

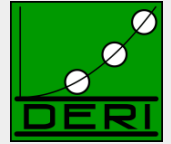
Publishing Statistics with the Data Cube Vocabulary

Richard Cyganiak

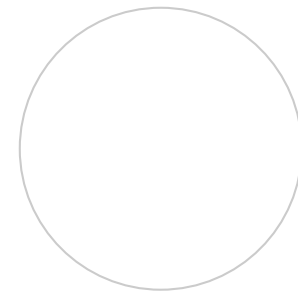
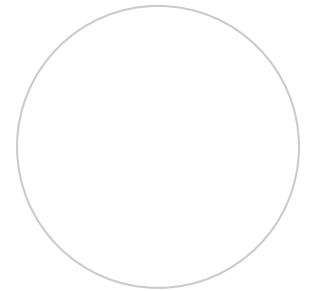
W3C GLD WG F2F1, 29 June 2011



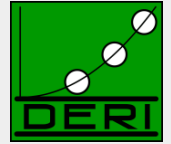
Statistical data



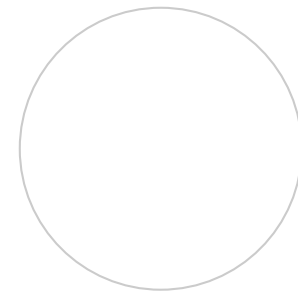
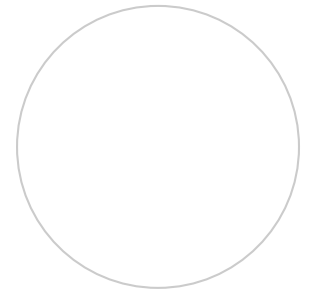
- Average income per household in 2002 by UK administrative area
- Number of migrants to Ireland by country of origin
- Goods loaded/unloaded by port
- Census data



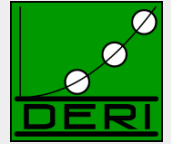
Characteristics of Statistical data



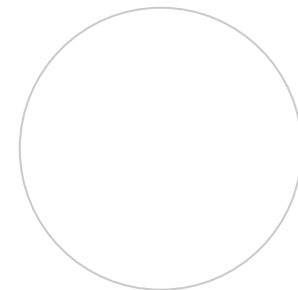
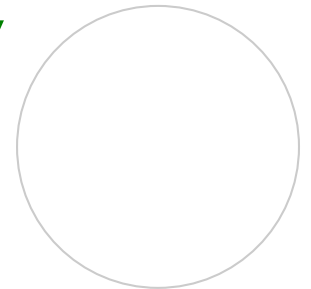
- Aggregate data
- Numeric
- Time series
- Multi-dimensional tables, cubes



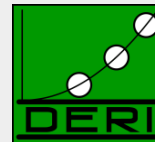
Statistical data is attractive



- Readily available
- High quality
- From data-savvy source
- Valuable background for all sorts of mashups/
analysis

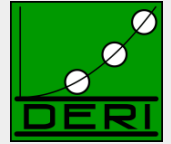


A data cube: life expectancy



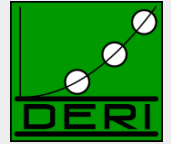
	2004-6		2005-7		2006-8	
	Male	Female	Male	Female	Male	Female
Newport	76.7	80.7	77.1	80.9	77.0	81.5
Cardiff	78.7	83.3	78.6	83.7	78.7	83.4
Monmouthshire	76.6	81.3	76.5	81.5	76.6	81.7
Merthyr Tydfil	75.5	79.1	75.5	79.4	74.9	79.6

- Dimensions: time, geography, sex
- Measure
- Attributes: Unit of measurement; accuracy; ...

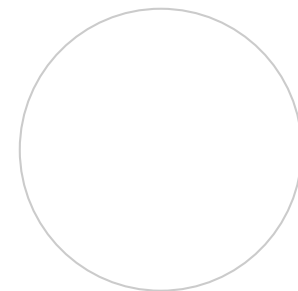
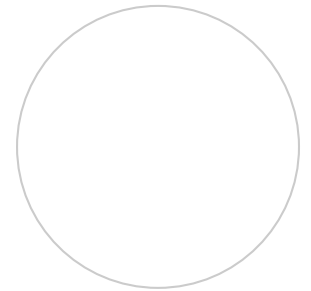


Data Cube Vocabulary

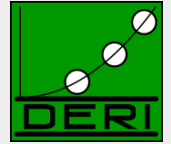
Based on SDMX



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

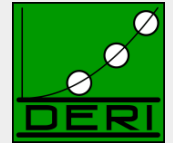


SDMX Users



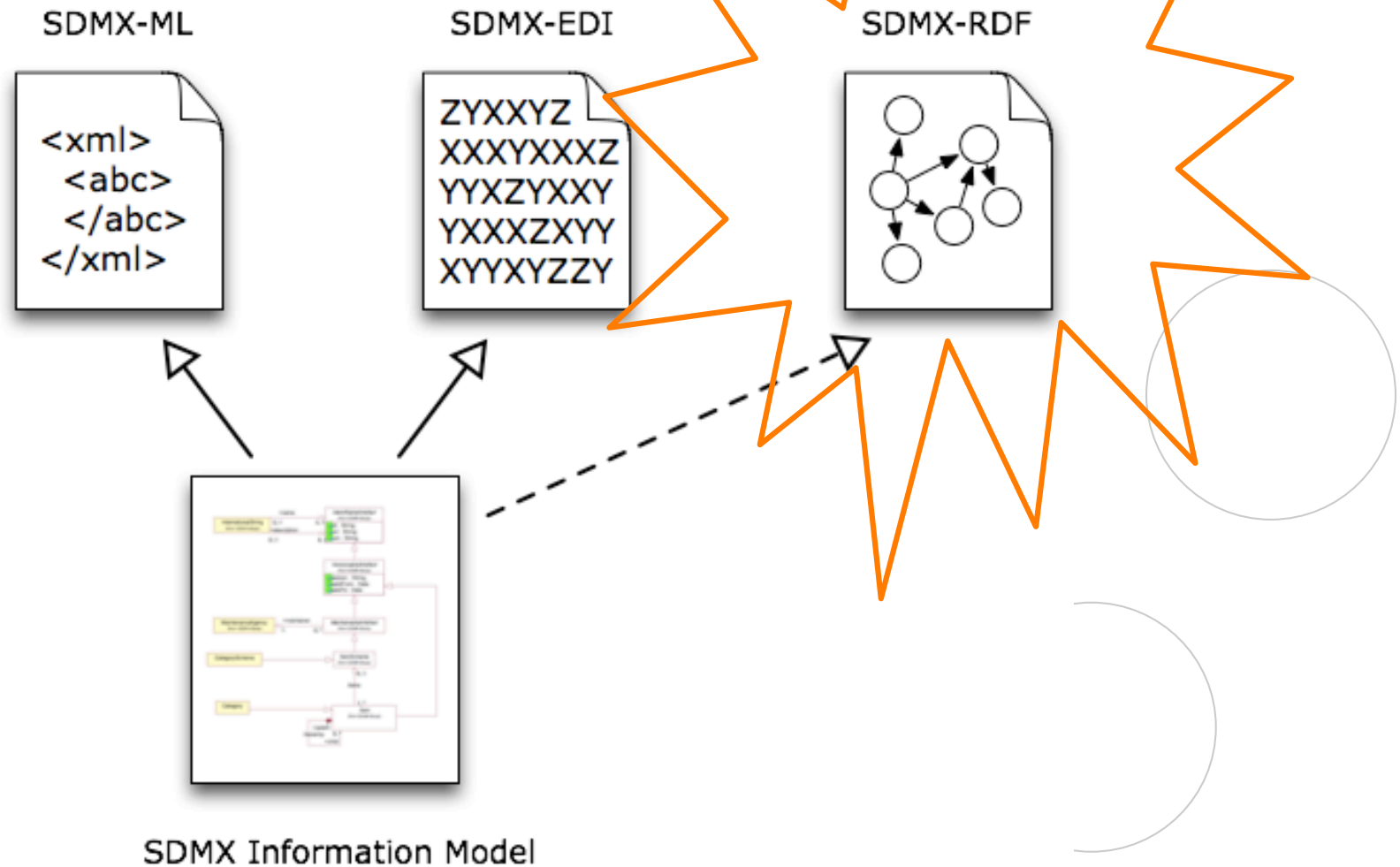
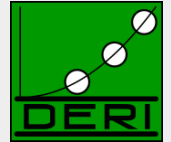
- U.S. Federal Reserve Board
- European Central Bank
- Eurostat
- WHO
- IMF
- World Bank
- OECD, UN and Eurostat expect publishers of national statistics to report in SDMX

SDMX in XML

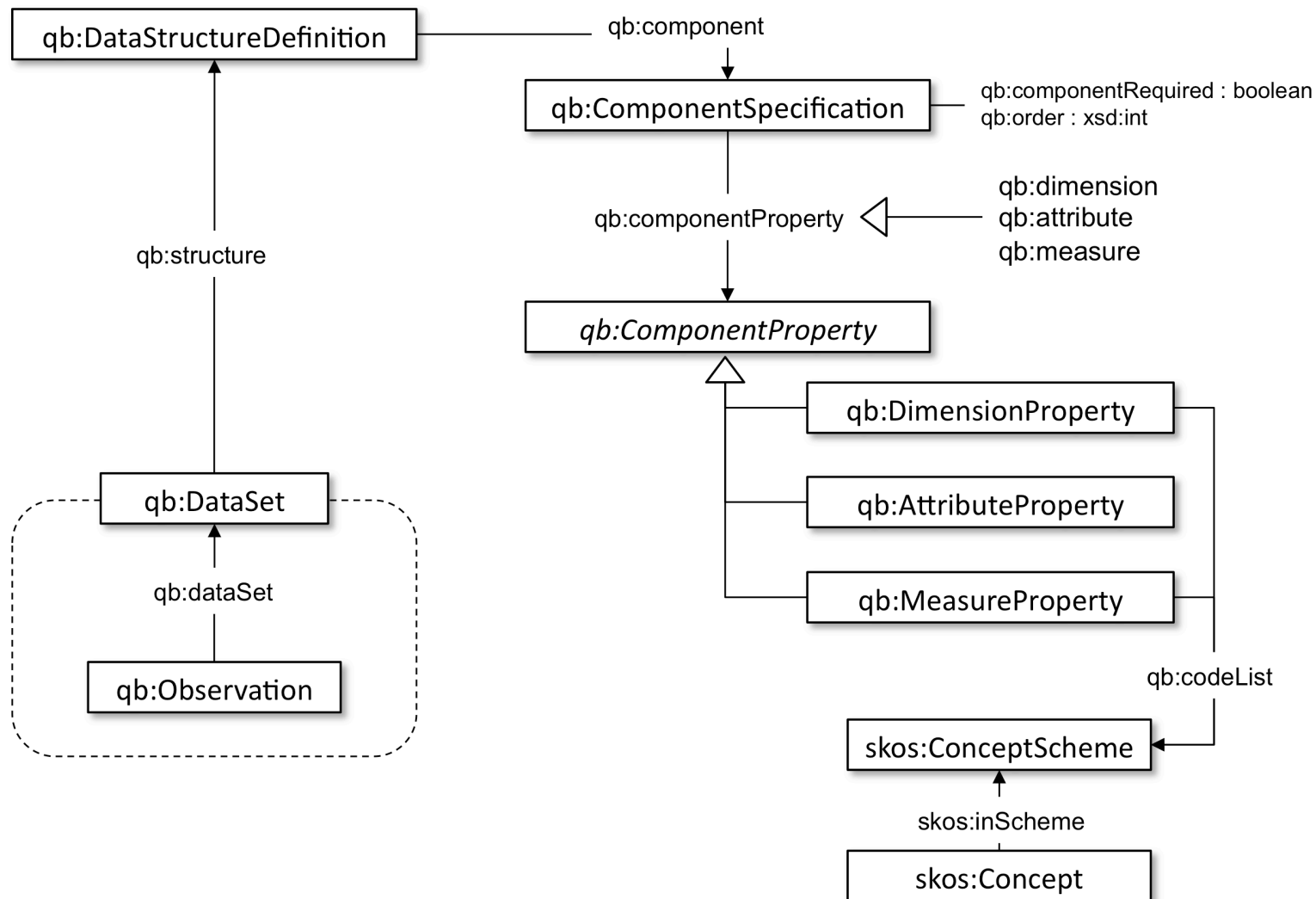
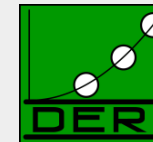


```
<DataSetAction>Append</DataSetAction>
<Extracted>2001-03-11T09:30:47-05:00</Extracted>
<ReportingBegin>2000-01-01T00:00:00</ReportingBegin>
<ReportingEnd>2000-12-01T00:00:00</ReportingEnd>
</Header>
<bisc:DataSet>
  <bisc:SiblingGroup VIS_CTY="MX" JD_TYPE="P" JD_CATEGORY="A" AVAILABILITY="".
  <bisc:SiblingGroup 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"/>
```

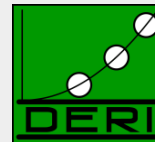

SDMX: Information model + syntaxes



Data Cube Vocabulary (core)



Draft Specification



The RDF Data Cube vocabulary

Last update:

2010-07-14

Editors:

Richard Cyganiak ([DERI, NUI Galway](#))

Dave Reynolds ([Epimorphics Ltd](#))

Jeni Tennison ([TSO](#))

Abstract

There are many situations where it would be useful to be able to publish multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts. The Data Cube vocabulary provides a means to do this using the W3C [RDF](#) (Resource Description Framework) standard. The model underpinning the Data Cube vocabulary is compatible with the cube model that underlies [SDMX](#) (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations. The Data Cube vocabulary is a core foundation which supports extension vocabularies to enable publication of other aspects of statistical data flows.

Status of this document

This is an editor's draft without any formal standing. It is not endorsed by any organisation. Anything in this document is still subject to change at this point. The editors seek feedback on the document. Please send any comments to the [project's Google Group](#).

Table of Contents

1 Introduction

- Data Cube: Design is done, spec written
- Need more tools
 - Stats2RDF: OntoWiki plugin for CSV => Data Cube
- Several deployed datasets
 - data.gov.uk datasets
 - 2006 Ireland census; 2011 in preparation
 - Italy: istat immigration
 - Enakting
 - ...
- Exploring usage for OLAP-style data warehouses
- Exploring usage for DDI