imaging **7**(8) (2021). https://doi.org/10.3390/jimaging7080123, https://www.mdpi.com/2313-433X/7/8/123
9. Craig, K., Hale, D., Grainger, C., Stewart, M.E.: Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. Metacognition and Learning **15**, 155–213 (2020)
10. Dijkshoorn, C., Jongma, L., Aroyo, L., van Ossenbruggen, J., Schreiber, G., ter Weele, W., Wielemaker, J.: The rijksmuseum collection as linked data. Semantic Web – Interoperability, Usability, Applicability **9**(2), 221–230 (Jan 2018). https://doi.org/10.3233/SW-170257
11. Ericsson, K.A., Simon, H.A.: Verbal reports as data. Psychological review **87**(3), 215 (1980)
12. Europeana Foundation: Europeana data model (edm) v5.2.8, https://pro.europeana.eu/page/edm-documentation
13. on FRBR/CRM Dialogue, I.W.G., Bekiari, C., Doerr, M., Le Boeuf, P., Riva, P.: Definition of frbroo: A conceptual model for bibliographic information in object-oriented formalism. Tech. rep., International Federation of Library Associations and Institutions (IFLA) (Mar 2017), https://repository.ifla.org/handle/123456789/659
14. Institute, G.R.: Art architecture thesaurus. Online (2024), available from: http://www.getty.edu/research/tools/vocabularies/aat/index.html [Accessed 13th March 2024]
15. Institute, G.R.: Union list of artist names. Online (2024), available from: http://www.getty.edu/research/tools/vocabularies/ulan/index.html [Accessed 13th March 2024]
16. Ishikawa, S., Sugiura, K.: Affective image captioning for visual artworks using emotion-based cross-attention mechanisms. IEEE Access **11**, 24527–24534 (2023). https://doi.org/10.1109/ACCESS.2023.3255887
17. Korro Bañuelos, J., Rodríguez Miranda, Á., Valle-Melón, J.M., Zornoza-Indart, A., Castellano-Román, M., Angulo-Fornos, R., Pinto-Puerto, F., Acosta Ibáñez, P., Ferreira-Lopes, P.: The role of information management for the sustainable conservation of cultural heritage. Sustainability **13**(8), 4325 (2021)

18. Liu, F., Zhang, M., Zheng, B., Cui, S., Ma, W., Liu, Z.: Feature fusion via multi-target learning for ancient artwork captioning. Information Fusion **97**, 101811 (2023). https://doi.org/https://doi.org/10.1016/j.inffus.2023.101811, https://www.sciencedirect.com/science/article/pii/S1566253523001203
19. Lu, Y., Guo, C., Dai, X., Wang, F.Y.: Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training. Neurocomputing **490**, 163–180 (2022). https://doi.org/https://doi.org/10.1016/j.neucom.2022.01.068, https://www.sciencedirect.com/science/article/pii/S092523122200087X
20. Lu, Y., Guo, C., Dai, X., Wang, F.Y.: Artcap: A dataset for image captioning of fine art paintings. IEEE Transactions on Computational Social Systems **11**(1), 576–587 (2024). https://doi.org/10.1109/TCSS.2022.3223539
21. Manaf, S.A., Nordin, M.J.: Review on statistical approaches for automatic image annotation. In: 2009 International Conference on Electrical Engineering and Informatics. vol. 01, pp. 56–61 (2009). https://doi.org/10.1109/ICEEI.2009.5254815
22. Nicholas J. Car, Timo Homburg, Matthew Perry, John Herring, Frans Knibbe, Simon J.D. Cox, Joseph Abhayaratna, Mathias Bonduel: OGC GeoSPARQL - A Geographic Query Language for RDF Data. OGC Implementation Standard OGC 22-047, Open Geospatial Consortium (2023), http://www.opengis.net/doc/IS/geosparql/1.1
23. Noble, H., Mitchell, G.: What is grounded theory? Evidence-Based Nursing **19**(2), 34–35 (2016). https://doi.org/10.1136/eb-2016-102306, https://ebn.bmj.com/content/19/2/34
24. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. International Journal on Semantic Web and Information Systems (IJSWIS) **10**(2), 7–34 (2014)
25. Prokop, M., Pilař, L., Tichá, I.: Impact of think-aloud on eye-tracking: A comparison of concurrent and retrospective think-aloud for research on decision-making in the game environment. Sensors **20**(10), 2750 (2020)
26. Raimond, Y.: The event ontology, https://motools.sourceforge.net/event/event.html
27. Sartini, B., Baroncini, S., van Erp, M., Tomasi, F., Gangemi, A.: Icon: An ontology for comprehensive artistic interpretations. J. Comput. Cult. Herit. **16**(3) (aug 2023). https://doi.org/10.1145/3594724, https://doi.org/10.1145/3594724
28. Sheng, S., Moens, M.F.: Generating captions for images of ancient artworks. In: Proceedings of the 27th ACM International Conference on Multimedia. p. 2478–2486. MM '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3343031.3350972, https://doi.org/10.1145/3343031.3350972
29. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M.: The NeOn Methodology for Ontology Engineering, pp. 9–34. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24794-1_2, https://doi.org/10.1007/978-3-642-24794-1_2
30. Szekely, P., Knoblock, C.A., Yang, F., Zhu, X., Fink, E.E., Allen, R., Goodlander, G.: Connecting the smithsonian american art museum to the linked data cloud. In: Cimiano, P., Corcho, O., Presutti, V., Hollink, L., Rudolph, S. (eds.) The Semantic Web: Semantics and Big Data. pp. 593–607. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
31. Veenman, M.V., Elshout, J.J., Groen, M.G.: Thinking aloud: Does it affect regulatory processes in learning? Tijdschrift voor Onderwijsresearch (1993)
32. VRA Core: Vra core schemas, https://www.loc.gov/standards/vracore/schemas.html

33. Whitaker, A., Bracegirdle, A., De Menil, S., Gitlitz, M.A., Saltos, L.: Art, antiquities, and blockchain: new approaches to the restitution of cultural heritage. International Journal of Cultural Policy **27**(3), 312–329 (2021)
34. World Wide Web Consortium, https://dom.spec.whatwg.org/: Document Object Model (DOM) — Living Standard (2024), accessed: 2024-03-15
35. Zheng, B., Liu, F., Zhang, M., Zhou, T., Cui, S., Ye, Y., Guo, Y.: Image captioning for cultural artworks: a case study on ceramics. Multimedia Systems **29**(6), 3223–3243 (Dec 2023). https://doi.org/10.1007/s00530-023-01178-8, `https://doi.org/10.1007/s00530-023-01178-8`