Representation of Multi-Dimensional Quantities, Time-Series, and More*

Maxime Lefrançois

École Nationale Supérieure des Mines, FAYOL-ENSMSE, Laboratoire Hubert Curien, F-42023 Saint-Étienne, France

maxime.lefrancois@emse.fr

Abstract. There are so many domains that need ontologies of quantities, units of measure, and to represent multi-dimensional quantities. Currently, no formal RDF model has been proposed that enable to describe systems that are multiple quantified, and quantities having values that evolve on multi-dimensional domains. Quantities may be probabilistic distributions, points, or intervals.

Keywords: literals, datatypes, RDF, linked data

1 Introduction

This paper introduces the ontology of quantities that has been developed for the ITEA 2 SEAS project. With respect to other ontologies that describe quantities, this ontology shows several important contributions:

- it enables to represent quantities defined on algebraic rings: booleans, reals, or even complex quantities, which is of utmost importance for scientific domains.
- it uses representation of Units as custom datatypes that can be recognized on the fly by RDF processors.
- its semantics is well defined, and grounded on probability theory;
- it enables to represent time-series, or actually a generalization for series of any quantity;
- it enables to represent aggregation of quantities;
- it enables to represent growth rate, derivatives, sum, and jumps of quantities with respect to other quantities;
- the part of the ontology that represent specific dimensions and quantities needed for a domain (Smart Grids in the case of the SEAS project), is automatically generated by a JSON configuration file. This enables to ease the understanding and the maintenance of this ontology by domain experts;
- the web platform where it is published enables any project to post configuration files, and generates a new corresponding part of the ontology on the fly. This makes this ontology extensible for representing new units, or quantities needed for other domains.

Note 1. The SEAS project imposes that dimensions with the rings over the following sets be possible: $\{true, false\}$, \mathbb{N} , \mathbb{Z} , \mathbb{R} , \mathbb{C} .

^{*} This work has been supported by ITEA2 project SEAS 12004.

2 Preliminaries

2.1 Examples

We use the Turtle syntax in RDF snippets, with prologue:

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix lindt: <http://purl.org/NET/lindt#>.
@base <http://w3id.org/fs/concepts/> .
```

2.2 RDF abstract syntax

Let I, B, and L, be the disjoint sets of *IRIs*, *blank nodes* (later called *bnodes*), and *literals*, respectively. The set of *RDF terms* is $T = I \cup B \cup L$. Let V be the set of *variables*, disjoint from all the other sets. A *generalized triple pattern* (later simply called an *triple pattern*) is an element of $(T \cup V)^3$ and the set of all triple patterns is \mathcal{T} . A (generalized) *Basic Graph Pattern* (*BGP*) is a finite set of triple patterns.

A triple pattern (resp., BGP) that does not contain any variable is called a *generalized RDF triple* (resp., *generalized RDF graph*). An instance of a BGP G is an BGP G' where all the variables of G have been replaced by terms in \mathbf{T} . The set of instances of a BGP G is written inst(G). By extension, inst(x) is defined when x is a triple, or an element of $(\mathbf{T} \cup \mathbf{V})$. The set of all generalized RDF graphs is G.

An RDF triple (resp., RDF graph) that does not contain any bnode is called a *ground triple* (resp., *ground graph*). A grounding of an RDF graph G is an RDF graph G' where all the bnodes of G have been replaced by terms in $I \cup L$. The set of groundings of a graph G is written grnd(G). By extension, grnd(x) is defined when x is a triple.

2.3 RDF semantics

RDF 1.1 Semantics [?] defines several formal semantics for RDF graphs, each of which offers a different expressive power. The most generic of them is the formal semantics that defines *simple entailment*:

Definition 1 (Simple interpretation). *A* simple interpretation *is a tuple* $\langle \Delta_R, \Delta_P, \llbracket \cdot \rrbracket, \cdot^{\tau} \rangle$ *such that:*

- Δ_R is a nonempty set, called the domain of \mathcal{I} , and whose elements are called resources;
- Δ_P is a nonempty set whose elements are called properties;
- $\llbracket \cdot \rrbracket$ is a mapping from Δ_P to $2^{\Delta_R \times \Delta_R}$;
- $^{\mathcal{I}}$ is a partial mapping from $\mathbf{I} \cup \mathbf{L}$ to Δ_R such that the restriction to \mathbf{I} is total.

Definition 2 (Simple model of RDF graphs). A simple interpretation $\mathcal{I} = \langle \Delta_R, \Delta_P, \llbracket \cdot \rrbracket, \cdot^{\mathcal{I}} \rangle$ is a simple model of a ground triple $t = \langle s, p, o \rangle$, noted $\mathcal{I} \models_{\mathsf{simple}} t$, iff $\langle s^{\mathcal{I}}, o^{\mathcal{I}} \rangle \in \llbracket p^{\mathcal{I}} \rrbracket$. \mathcal{I} is a simple model of a ground graph G (also noted $\mathcal{I} \models_{\mathsf{simple}} G$) iff it is a simple model of all triples in G. Finally, \mathcal{I} is a simple model of an RDF graph H iff there exists a grounding H' of H such that $\mathcal{I} \models_{\mathsf{simple}} H'$ (again, this is noted $\mathcal{I} \models_{\mathsf{simple}} H'$).

Definition 3 (Simple entailment). An RDF graph G is said to simply entail another RDF graph H if all simple models of G are simple models of H.

Also we will introduce the *D*-entailment regime, and the *D*-RDFS-entailment regime. (TODO)

- Δ_C is a nonempty set whose elements are called *classes*; $[\![\cdot]\!]_C$, the *class extension*, is a mapping from Δ_C to 2^{Δ_R} .

3 **RDF-based Semantics**

Dimensions and units

Definition 4 (Dimensions). The set of dimensions is noted \mathcal{D} . Every dimension $D \in \mathcal{D}$ is associated with a set set(D) of values.

Example of dimensions include time, length, temperature.

The RDF representation of the set of dimensions is the class $\langle \text{Dimension} \rangle^{\mathcal{I}} \in \Delta_C$:

$$\llbracket \langle \text{Dimension} \rangle^{\mathcal{I}} \rrbracket_{\mathcal{C}} = \mathcal{D} \tag{1}$$

Property <unit>^I links a dimension to a datatype, whose value space is a subset of set(D). The RDF Graph below represents that any of the datatypes <joule> $^{\mathcal{I}}$, <wattHour>^I, or <electronVolt>^I may be used to describe a value of the energy dimen-

```
<EnergyDimension> a <Dimension> ;
 <unit> <joule>, <wattHour>, <electronVolt> .
```

When appropriate, datatypes should be one of the XML Schema datatypes. Else, it should be a LINDT¹ custom datatype, which may be recognized on the fly by RDF processors and SPARQL engines. A LINDT datatype is a datatype at the URL of which one may retrieve a RDF Document that describes it, and a Javascript script that can be used to process literals with this datatype. Example of resources in the LINDT vocabulary include:

- class lindt:TotallyOrdered specifies that the Javascript script implements function compare, to compare two literals with this datatype;
- class lindt: Group^T, specifies that the Javascript script implements functions add and substract to add and substract two literals with this datatype, and additiveIdentity to retrieve the additive identity element;
- class lindt:Ring^I, specifies that the Javascript script also implements functions multiply and divide to multiply and divide two literals with this datatype, and multiplicativeIdentity to retrieve the multiplicative identity element.

We define important subclasses of $<Dimension>^{\mathcal{I}}$:

¹ LINDT - Linked Datatypes - http://purl.org/NET/lindt - work in progress.

- <GroupDimension>^I is the class of dimensions that are also groups, i.e., unit datatypes of the dimension should be group, and coherent (todo: define coherence). Property <additiveIdentityValue>^I links a group dimension to an element that must be the additive identity element of any of its unit datatypes. For some group dimensions, the datatype to describe values is different from the datatype to describe the delta between these values. In such cases, we use property <deltaUnit>^I for the latter. The canonical example is dimension <TimeDimension>^I, that uses xsd:dateTime for values and xsd:duration for deltas:

```
<TimeDimension> <unit> xsd:dateTime ; <deltaUnit> xsd:duration .
```

Also, property <additiveIdentityValue> $^{\mathcal{I}}$ is canonically an instance of the <deltaUnit> $^{\mathcal{I}}$.

- <TotallyOrderedGroupDimension>^I is the class of dimensions that are also totally ordered groups, i.e., unit datatypes of the dimension should be totally ordered, and coherent (todo: define coherence). Property <nullvalue>^I links a totally ordered group dimension to an element that must be the null element of any of its unit datatype. For example, even though <TemperatureDimension>^I may be defined as being an instance of <TotallyOrderedGroupDimension>^I, it does not have a null value as this value is different for <degreeCelsius>, <degreeKelvin>, and <degreeFarenheit>.

We also anticipate cross dimension operations. Property <derivativeOfProperty> $^{\mathcal{I}} \in \Delta_P$ links a dimension to a property. The semantics of RDF Graph below is that dimension <PowerDimension> $^{\mathcal{I}}$ is the derivative of <EnergyDimension> $^{\mathcal{I}}$ with respect to <TimeDimension> $^{\mathcal{I}}$.

```
<TimeDimension> <derivativeOfProperty> <timeDerivativeOf> . <PowerDimension> <timeDerivativeOf> <EnergyDimension> .
```

Conversely, we define property <primitiveOfProperty> $^{\mathcal{I}} \in \Delta_P$, that links a dimension to a property. The semantics of RDF Graph below is that dimension <EnergyDimension> $^{\mathcal{I}}$ is the primitive of <PowerDimension> $^{\mathcal{I}}$ with respect to <TimeDimension> $^{\mathcal{I}}$.

```
<TimeDimension> <primitiveOfProperty> <timePrimitiveOf> .<br/><EnergyDimension> <timePrimitiveOf> <PowerDimension> .
```

We extend the LINDT vocabulary to represent this at the level of custom datatypes:

```
<watt> lindt:multiplyArgType xsd:duration ; lindt:multiplyReturnType <wattHour> .
```

This RDF Graph suggests that the Javascript object retrievable from Javascript script at URL <watt> implements the **multiply** function, which takes two lexical forms as first and second parameters, accepts xsd:duration as third parameter (i.e., the datatype of the second lexical form), and accepts <watthour> as fourth parameter (i.e., the datatype of the returned lexical form). Similarly, RDF Graph:

```
<wattHour> lindt:divideArgType xsd:duration ; lindt:divideReturnType <watt> .
```

suggests that the Javascript object retrievable from Javascript script at URL <watthour>implements the **divideBy** function, which takes two lexical forms as first and second parameters, accepts xsd:duration as third parameter (i.e., the datatype of the second lexical form), and accepts <watt> as fourth parameter (i.e., the datatype of the returned lexical form).

3.2 Distributions

Definition 5 (**Distributions on a dimension**). The set of distributions on a dimension $D \in \mathcal{D}$, noted $\operatorname{distr}(D) = 2^{\operatorname{set}(D)}$, is the set of subsets of $\operatorname{set}(D)$.

Let be a dimension $?\dim^{\mathcal{I}} = D \in \mathcal{D}$. Then the following RDF graph defines that class $?dDistributionClass^{\mathcal{I}} \in \Delta_C$ is the class of distributions over D. Then for any distribution $?d^{\mathcal{I}} = \delta \in \operatorname{distr}(D)$ of $?dDistributionClass^{\mathcal{I}} \in \Delta_C$, property $<\dim^{\mathcal{I}} \in \Delta_P$ links $?d^{\mathcal{I}}$ to $?\dim^{\mathcal{I}}$.

```
?dim <distributionsClass> ?dDistributionClass .
?d a ?dDistributionClass ; <dimension> ?dim .
```

Let be a value on dimension D, $v^{\mathcal{I}} = v \in \operatorname{set}(D)$. Then {?d <contains> ?v} (resp. {?d <excludes> ?v}) represents $v \in \delta$ (resp. $v \notin \delta$). For conciseness, we note:

$$\{?d < contains > ?v\} \Leftrightarrow v \in \delta$$
 (2)

$$\{?d < excludes > ?v\} \Leftrightarrow v \notin \delta$$
 (3)

One could fully characterize distribution δ with an infinite number of triples. Now we introduce a set of classes and properties that enable to describe the most classical distributions in a limited number of triples, and ease the writing of RDF snippets by domain experts in the most simple use cases. These resources were chosen during requirement capturing sessions organized in December 2015 for the SEAS project².

$$\{?d \ a \ \langle FullDistribution \rangle\} \ \Leftrightarrow \delta = set(D) \tag{5}$$

$$\{?d < value > ?v\} \Leftrightarrow \delta = \{v\}$$
 (6)

{?d1 :d2}
$$\Leftrightarrow \delta_1 \subseteq \delta_2$$
 (7)

{?dl :d2}
$$\Leftrightarrow \delta_1 = \delta_2$$
 (8)

?d
$$\Leftrightarrow \delta = \delta_1 \cap \ldots \cap \delta_n$$
 (9)

?d (?d1 ... ?dn)
$$\Leftrightarrow \delta = \delta_1 \cup ... \cup \delta_n$$
 (10)

For dimension D that are totally ordered by \leq , let functions \min represent the potential greatest lower bound, \max represent the potential least upper bound, and \max represent the delta between the \max and the \min , if both exist. These functions are coupled: $\forall \delta \in \operatorname{distr}(D), \min(\delta) + \operatorname{rng}(\delta) = \max(\delta)$.

Let
$$?d1^{\mathcal{I}} = \delta_1 \in \operatorname{distr}(D), \dots ?dn^{\mathcal{I}} = \delta_n \in \operatorname{distr}(D)$$
:

² SEAS Requirement capturing Workshop - http://data.the-smart-energy.com/ workshop/2015/12/

{?d ?v}
$$\Leftrightarrow (\forall x)[x < v \Rightarrow x \notin \delta]$$
 (11)

{?d ?v}
$$\Leftrightarrow (\forall x)[x > v \Rightarrow x \notin \delta]$$
 (12)

$$\{?d < \min(\delta) = v \} \Leftrightarrow \min(\delta) = v$$
 (13)

$$\{?d < maximum > ?v\} \Leftrightarrow max(\delta) = v \tag{14}$$

$$\{?d < range > ?v\} \Leftrightarrow rng(\delta) = v \tag{15}$$

$$\{?d \ a \ \text{minimumIncluded} \ ?v\} \Leftrightarrow ?d \ \text{minimum} \ ?v; \ \text{includes} \ ?v$$
 (17)

$$\{?d \ a \ minimumExcluded>\ ?v\} \Leftrightarrow ?d \ minimum>\ ?v; \ excludes>\ ?v$$
 (18)

$$\{?d \ a \ \text{maximumIncluded} > ?v\} \Leftrightarrow ?d \ \text{maximum} > ?v; \ \text{sincludes} > ?v.$$
 (19)

$$\{?d \ a \ \text{maximumExcluded} > ?v\} \Leftrightarrow ?d \ \text{maximum} > ?v; \ \text{excludes} > ?v$$
 (20)

3.3 Facets, F-points, and F-spaces

Definition 6 (Facets). The set of facets is noted \mathcal{F} . Any facet $f \in \mathcal{F}$ is assigned a dimension by mapping dim.

The set of facets is represented in RDF by class <Facet $>^{\mathcal{I}} \in \Delta_C$, and the mapping dim is represented by property <facetDimension $>^{\mathcal{I}} \in \Delta_P$:

$$\llbracket \langle \text{Facet} \rangle^{\mathcal{I}} \rrbracket_{\mathcal{C}} = \mathcal{F} \tag{21}$$

$$\llbracket \langle \text{facetDimension} \rangle^{\mathcal{I}} \rrbracket = \dim$$
 (22)

The following RDF Graph illustrates the difference between dimensions and facets:

```
<PowerDimension> a <Dimension> .
<ConsumptionPowerFacet> a <Facet> ; <facetDimension> <PowerDimension> .
<StoringPowerFacet> a <Facet> ; <facetDimension> <PowerDimension> .
<ProductionPowerFacet> a <Facet> ; <facetDimension> <PowerDimension> .
```

Definition 7 (Faceted Point (F-point), and its Facets). The set of Faceted Points, or F-points, Ω , is equal to the Cartesian product of the dimension sets of each facet in \mathcal{F} , augmented with a special "nothing" element \bullet :

$$\Omega = \underset{f \in \mathcal{F}}{\times} \left(\operatorname{set}(\dim(f)) \cup \{\bullet\} \right)$$

Element f of a F-point $v \in \Omega$ is named facet f, and is noted: $\pi_f v$. If $\pi_f v = \bullet$, then there is conceptually no facet f of v.

Element • helps to cope with the RDF open world assumption. Without further knowledge, any F-point has on any facet either a value of $\operatorname{set}(\dim(f))$, or no value at all. For instance, some MD-Value may have facet ${\operatorname{ConsumptionPowerFacet}}^{\mathcal{I}}$ "3.4"^^<watt>^ ${\mathcal{I}}$, and facet ${\operatorname{CimeFacet}}^{\mathcal{I}}$ "2016-02-25T12:00:00Z"^^xsd:dateTime ${\mathcal{I}}$, but no facet ${\operatorname{CwaterDepthFacet}}^{\mathcal{I}}$. We do not introduce direct RDF representation of F-points.

Definition 8 (Faceted Space (F-spaces)). The set of Faceted Spaces, or F-spaces, $\Sigma = 2^{\Omega}$, is the set of subsets of Ω . A F-space $\tau \in \Sigma$ is characterized by the set of F-points it contains.

The set of F-spaces is represented by RDF class <FacetedSpace $>^{\mathcal{I}} \in \Delta_C$:

Then we introduce the following properties, defined for any set of F-spaces $2t^{\mathcal{I}} = \tau$, $2t1^{\mathcal{I}} = \tau_1, \dots 2tn^{\mathcal{I}} = \tau_n \in \Sigma$:

$$\{?t1 < equivalentSpace > ?t2\} \Leftrightarrow \tau_1 = \tau_2$$
 (25)

{?t (?t1 ... ?tn)}
$$\Leftrightarrow \tau = \tau_1 \cap \ldots \cap \tau_n$$
 (26)

Definition 9 (**Projection of a F-space on a facet**). The projection of a F-space $\tau \in \Sigma$ on facet $f \in \mathcal{F}$ is noted $\pi_f \tau$:

$$\pi_f \tau = \{ \pi_f \mathbf{v} | \mathbf{v} \in \tau \} \setminus \bullet \in \operatorname{distr}(\dim(f))$$
 (28)

For any facet $?f^{\mathcal{I}} = f \in \mathcal{F}$, triple pattern $\{?f < projection Property > ?f Projection \}$ defines $?f Projection^{\mathcal{I}}$ as the property that links a F-space $?t^{\mathcal{I}} = \tau \in \Sigma$ to its projection $?d^{\mathcal{I}} = \delta \in \operatorname{distr}(\dim(f))$ on facet f:

{?t ?fProjection ?d}
$$\Leftrightarrow \pi_f \tau = \delta$$
 (29)

For sake of conciseness, we also enable the definition of other properties and classes as shortcuts to describe the projections on facet f directly at the level of the F-space:

Then for any $2v^{\mathcal{I}} = v \in \operatorname{set}(\dim(f))$ and $2t^{\mathcal{I}} = \tau$, $2t^{\mathcal{I}} = \tau_1$, $2t^{\mathcal{I}} = \tau_2 \in \Sigma$:

3.4 MD-Space series

Definition 10 (Restriction of a F-space with respect to a distribution along a facet).

The restriction of a F-space τ with respect to a distribution $\delta \in \operatorname{distr}(\dim(f))$ along facet f is noted $\sigma_f^{\delta}\tau$, and is defined as follows:

$$\sigma_f^{\delta} \tau = \{ \boldsymbol{v} | \boldsymbol{v} \in \tau \text{ and } \pi_f \boldsymbol{v} \in \delta \}$$
 (43)

Let be two F-spaces $?t1^{\mathcal{I}} = \tau_1, ?t2^{\mathcal{I}} = \tau_2 \in \Sigma$, a facet $?f^{\mathcal{I}} = f \in \mathcal{F}$, and a distribution on the dimension of that facet $?d^{\mathcal{I}} = \delta \in \operatorname{distr}(\dim(f))$. Then,

Multiple restrictions can be associated with a $Restriction^{\mathcal{I}}$. This is slicing. We define series of F-spaces, that enable to decompose the description of a F-space in a series of its restrictions along a common facet.

Let be a facet $?f^{\mathcal{I}} = f \in \mathcal{F}$, and a F-space $?t^{\mathcal{I}} = \tau \in \Sigma$. Then the following RDF graph defines:

- property ?fSeries $^{\mathcal{I}} \in \Delta_P$ links F-spaces to series of their restrictions along facet f;
- $?ts^{\mathcal{I}} \in \Delta_P$ is a series of restrictions of F-space τ along facet f;
- F-spaces $\tau_1, \ldots \tau_n \in \Sigma$ are the individual F-spaces in the series.

```
?f <seriesProperty> ?fSeries .
?t ?fSeries ?ts .
?ts <series> ( ?t1 ... ?tn ) .
```

Every τ_k is the restriction of F-space τ with respect to $\pi_f \tau_k$ along facet f:

$$\tau_k = \sigma_f^{\pi_f \tau_k} \tau \tag{45}$$

(60)

The description of individual projections $\pi_f \tau_k$ may be constrained at the level of the series using the following properties and classes:

{?ts a <EndExcludedSeries>} $\Leftrightarrow (\forall 1 \leq k \leq n) [\max(\pi_f \tau_k) \notin \pi_f \tau_k]$