

DOREMUS to Schema.org: Mapping a Complex Vocabulary to a Simpler One

Pasquale Lisena¹ and Raphaël Troncy¹

EURECOM, Sophia Antipolis, France
{pasquale.lisena|raphael.troncy}@eurecom.fr

Abstract. Librarians and music professionals often use complex models and ontologies such as FRBRoo to represent music metadata. As a consequence, this metadata is not easily consumable by general search engines or external web applications. This paper presents a methodology, composed of a set of recipes, for mapping a complex ontology to a simpler model, namely Schema.org.

1 Music Information and Structured Data

Search engines and web applications display more and more knowledge panels alongside raw search results. This is made possible thanks to the adoption of some form of *Structured Data markup*, like Schema.org [4] that allows to describe the content of a web page in a machine-understandable way. In particular, Schema.org offers classes such as `CreativeWork` or `Event` and music-specific subclasses such as `MusicComposition` or `MusicEvent`.

Musical works are complex objects that require complex ontologies for expressing the richness of the information. FRBRoo is an ontology for describing bibliographic information [2]. The central feature of this ontology is the presence of the triplet Work-Event-Expression which considers that any artistic Work, only exists through an Event of creation, that realizes the Work itself into an Expression. The DOREMUS ontology¹ [1] extends FRBRoo with music-specific classes and properties like the key, the genre or the casting that could interpret a work. Because of this complexity, the consumption of these data by search engines and external applications is not easy. Nogales et al. proposed a mapping between the classes and properties of Schema.org and the ones of different vocabularies which have exactly the same name (or synonyms) [5]. Godby proposes to map each level of the chain Work - Expression - Manifestation - Item with a entity of the CreativeWork class [3]. A limit of this strategy is that the information is split up among different objects.

In this paper, we provide a methodology composed of four recipes for translating a complex ontology like DOREMUS into Schema.org. This method is based on the observation of the graph and it assumes sufficient knowledge of the source and target models. As example, we will represent Beethoven's *Sonata "Quasi una Fantasia"*, described in DOREMUS in Figure 1 using Schema.org².

¹ <http://www.doremus.org>

² In the following we use respectively the prefixes `mus:` and `sdo:`

Identify similar properties. For each class mapped, we must align their properties. The criteria are again: have similar names, have similar descriptions and have similar expected values. Each mapped property could have as value a literal (e.g. key and genre) or another class (e.g. the composer is a Person). In the latter case, if we have not previously mapped it, we consider this class as a new input for the first steps, until every node in the graph is reached.

Simplify the graph. Different DOREMUS classes can be mapped to the same Schema.org type: e.g. both Works and Expressions are mapped to `sdo:MusicComposition`. Merging these nodes can produce the advantage of a simpler model, in which the information is distributed in as less nodes as possible. Two nodes are good candidates for the merging when: 1. They have the class or a superclass in common; 2. The direct connection with a specific class is realized with the same property; 3. They are directly connected; 4. They have no conflicting properties. The `mus:F1_Work` and `mus:F2_Expression` classes aligned with `sdo:MusicComposition` satisfy all these criteria. The difference between Work and Expression as defined by FRBRoo has also been considered as tiny from the point of view of Schema.org [3].

Redundant nodes are substituted with a new node with: 1. The same class of the original one (or the most specific among the two); 2. The sum of their properties. The result of this phase is a new graph as depicted in Figure 2.

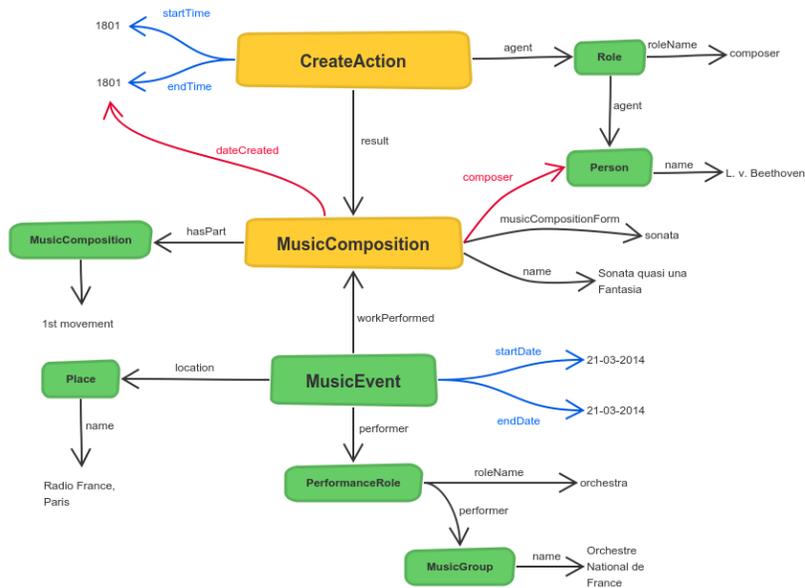


Fig. 2. Graph of Beethoven's *Sonata "Quasi una Fantasia"* in Schema.org.

3 Evaluation and Future Work

The evaluation of the goodness of the mapping is a long term goal, that can be realized only when data will be more easily consumed (e.g. by search engines). For having a quick feedback, we prepared a JSON-LD version of the *Sonata Quasi Una Fantasia* available at <https://goo.gl/zzXVWh>. Both the Structured Data Testing Tool by Google³ and the Structured Data Linter⁴ show the structure of the Schema.org graph like a tree, with 76 statements correctly recognized. The Structured Data Linter offers in addition a preview of the result as they could be shown in a SERP. For having a stronger visual feedback, we developed a lightweight web application, named *Schema.org Visualizer*⁵, which consumes the data in JSON-LD and shows a knowledge cards for the described content. The result for our example is available at <https://goo.gl/mCAzsw>.

We are planning to test our recipes on the ELI ontology⁶ which is also based on FRBRoo. The goal is to compare the result of the mapping to the one which has recently been hand-made designed for being considered as a Schema.org extension⁷. Future work includes an implementation strategy for this methodology, that will also highlight the content excluded from the mapping, in order to better evaluate the possibility of presenting a Schema.org extension.

Acknowledgments. This work has been partially supported by the French National Research Agency (ANR) within the DOREMUS Project, under grant number ANR-14-CE24-0020.

References

1. M. Achichi, R. Bailly, C. Cecconi, M. Destandau, K. Todorov, and R. Troncy. DOREMUS: Doing Reusable Musical Data. In *14th International Semantic Web Conference (ISWC)*, 2015.
2. M. Doerr, C. Bekiari, and P. LeBoeuf. FRBRoo: a conceptual model for performing arts. In *Annual Conference of CIDOC*, pages 6–18, 2008.
3. C. J. Godby. *The Relationship between BIBFRAME and OCLC's Linked-Data Model of Bibliographic Description: A Working Paper*. OCLC Research, 2013.
4. R. Guha, D. Brickley, and S. MacBeth. Schema.org: Evolution of Structured Data on the Web. *Communications of the ACM*, 13(9):44–51, 2016.
5. A. Nogales, M.-A. Sicilia, S. Sánchez-Alonso, and E. Garcia-Barriocanal. Linking from Schema.org microdata to the Web of Linked Data: An empirical assessment. *Computer Standards & Interfaces*, 45:90–99, 2016.

³ <https://search.google.com/structured-data/testing-tool>

⁴ <http://linter.structured-data.org/>

⁵ <https://github.com/pasqLisena/schema-visualizer>

⁶ <http://publications.europa.eu/mdr/eli/>

⁷ <https://github.com/schemaorg/schemaorg/issues/1156>