Exploring a Semantic Framework for Integrating DPM, XBRL and SDMX Data

Roberto García[1] and Rosa Gil[1]

[1] Universitat de Lleida, 25003 Lleida, Spain
rgarcia@diei.udl.cat, rgil@diei.udl.cat

Abstract. This work is a preliminary proposal for the integration of financial data using semantic technologies. The input data formats under consideration are XBRL, including the Data Point Model (DPM), and the Statistical Data and Metadata eXchange (SDMX). The proposal is to map this data to a common framework based on semantic technologies. Besides, the transformation should also take into account the different mechanisms to structure these data formats, like Data Structure Definitions (DSD) for SDMX, XBRL Taxonomies or Data Dictionaries in the case of DPM.

Keywords: Financial Data, XBRL, DPM, SDMX, RDF, Ontology, Semantic Web.

1 Introduction

Together with the proliferation financial data and formats used to publish it, there is an increased need for ways to integrate it. A promising candidate to facilitate this integration, because it moves the effort one level up to the level of intended meanings instead of trying to deal with syntax subtleties, is semantic technologies. For the integration, we consider as sources data based on XBRL, including the Data Point Model (DPM), and SDMX. In addition to mapping them to a semantic space, we also consider the different approaches used to structure data based on the previous formats, so it is easier to capture their semantics. Concretely, Data Structure Definitions (DSD) for SDMX, XBRL Taxonomies and DPM Data Dictionaries, as shown in Fig. 1.

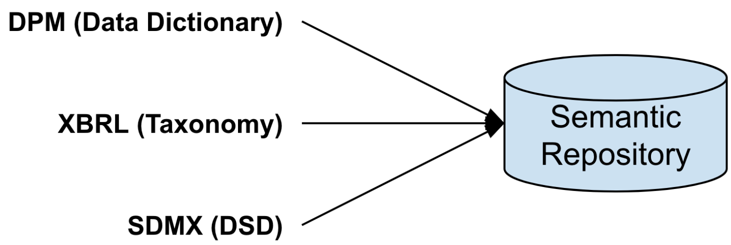

Fig. 1. Semantic integration of DPM, XBRL and SDMX.

To choose a semantic framework to integrate the data, it has been vital to consider the multidimensional nature of the data, far beyond 2D data available from spreadsheets. This way it will make possible to avoid having to encode hidden dimensions into footnotes, attachments, etc. For instance, DPM allows multiple dimensions per data point. Moreover, dimensions might be hierarchically organised (like geographical administrative divisions).

After considering different alternatives to perform the integration of the different sources of data using semantic technologies, like existing approaches to mapping XBRL to RDF (García & Gil, 2010), the proposal is to use the existing RDF Data Cube Vocabulary (Cyganiak & Reynolds, 2014) as the semantic framework for data integration.

Consequently, the idea is to map DPM, XBRL and SDMX to the RDF Data Cube Vocabulary. This vocabulary is capable of representing multidimensional data; in fact, it was initially conceived as the Semantic Web vocabulary to represent statistical data, and particularly SDMX data.

Another benefit of the RDF Data Cube Vocabulary is that it is a W3C Recommendation and consequently a Web standard. Moreover, there have been previous attempts to represent XBRL data using it (Kämpgen et al., 2014) that showed it is feasible to use it to express financial data.

Next, Section 1 provides an overview of the RDF Data Cube Vocabulary and, then, Section 2 presents an example of modelling of DPM and XBRL data using the RDF Data Cube Vocabulary.

1.1 Overview of the RDF Data Cube Vocabulary

In the RDF Data Cube Vocabulary, a dataset comprises a collection of observations made at some points across some logical space. The collection can be characterized by a set of dimensions that define what the observation applies to (e.g. time, area, gender). It also includes metadata describing what has been measured (metrics like economic activity, population), how it was measured and how the observations are expressed (attributes like units, multipliers, status). The dataset is thus structured as a multidimensional space, or hypercube, indexed by those dimensions.

A cube is organized according to a set of dimensions, attributes and measures:

- The dimension components serve to identify the observations. A collection of values for all the dimension components is sufficient to determine a single observation. Examples of dimensions include the time to which the observation applies or a geographic region which the observation covers.
- On the other hand, the measure components represent the phenomenon being observed.
- Finally, the attribute components allow us to qualify and interpret the observed value(s). They enable specification of the units of measure, any scaling factors and metadata such as the status of the observation (e.g. estimated, provisional).

It is frequently useful to group subsets of observations within a dataset. In particular, to fix all but one (or a small subset) of the dimensions and be able to refer to all observations with those dimension values as a single entity. We call such a selection a slice through the cube. Fig. 2 shows the concepts and relationships that constitute the RDF Data Cube Vocabulary.

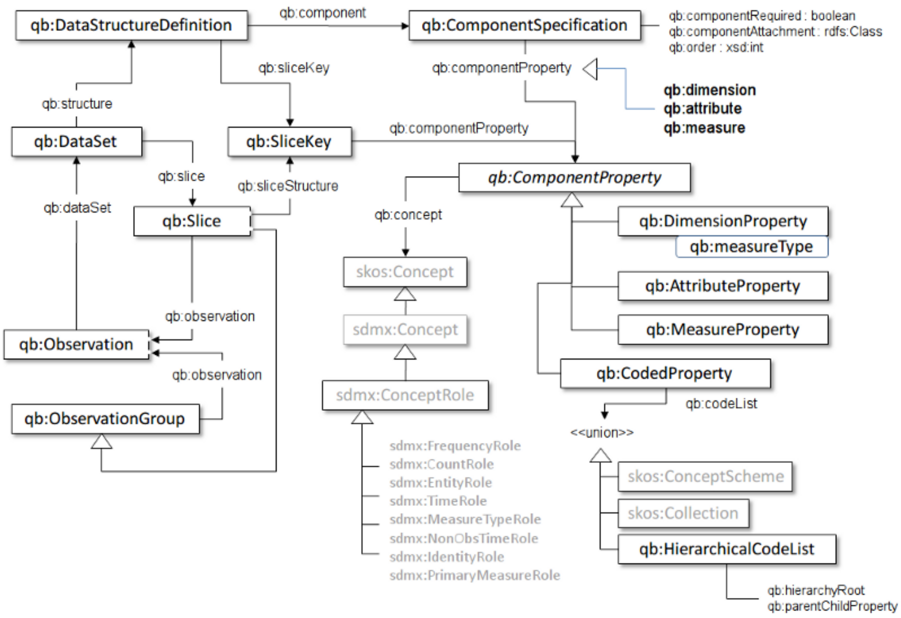

Fig. 2. RDF Data Cube Vocabulary key terms and their relationships

2 Modelling DPM and XBRL using RDF Data Cubes

This section illustrates the capabilities of the RDF Data Cube Vocabulary to model DPM and XBRL financial data using the example presented next. The DPM examples are based on the taxonomy "FINancial REPorting 2016-A Individual (2.1.5)", authored by EBA using DPM 2.5 and based on table "Balance Sheet Statement: Assets (F_01.01)", row "Total assets" and column "Carrying amount".

The Data Point is:

- Metric: eba_mi53 - Carrying amount
- Dimension 1: BAS - Base
- Dimension 1 Value: x6 - Assets
- Dimension 2: MCY - Main Category
- Dimension 2 Value: x25 - All assets

Moreover, the entity has LEI and the point in time is 2017-07-01. The previous Data Point is represented using XBRL as the instance presented in Table 1.

Table 1. XBRL representation of the Data Point

```
<xbrli:context id="c1">
   <xbrli:entity>
      <xbrli:identifier scheme="http://standards.iso.org/iso/17442">
      entity-1</xbrli:identifier>
   </xbrli:entity>
   <xbrli:period>
      <xbrli:instant>2017-07-01</xbrli:instant>
   </xbrli:period>
   <xbrli:scenario>
      <xbrldi:explicitMember dimension="eba_dim:BAS">eba_BA:x6
      </xbrldi:explicitMember>
      <xbrldi:explicitMember dimension="eba_dim:MCY">eba_MC:x25
      </xbrldi:explicitMember>
   </xbrli:scenario>
</xbrli:context>
<eba_met:mi53 unitRef="uEUR" decimals="-3" contextRef="c1">1</eba_met:mir3>
```

The previous Data Point (or the corresponding XBRL instance) can be mapped using the capabilities of the RDF Data Cube Vocabulary to the Observation presented in Table 2.

Table 2. RDF Data Cube Vocabulary representation of the Data Point and XBRL instance

```
ex:dst-1/obs-1 a qb:Observation;
    qb:dataSet  ex:dtst-1 ;
    xbrli:entity lei:549300N33JQ7EG2VD447 ;
    sdmx-dim:refTime "2017-07-01"^^xsd:date ;
    eba_dim:BAS eba_BA:x6 ;
    eba_dim:MCY eba_MC:x25 ;
    eba_met:mi53 "1"^^xsd:int ;
    sdmx-att:decimals "-3"^^xsd:int ;
    sdmx-att:currency currency:EUR .
```

The representation using the RDF Data Cube Vocabulary uses the concepts and relationships related with dimensions to model the dimensions in the Data Point, while the ones related with metrics are used for the Data Point's metric. The time point and the referred entity are also modelled using concepts and relationships related to dimensions. Finally, attributes are used for decimals and currency. Fig. 3 show a graphical representation of the Observation in Table 2.

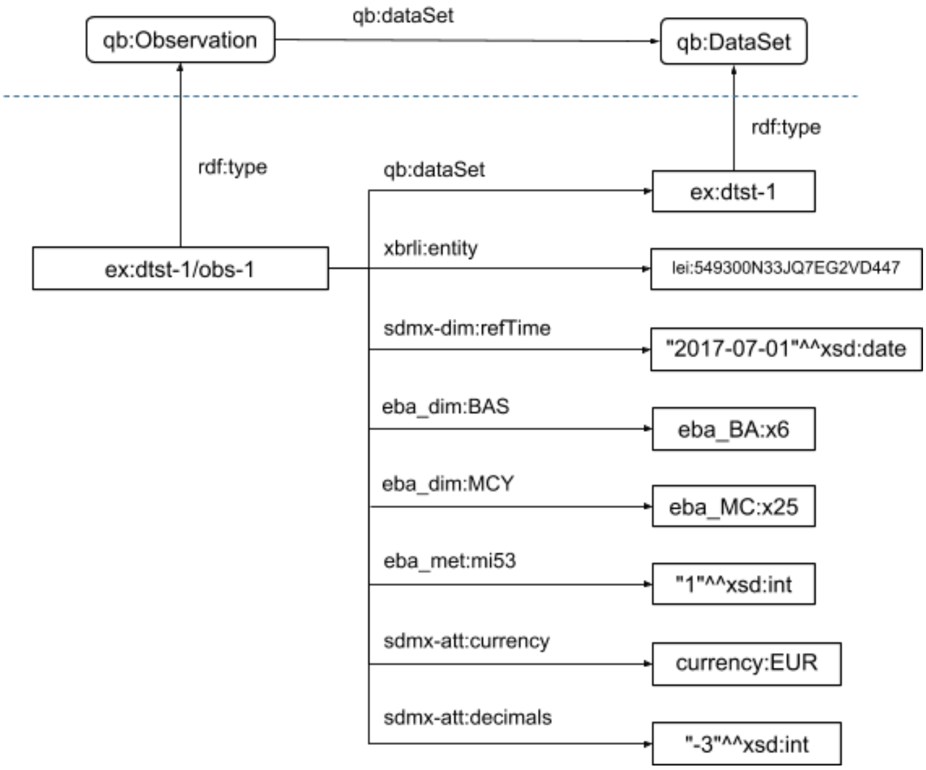

Fig. 3. Graph view of the RDF Data Cube Observation in Table 2

Besides of being capable of modelling a Data Point (or XBRL instance), the RDF Data Cube Vocabulary provides mechanisms for modelling how the dimensions, metrics and attributes are structured. This way, it is possible to capture the information in DPM Data Dictionaries or XBRL Taxonomies.

Fig. 4 presents the RDF Data Cube concepts and relationships responsible for capturing the structure of a dataset. Each dataset is related to a Data Structure Definition that defines its components, the measures, dimensions and attributes, which are Component Properties, a kind of RDF Property.

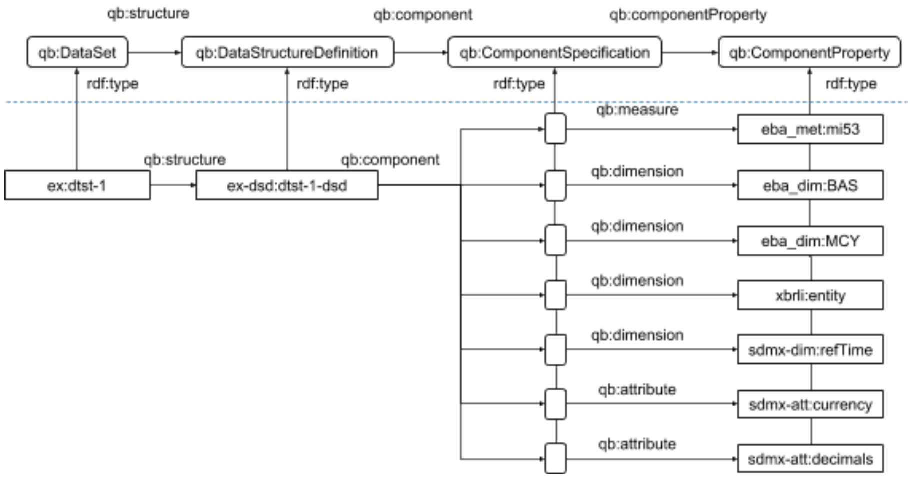

Fig. 4. Data Structure Definition for the example observation dataset

The Data Structure Definition also defines the values that the measures, dimensions and attribute properties link to, their ranges, as shown in Fig. 5. Ranges can be data types, like dates or integer, or defined based on taxonomies, as shown in Fig. 6.

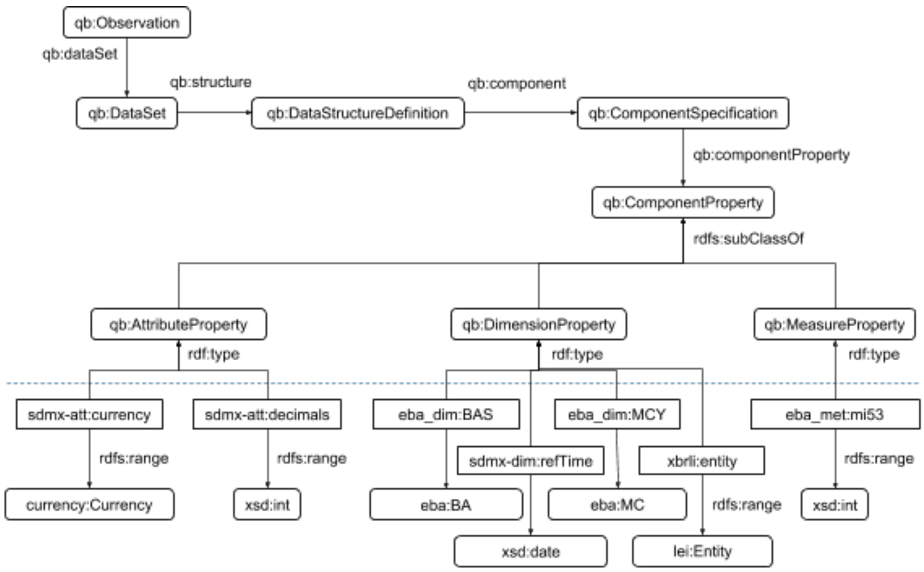

Fig. 5. Definition of attribute, dimension and measure properties and their ranges

Fig. 6 provides an example of range definition for one of the dimension properties in Fig. 5. Concretely, the range of the property *eba_dim:BAS*, which is eba:BA, is defined as both a Code List and a Concept with a set of members, which can be hierarchically organised.

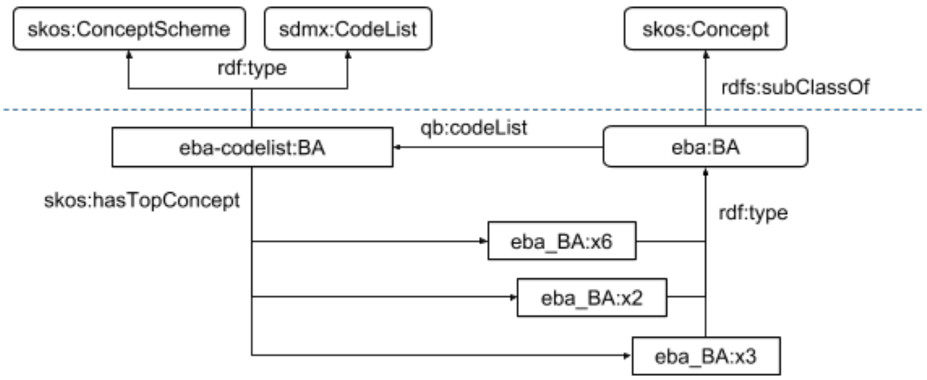

Fig. 6. Definition of the eba:BA domain as a Code List and a set of concepts

3 Conclusions and Future Work

As it has been shown through the exercise of modelling DMP and XBRL using the RDF Data Cube Vocabulary in Section 2, it is possible to capture information at the Data Point (XBRL Instance) and Data Dictionary (XBRL Taxonomy) using this vocabulary. As this vocabulary is also capable of modelling SDMX statistical data, because it was initially designed with this aim, it constitutes a semantic framework where information from all these sources can be integrated.

Moreover, as working with semantic technologies, it becomes easier to automatise the process of integrating different data structures (working at the level of dictionaries and taxonomies) and making all the information interoperable.

To do so, a more systematic analysis of how the different constructs in the DPM Dictionaries and XBRL Taxonomies can be mapped to the RDF Data Cube Data Structure Definitions (DSDs) is necessary. This effort should be followed by an analysis of their underlying meaning and the formalisation of the semantic relationships among the concepts and relationships defined in the DPM Dictionaries, XBRL Taxonomies and SDMX DSDs. For instance, formalise the equivalence between the concepts related to currency values in all them so they can be queried transparently using semantic requests.

Additionally, it will be possible to benefit from existing efforts to unify these dictionaries and taxonomies, like the ECB Single Data Dictionary (SDD) that can also be formalised using semantic technologies and become the hub where the other can be integrated using semantic relationships.

Acknowledgements

We would like to thank Ignacio Boixo for his guidance during the conception of this experiment. His feedback helped us better identify the current issues in the financial data integration space and then shape the proposal in a way that better suits these needs.

References

Cyganiak, R., & Reynolds, D. (2014). The RDF Data Cube Vocabulary. W3C Recommendation, 16 January 2014. Available from: http://www.w3.org/TR/vocab-data-cube/

García, R., & Gil, R. (2010). Linking XBRL financial data. In Linking enterprise data (pp. 103-125). Springer US.

Kämpgen, B., Weller, T., O'Riain, S., Weber, C., & Harth, A. (2014, May). Accepting the XBRL challenge with linked data for financial data integration. In European Semantic Web Conference (pp. 595-610). Springer.