

# DDI-RDF Discovery Vocabulary

A Metadata Vocabulary for Documenting Research and Survey Data

## Linked Data on the Web (LDOW 2013)

14.05.2013

**Thomas Bosch**

GESIS, Germany

[thomas.bosch@gesis.org](mailto:thomas.bosch@gesis.org)

**Arofan Gregory**

Open Data Foundation, USA

[agregory@opendatafoundation.org](mailto:agregory@opendatafoundation.org)

**Richard Cyganiak**

DERI, Ireland

[richard.cyganiak@deri.org](mailto:richard.cyganiak@deri.org)

**Joachim Wackerow**

GESIS, Germany

[joachim.wackerow@gesis.org](mailto:joachim.wackerow@gesis.org)

## Outline

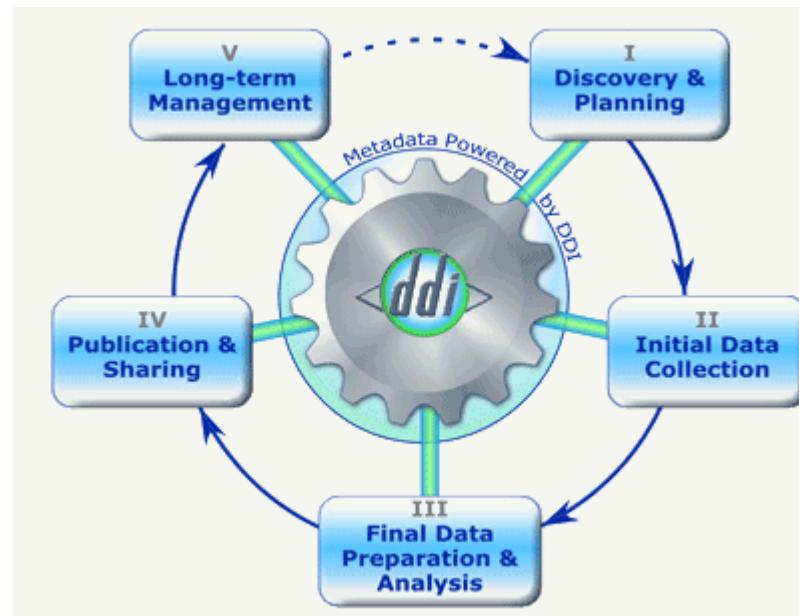
- What is DDI?
- Motivation
- Relationships to Vocabularies
- DDI-RDF Discovery Vocabulary
- Conceptual Model

## What is DDI?

- DDI (**Data Documentation Initiative**)
- DDI is an established **international standard** for the documentation and management of data from the social, behavioral, and economic sciences
- DDI is a **data model** for describing statistical data
  - data collected for research and official statistics
- The DDI Alliance
  - International consortium of > 35 member institutions
  - produces and maintains the DDI

## What is DDI?

- DDI supports the entire **research data lifecycle**  
→ Secondary analysis results can be reproduced



## What is DDI?

- DDI focuses on the documentation of **microdata**
- DDI also supports **aggregated data**
- DDI-C (Codebook)
  - general information about a study
  - data dictionary
- DDI-L (Lifecycle)
  - description of more complex multi-wave studies
  - throughout the data lifecycle

## What is DDI?

- Structured high quality metadata enable secondary analysis without the need to contact the primary researcher
- DDI enables the reuse of metadata of existing studies for designing new studies
- DDI is currently specified using **XML Schemas**
  - XML Schemas are organized in multiple modules corresponding to the individual stages of the research data lifecycle
  - XML Schemas comprehend over 800 XML elements

## Motivation for the DDI Community

- **publish** microdata (data sets representing microdata)
- **increase visibility** of microdata
- **increase use** of microdata
- **discover** microdata
- **enable inferencing** on microdata
- harmonize microdata (make microdata comparable)
- RDF tools can process DDI-RDF

## Motivation for the LD Community

- an **ontology describing the statistical domain** is now available
- **publish microdata**
- **publish metadata** on microdata
- metadata about already published but under-documented microdata can be published
- RDF tools can **process** DDI-RDF
- to **link** microdata to other microdata
  - making the data and the results of research (e.g. publications) more closely connected

# Relationships to Vocabularies

- **DCMI Metadata Terms**
  - are used for citation purposes
- **Simple Knowledge Organization System (SKOS)**
  - is used for creating hierarchies of concepts similar to thesauri and classification systems
- **SKOS Extension (XKOS)**
  - a vocabulary which extends SKOS to allow for a more complete description of formal statistical classifications
  - planned for publication 2013 by the DDI Alliance
  - reference: <https://github.com/linked-statistics/xkos>

# Relationships to Vocabularies

- **Data Catalog Vocabulary (DCAT)**
  - W3C standard for describing catalogs of data sets
  - $\text{disco:LogicalDataSet} \sqsubseteq \text{dcat:Dataset}$
  - $\text{disco: DataFile} \sqsubseteq \text{dcat:Distribution}$
- **RDF Data Cube Vocabulary**
  - W3C standard for representing data cubes, i.e. multidimensional aggregate data
  - $\text{disco:aggregation} (\text{disco:LogicalDataSet}, \text{qb:DataSet})$
  - $\text{disco:inputVariable} (\text{qb:DataSet}, \text{disco:Variable})$
  - reference: <http://www.w3.org/TR/vocab-data-cube/>

## DDI-RDF Discovery Vocabulary

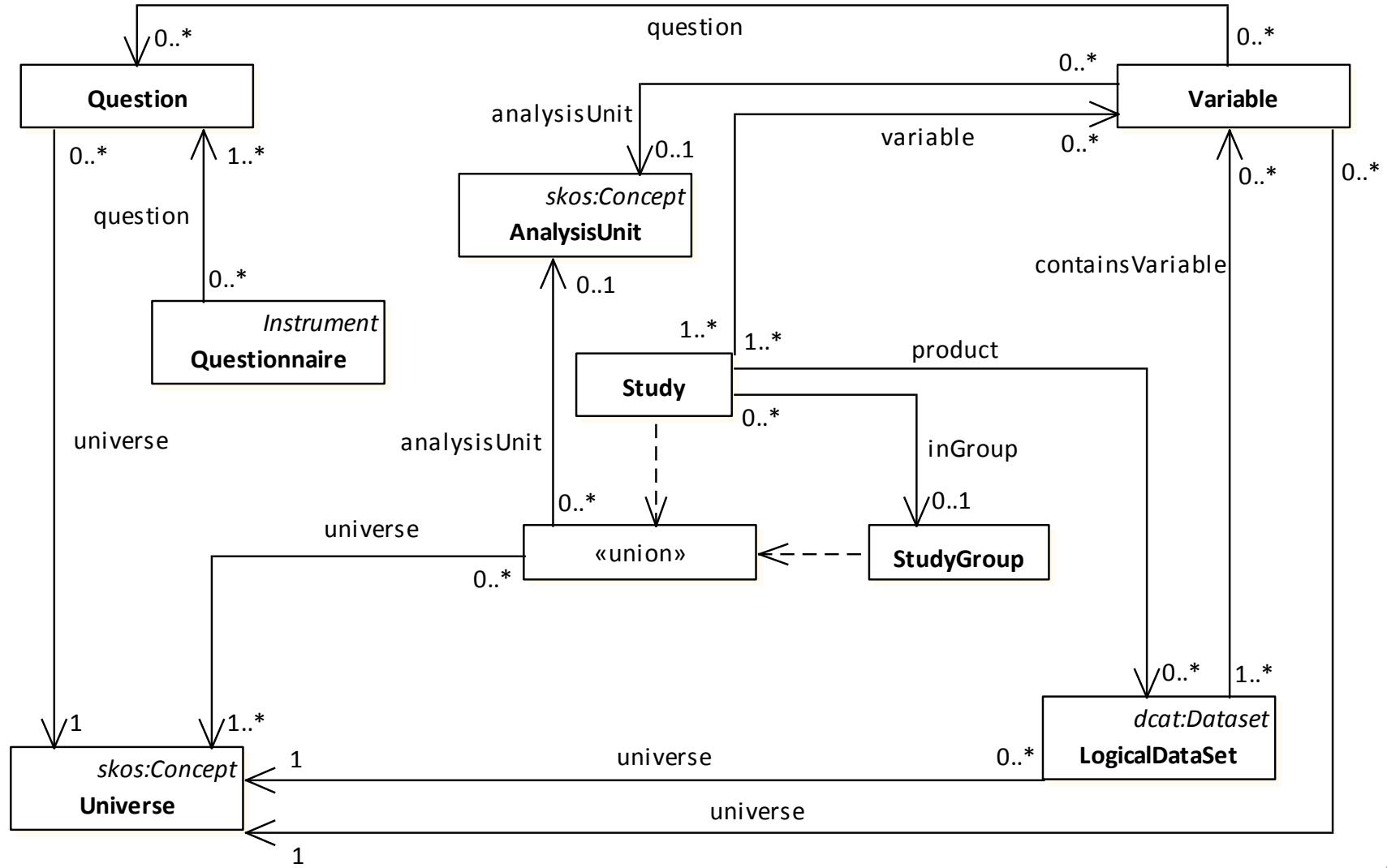
- contains only a small **subset of DDI-XML + additional axioms**
- The conceptual model is derived from **use cases** which are typical in the statistical community
- Statistical domain experts have formulated these use cases which are seen as most significant to solve frequent problems
- enables to
  - **publish**
  - **discover**

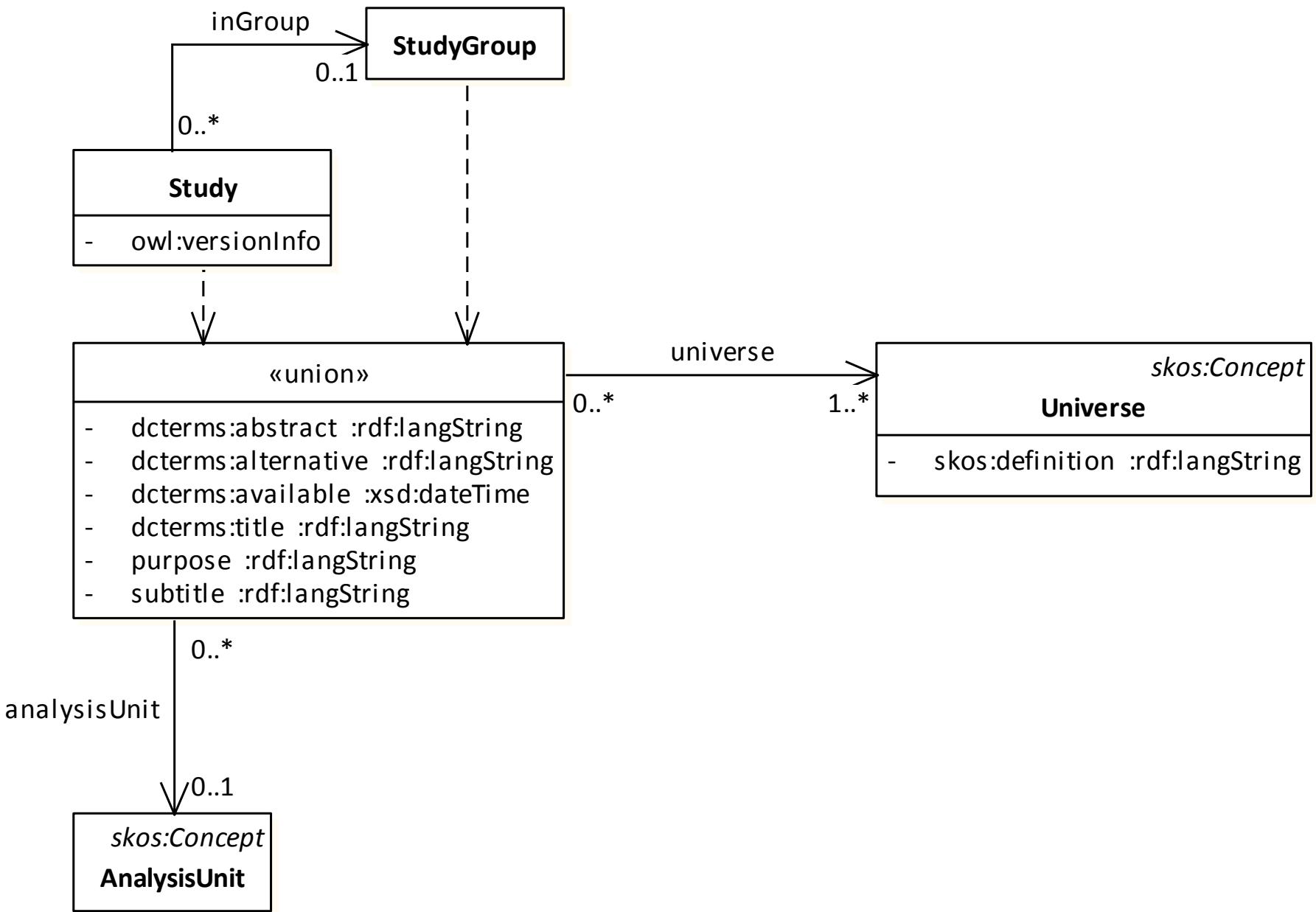
**microdata and metadata** about microdata (research and survey data) in the Web of Linked Data

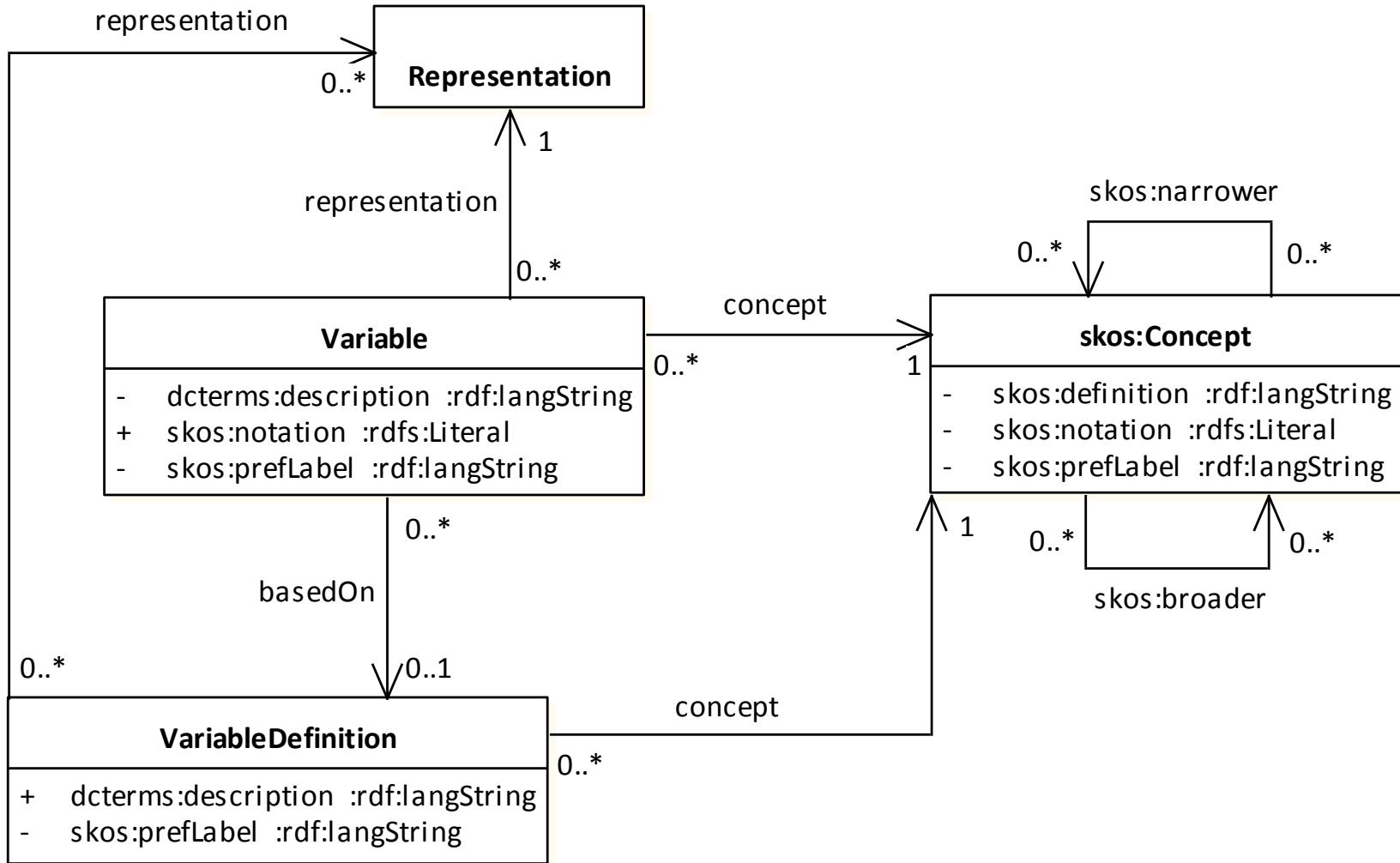
## DDI-RDF Discovery Vocabulary

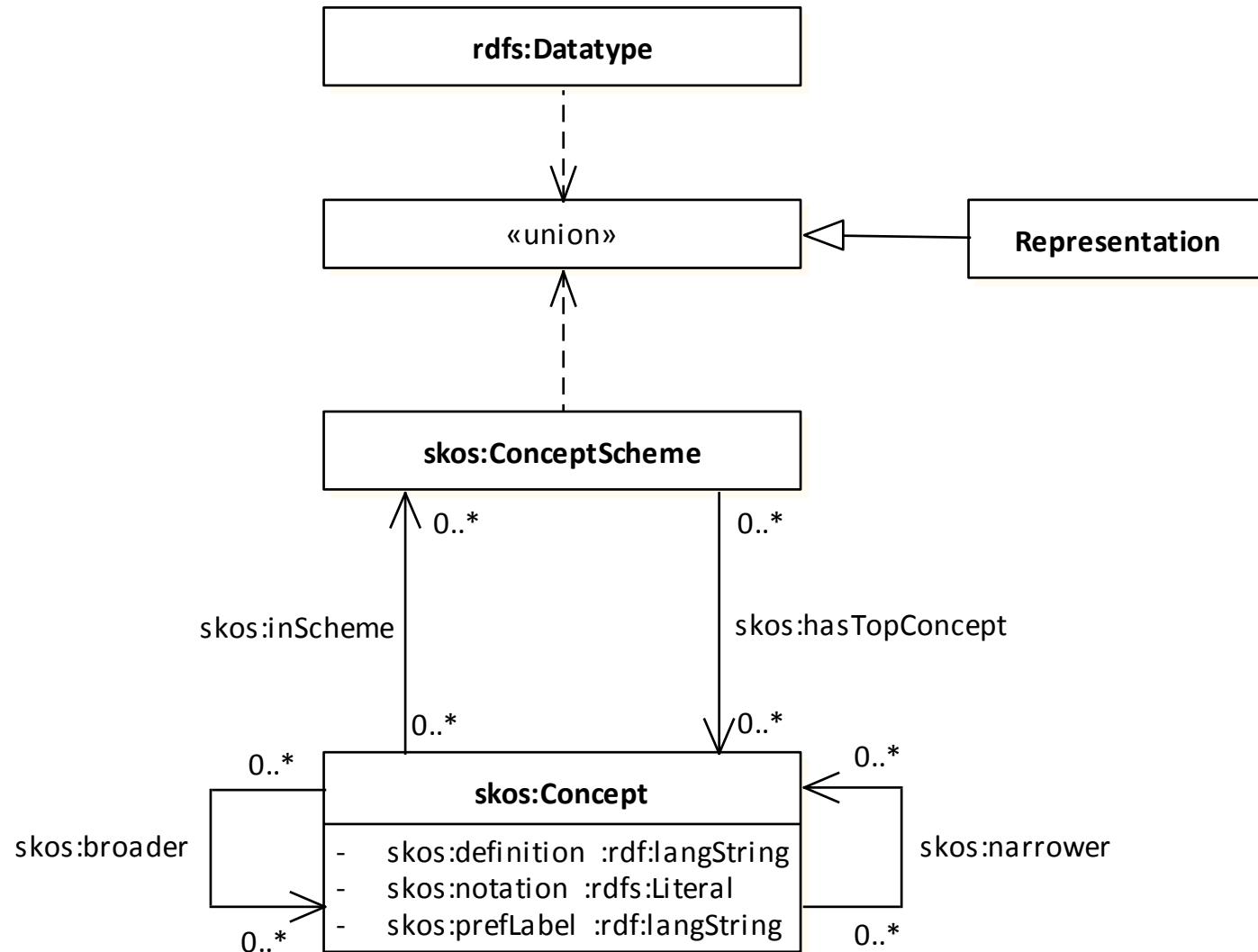
- **Availability of (meta)data**
  - Microdata may be available (typically as CSV files)
  - In most cases, metadata about microdata is NOT available
- contains major types of metadata of DDI-C and DDI-L
- **Mappings** from DDI-XML to DDI-RDF
- **No straightforward Mapping** from DDI-RDF to DDI-XML
  - enables better support for the LD community
  - partly no corresponding constructs in DDI-XML
- 26 **experts from the statistics and the Linked Data community** of 12 different countries have contributed

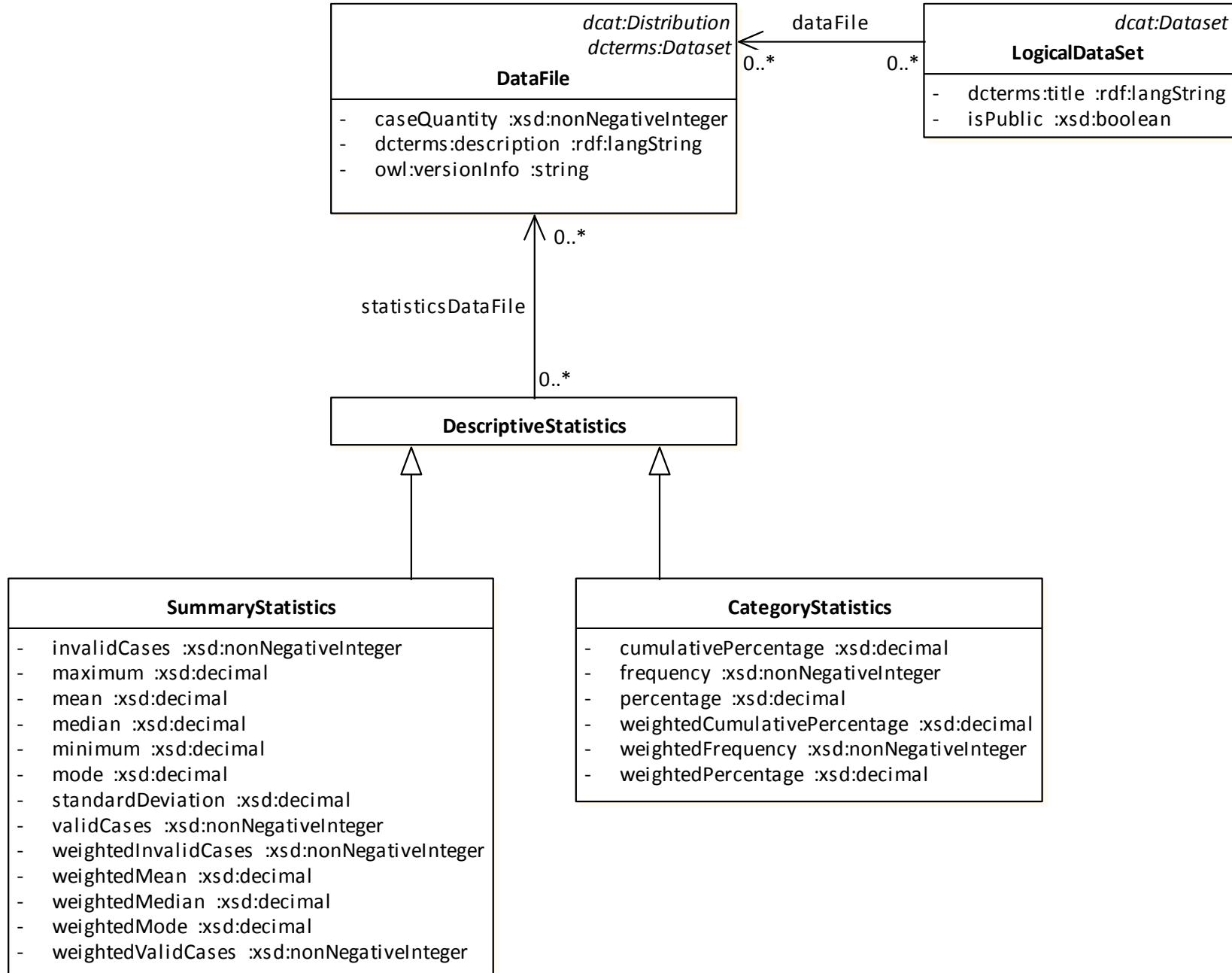
# Conceptual Model

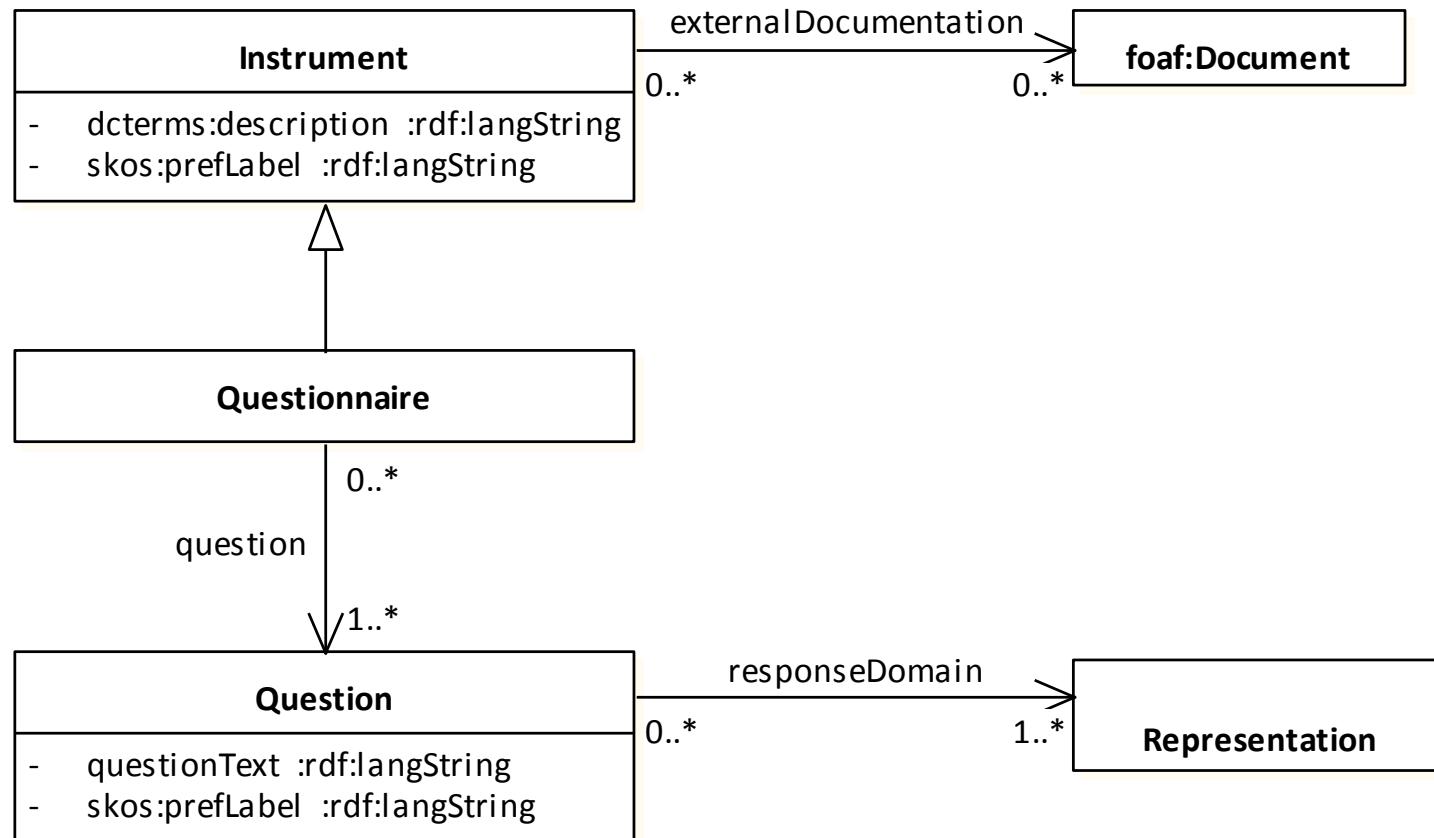












# Thank you for your attention...

- Unofficial draft [planned as specification by DDI Alliance by 2013]  
<http://rdf-vocabulary.ddialliance.org/discovery>
- Specification (current state) on GitHub repository  
<https://github.com/linked-statistics/disco-spec>
- Scenarios for the DDI-RDF Discovery Vocabulary [in preparation]  
<http://dx.doi.org/10.3886/DDISemanticWeb02>

Thomas Bosch  
GESIS - Leibniz Institute for the Social Sciences

thomas.bosch@gesis.org  
boschthomas.blogspot.com  
<https://github.com/boschthomas/PhD>



## Acknowledgements

26 experts from the statistical community and the Linked Data community coming from 12 different countries contributed to this work. They were participating in the events mentioned below.

- 1<sup>st</sup> workshop on 'Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web' at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in September 2011
- Working meeting in the course of the 3rd Annual European DDI Users Group Meeting (EDDI11) in Gothenburg, Sweden in December 2011
- 2<sup>nd</sup> workshop on 'Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web' at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in October 2012
- Working meeting at GESIS - Leibniz Institute for the Social Sciences in Mannheim, Germany in February 2013