Semantic Statistics: Bringing together SDMX and SCOVO

LDOW 2009 | Richard Cyganiak, Jeni Tennison, Arofan Gregory, Wolfgang Halb, Simon Field

Statistical data

- Average income per household in 2002 by UK administrative area
- Number of migrants to Ireland by country of origin
- Goods loaded/unloaded by port

Characteristics of Statistical data

- Aggregate data
- Numeric
- Time series
- Multi-dimensional tables, cubes
- Evidence for policy-making
- Open data!

Office for National Statistics workshop

- Workshop in Feb 2010
- ONS asks what to do about linked data
- Meeting of communities
- Use SCOVO? Use SDMX?

SCOVO

The Statistical Core Vocabulary (scovo)

This version:

http://purl.org/NET/scovo

Editors (alphabetical):

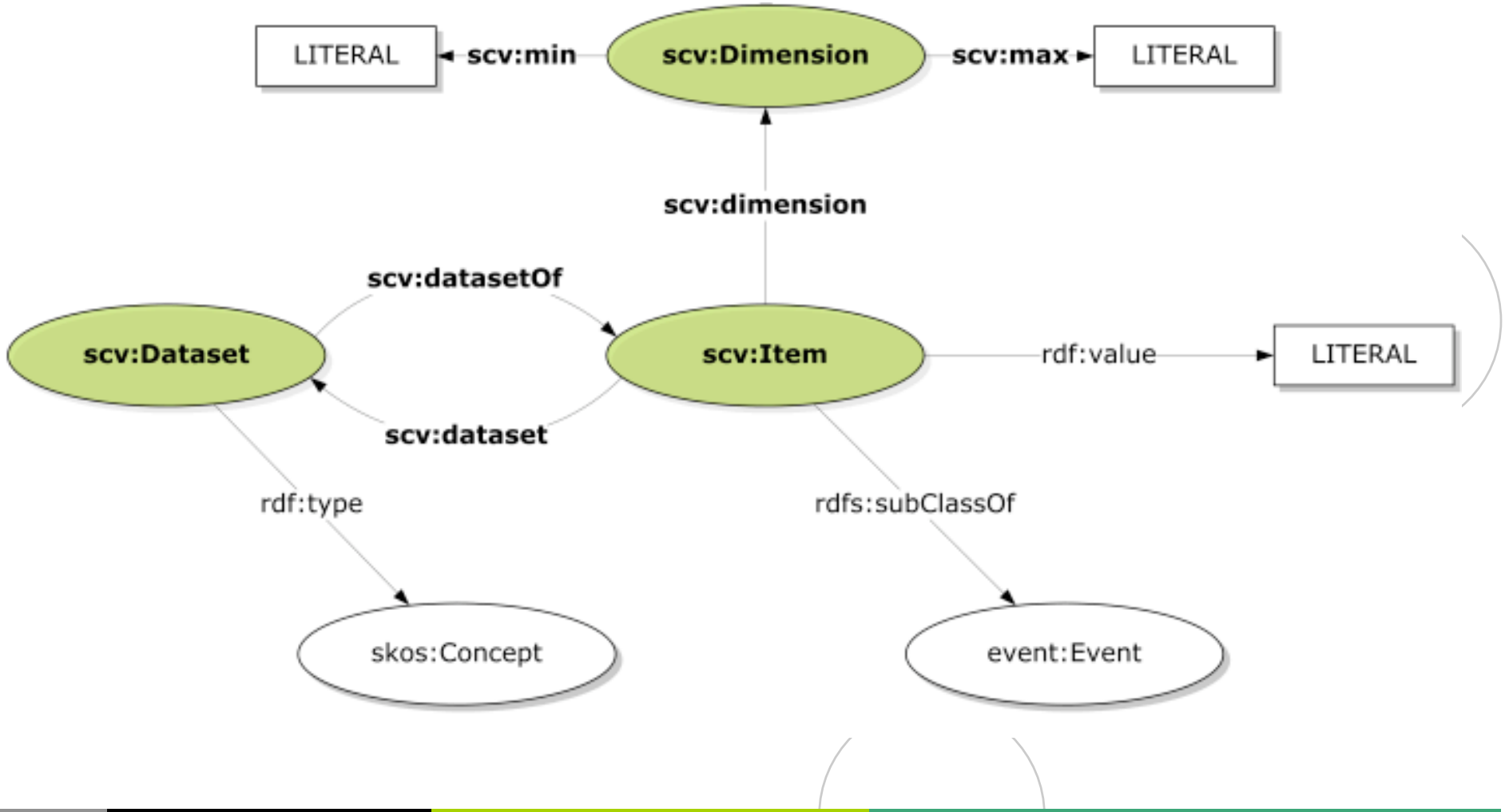
Ayers, Danny (Talis)
Feigenbaum, Lee (Cambridge Semantics)
Halb, Wolfgang (JOANNEUM RESEARCH)
Hausenblas, Michael (JOANNEUM RESEARCH)
Heath, Tom (Talis)
Raimond, Yves (Centre for Digital Music, Queen Mary, University of London)

Abstract

This document specifies an [RDF-Schema] vocabulary for representing statistical data on the Web. It is normatively encoded in [XHTML+RDFa], that is embedded in this page.

Table of Contents

SCOVO Overview



SCOVO use

- Riese (Eurostat)
- voiD, RDFstats
- LOIUS (Italian university statistics)
- German Environment Speciman Bank
- UK government data

SDMX

- Statistical Data and Metadata Exchange
- Development started in 2001
- EDIFACT and XML

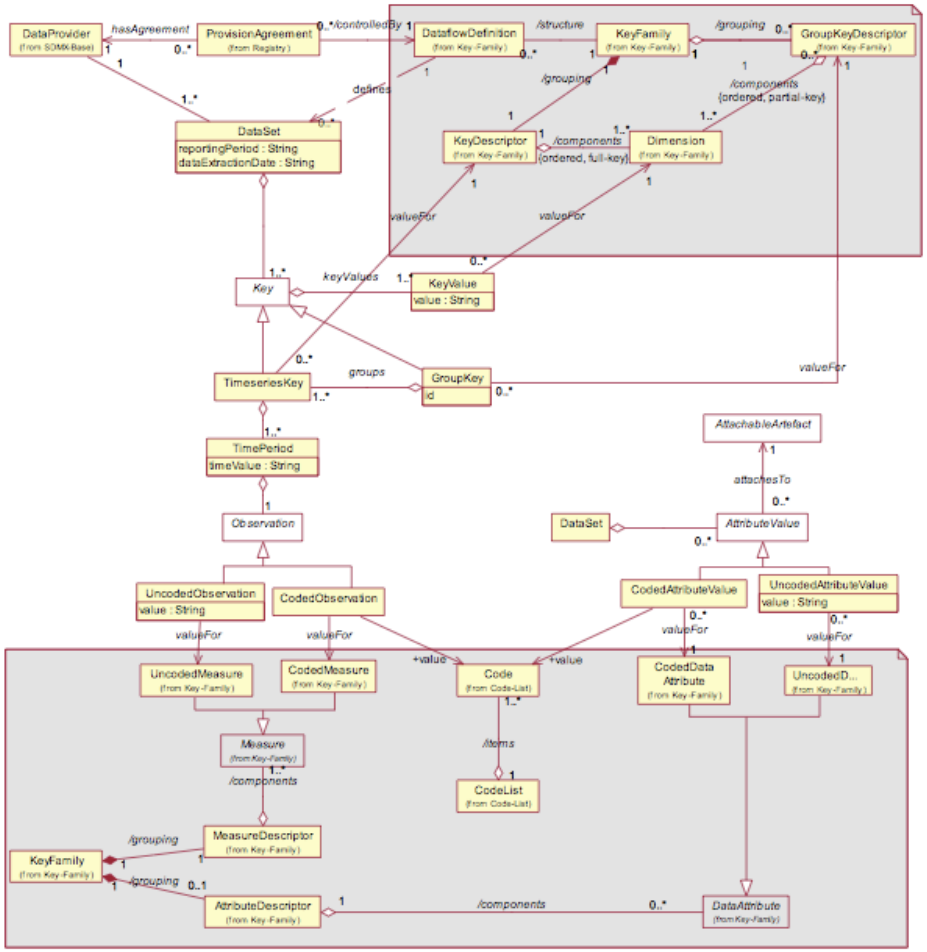
SDMX Users

- U.S. Federal Reserve Board
- European Central Bank
- Eurostat
- WHO
- IMF
- World Bank
- OECD, UN and Eurostat expect publishers of national statistics to report in SDMX

SDMX in XML

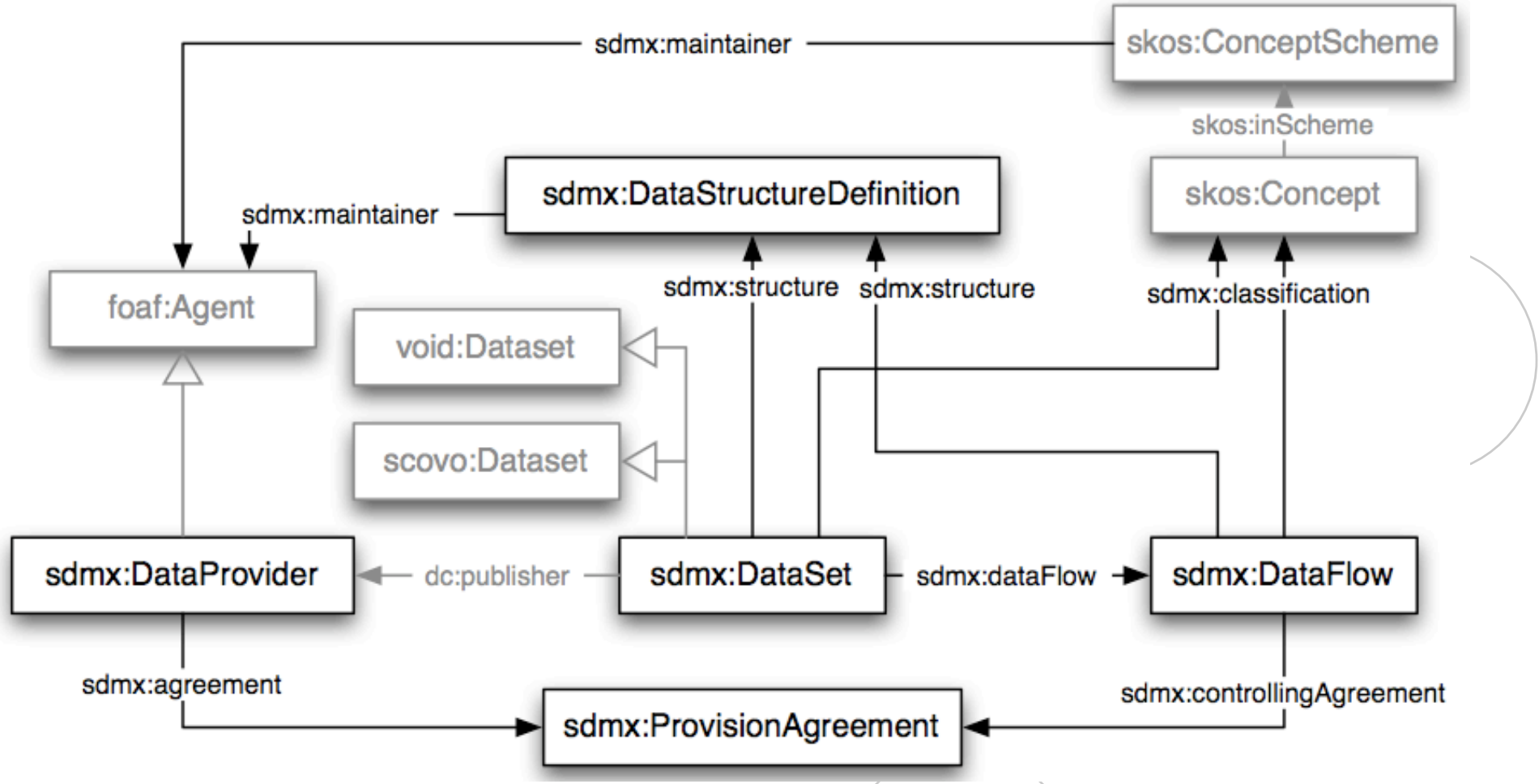

```
        <DataSetAction>Append</DataSetAction>
        <Extracted>2001-03-11T09:30:47-05:00</Extracted>
        <ReportingBegin>2000-01-01T00:00:00</ReportingBegin>
        <ReportingEnd>2000-12-01T00:00:00</ReportingEnd>
</Header>
<bisc:DataSet>
        <bisc:SiblingGroup  VIS_CTY="MX" JD_TYPE="P" JD_CATEGORY="A" AVAILABILITY="
        <bisc:SiblingGroup VIS_CTY="MX" JD_TYPE="P" JD_CATEGORY="B" AVAILABILITY="A
        <bisc:Series FREQ="M" COLLECTION="B" TIME_FORMAT="P1M" VIS_CTY="MX" JD_TYPE
                <bisc:Obs TIME_PERIOD="2000-01" OBS_VALUE="3.14" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2001-02" OBS_VALUE="2.29" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-03" OBS_VALUE="3.14" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-04" OBS_VALUE="5.24" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-05" OBS_VALUE="3.14" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-06" OBS_VALUE="3.78" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-07" OBS_VALUE="3.65" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-08" OBS_VALUE="2.37" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-09" OBS_VALUE="3.14" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-10" OBS_VALUE="3.17" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-11" OBS_VALUE="3.34" OBS_STATUS="A"/>
                <bisc:Obs TIME_PERIOD="2000-12" OBS_VALUE="1.21" OBS_STATUS="A"/>
        </bisc:Series>
        <bisc:Series FREQ="A" COLLECTION="B" TIME_FORMAT="P1Y" VIS_CTY="MX" JD_TYPE
                <bisc:Obs TIME_PERIOD="2000-01" OBS_VALUE="3.14" OBS_STATUS="A"/>
```

SDMX information model

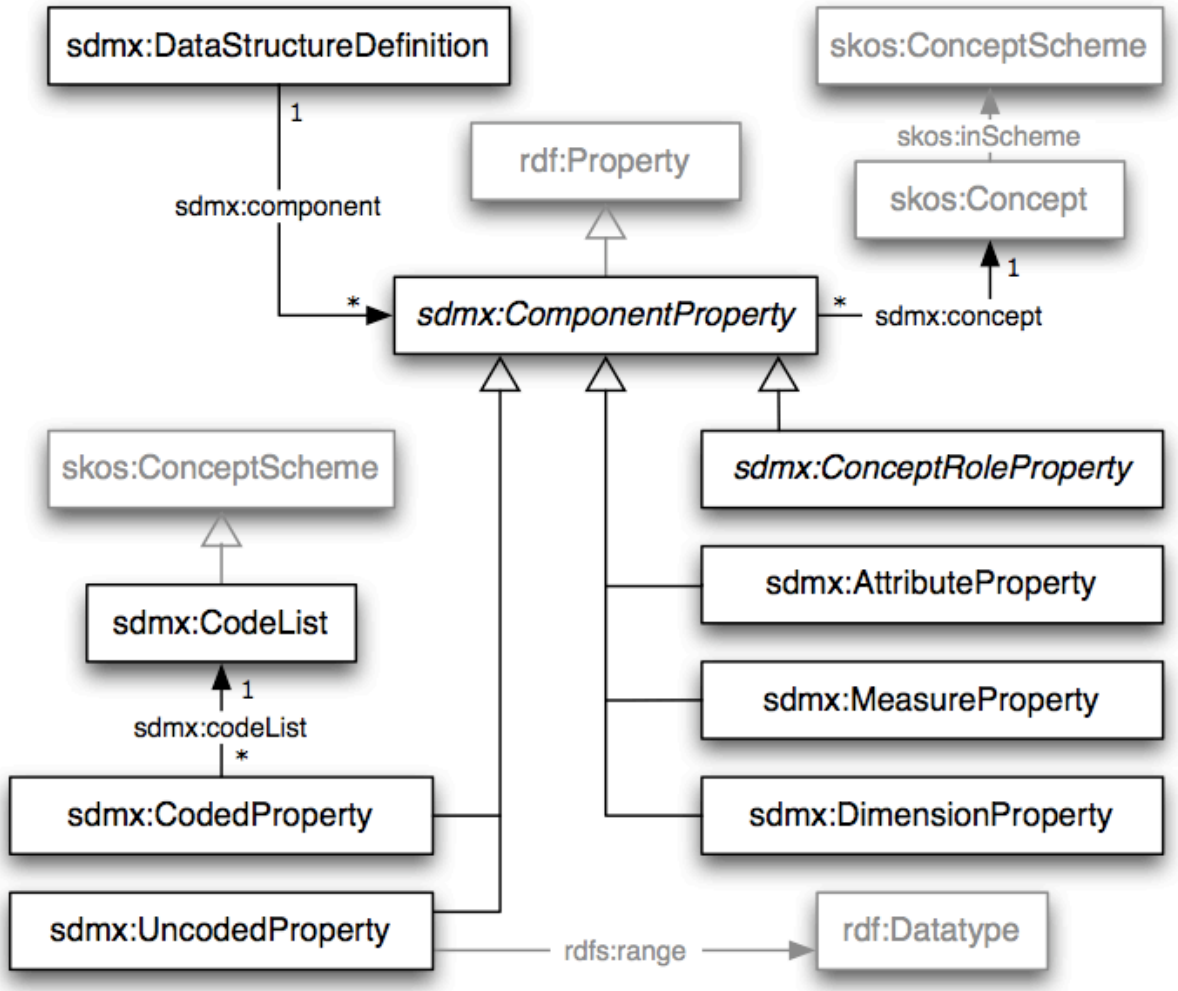


Let's express the SDMX information model in RDF

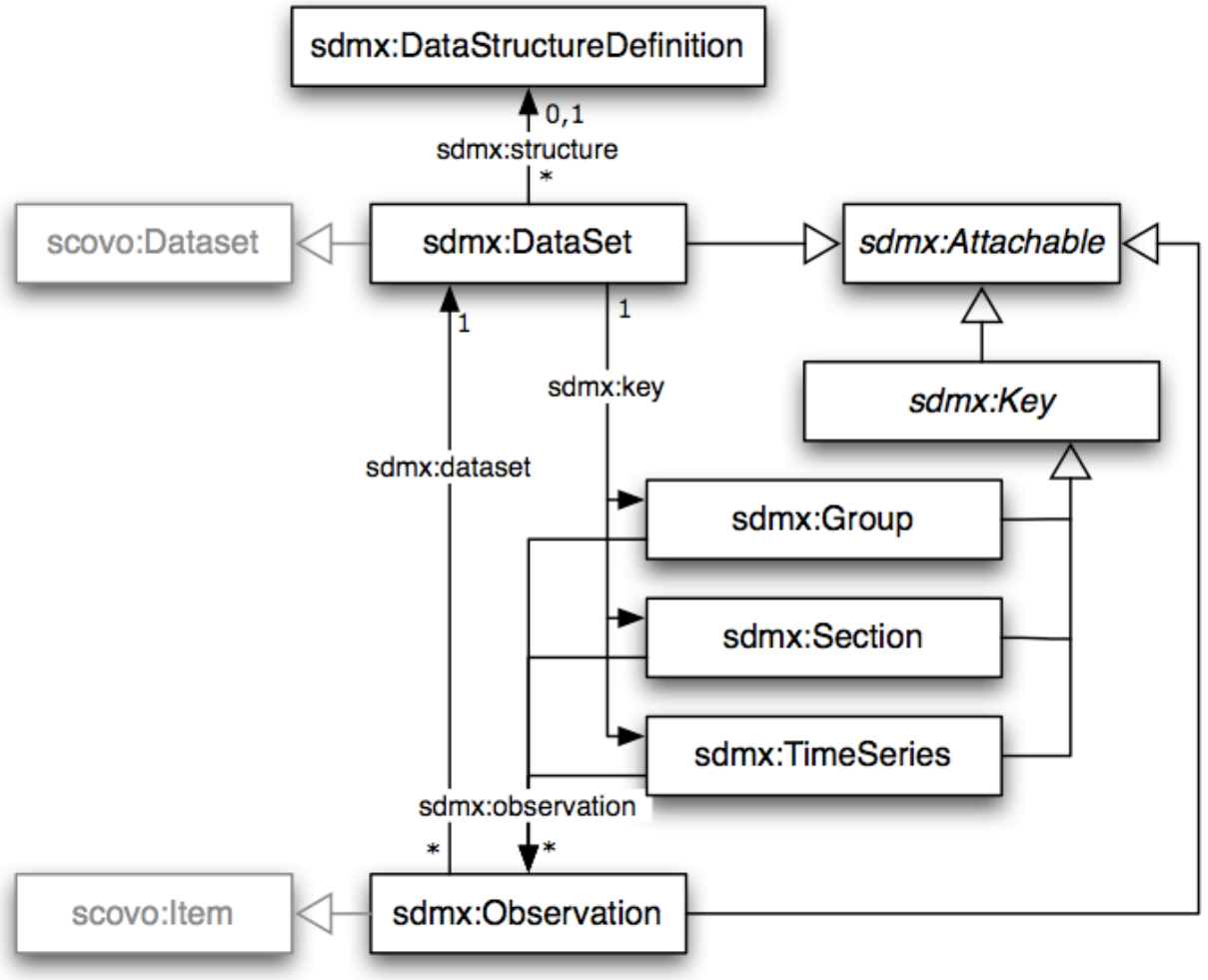
SDMX-RDF, part 1



SDMX in RDF, part 2



SDMX in RDF, part 3



Differences from SCOVO

- Distinguishing Dimensions and Attributes
- Data Structure Definitions
- Code Lists (SKOS)
- Re-usable Concepts
- Time Series and Groups within a dataset

Ongoing work

- Spec
- PESA data (Epimorphics)
- XSLT for SDMX-ML (Jeni)
- CSO statistics, PC-Axis (DERI)
- Meeting in July

Observations

- Domain standards capture domain expertise
- Unlocking valuable data
- RDF tooling is pretty good
- RDF advantage: re-use of standardization efforts (Dublin Core, SKOS, voiD)

Learn more

- Paper
  - http://richard.cyganiak.de/2008/papers/sdmx-ldow2010.pdf
- Google Group
  - Discussion list
  - http://groups.google.com/group/publishing-statistical-data
- Google Code project
  - Issue tracker
  - Spec drafts
  - http://code.google.com/p/publishing-statistical-data