Micro-ontology and Mapping to Schema.org

Domain: Public Art and Monuments

Group 2: Julie Brown, Danielle Green, Danielle Galván Gomez, Sara Hruska, and Melissa Runnels

Date: December 10, 2023

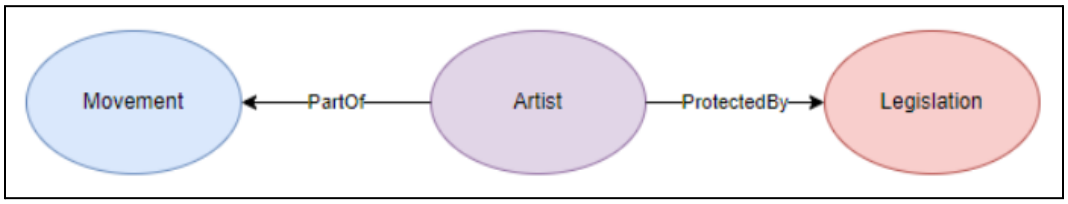

Figure 1: Knowledge graph illustrating team-determined labels and relationships in the Public Art and Monuments domain

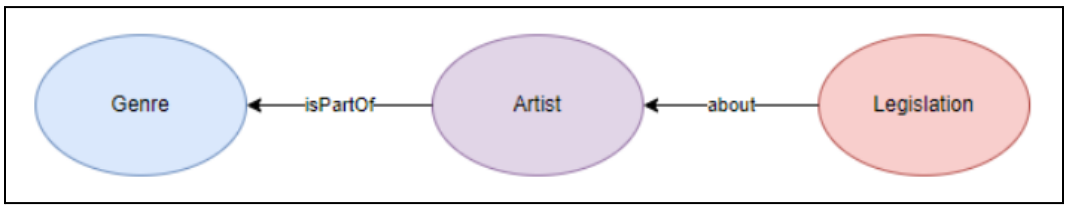

Figure 2: Knowledge graph illustrating schema.org labels and relationships derived from team-determined terms for the Public Art and Monuments domain represented in Figure 1

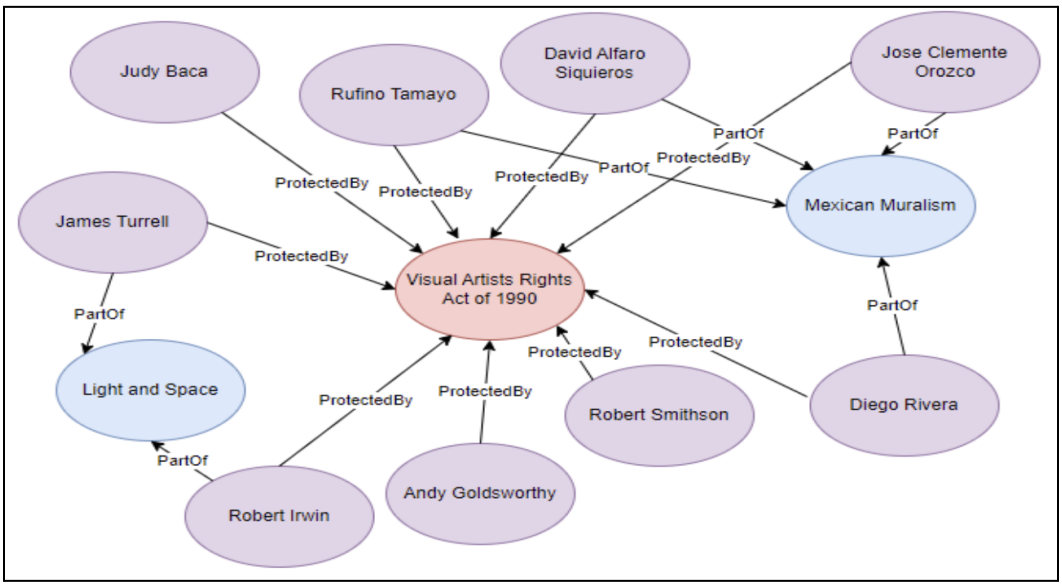

Figure 3: Micro-ontology using named entities from the Public Art and Monuments domain

Schema.org Analysis

Once our team aligned on our class names and relationships (see Figure 1), we navigated to schema.org to begin identifying matching terms. We started by focusing on class names, or the labels within circles, in Figure 1. We used a combination of searching and browsing techniques, and started to determine analogous class labels in the diagram from left to right. Beginning with “Movement” proved quite challenging. Scanning through search results quickly showed the available terms had different meanings than the Art Movements we were looking to represent. On the other hand, a simple search, “Artist,” immediately returned a result with a description that matched our thesaurus’ intention. Given the mixed results using the search technique, we moved on to exploring via site Docs and Schemas. Both entry points had links to the full hierarchy, and Schemas also included a list of commonly used “types” (or “classes”). In these views, as well as within the page for Artist, it was easier to identify the classes of “CreativeWork” and “VisualArtwork” and browse through terms more likely to be related to our domain. It was straightforward to identify “Legislation” and required more effort and reading Descriptions to determine “Genre” as an alternative for “Movement.” We then transitioned to identifying the relationships, or labels on arrows in Figure 1, again working left to right and using a mix of search and browse techniques. Determining “isPartOf” for “PartOf” was easy, while identifying “about” for “ProtectedBy” required deeper readings and group discussion of the Descriptions of Properties from Legislation and CreativeWork to align on the best fit. We were then able to update our knowledge graph with the schema.org labels, as shown in Figure 2. Reference Figure 3 for an example of the micro-ontology using named entities from our domain.

Our Challenges

Although we had sixteen named entities across five broader categories or classes in our thesaurus, we found it difficult to establish relationships between these entities across classes. We believe this was a challenge because a majority of our thesaurus construction focused on creating hierarchical and equivalence relationships within specific and carefully defined facets. So while much of our terminology was relevant and exhaustive in representing our domain, there were few common attributes shared between named entities within different faceted categories. While it was possible to create triples quite easily from our named entities (ex. Artist isPartOf Art Movement), it was difficult to then create a more complex knowledge graph that represented more complex semantic linkages. For example, we struggled to connect any named entities to Native American Graves Protection and Repatriation Act (NAGPRA), although this might have been resolved by including dates, Indigenous artists, or Indigenous artworks in our thesaurus. Even after we decided on our descriptions and attributes between classes, mapping our knowledge graphs to schema.org was a challenge because we could not find an exact match to some of our terminology. Part of this is that our thesaurus was made for human users and indexers, while schema.org is focused on labels in languages used to create Semantic Web ontologies for computers. For example, we wanted to show that visual artists are ProtectedBy the Visual Artist Rights Act of 1990 (legislation), but there was no predicate available that could communicate the nuanced term, so we went with “about” which is technically correct but dilutes the connotations of the relationship we were trying to represent. Ultimately, the challenge with mapping our knowledge graph and deciding on labels in schema.org was that our

thesaurus is for a human audience to understand how to navigate and use terms, whereas schema.org is for communicating concepts to a computer that will need to learn how to respond to terms when certain conditions are met.

What We Learned

Public Art and Monuments is a complicated domain for which there is not a dedicated controlled vocabulary. It became apparent as we built out our taxonomy and this micro-ontology that because of its complexity, the domain requires a large, sophisticated vocabulary in order to be both adequately specific and exhaustive. As discussed in the previous section, our thesaurus lacks enough terms to represent some relationships accurately in the ontology – and this illuminates unintentional bias and subsequent lacunae in our vocabulary. For instance, we noticed this gap in our development of the thesaurus when we found we could not create triples connected to NAGPRA. This shows that constructing an ontology and then plugging specific data into it is a useful tool for “stress-testing” for gaps and biases in a controlled vocabulary.

Our thesaurus, even with its gaps, is useful across a range of documents and is relatively easy to use in its current state, opinions shared by our third-party evaluators. However, the process of mapping our micro-ontology to schema.org showed that the standardized metadata for the web lacks sufficient terminology for our domain, even given the tininess of our ontology. This suggests that our thesaurus fills a need and could readily be improved and expanded to create a truly comprehensive thesaurus in its domain, which could also serve as a foundation to develop a better metadata vocabulary for the domain in schema.org.