Ranging the Description Ontology Design Pattern with Literal Values

http://ontologydesignpatterns.org/wiki/Submissions:Description_in_Range

Silvio Peroni[1], Giorgia Lodi[2], and Aldo Gangemi[2]

[1] DASPLab, DISI, University of Bologna, Bologna, Italy
[2] STLab, ISTC, CNR, Rome, Italy
silvio.peroni@unibo.it, giorgia.lodi@istc.cnr.it, aldo.gangemi@istc.cnr.it

Abstract. This paper describes a specialization of the Description Ontology Design Pattern (ODP), i.e., the Description In Range ODP, that allows one to range the conceptualization of a descriptive context within specific borders defined by means of literal values. The specialization emerged within an e-government Linked Open Data (LOD) project named FOOD, where it was required to model in OWL descriptions of raw material and physical, chemical, organoleptic characteristics of European Union (EU) quality schemes for agricultural and food products – i.e., PDO (Protected Designation of Origin), PGI (Protected Geographical Indication) and TSG (Traditional Speciality Guaranteed).

Keywords: Linked Open Data, Ontology Design Patterns - Description Pattern - OWL

1 Introduction

PDO (Protected Designation of Origin), PGI (Protected Geographical Indication) and TSG (Traditional Speciality Guaranteed) are European Union (EU) quality schemes assigned to agricultural and food products that are produced, arranged and distributed to customers following strict and well defined rules. The rules are usually specified in policy documents (or product specifications), official published in EU Member States' Gazettes, and available in all the EU languages within the DOOR European system [1].

Usually available in PDF, policy documents contain unstructured data (mainly free text) that describes production places, raw materials and their specifications, physical, chemical, microbiological or organoleptic characteristics, production methods, logos details, related to food and agricultural products such as wine, cheese, oil, fish and so on. This data, if published as (linked) open data, can contribute to unlock a great value for the agriculture sector. In fact, their availability in an interoperable and machine readable form can facilitate:

- the standardization of the policy documents definition process (currently, in our country, the Ministry of Agriculture still publishes extremely heterogeneous policy documents, even for products of the same category);
- the development of applications and services for the agriculture domain, which may span from food frauds detection, production and distribution traceability of quality products, to the definition of wine and food tours for a country, to cite a few examples.

Within this context, we carried out a national project named FOOD (FOod in Open Data) [4]. FOOD aims at publishing the linked open data (LOD) of Italian quality products PDO, PGI and STG, compliant with OWL ontologies we developed for describing such domain. In the project, we adopted a pattern-based ontology engineering approach for the development of such OWL ontologies, and we extensively reused well-known Ontology Design Patterns (ODPs).

During the modelling phase, it also emerged the need of introducing a variant of the existing Description pattern [5]. This pattern turned out ideal for representing descriptive contexts for raw materials used in products' production process, and for expressing physical, chemical, microbiological or organoleptic characteristics of such products. However, its current form did not allow us to range the description with literal values. For instance, it was required us to express that the PDO wine "Amarone della Valpolicella" should be made by using grapes based on a specific ampelographic composition consisting of "Corvina and Cervone" grape from 45% to 95%, Rondinella grape from 5% to 30% and other grapes up to at most 25%. This kind of descriptions were not directly expressible by using the aforementioned pattern, and thus extensions had to be thought.

This paper introduces the *Description in Range* (*DIR*) ontology design pattern that specializes the Description ODP in order to express the aforementioned requirements, so as to range the conceptualization of a descriptive context within borders represented by literal values.

The rest of this paper is structured as follows. In Section 2 and Section 3 we introduce a general and detailed description of the Description in Range ODP. In Section 4 we discuss some use cases in order to show the applicability of the introduced ODP variant to real agricultural and food data. Finally, in Section 5 we conclude the paper.

2 General Description

The ontology design pattern *Description in Range* (*DIR*) we introduce in this paper extends the Description ODP [5] so as to enable the specification of additional (textual) bounds to constrain the concept(s) used in a description. The main idea behind this pattern is that an entity can be described in terms of some concept according to particular textual values. For instance, we can say that:

- the colour (the concept) of a particular wine (the entity being described) is ruby (our value that restricts the bound of the concept in consideration);

- the diameter (the concept) of a particular cheese (the entity being described) can vary from 24 to 36 centimetres (two distinct values that delimit the lower and upper bounds of the concept, respectively).

At the same time, we did not want to constrain too much the various bound values; thus, we preferred to leave them specifiable as free text literals (that may also include units, e.g. "26 cm").

2.1 Intent

DIR allows one to range the conceptualization of a descriptive context within specific borders defined by means of literal values.

2.2 Competency Questions

Since DIR extends the Description ODP, it also inherits all the competency questions that have been specified in that pattern. In addition, it also allows one to answer to additional competency questions introduced as follows:

1. What are the descriptions related to a particular entity?
2. Which entities are described in terms of a particular concept?
3. How have the concepts involved in the description of a certain entity been restricted within particular bounds?

2.3 About the use of literals for upper-/lower-bound values

The choice of using literals for defining the various bound values is not actually the best possible option from a pure ontological point of view. In fact, values and their related units could be described by a more complex object, say "Value-WithUnit", that explicitly refers to the literal value (e.g. "26") and the related unit (e.g. the individual for indicating centimetres), separately.

However, this choice was driven by the actual use-case we had to deal with, i.e. the automatic and manual extraction of significant bunches of text from the policy documents of Italian products we processed, as largely explained in [4]. While at a first reading the main parts of the documents had a pretty similar organisation, there was a rather high degree of heterogeneity among documents describing different kinds of products (e.g., wines vs. bread), as well as among documents describing products of the same type.

Thus, it was possible that bound values were:

- described clearly (e.g. "the diameter must be included between 24 and 32 centimetres");
- defined by means of multiple units (e.g. "the length must be more than 57 centimetres and less than 2 metres");
- described by means of words instead of arabic numbers (e.g. "fifty-five" instead of "55");
- introduced by using acronyms for units (e.g. "from 67cm to 1.23m");

– specified without any clear bound value (e.g. "the color must be intense red");
– any arbitrary combination of the aforementioned points.

In the light of all these variables, also considering the time required for obtaining the final result by the institutions involved in the FOOD project, the development of scripts for the automatic extraction of relevant text from such documents in an appropriate way, i.e. by disentangling all the ambiguities introduced above, was not manageable.

In addition, the data extraction performed by humans could be problematic if all these issues had to be handled properly, i.e. by disambiguating all the cases. The people involved in the extraction process were domain experts from the Ministry of Agriculture and from the Italian Digital Agency. However they were not necessarily aware of all these alternative ways of expressing such data in the processed documents at the time of the extraction, and might not agree on how they should be disentangled.

Finally, there was an additional barrier that prevented us from guiding human experts in describing such data properly, i.e. the interface used for harvesting them. While we proposed to develop a Web-based interface for allowing experts to insert such data within a well-prepared Web form, they eventually decided to use Excel (and even without any particular feature, e.g. the use of macros) due to their previous knowledge in using such application. A post-processing phase of all these Excel data – that refers to more than eight hundred policy documents, each describing several distinct products of the same quality scheme – was not possible in reasonable time, either.

Thus, taking into account this working context, the choice of simplifying the specification of values+units values in single literals was, actually, a winning strategy for guaranteeing the finalisation of the project. In fact, it allowed us to address the issue of extracting, mainly manually, a large amount of descriptions from all the Italian policy documents. The ontologies we developed were clearly driven by this choice, and the DIR pattern we extracted from them follows the same permissive approach.

3 Elements and OWL formalization

The Description in Range (DIR) ontology design pattern, summarised in Fig. 1, includes two main classes:

– class *Entity* – this class describes all the entities that can be described by a certain description (class *description:Description*);
– class *DescriptionInRange* – this class is subclass of *description:Description* and allows for ranging the involved concepts (through the property *description:usesConcept*) according to particular literal values.

DIR also specifies two object properties, i.e., *isDescribedBy* and its inverse *describes*, that allow one to link, respectively, an entity to a description describing it and vice versa.

Finally, the pattern also defines a data property, i.e., *isRangedByValue*, that allows one to specify a particular value associated with an individual of a *DescriptionInRange*. This property has four additional data properties for permitting the specification of lower and upper bounds, i.e.,:

– data property *hasInclusiveLowerBoundValue* – it allows one to specify any value with an inclusive lower bound for a certain amount of a concept as conceived in a particular description (e.g., at least 25cm);
– data property *hasNonInclusiveLowerBoundValue* – disjoint from the previous property, it allows one to specify any value with an non-inclusive lower bound for a certain amount of a concept as conceived in a particular description (e.g., greater than 25cm);
– data property *hasInclusiveUpperBoundValue* – it allows one to specify any value with an inclusive upper bound for a certain amount of a concept as conceived in a particular description (e.g., at most 25cm);
– data property *hasNonInclusiveUpperBoundValue* – disjoint from the previous property, it allows to specify any value with an non-inclusive upper bound for a certain amount of a concept as conceived in a particular description (e.g. less than 25cm).

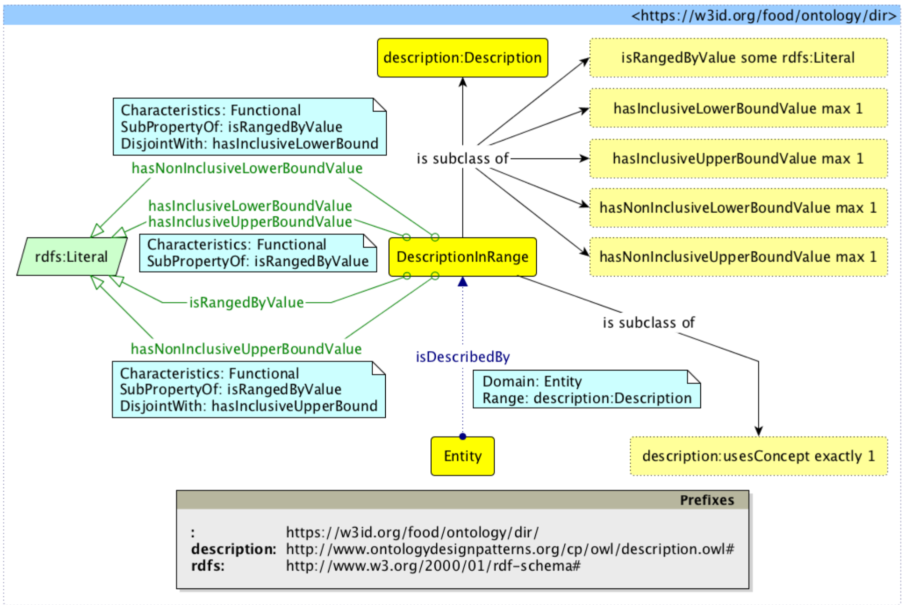

Fig. 1. A Graffoo diagram [2] introducing the Description In Range ontology design pattern.

3.1 OWL axiomatization

The Graffoo diagram introduced in Fig. 1 is represented in Manchester Syntax as follows. Firstly, the ontology describing the pattern and the two main classes are defined, i.e. *DescriptionInRange* and *Entity*. In particular, the former is defined in terms of several restrictions that also involve concepts introduced in the Description ODP [5]. These are shown as follows:

```
Ontology:
  <https://w3id.org/food/ontology/dir>
  <https://w3id.org/food/ontology/dir/1.0.0>
  Import:
    <http://www.ontologydesignpatterns.org/cp/owl/description.owl>

Class: dir:DescriptionInRange
  SubClassOf:
    description:Description,
    description:usesConcept exactly 1 description:Concept,
    dir:isRangedByValue some rdfs:Literal,
    dir:hasInclusiveLowerBoundValue max 1 rdfs:Literal,
    dir:hasInclusiveUpperBoundValue max 1 rdfs:Literal,
    dir:hasNonInclusiveLowerBoundValue max 1 rdfs:Literal,
    dir:hasNonInclusiveUpperBoundValue max 1 rdfs:Literal

Class: dir:Entity
```

The aforementioned two classes are linked by means of a particular object property (and its inverse), i.e. *isDescribedBy* (and *describes*), introduced in DIR. This property allows one to specify a description associated with a certain entity (and vice versa), as shown in the following excerpt:

```
ObjectProperty: dir:isDescribedBy
  Domain: dir:Entity
  Range: description:Description
  InverseOf: dir:describes

ObjectProperty: dir:describes
```

It is worth noting that the particular concept considered in any individual of the class *DescriptionInRange* is actually specified by means of a property defined in the Description ODP, i.e. *description:usesConcept*. Thus, the values related to the involved concept ,within a particular description, are specifiable by means of the data property *isRangedByValue*, defined as follows:

```
DataProperty: dir:isRangedByValue
  Domain: dir:DescriptionInRange
  Range: rdfs:Literal
```

Finally, the aforementioned data property is used as super-property by other four properties for specifying (inclusive and non-inclusive) lower and upper bound values related to a certain *DescriptionInRange*. They are defined as follows:

```
DataProperty: dir:hasInclusiveLowerBoundValue
  Characteristics: Functional
  SubPropertyOf: dir:isRangedByValue
  DisjointWith: dir:hasNonInclusiveLowerBoundValue

DataProperty: dir:hasInclusiveUpperBoundValue
  Characteristics: Functional
  SubPropertyOf: dir:isRangedByValue
```

```
  DisjointWith: dir:hasNonInclusiveUpperBoundValue

DataProperty: dir:hasNonInclusiveLowerBoundValue
  Characteristics: Functional
  SubPropertyOf: dir:isRangedByValue
  DisjointWith: dir:hasInclusiveLowerBoundValue

DataProperty: dir:hasNonInclusiveUpperBoundValue
  Characteristics: Functional
  SubPropertyOf: dir:isRangedByValue
  DisjointWith: dir:hasInclusiveUpperBoundValue
```

3.2 Alignment with DOLCE

The first set of ontology design patterns, originally introduced in the Ontology Design Patterns portal (`http://ontologydesignpatterns.org`), were actually a modularised version of DOLCE. Thus, since DIR extends one of these patterns, i.e. the Description ODP, we decided to provide an explicit alignment to DOLCE Ultra-Lite (available at `http://www.ontologydesignpatterns.org/ont/dul/DUL.owl`). In particular:

- the class *DescriptionInRange* is defined as subclass of *dul:Description*;
- the class *Entity* is defined equivalent to *dul:Entity*;
- the object property *isDescribedBy* is defined equivalent to *dul:isDescribedBy*, and a similar alignment is specified for its inverse, i.e., from *describes* to *dul:describes*;
- the data property *isRangedByValue* is defined as sub-property of *dul:hasDataValue*.

Due to the aforementioned alignments, the class *description:Description* can be inferred as equivalent to *dul:Description*.

3.3 Alternative modelling

Alternative proposals can be considered for modelling this kind of data that DIR aims at describing. One proposal would be to define values as complex objects (e.g. individuals of a class *ValueWithUnit*) composed by, for instance, a number (the data property *hasNumericValue*) and a unit of measure (the object property *hasUnit* linking to an individual of a class *UnitOfMeasure*). This type of modelling would allow us to also add additional constraints between lower- and upper-bound values, for instance in order to avoid that lower-bound values are greater than upper-bound values.

While existing ontologies have been already developed for this purpose, such as the *Quantities, Units, Dimensions and Data Types* ontology (available at `http://www.qudt.org`), as already mentioned in Section 2.3, this possibility was not practically feasible for the empirical and real-case scenario we had to deal with.

However, we believe it could be possible to study conversion strategies – e.g. by applying Natural Language Processing approaches to such upper-/lower-bound values expressed compliantly with DIR – that enable the translation of DIR-based data by means of more appropriate and structured ontologies.

Listing 1.1. Raw material description for “Amarone della Valpolicella”.

```
prod:wine-amarone-della-valpolicella-classic-red-classic
  upper:hasDescription
    draw:wine-amarone-della-valpolicella-classic-red-classic-rawmaterial-1 ,
    ... .

draw:wine-amarone-della-valpolicella-classic-red-classic-rawmaterial-1
  a upper:DescriptionOfRawMaterial;
  upper:hasRawMaterial raw:grape-corvina, raw:grape-corvinone ;
  upper:hasMinimumValue "45%" ;
  upper:hasMaximumValue "95%" .
```

4 Use cases

We present two use cases in which we applied the Description in Range ontology design pattern.

4.1 Raw materials

The first use case recalls the example reported in Section 1, and it regards the descriptive context for raw materials of quality scheme products. Thus, let us consider the real example of the Italian wine “Amarone della Valpolicella”. In order to obtain the registered quality scheme PDO, it has to comply with a set of rules including grape composition rules. In particular, it has to be made with a specific ampelographic composition that indicates the different types of grapes that contribute to form the wine, together with their percentage ranges. In the case of “Amarone della Valpolicella”, it must be composed (among the others) by Corvina and Corvinone grapes from 45% to 95%.

Listing 1.1 represents the Turtle excerpt, described by means of the ontologies we have developed in the context of the FOOD project, of part of the raw material description for “Amarone della Valpolicella”. The excerpt includes a description of raw material that specifies two grapes together, i.e., grape “Corvina” and grape “Corvinone”, with two boundaries of the percentage range within which the two grapes contribute to the production of “Amarone della Valpolicella” classic red.

As shown, the wine “Amarone della Valpolicella” (the DIR class *Entity*) has a specific description (represented using the property *upper:hasDescription* we defined in the policy documents ontology named upper [3], that is aligned with the DIR property *isDescribedBy*) that is an individual of the class *upper:DescriptionOfRawMaterial* (the DIR class *DescriptionInRange*). This class allows us to define the *upper:RawMaterial*, linked through the property *upper:hasRawMaterial* (the Description ODP property *description:usesConcept*), and its associated ranges expressed by means of the properties *upper:hasMinimumValue* (the DIR property *hasInclusiveLowerBoundValue*) and *upper:hasMaximimValue* (the DIR property *hasInclusiveUpperBoundValue*).

Listing 1.2. Characteristics description of "Parmigiano Reggiano" PDO when commercialised to consumers.

```
prod:cheese-parmigiano-reggiano-hard
  upper:hasDescription
    dchar:cheese-parmigiano-reggiano-hard-characteristic-3 ,
    dchar:cheese-parmigiano-reggiano-hard-characteristic-5 ,
    ...

dchar:cheese-parmigiano-reggiano-hard-characteristic-3
  a upper:DescriptionOfCharacteristic ;
  upper:hasCharacteristic char:diameter-plain-faces ;
  upper:hasMinimumValue "35 cm" ;
  upper:hasMaximumValue "45 cm" .

dchar:cheese-parmigiano-reggiano-hard-characteristic-5
  a upper:DescriptionOfCharacteristic ;
  upper:hasCharacteristic char:external-aspect ;
  upper:hasValue "natural straw-coloured rind" .
```

4.2 Characteristics

A second use case shows how the same design pattern has been adopted when describing physical, chemical, microbiological or organoleptic characteristics of the EU quality schemes products when they are commercialized. In this case, we consider the real example of the Italian cheese "Parmigiano Reggiano", which is registered as PDO. Its policy document indicates that, for consumers, it must present a set of characteristics among which the "diameter of the plain faces", which should be within the range "35-45 cm", and the "external aspect" that should be "natural straw-coloured rind".

Listing 1.2 shows the application of the Description in Range ODP when a description of the characteristics of the quality products are to be represented. The example is very similar to the one in Listing 1.1 but involves the classes *upper:DescriptionOfCharacteristics* (the DIR class *DescriptionInRange*), the property *upper:hasCharacteristic* (the Description ODP property *description:usesConcept*), and datatype properties *upper:hasMinimumValue*, *upper:hasMaximumValue* and *upper:hasValue* (the DIR property *isRangedByValue*), the latter one used for representing the bound of that characteristics by means of a single value.

5 Conclusions

In this paper we described a specialization of the Description ontology design pattern that emerged in the context of the definition of OWL ontologies for an Italian e-government project named FOOD. FOOD's main objective was to make available in LOD the data contained in policy documents of EU quality schemes for agricultural and food products (i.e., PDO, PGI, TSG).

The specialization we defined is named Description in Range and it can be used to range descriptive contexts with boundaries that are represented by literal values. In this paper, we presented two specific use cases where the specialization can be effectively used; namely, when modelling raw material composition of the EU quality schemes products, and when representing their physical, chemical, microbiological or organoleptic characteristics.

References

1. European Commission, The DOOR system, `http://goo.gl/VBSNQu`, 2016
2. Falco, R., Gangemi, A., Peroni, S., Vitali, F. (2014). Modelling OWL ontologies with Graffoo. In Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (Eds.), The Semantic Web: ESWC 2014 Satellite Events, Lecture Notes in Computer Science 8798: 320–325. Berlin, Germany: Springer. DOI: 10.1007/978-3-319-11955-7_42
3. FOOD project, Silvio Peroni, Giorgia Lodi, Aldo Gangemi, Luigi Asprino "Ontology of the policy documents", `http://w3id.org/food/ontology/disciplinare-upper`, 2015
4. Silvio Peroni, Giorgia Lodi, Luigi Asprino, Aldo Gangemi, "FOOD: FOod in Open Data", *to appear in the proceedings of the 15thInternational Conference on Semantic Web (resource track),* 17-21 October 2016, Kobe, Japan.
5. The Description Ontology Design Pattern, `http://ontologydesignpatterns.org/wiki/Submissions:Description`, 2016