Linked Data Final Project

Mapping Linked Art (linked.art) to Schema.org
Craig Parry and Jon Sykora
May 3rd, 2021

1 Introduction

In this project we attempt to identify and build a tool for the mapping of Linked Art (linked.art) to Schema.org. Schema.org is one of the largest and most comprehensive schemas for semantic web data annotation, provided by the collaboration of Google, Microsoft and Yandex. Linked Art is a model that is created by a community of museums and artists that is built to try to provide a comprehensive schema to describe content and artifacts of concern to the museum community. The Georgia O'Keefe museum follows the Linked Art model with slight modifications noted in their documentation. The purpose of their interest in semantic web data and linked data models is to connect their data to other museums across the internet. Connecting the Linked Art model will provide their data with deeper query opportunities and a broader reach of their data to other databases that use the Linked Art model for annotation. Further the goal of mapping to a larger schema, Schema.org, will extend the reach of their data to more models across the internet that either are mapped to Schema.org or use their model for their annotations. This project will attempt to define a mapping for a portion of the Linked Art model to Schema.org and build a tool to map JSON-LD annotated data, which can be used as a guide to map the rest of the Linked Art model.

The paper is organized as follows: Section 2 contains the project proposal that has been approved for use. Section 3 presents a summary of the information discovered while researching related topics for the project. Section 4 contains the results and deliverables for the project, laid out in subsections for clarity and to separate the results into categorical data. Section 5 concludes the paper with a discussion of the results, and a reflection on the role of the relational model in the context of the current status of the Semantic Web. Section 6 lists resources we used in creating our project.

2 The Project Proposal

Through the Georgia O'keeffe linked data presentation we learned more about the structure of museum linked data on the semantic web. They used a schema from the Linked Art project to create their RDF annotations and created their json-ld annotations from that schema. One of the ongoing goals described by the presentation is to extend the reach of Georgia O'keeffe museum data to other collections across the web. So that their data is open and accessible and can be connected to the semantic web for anyone's use and to make new connections to new knowledge bases.

Our idea to help extend the reach of their linked data is to map their data to a more general schema. The Linked Art project was created for museums and has some distinctions from Schema.org, which is another schema used for annotations. The Schema.org annotations are used by search engines such as Google's search engine optimizer and is used for the structured data in Google's knowledge graph. So if we

build or apply a mapping to extend their structured data to the Schema.org annotations, then the Okeeffe data will be more reachable by others through the semantic web.

To reach our goal we need to complete the following steps:
1. Research a current mapping of Linked Art to Schema.org
2. Describe and document a mapping of Linked Art to Schema.org (if necessary)
3. Build a demo to map a sample of semantically annotated data from Okeeffe Museum to Schema.org annotations
4. Extend the demo to all labels categorized by Linked Art that have a mapping in Schema.org

The timeline for this project would be to expect the research and mapping (steps 1 & 2) to take 1-2 weeks. Leaving 1-2 weeks to build a demo and extend that demo as far as possible (steps 3 & 4).

3 Summary of Findings

The mapping of our examples showed that the mapping will be achievable with manual labor to hand-match properties with semantically similar descriptions, consolidating types into the Schema.org hierarchy and decisions as to which properties can be ignored. To extend this mapping to the entirety of the Linked Art model would be difficult to automate efficiently, but if the groundwork of mapping properties is created a similar scripting method can be used to map every type to the Schema.org model. The types and properties themselves do have a high degree of correlation between the two schemas which is promising for the desired linking.

Some of the decisions made by the Linked Art community were made to try to add expressiveness that the art and museum community would like to add to their linked data. Such as the classification of artwork and making the distinction of a physical piece of artwork and the conceptual property of that piece of art as two separate objects. These details are lost in the translation to Schema.org, but some are redundant such as multiple instances of the same object. Schema.org uses very effective design and thorough semantics to allow for a very simple to understand implementation using a hierarchical design. On the other hand, the Linked Art version has made a verbose implementation that is cluttered by the use of relationship identifiers. This has the direct effect of Linked Art deciding to create their model without the strict use of theoretical semantics to define their schemas, but to define their schema by the description of that schema in a human readable form in their JSON-LD objects.

4 Detailed List and Description of Results and Deliverables

This section will describe the results of our research on the schemas provided by Linked Art and Schema.org, as well as describe the mapping tool built to demonstrate a

mapping between the two schemas. Subsection 4.1 describes the Linked Art schema. Subsection 4.2 describes the Schema.org schema. Subsection 4.3 explains the reasoning for the mapping. The description and examples of the mapping that the tool provides are explained in subsection 4.4. Subsection 4.5 discusses some key differences between the schema's JSON-LD implementations.

4.1 Description of Linked Art (linked.art)

Linked Art (linked.art) is a community project founded to design and implement a shared model based on Linked Open Data (LOD) with the intent to describe works of art and artifacts for the art community. The purpose of Linked Open Data is to structure web data to be more universally accessible, allowing it to be linked to related data across the web. They emphasize that the purpose of using LOD is to make their data usable, so they designed their framework towards being accessible for consumers, mainly developers, therefore they implemented their project based on what they call Linked Open Usable Data (LOUD) principles.

Their main consumer would be developers, therefore they focused their efforts on the usability of their API. The API for LOD uses http and is dependent on complete ontologies and semantics for theoretical correctness, which they deemed less than practical. As a result, they implemented their API in JSON-LD which allowed for some mappings of ontological constructs within the JSON object structure, without needing the exact precision that LOD and ontologies require. This was an attempt to make LOUD more usable by allowing developers more freedom, less constraint and using JSON (a popular framework for developers).

LOUD principles include the right abstraction for the audience, which allows developers to use the most useful abstraction or the level of abstraction that their needs require. Rather than using the purest form of the abstracted schema that would be necessary for a theoretical ontology. Few barriers to entry allows developers to quickly understand how to get started and use the model with their data. The model should be comprehensible by introspection. The developer can understand the data by looking at the JSON-LD and is not required to have knowledge of the complete ontology and vocabularies to understand the structure of the data. Documentation with working examples should be provided to give context to the developer, for pattern correctness checks and to address rules that cannot be intuited by the developer. There are few exceptions and many consistent patterns, so the number of rules the developer needs to learn to handle exceptions is reduced.

5

4.2 Description of Schema.org

Schema.org is a collaborative effort to support structured data on the internet, web pages, email messages and future mediums. Their vocabulary is implemented to support RDFa, Microdata and JSON-LD encodings. The vocabulary expresses entities, relationships between entities, actions and can be extended through the documentation of their model. The Schema.org model has been used to markup over 10 million site's web pages and email messages. Schema.org was founded by Google, Microsoft and Yandex and their vocabularies are developed by an open community process. Google, Microsoft and Yandex use the schema to help enrich and extend their applications including Google's Search Engine Optimizer. They make use of JSON-LD marked data and a knowledge graph created with that data to enhance their search engines. The shared vocabulary allows for the maximum benefit for development of the schema.

Their philosophy is to mark up as much content as possible so long as the content is not hidden from viewers. The vocabulary used by Schema.org allows for a very general description of an object. All objects could be classified as a "Thing" with the properties name, description, url, and image. The objects can be classified as a more specific type which inherits its parent type's properties. Their schema is arranged in a hierarchical manner to represent parent-child relationships and their extensions. Their vocabulary currently has 779 types and 1390 properties.

4.3 Description of the reason for mapping the two schemas

The Georgia O'keeffe Museum uses the Linked Art (linked.art) model for the implementation of their linked data with slight modifications. In order to have the museum's data reach as much external data as possible, it would need to be compatible with Schema.org. The Linked Art model does not require the same precision as Schema.org when describing linked data. The Schema.org ontologies and vocabularies aim for theoretical correctness when describing the semantics of data described by their schema. This creates a gap in the two schemas where properties and types described by Schema.org either do not match in naming conventions or are not described at all by the Linked Art schema and vice versa. Mapping Linked Art to Schema.org will allow for the data described by the Linked Art schema to be linked with data that is described by Schema.org, allowing it to take advantage of Google's search engine optimizer. This will extend the reach of the data described by Linked Art and allow the data to be included in queries involving types and properties that are described by the Schema.org schema, as well as use tools built with Schema.org vocabularies and ontologies. Mapping these schemas will allow for future users of Linked Art to extend their data to Schema.org. As well as allow for data that is currently mapped by Linked Art to extend the reach of their data without requiring Linked Art and its adopters to radically change their existing data.

4.4 Mapping Linked Art's VisualItem to Schema.org's VisualArtWork

        To demonstrate a mapping of the two schemas we have created a parser that will take a Linked Art encoded JSON-LD for the VisualItem or HumanMadeObject types, and map it to Schema.org's VisualArtwork type. The example mapping in table 1 is of a Linked Art JSON-LD object embedded in an html script representing a VisualItem with highlights to visualize how properties map to Schema.org encoding JSON-LD. The json format that they use subtypes items by defining their relationships to their parent types and related types by the use of relationship identities such as "classified_by", "identified_by", "referred_to_by", etc. They also make use of linking the getty vocabulary ids for each type as well.

Linked.art VisualItem JSON-LD example	Schema.org mapped JSON-LD example

<pre><script type="application/ld+json">
{
  "@context": "https://linked.art/ns/v1/linked-art.json",
  "id": "https://linked.art/example/VisualItem/0",
  "type": "VisualItem",
  "_label": "Appearance of Miro's The Farm",
  "classified_as": [
    {
      "id": "http://vocab.getty.edu/aat/300021495",
      "type": "Type",
      "_label": "Cubist",
      "classified_as": [
        {
          "id": "http://vocab.getty.edu/aat/300015646",
          "type": "Type",
          "_label": "Style"
        }
      ]
    }
  ],
  "identified_by": [
    {
      "type": "Name",
      "content": "Appearance of The Farm"
    }
  ],
  "referred_to_by": [
    {
      "type": "LinguisticObject",
      "classified_as": [
        {
          "id": "http://vocab.getty.edu/aat/300435416",
          "type": "Type",
          "_label": "Description",
          "classified_as": [
            {
              "id":
"http://vocab.getty.edu/aat/300418049",
              "type": "Type",
              "_label": "Brief Text"
            }
          ]
        }
      ],</pre>
	<pre><script type="application/ld+json">
	{
	  "@context": "https://schema.org",
	  "@type": "VisualArtwork",
	  "name": "The Farm",
	  "alternateName": "Appearance of the Farm",
	  "mainEntity": "Tree",
	  "dateCreated": "1922",
	  "contentLocation": "Mont-roig del Camp",
	  "url": "https://www.nga.gov/collection/art-object-page.69660.html"
	  "description": "A brilliant amalgamation of an intense, even primitive, realism with the formal vocabulary of cubism.",
	  "creator": [
	    {
	      "@type": "Person",
	      "name": "Juan Miro"
	    }
	  ]
	}</pre>

```json
      "content": "A brilliant amalgamation of an intense,
even primitive, realism with the formal vocabulary of
cubism."
    }
  ],
"represents_instance_of_type": [
    {
      "id": "http://vocab.getty.edu/aat/300132410",
      "type": "Type",
      "_label": "Tree"
    }
  ],
"shown_by": [
    {
      "id": "https://www.nga.gov/collection/art-object-
page.69660.html",
      "type": "HumanMadeObject",
      "_label": "The Farm"
    }
  ],
  "created_by": {
    "type": "Creation",
    "timespan": {
      "type": "TimeSpan",
      "_label": "1921-1922",
      "begin_of_the_begin": "1921-01-01T00:00:00Z",
      "end_of_the_end": "1922-12-31T23:59:59Z"
    },
    "carried_out_by": [
      {
        "id": "http://vocab.getty.edu/ulan/500014094",
        "type": "Person",
        "_label": "Juan Miro"
      }
    ]
  },
  "represents": [
    {
      "id": "http://vocab.getty.edu/tgn/7300934",
      "type": "Place",
      "_label": "Mont-roig del Camp"
    }
  ]
}
```
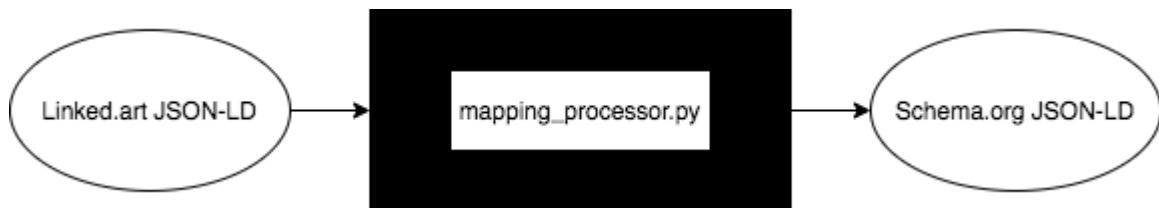
Table 1: Mapped VisualItem JSON-LD Example

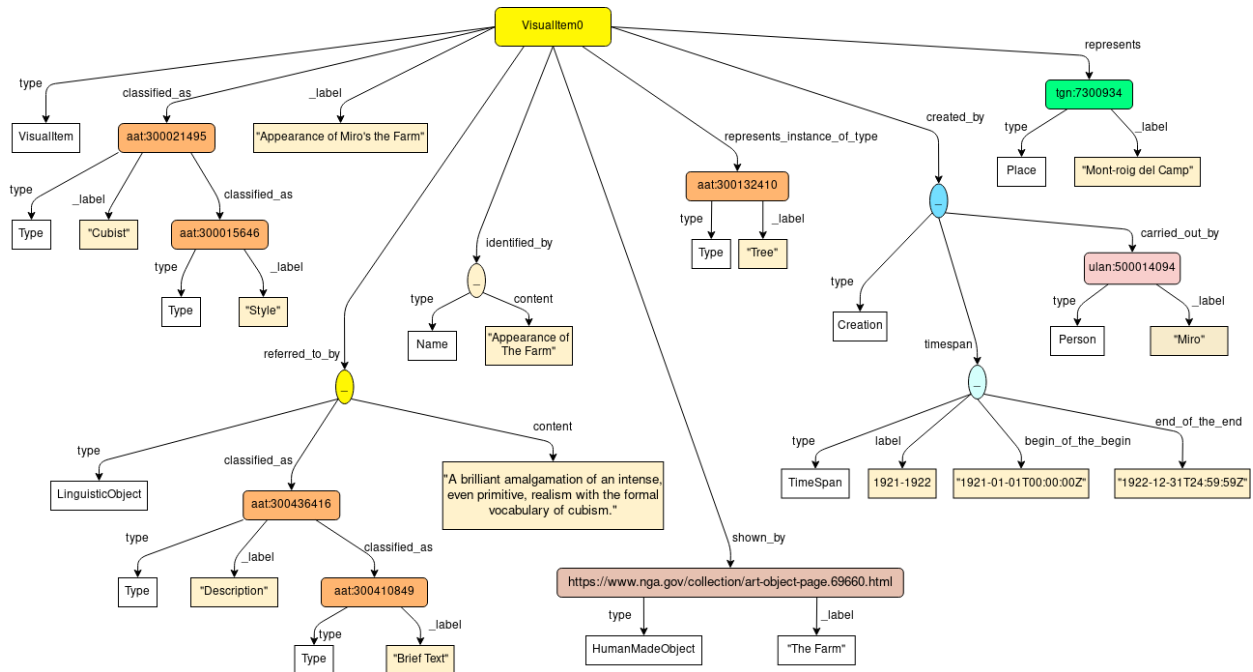Figure 1: Linked Art VisualItem schema diagram

Figure 2: Mapping blackbox diagram

Figure 1 gives a visual representation of the schema that represents the Linked Art VisualItem JSON-LD example provided in table 1. Mapping this example of Linked Art JSON-LD involves parsing through the JSON object to identify types that correspond to properties included in the Schema.org schema, filtering out identifiers not used by Schema.org and ignoring the cases where there is no appropriate match. The product of this mapping will result in a new JSON-LD object embedded in a html script that follows the Schema.org schema as seen in table 1 and the corresponding schema diagram in figure 3. Table 2 represents an example of a HumanMadeObject type from Linked Art mapped to a Schema.org VisualArtwork type through the same process.
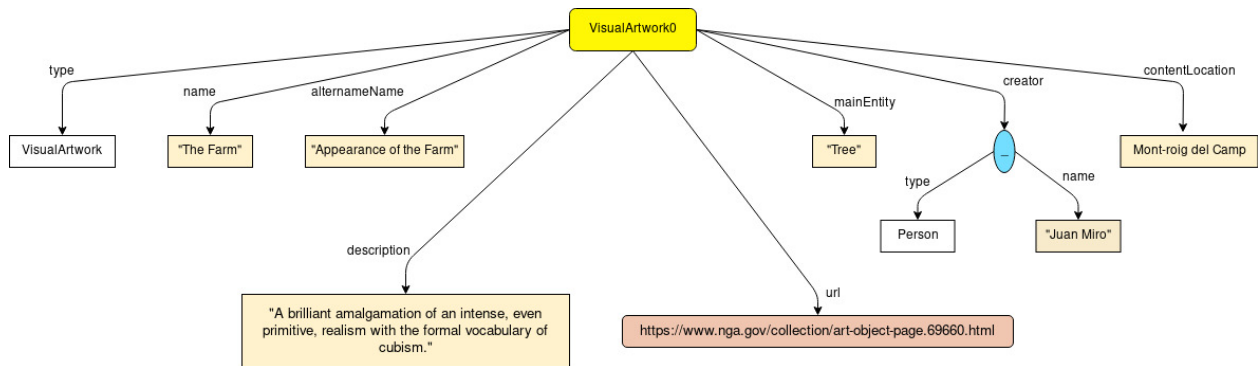

Figure 3: Schema.org VisualArtwork schema diagram

9

Linked Art HumanMadeObject JSON-LD example	Schema.org mapped JSON-LD example

<pre><script type="application/ld+json">
{
  "@context": "https://linked.art/ns/v1/linked-art.json",
  "id": "https://linked.art/example/object/0",
  "type": "HumanMadeObject",
  "_label": "The Farm",
  "classified_as": [
    {
      "id": "http://vocab.getty.edu/aat/300033618",
      "type": "Type",
      "_label": "Painting",
      "classified_as": [
        {
          "id": "http://vocab.getty.edu/aat/300435443",
          "type": "Type",
          "_label": "Type of Work"
        }
      ]
    },
    {
      "id": "http://vocab.getty.edu/aat/300133025",
      "type": "Type",
      "_label": "Artwork"
    }
  ],
  "identified_by": [
    {
      "type": "Name",
      "classified_as": [
        {
          "id": "http://vocab.getty.edu/aat/300404670",
          "type": "Type",
          "_label": "Primary Name"
        }
      ],
      "content": "The Farm",
      "language": [
        {
          "id": "http://vocab.getty.edu/aat/300388277",
          "type": "Language",
          "_label": "English"
        }
      ]
    },
    {
      "type": "Identifier",
      "classified_as": [
        {
          "id": "http://vocab.getty.edu/aat/300312355",
          "type": "Type",
          "_label": "Accession Number"
        }
      ],
      "content": "1987.18.1"
    }
  ],
  "referred_to_by": [
    {</pre>
	<pre><script type="application/ld+json">
	{
	  "@context": "https://schema.org",
	  "@type": "VisualArtwork",
	  "name": "The Farm",
	  "alternateName": "Appearance of the Farm",
	  "artForm": "Painting",
	  "dateCreated": "1922",
	  "locationCreated": "Mont-roig del Camp, Spain",
	  "sameAs": "https://www.wikidata.org/wiki/Q1192436",
	  "url": "https://www.nga.gov/collection/art-object-page.69660.html",
	  "description": "A brilliant amalgamation of an intense, even primitive, realism with the formal vocabulary of cubism.",
	  "creator": [
	    {
	      "@type": "Person",
	      "name": "Juan Miro"
	    }
	  ],
	  "height": "123.8 centimeters",
	  "width": "141.3 centimeters",
	  "artworkSurface": "Canvas",
	  "artMedium": "Oil Paint"
	}</pre>

```json
        "type": "LinguisticObject",
        "classified_as": [
          {
            "id":
"http://vocab.getty.edu/aat/300435416",
            "type": "Type",
            "_label": "Description",
            "classified_as": [
              {
                "id":
"http://vocab.getty.edu/aat/300418049",
                "type": "Type",
                "_label": "Brief Text"
              }
            ]
          }
        ],
        "content": "A brilliant amalgamation of an
intense, even primitive, realism with the formal
vocabulary of cubism."
      }
    ],
    "dimension": [
      {
        "type": "Dimension",
        "classified_as": [
          {
            "id":
"http://vocab.getty.edu/aat/300055644",
            "type": "Type",
            "_label": "Height"
          }
        ],
        "value":  123.8,
        "unit": {
          "id": "http://vocab.getty.edu/aat/300379098",
          "type": "MeasurementUnit",
          "_label": "centimeters"
        }
      },
      {
        "type": "Dimension",
        "classified_as": [
          {
            "id":
"http://vocab.getty.edu/aat/300055647",
            "type": "Type",
            "_label": "Width"
          }
        ],
        "value": 141.3,
        "unit": {
          "id": "http://vocab.getty.edu/aat/300379098",
          "type": "MeasurementUnit",
          "_label": "centimeters"
        }
      }
    ],
    "made_of": [
      {
        "id": "http://vocab.getty.edu/aat/300012363",
        "type": "Material",
        "_label": "Canvas"
      },
      {
        "id": "http://vocab.getty.edu/aat/300015050",
        "type": "Material",
        "_label": "Oil Paint"
```

```json
      }
    ],
    "shows": [
      {
        "id": "https://www.nga.gov/collection/art-object-page.69660.html",
        "type": "VisualItem",
        "_label": "Appearance of the Farm"
      }
    ],
    "current_owner": [
      {
        "id": "https://www.nga.gov/",
        "type": "Group",
        "_label": "National Gallery of Art"
      }
    ],
    "equivalent": [
      {
        "id": "https://www.wikidata.org/wiki/Q1192436",
        "type": "HumanMadeObject",
        "_label": "The Farm"
      }
    ],
    "produced_by": {
      "type": "Production",
      "timespan": {
        "type": "TimeSpan",
        "identified_by": [
          {
            "type": "Name",
            "classified_as": [
              {
                "id":
"http://vocab.getty.edu/aat/300404669",
                "type": "Type",
                "_label": "Display Title"
              }
            ],
            "content": "1921-1922"
          }
        ],
        "begin_of_the_begin": "1921-01-01T00:00:00Z",
        "end_of_the_end": "1922-12-31T23:59:59Z"
      },
      "took_place_at": [
        {
          "id": "http://vocab.getty.edu/tgn/7000457",
          "type": "Place",
          "_label": "Mont-roig del Camp, Spain"
        }
      ],
      "carried_out_by": [
        {
          "id": "http://vocab.getty.edu/ulan/500010879",
          "type": "Person",
          "_label": "Juan Miro"
        }
      ]
    }
}
```

Table 2: Mapped HumanMadeObject JSON-LD Example

4.5 Description of JSON-LD and differences

There are some key differences that needed to be handled when mapping the Linked Art schema to the Schema.org schema. Schema.org does not contain a property for "Style", which is a property described in Linked Art's VisualItem. In the example in figure 1, the "Style" is classified as "Cubist", but was ignored for our translation to Schema.org. Linked.arts' VisualItem contained properties that seemed to describe the object, and did not contain facts about the object itself. Linked.art's HumanMadeObject contains properties that describe the materials used (i.e. oil paint, watercolor, wood), dimensions and other details related to the physical object that is the art. All of the properties in VisualItem and HumanMadeObject can be translated to equivalent properties in Schema.orgs VisualArtwork with the exception of "Style". The format of the Schema.org JSON-LD is also much simpler while containing most of the same information about a piece of art. Another observation to note is the degree of redundancy between Linked.art's VisualItem and HumanMadeObject. Which can be seen from Tables 1 and 2, Linked.art's HumanMadeObject contains the same information as VisualItem, using different properties, with the exception of the mainEntity and contentLocation properties.

The Georgia O'Keeffe museum's open data model implementation differs slightly from the Linked Art model. When using the tool to extend the mapping of Linked Art to Schema.org, to the Georgia O'Keeffe data, these differences must be accounted for by scripting or manually changing these properties where appropriate. The O'Keeffe model uses the type ManMadeObject instead of HumanMadeObject. The type-of-types pattern is not used by the O'keeffe model, but this can be ignored. The "label" property must be changed to "_label to match the Linked Art model. LinguisticObject and Identifier properties must be changed from "value" to "content". Human-readable timespan dates are expressed using the "label" property instead of the identifier (relationship) TimeSpan.

5 Conclusion and Discussion

Through our research to understand the Linked Art model, the Schema.org model and strategies used for mapping we were able to design and demonstrate a mapping for these models. Our mapping relies on understanding the semantics of the types and properties of each model to successfully translate any particular schema. Which is a major limiting factor to scaling the automation of the mapping. Naming conventions and semantic meaning of the language used is as necessary to understand whether two properties or types should be linked as similar objects. The Linked.art model is a reasonable size to map further and make fully compatible with Schema.org, but this would still require every type to be hand mapped and a similar script to be

written for every type that needs translation. This mapping strategy could also be used for similar data models, but also with the same limitation as for any nuances expressed in the model may need exceptions to handle them. Attempting to bridge this gap between Linked.art and Schema.org really emphasizes the need for a standard to be followed by the semantic web. The more models and standards that are introduced, the more work is created for community projects to extend their work to other models.

When working with semantic web data it makes sense for there to be a unified structure or standard, so that there is as much compatibility as possible for linking data across the web. The more deviants that there are from any standard, the harder it becomes to link data based on lexicographical and context semantics. Differing expressions of relational structure can cause issues as we start building models that differ, but describe similar objects. A key difference between the models we used and the relational model is the use of foreign and primary keys. Neither Schema.org or Linked.art used primary keys in their models as they are able to leverage item identifiers used by vocabularies and type matching conventions. The relational model also relies on encapsulating all constraints and relationships within its miniworld and describing the whole Internet of Things is a massive undertaking. So semantic web models that concisely describe attributes and relations seem to be more effective than a model that is overly verbose and nuanced.

One of the main examples of the pitfalls of the relational model and the semantic web that we encountered was Linked Art's use of identifiers to describe relationships between their types and properties, which emulates the relational model. The intent of their model was to describe the data in a way that the users (developers) do not need to understand RDF graphs, vocabularies or theoretical semantics to use the data. By using these relationship identifiers they could create their model, define relationships and types all within their JSON-LD object. In contrast to the model provided by Linked Art, the Schema.org model uses a hierarchical model where every type inherits properties from their parent type and "Thing" is the parent of every type. The properties themselves can represent relationships to other objects that fit the type of that property. From these type descriptions RDF, microdata and JSON-LD formats can be annotated from the properties described by their type. There is a large amount of overlap with the properties of similar types due to the inheritance of their parent types' properties.

The Linked Art model's use of relationship identifiers complicates the model by providing an extra layer of complexity by naming conventions to relationships and types-of-types relationships. Those complexities are addressed concisely by the Schema.org model by using inheritance and expected type constraints for properties of types. Their use of properties gives descriptive names to their properties and the type property can relate objects by taking advantage of the type constraint pointing to the expected object. Their JSON-LD format additionally introduces another hurdle by requiring parsing knowledge of relationship identifiers pointing to types with labels. Schema.org's

simplicity of naming conventions and typing are more effective. Although Linked.art's model and its implementation may be effective for expressing the nuances of their minworld, we believe that Schemas.org sets the standard for a concise implementation that expresses their schema and relations.

The next step for this project would be to extend this strategy for implementing a mapping to the rest of the Linked art types so that there is a complete mapping for the schema to Schema.org. As well as extend the script to handle any type that is available in the Linked art model. Without a complete mapping, only portions of the two schemas will be compatible to take advantage of Schema.org's linking capabilities.

6. Resources

1. Tonon, A., Felder, V., Difallah, D. E., & Cudré-Mauroux, P. (2016, October). Voldemortkg: Mapping schema. org and web entities to linked open data. In *International Semantic Web Conference* (pp. 220-228). Springer, Cham.

2. Linked Art. (n.d.). https://linked.art/.

3. *Welcome to Schema.org*. Schema.org - Schema.org. (n.d.). https://schema.org/.

4. Linked Data Review + Documentation. (n.d.). http://gokm-docs.okeeffemuseum.org/entity/Art.