EXPLORING A SEMANTIC FRAMEWORK FOR INTEGRATING DPM, XBRL AND SDMX DATA

Roberto García
Associate Professor
Universitat de Lleida
INTRODUCTION

- **Proliferation** financial data and available formats
- Increased need for ways to *integrate* it

- **Semantic Technologies:**
  - facilitate integration by moving effort to the level of meanings
  - instead of trying to deal with syntax subtleties

- Explore this alternative through a practical *experiment*
INTEGRATION SOURCES

• **Data sources:**
  - XBRL,
  - Data Point Model (DPM)
  - SDMX

• **Schema sources:**
  - XBRL Taxonomies,
  - DPM Data Dictionaries
  - SMX Data Structure Definitions (DSD)
CONCEPTUAL FRAMEWORK

• Consider the **multidimensional** nature of the data (e.g. DPM)
  • Far beyond 2D data available from **spreadsheets**
  • **Avoid** having to encode “hidden dimensions” into footnotes, attachments, etc.
  • **Dimensions** might be **hierarchically organised** (like geographical administrative divisions)

• Proposal: **RDF Data Cube Vocabulary** (based on semantic technologies, RDF & Web Ontologies)
  • Supports **multidimensional** data
  • **Based on SDMX** and the Semantic Web vocabulary for statistical data
  • Web standard (**W3C Recommendation**)

• Approach:
  • **Map DPM and XBRL** to the RDF Data Cube Vocabulary (example next)
    • SDMX trivially becomes RDF based on the Data Cube Vocabulary
DATA CUBE

Dataset a collection of observations

Dimensions identify an observation
e.g. observation time or a geographic region

Measures represent observed phenomenon

Attributes qualify / help interpret observations
e.g. units of measure, scaling factors or observation status (estimated, provisional,...)

Slice subsets observations by fixing all but one dimension (or a few)

https://www.slideshare.net/140er/lets-talkaboutstatisticaldatainrdf
RDF DATA CUBE

Dataset a collection of observations

Dimensions identify an observation
e.g. observation time or a geographic region

Measures represent observed phenomenon

Attributes qualify / help interpret observations
e.g. units of measure, scaling factors or observation status (estimated, provisional,...)

Slice subsets observations by fixing all but one dimension (or a few)

https://www.w3.org/TR/vocab-data-cube/#outline
MODELLING EXAMPLE

• **Data Point** example 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"

  • Metric: **eba_mi53 - Carrying amount** → Value: **1000 EUR**
  • Dimension 1: **BAS – Base** → Dimension 1 Value: **x6 - Assets**
  • Dimension 2: **MCY - Main Category** → Dimension 2 Value: **x25 - All assets**

• Plus entity with LEI **549300N33JQ7EG2VD447** and time **2017-07-01**
• XBRL representation of the Data Point

```xml
<xbrli:context id="c1">
  <xbrli:entity>
    <xbrli:identifier scheme="http://standards.iso.org/iso/17442">
      549300N33JQ7EG2VD447
    </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="EUR" decimals="-3" contextRef="c1">1</eba_met:mi53>
```
• RDF Data Cube Vocabulary representation of the Data Point and XBRL instance

```turtle
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 .
```
MODELLING EXAMPLE

- RDF Data Cube Vocabulary terms to model:
  - Observations linked to their dataset
  - Dimensions, including entities and time
  - Measures, including data type
  - Attributes, decimals and currency
• RDF Data Cube Vocabulary also to model how the dimensions, metrics and attributes are structured

• Capture
  • DPM Data Dictionaries
  • XBRL Taxonomies in a Data Structure Definition (DSD) linked to each dataset
• **DSD** also defines the types of the values of measures, dimensions and attributes (their **ranges**):
  - **Data types**
    (date, integer,...)
  - **Taxonomy terms**
Example: the range of the property `eba_dim:BAS` is `eba:BA`.
* `eba:BA` is defined as a **SDMX Code List** (and a semantic SKOS Concept Scheme) with members:
  * `eba_BA:x6`
  * `eba_BA:x2`
  * `eba_BA:x3`

(members can be hierarchically organised)
CONCLUSIONS

• Possible to use the RDF Data Cube Vocabulary to semantically model and integrate:
  • Data Point / XBRL Instance
  • Data Dictionary / XBRL Taxonomy
• Per design, also SDMX / DSD

• Semantic technologies facilitate the integration by operating at the level of dictionaries and taxonomies

• Facilitates multidimensional data management and multiple views on the same data
FUTURE WORK

• More **systematic analysis** of how the different constructs in the DPM Dictionaries and XBRL Taxonomies can be **mapped** to the RDF Data Cube DSDs (automation?)

• **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 of them so they can be queried transparently using semantic requests

• Additionally, possible to **benefit from existing efforts to unify these dictionaries and taxonomies**
  • **ECB Single Data Dictionary** (SDD) can also be formalised using semantic technologies and become the **hub for integration using semantic relationships**
THANK YOU

EUROFILING XBRL WEEK
WARSAW 28-30 MAY 2018