



Slideshare

<http://de.slideshare.net/boschthomas>

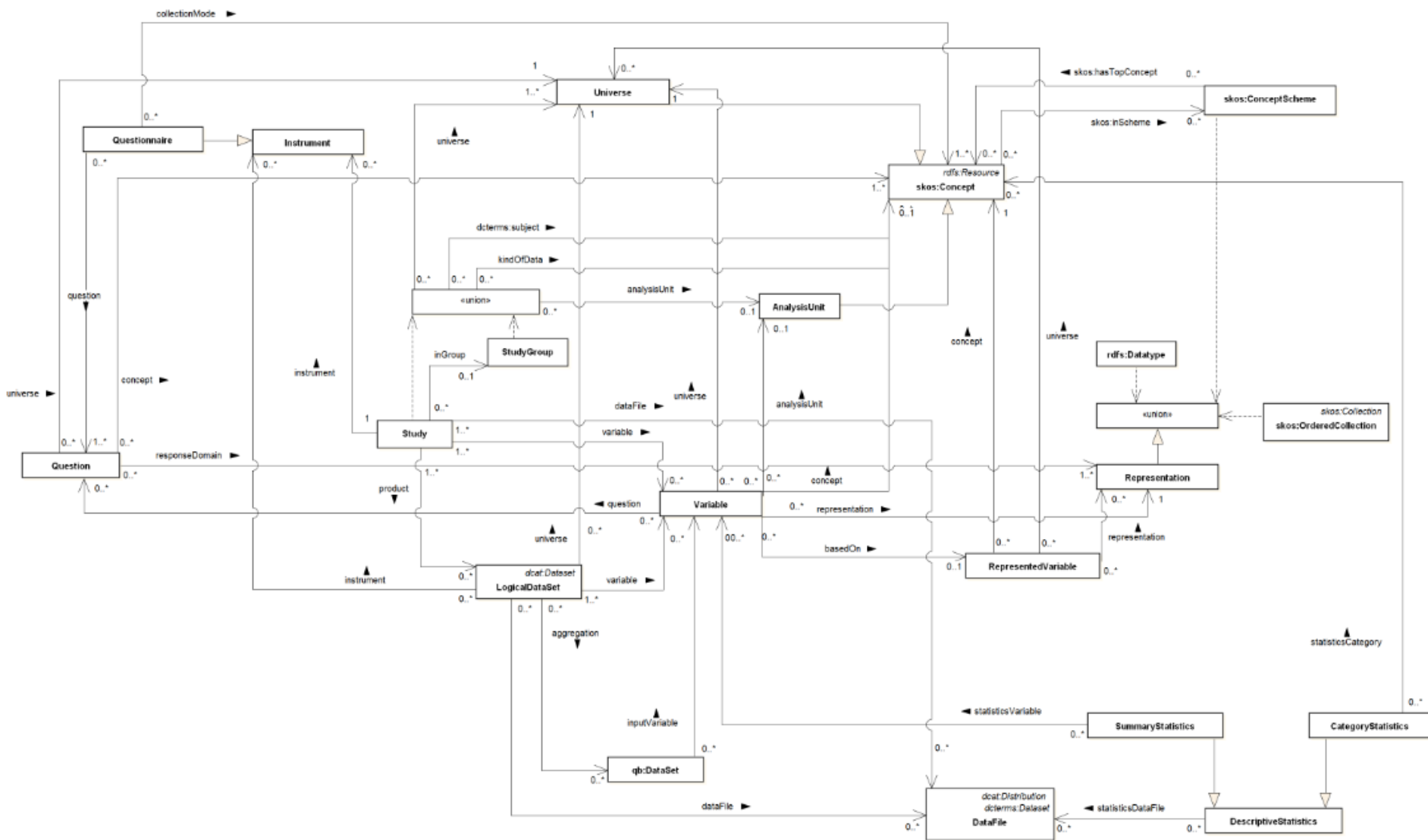
Questions?

Please don't hesitate!



as **big** as

DDI-L?



Disco Spec

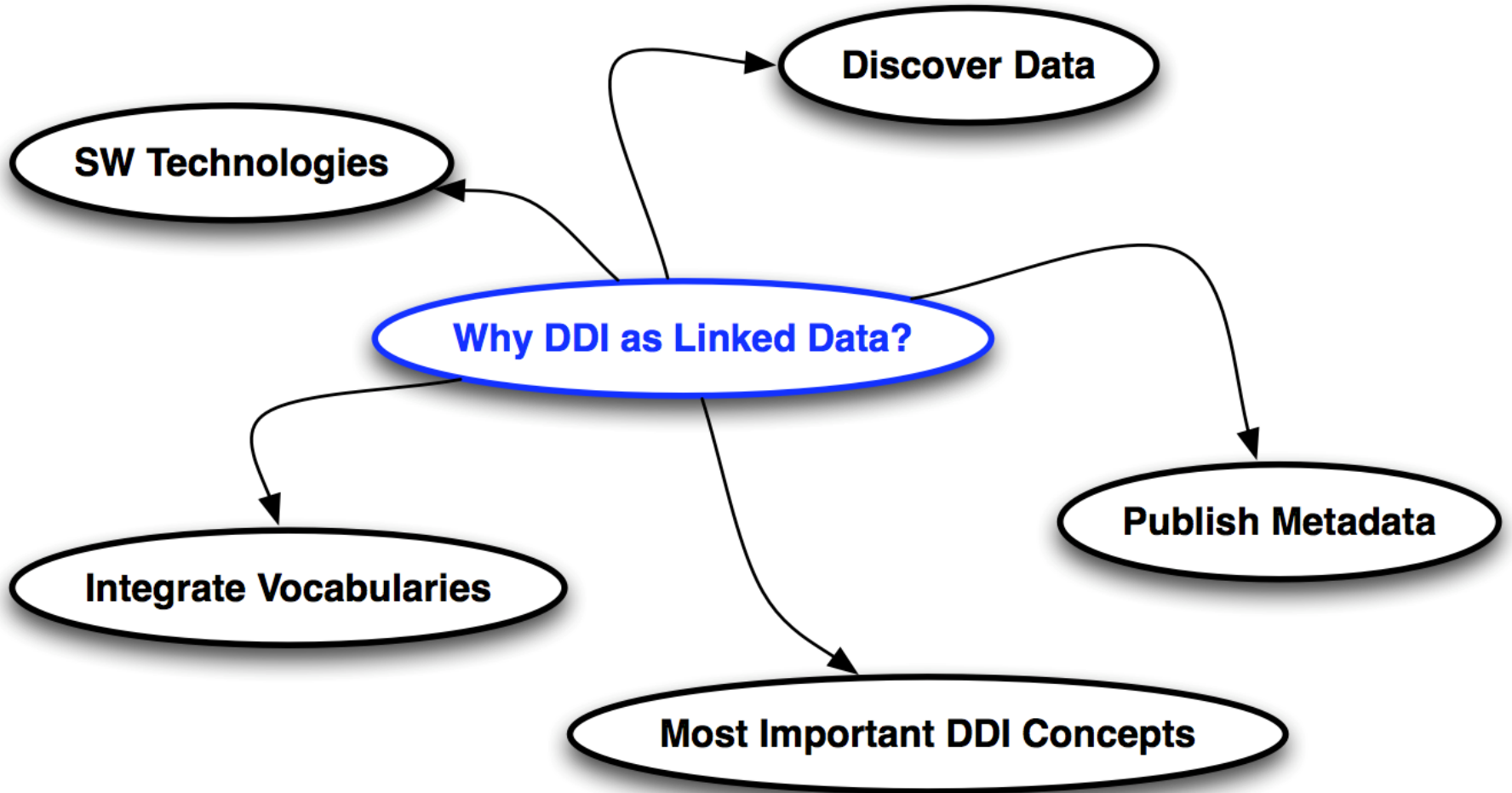
[https://github.com/linked-statistics/
disco-spec/blob/master/discovery.html](https://github.com/linked-statistics/disco-spec/blob/master/discovery.html)

Triple Store

<http://multiweb.gesis.org/openrdf-workbench/repositories/discotest/summary>

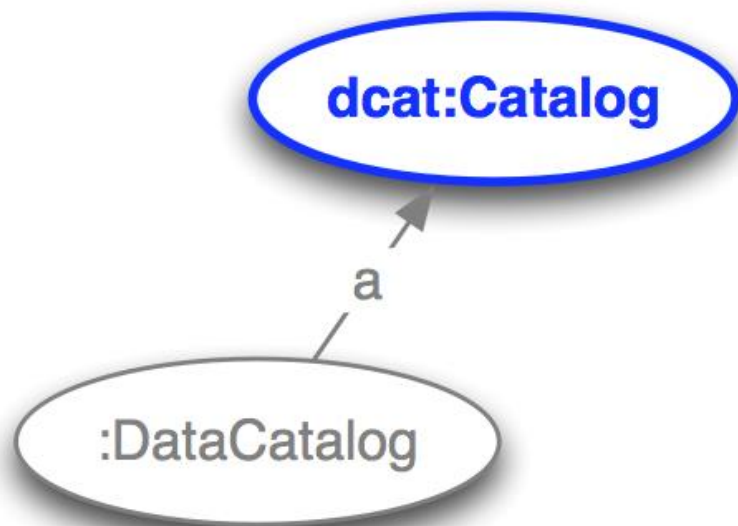
Why Disco?

Why DDI as Linked Data?

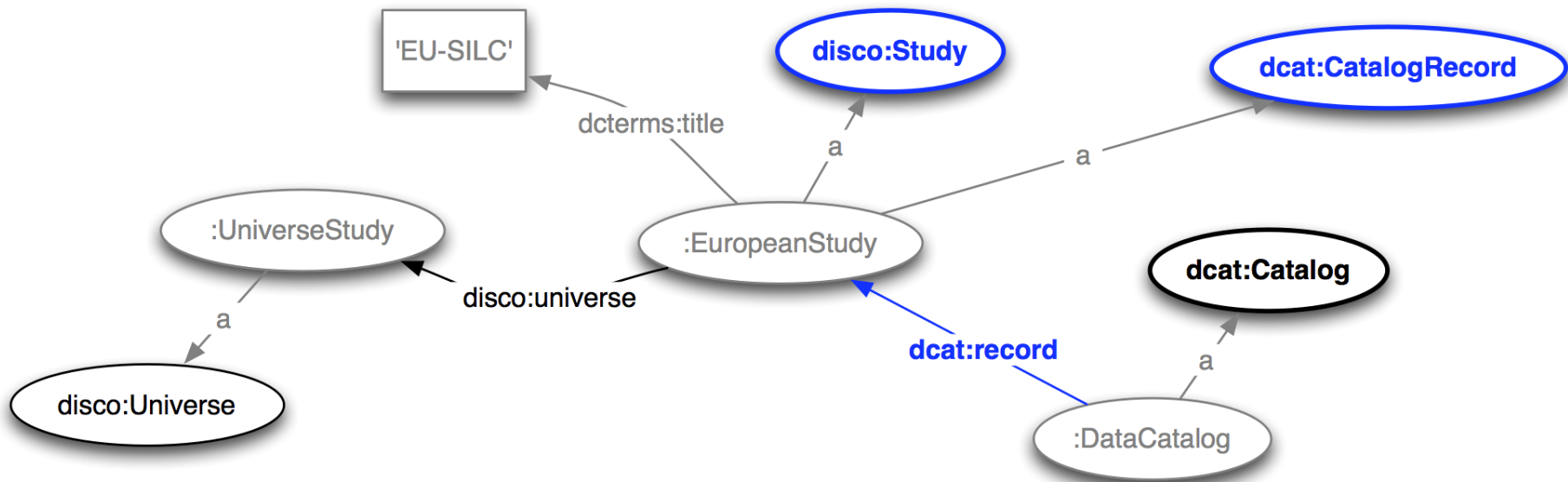


Use Case

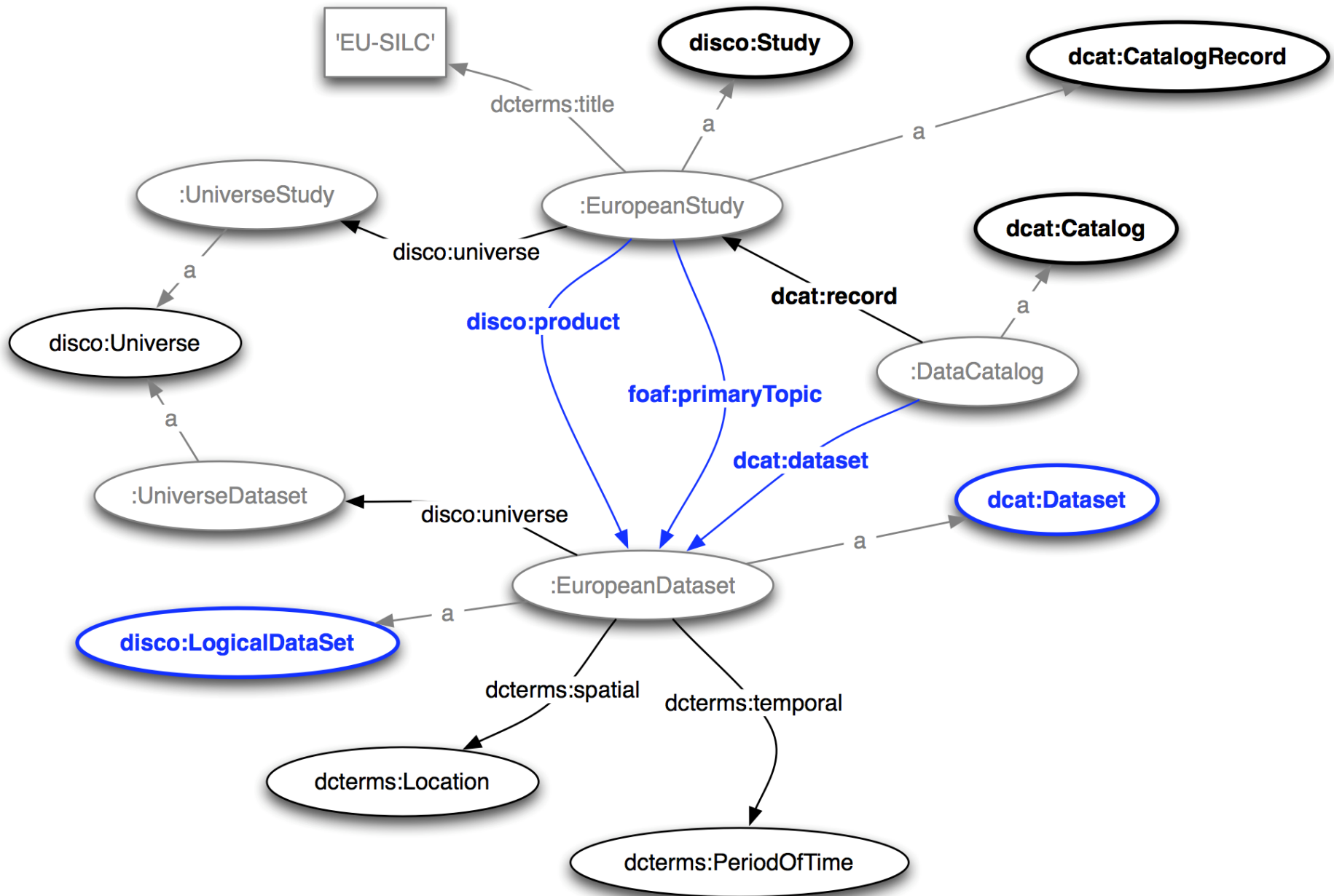
Where
to **search** for
data?



Which
microdata does exist
according to specific
metadata?



Which **datasets**
are **associated** with the
microdata?

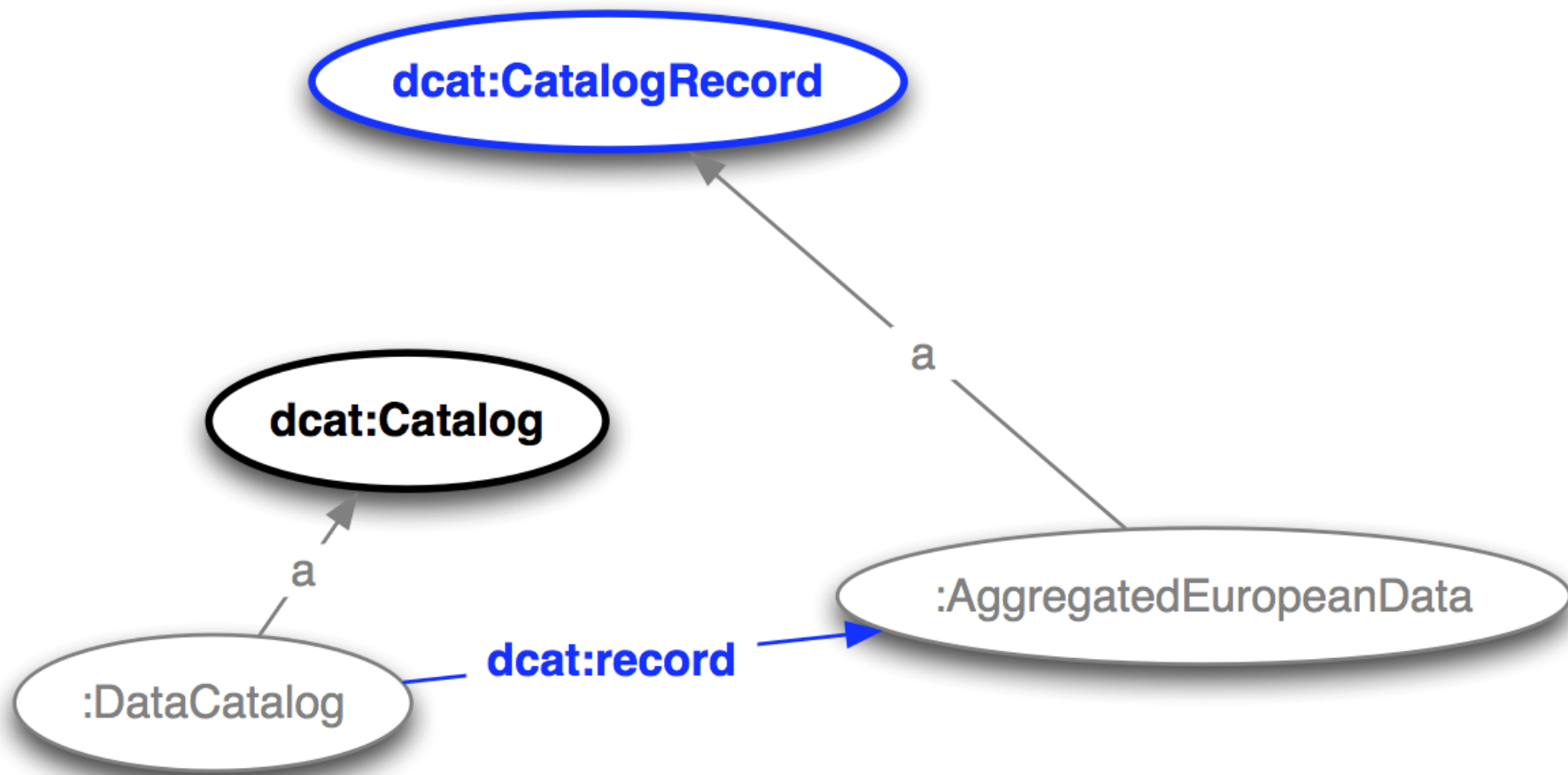


Which

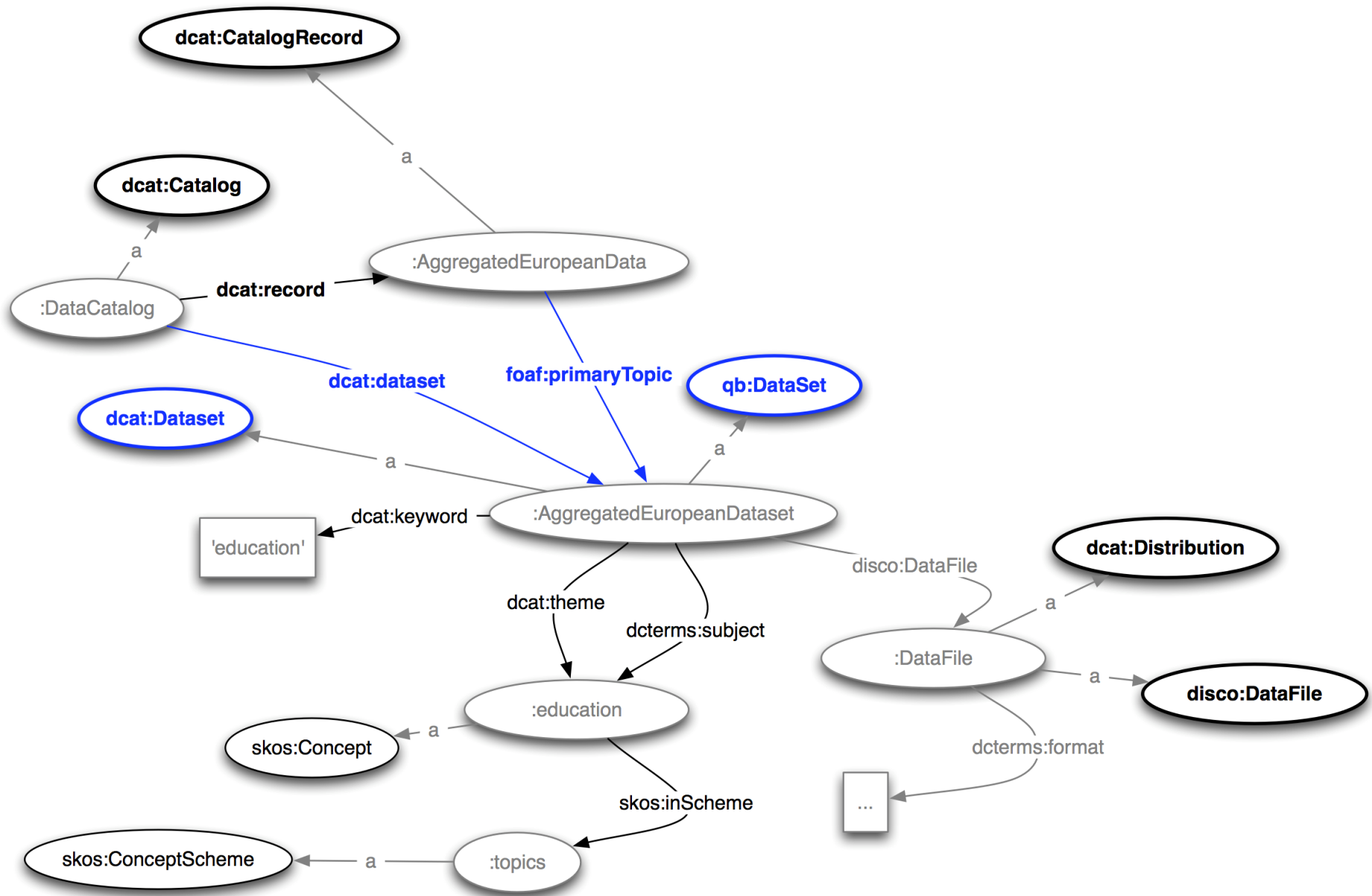
aggregated data

according to specific

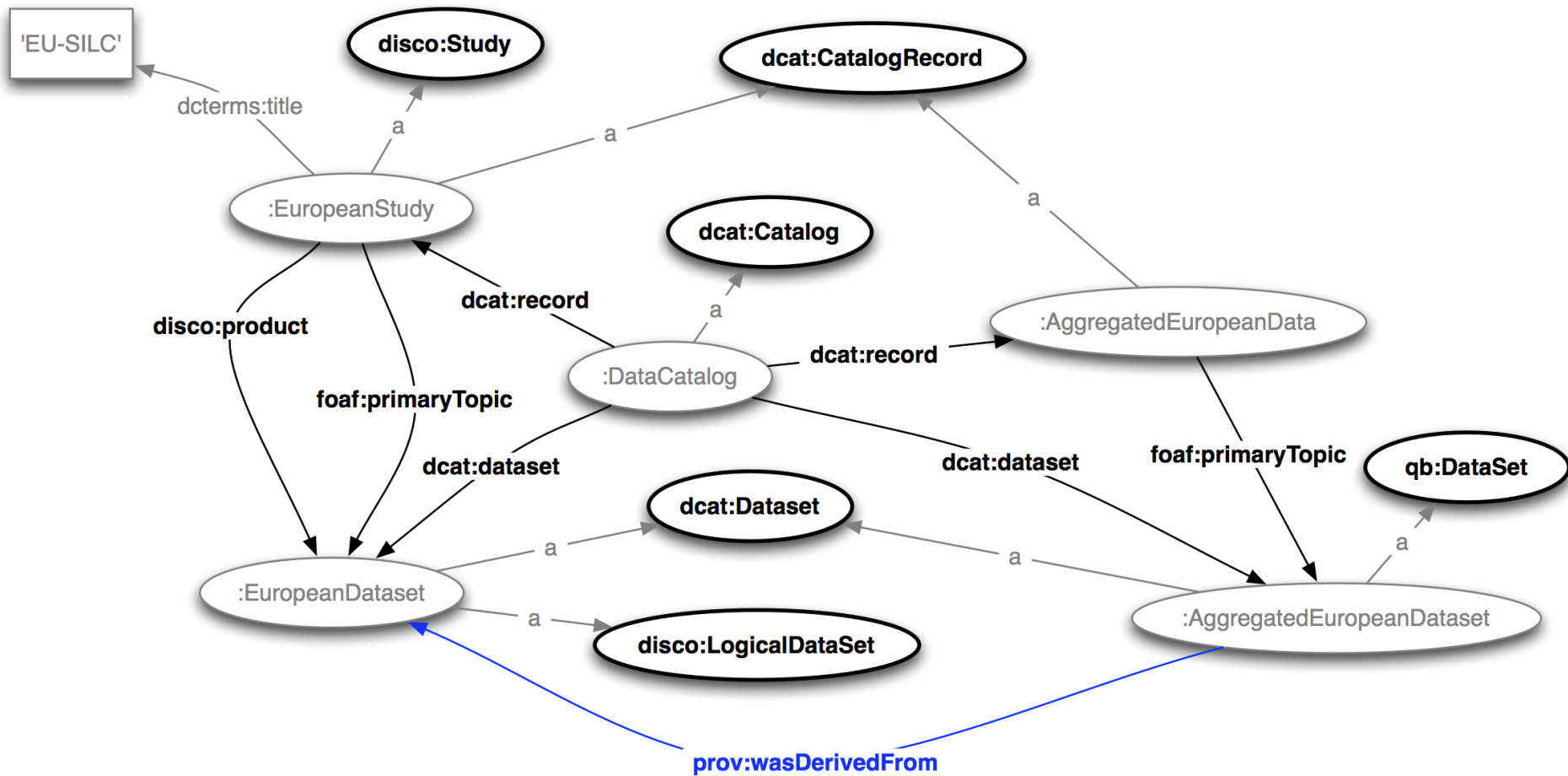
metadata does exist?



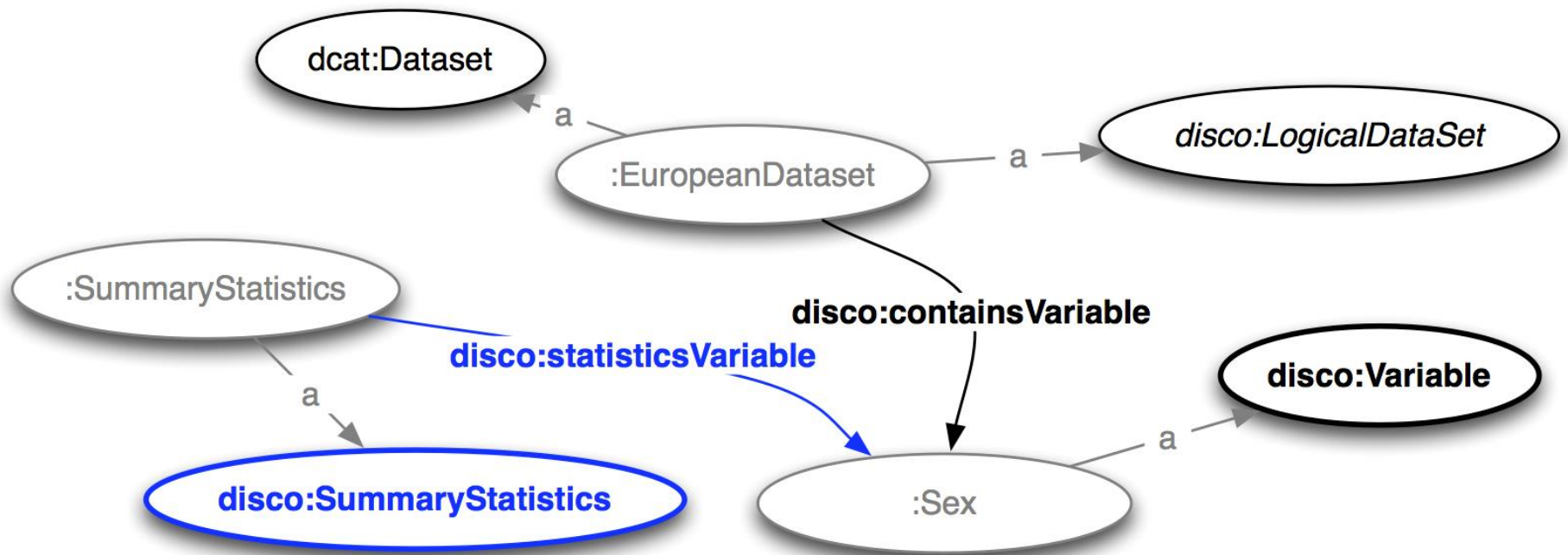
Which **datasets**
are **associated** with
aggregated data?



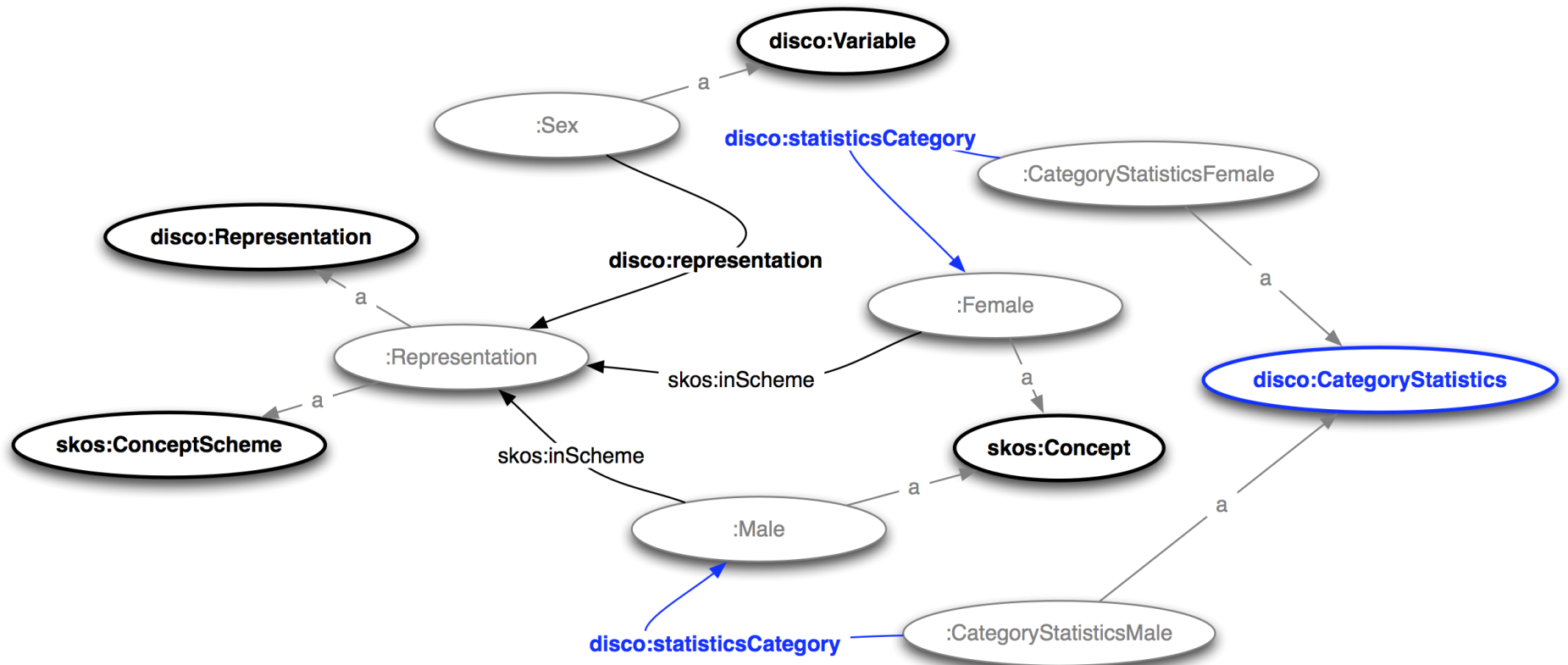
From which
microdata datasets is the
aggregated dataset
derived?



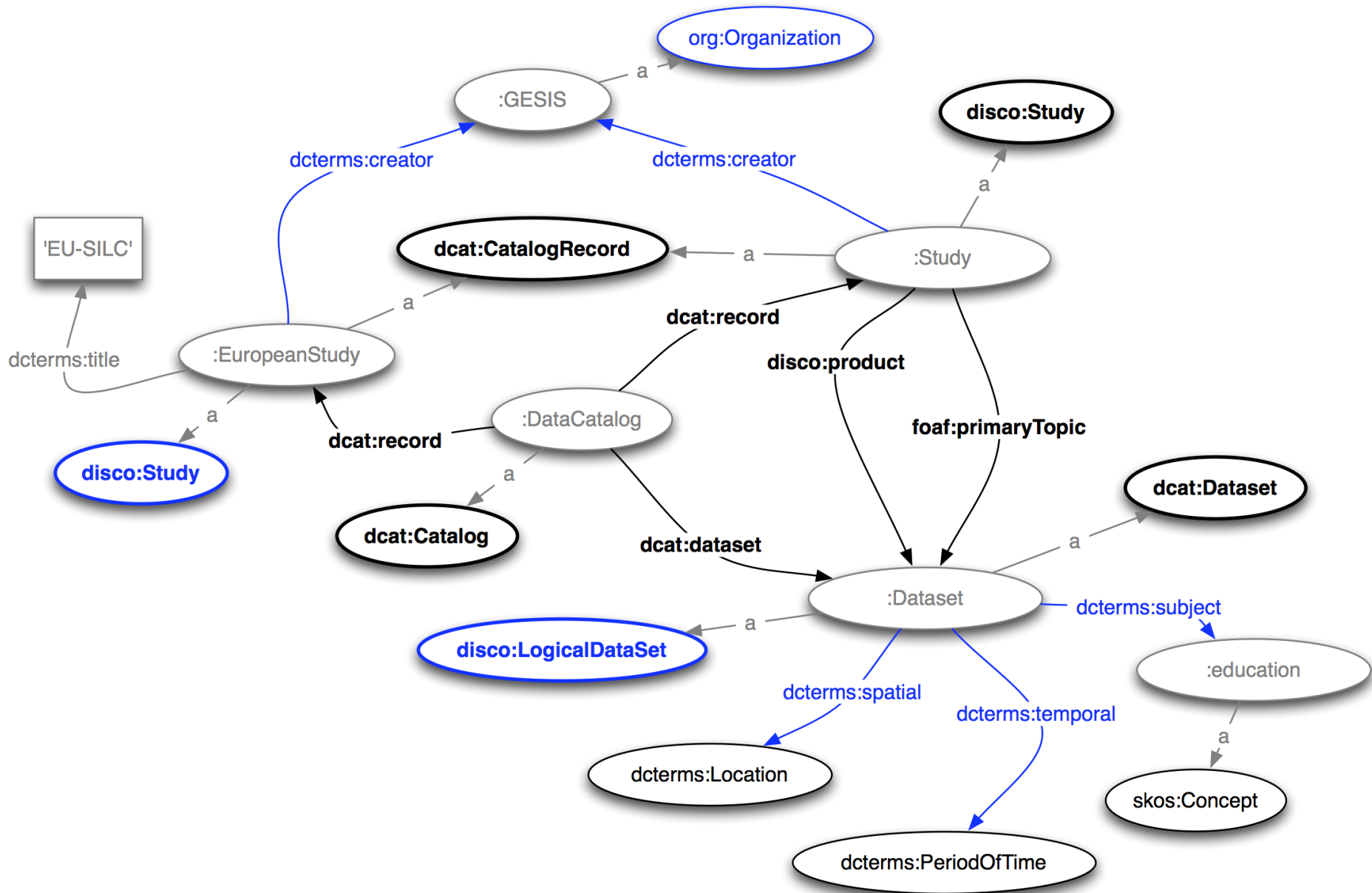
Which
summary
statistics does a
variable have?



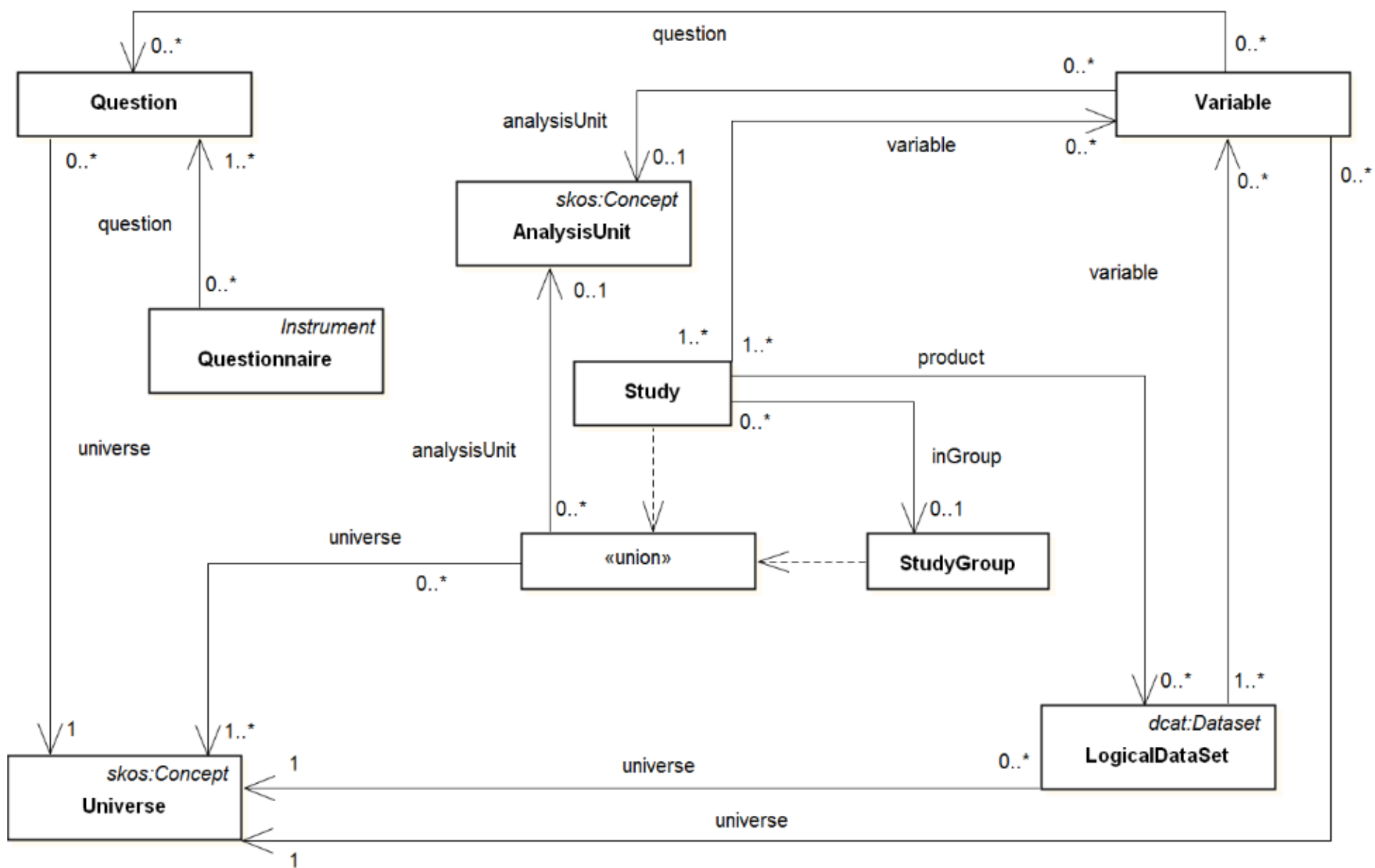
Which
category
statistics does a
variable representation have?



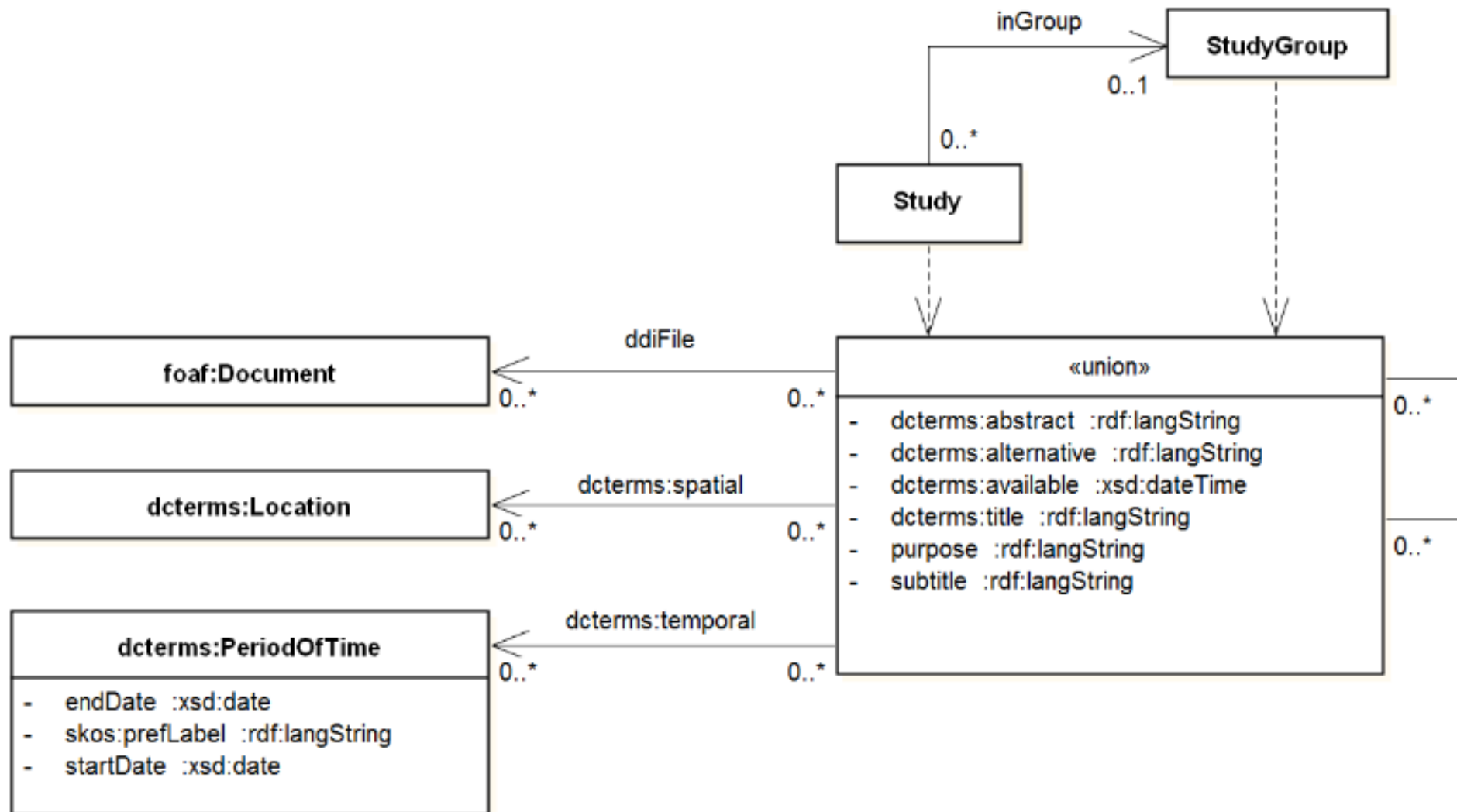
Which **microdata**
datasets are **created**
by the research institute
'GESIS'?

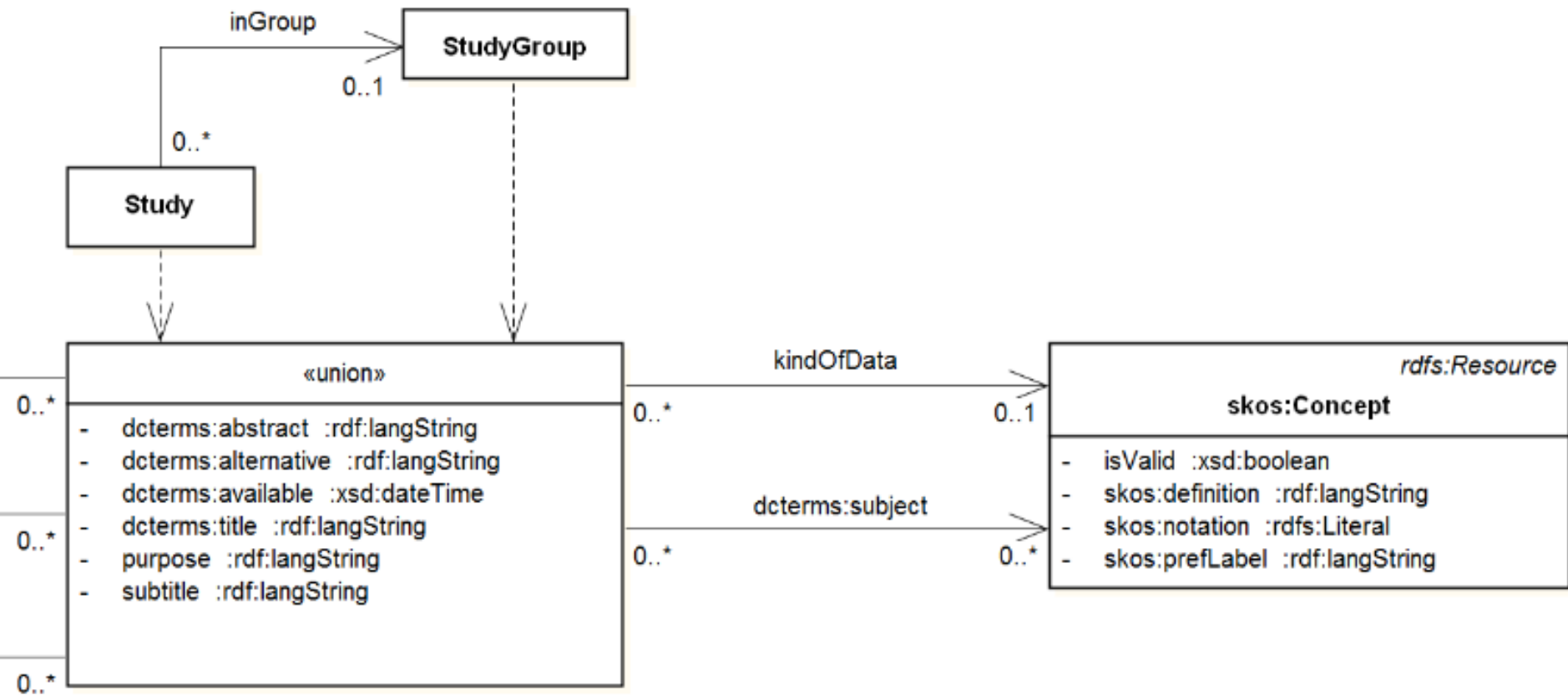


Overview



Series Studies







Series

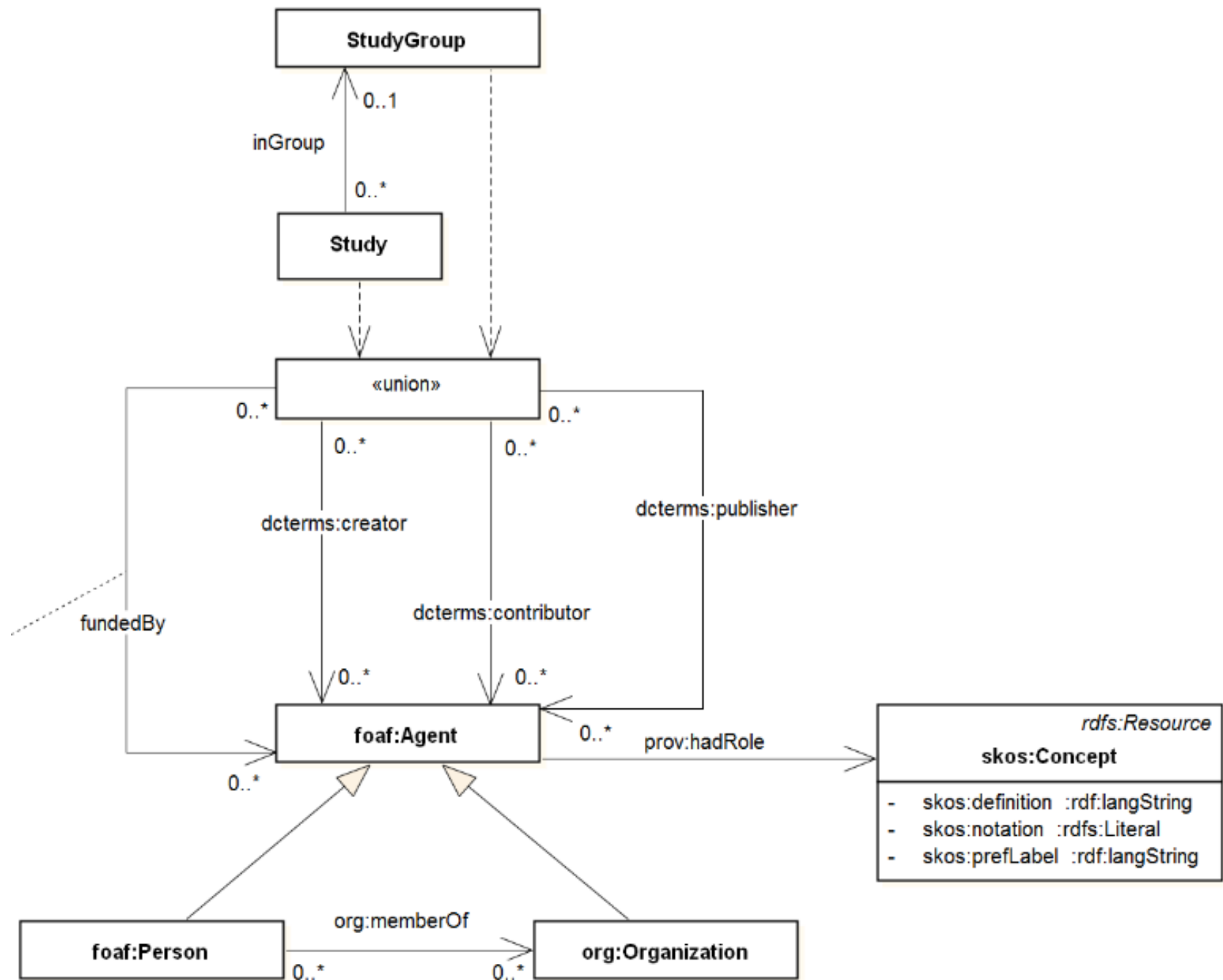
Series title: CIS



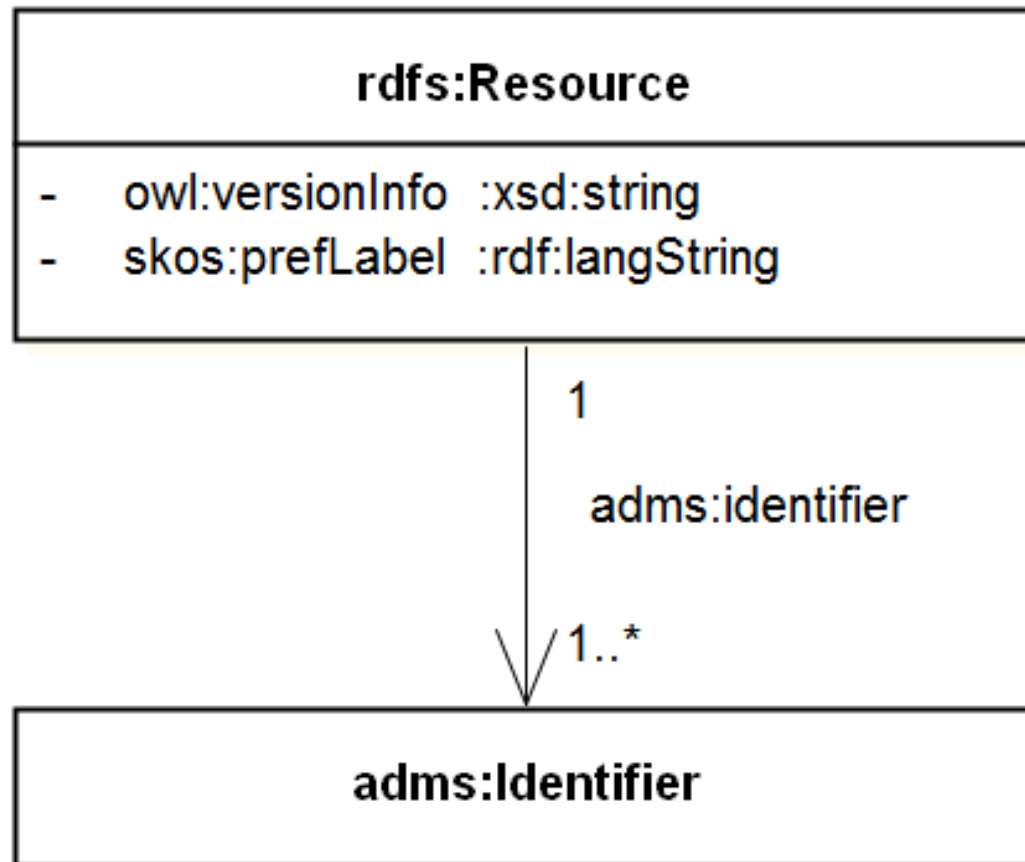
Study

Study title: EU-LFS 1991

Agents



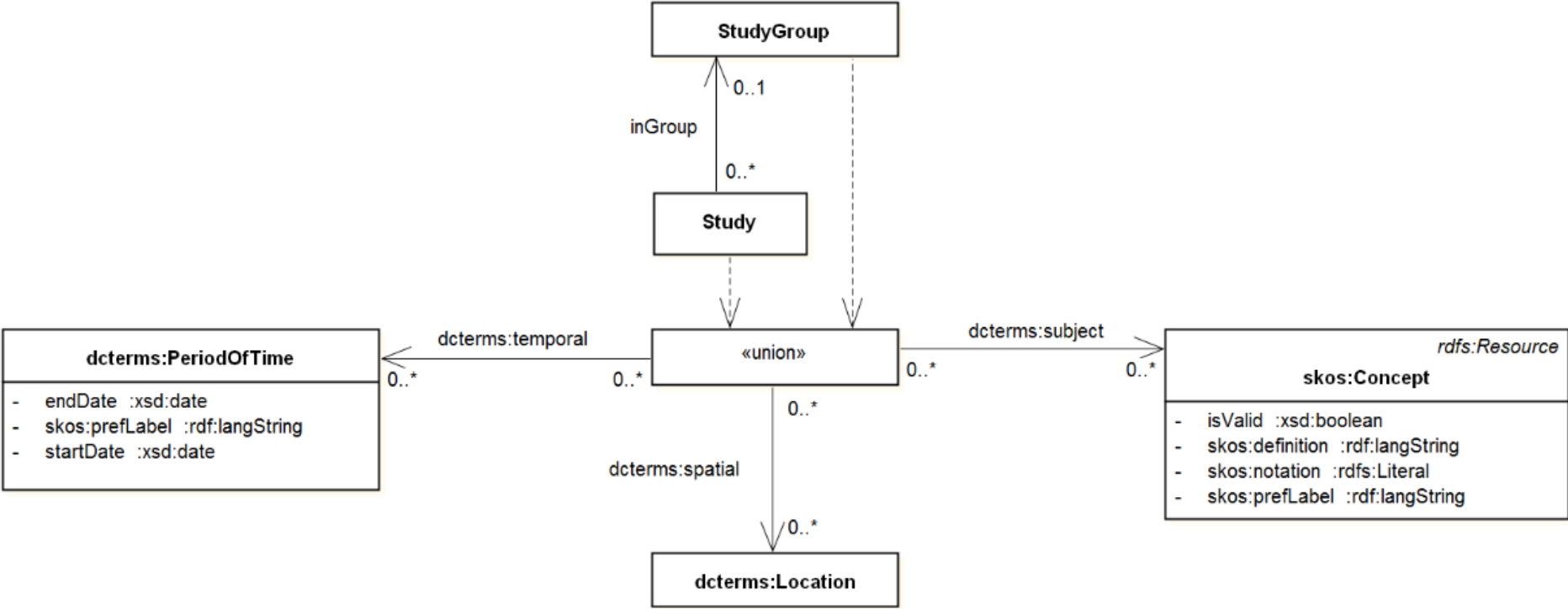
Identification Versioning

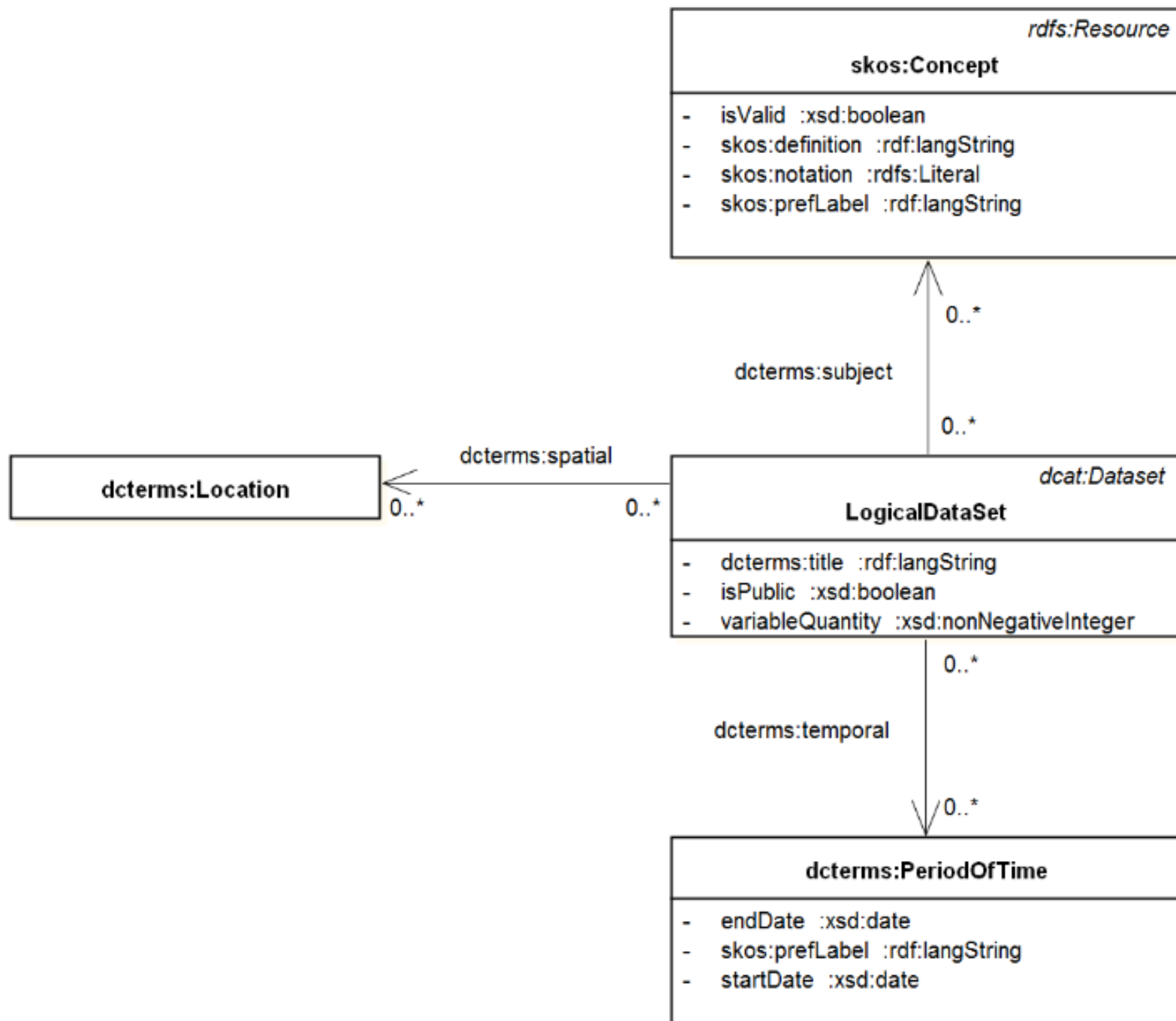


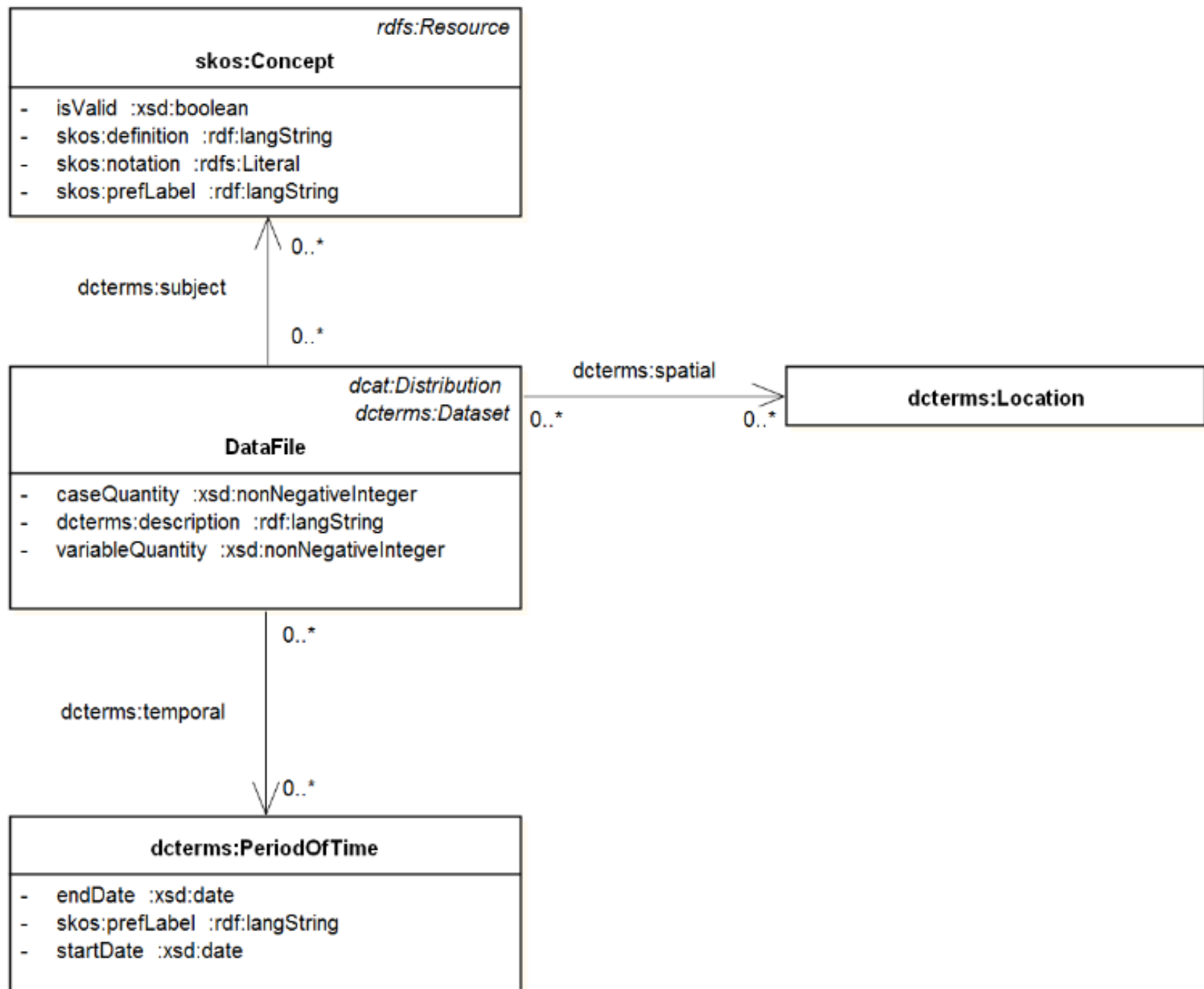
```
ddi:study
  a disco:Study;
  dcterms:title
    "National Population and
    Housing Census, 1980"@en;
adms:identifier [
  a adms:Identifier;
  skos:notation
    "us:ddi:us.mpc:ARG_1980_PHC_v01_A_IPUMS:1";
  adms:schemaAgency "DDI Alliance"@en.
].
```

```
ddi:study
  a disco:Study ;
  dcterms:creator [
    rdfs:label
      "Minnesota Population Center"@en ;
    skos:notation "MPC" ;
    adms:identifier [
      a adms:Identifier ;
      skos:notation "us.mpc" ;
      adms:schemaAgency
        "DDI Alliance"@en ] ] .
```

Coverage







Spatial Coverage


```
<urn:ddi:de.gesis:study_EU-SILC-2005:0.1>
  a disco:Study ;
  dcterms:title
    "EU-SILC 2005"@en ;
  skos:prefLabel "2005"@en ;
dcterms:spatial
  <http://sws.geonames.org/2782113> ,
  ... ,
  :AllCountriesOfStudy ;
```

```
:AllCountriesOfStudy  
  a dct:Location ,  
    missy:Country;  
  rdfs:label  
    "all countries of study";  
  missy:code "" .
```



Countries

Study: EU-LFS 2004

Temporal Coverage

```
<urn:ddi:de.gesis:study_EU-SILC-2005:0.1>  
  a disco:Study ;  
  dcterms:title "EU-SILC 2005"@en ;  
  skos:prefLabel "2005"@en ;  
  dcterms:temporal  
    <urn:ddi:de.gesis:0ba9b4f3-ec22-  
4471-8ffa-a38e8ada187a:0.1> ;
```

<urn:ddi:de.gesis:0ba9b4f3-ec22-4471-
8ffa-a38e8ada187a:0.1>

a **disco:PeriodOfTime** ;

dcterms:date

"Jan 1, 2005 12:00:00 AM"^^xsd:date .

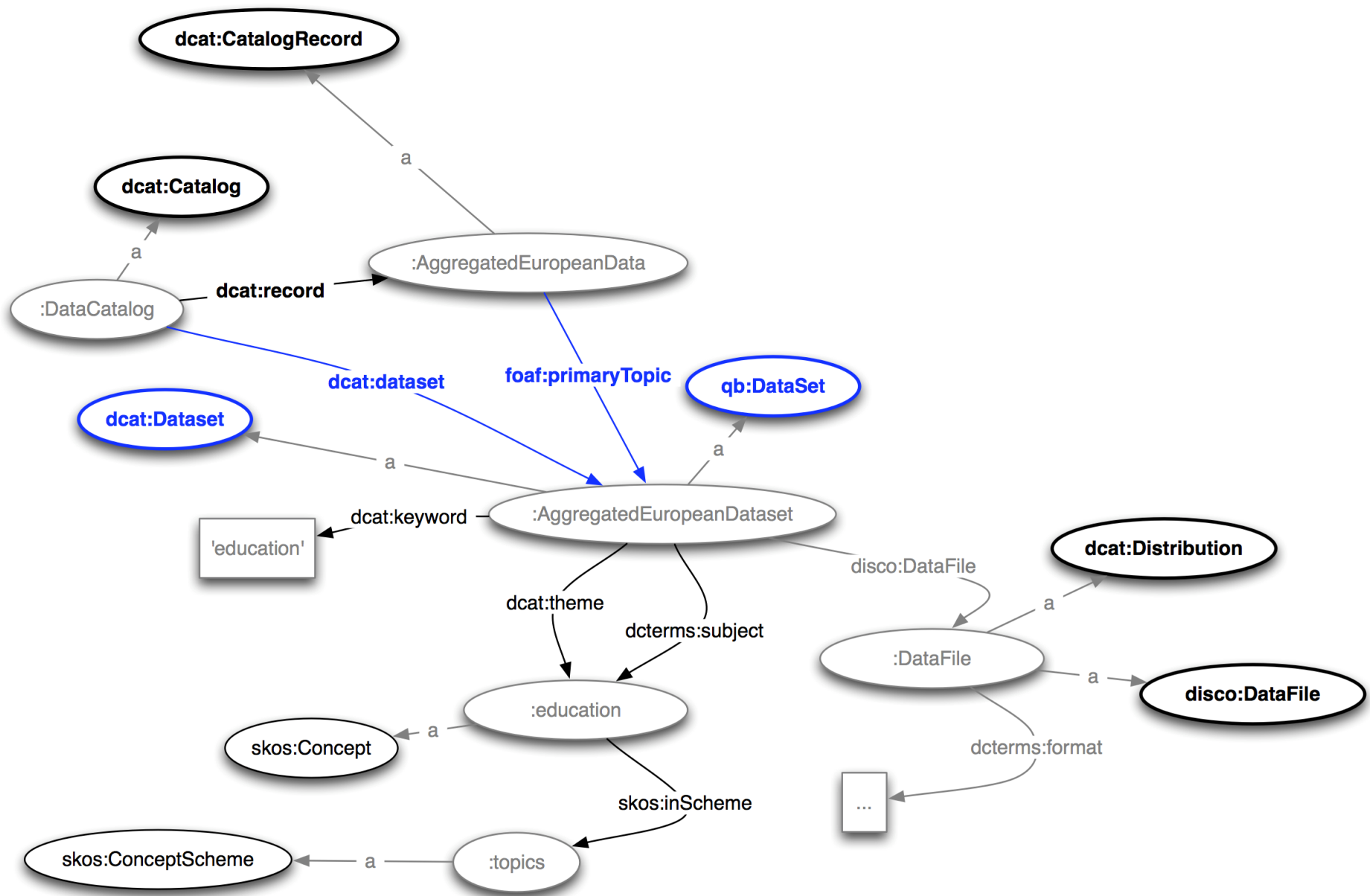


Year

Study title:

'Structure of Earnings Survey – 2006'

Topical Coverage



missy:PB100

a disco:Variable

dterms:subject [

a skos:Concept ;

skos:notation

"Quarter of the personal
interview"@en] .



Topical Coverage

Variable ID:

<urn:ddi:de.gesis:variable_EU-SILC-2010-panel-p-data-2010_rev2-PB020:0.1>

Variable name:

'PB020'

Thematic Classification

Thematische Gliederung

- + Demographie und Bevölkerung
- + Nationalität und Migration
- **Arbeitsmarkt und Erwerbsbeteiligung**
- + Unterhalt und Einkommen
- + Sozialversicherung und Vorsorge
- + Bildung und Qualifikation
- + Pendler
- + Privathaushalt und Familie
- + Wohnverhältnisse
- + Gesundheit
- + Stichprobe

[Gesamtübersicht / Arbeitsmarkt und Erwerbsbeteiligung](#) >> [Arbeitslosigkeit](#) >> [Meldung bei der Agentur für Arbeit: Erwerbstätige](#)

+ Erwerbsstatus

+ Erwerbsbeteiligung

- Arbeitslosigkeit

- Meldung beim Arbeitsamt

- Meldung bei der Agentur für Arbeit

- **Meldung bei der Agentur für Arbeit: Erwerbstätige**

- Meldung bei der Agentur für Arbeit: Nichterwerbstätige

- Arbeitslosigkeit

- Arbeitslosengeld oder -hilfe

+ Arbeitsuche

+ Vereinbarkeit von Beruf und Familie

+ Übergang vom Erwerbsleben in den Ruhestand

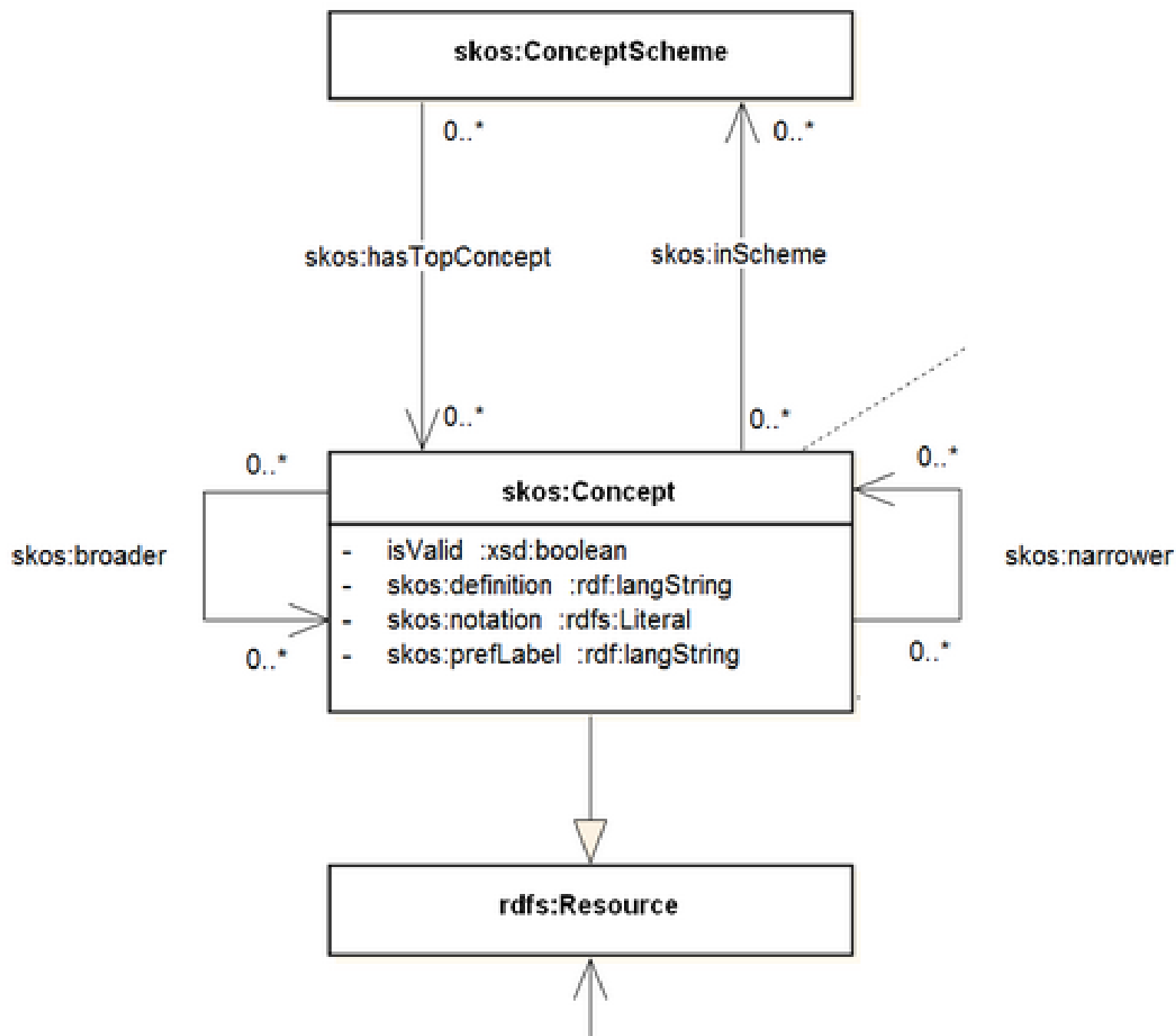


Veröffentlichungen zu diesem Thema

Mikrozensus: Erhebungszeitpunkte [Meldung bei der Agentur für Arbeit: Erwerbstätige](#)

2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996
EF225	EF225	EF225	EF225	EF225	EF225	EF225	EF214	EF214	EF214	EF214	EF214	EF214	EF214	EF214	EF214





Series-Level

```
:thematicClassification  
  a skos:ConceptScheme ;  
  skos:hasTopConcept  
    :concept1 ,  
    :concept2 ,  
    :concept3 .
```

Narrower Concepts

```
:superConcept
  a skos:Concept ;
  skos:notation
    "Demographic background"@en ;
skos:narrower
  :subConcept1 ,
  :subConcept2 .
```




Direct Broader Concepts

Concept:

'Country'@en



All (Direct + Indirect) Broader Concepts

Concept:

'Country'@en



Direct Narrower Concepts

Concept: "Type of cooperation"@en



All (Direct + Indirect) Narrower Concepts

Concept: "Type of cooperation"@en



Top Concepts

Series: EU-SILC

Thematic Classification:

<urn:ddi:de.gesis:7bb54a91-4b26-4f6e-a1b7-be48cb58be24:0.1>



2-Level Concepts

Series: EU-SILC

Thematic Classification:

<urn:ddi:de.gesis:7bb54a91-4b26-4f6e-
a1b7-be48cb58be24:0.1>



Lowest-Level Concepts (Leaf Concepts)

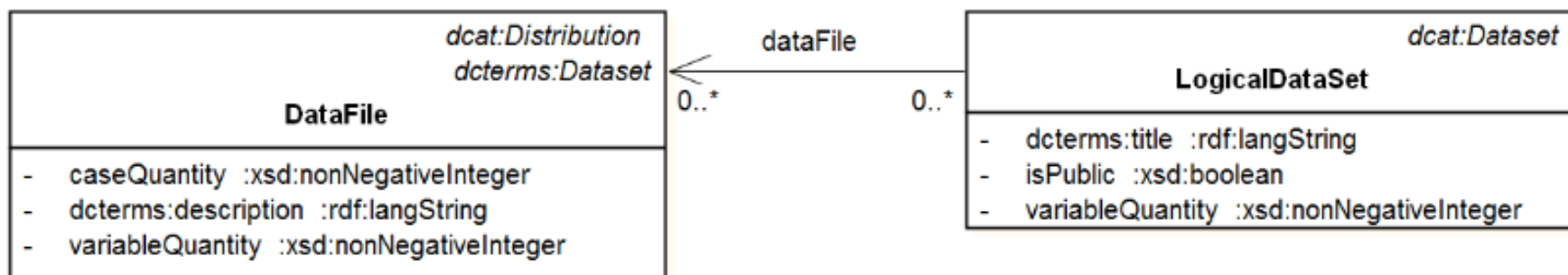
Series: EU-SILC

Thematic Classification:

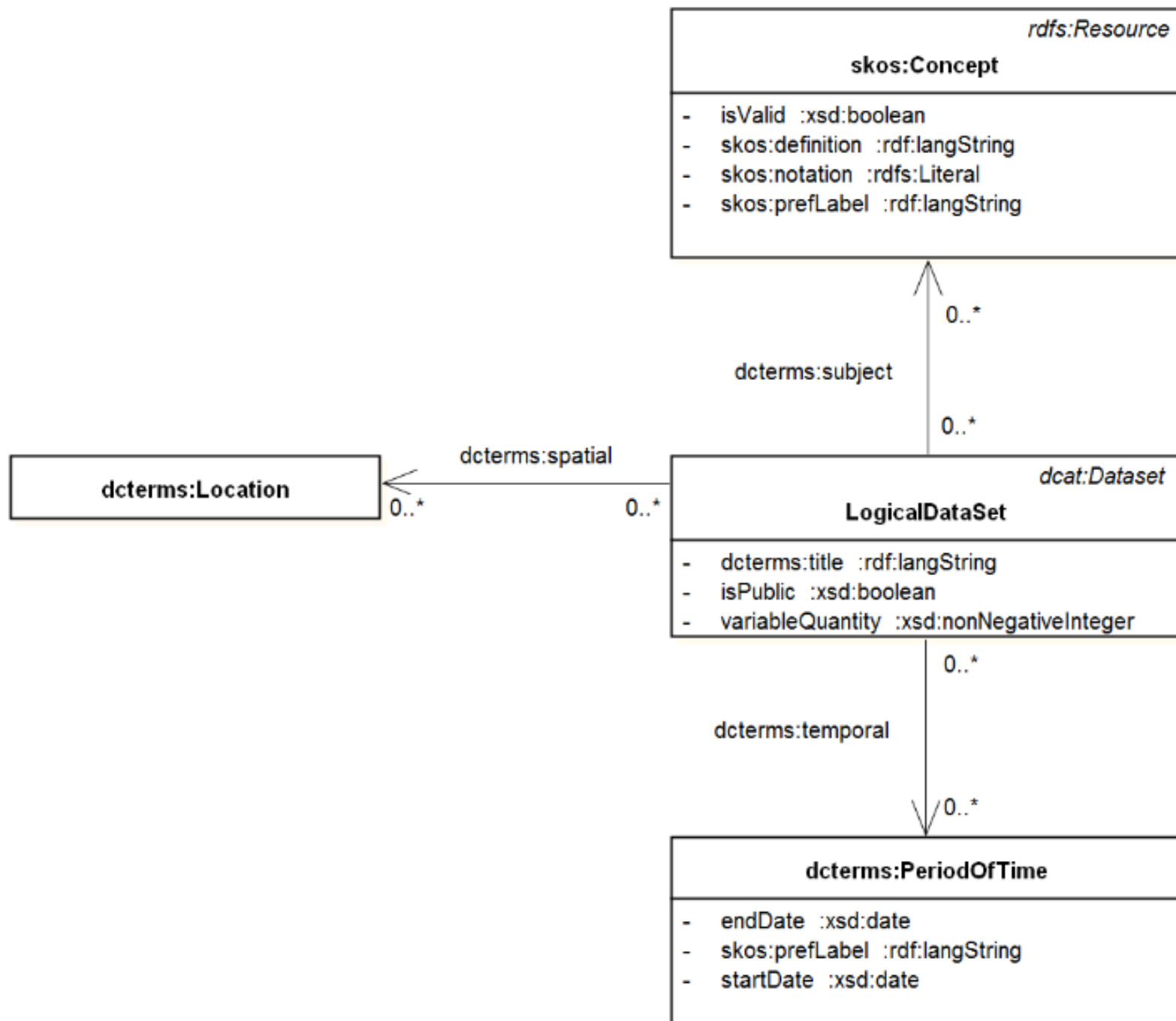
<urn:ddi:de.gesis:7bb54a91-4b26-4f6e-
a1b7-be48cb58be24:0.1>

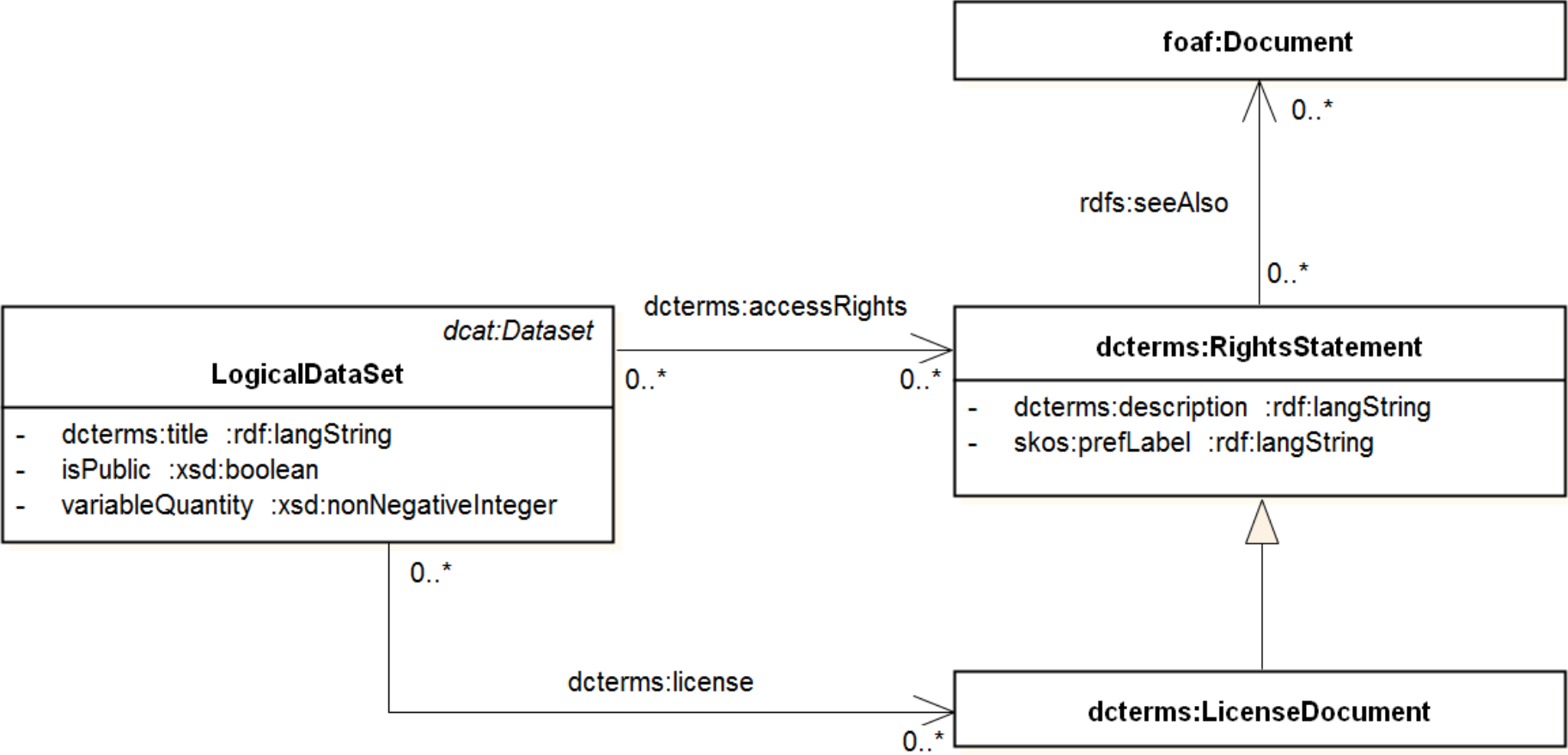
Data Sets

Data Files



Data Sets



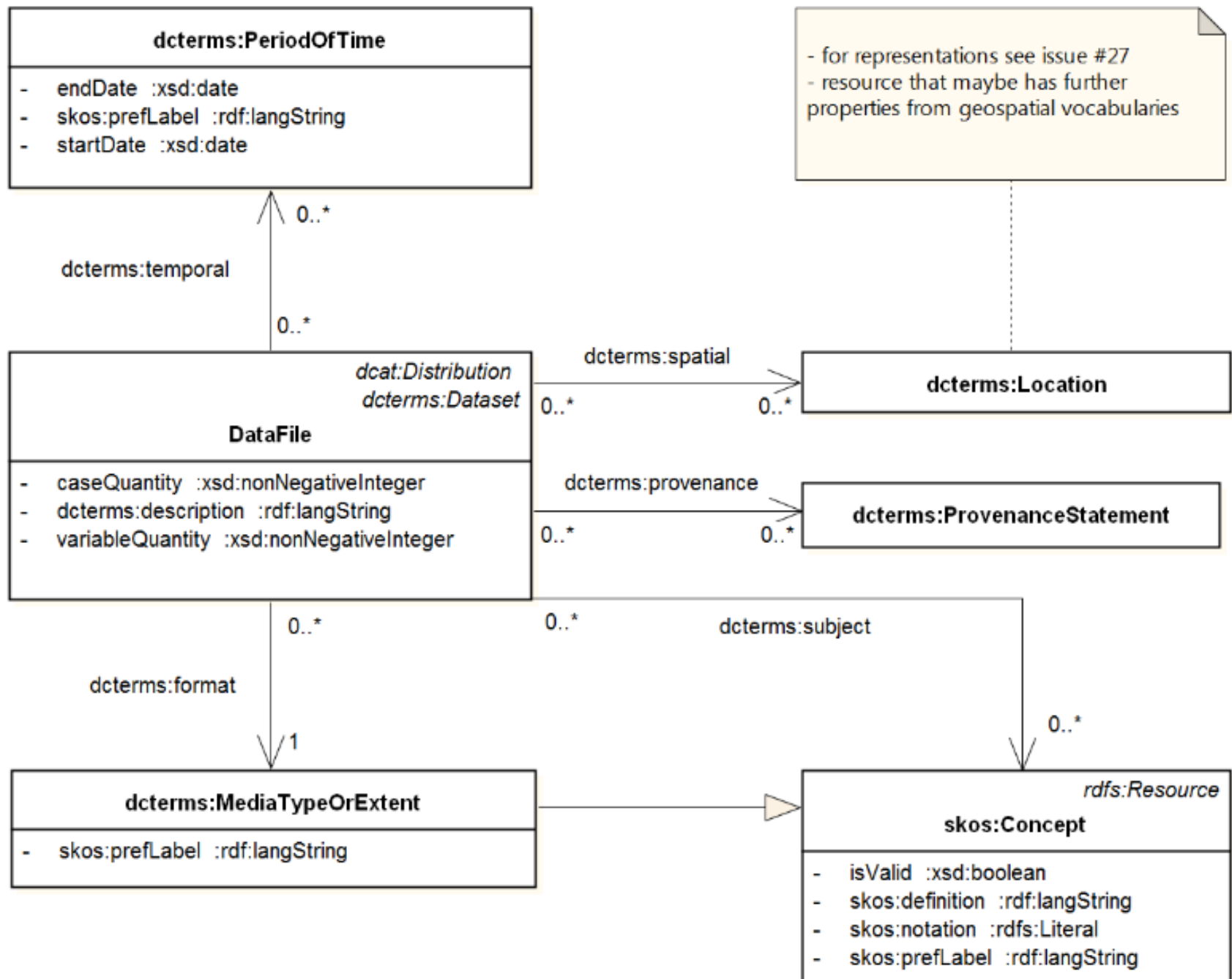




All Data Sets (IDs)

Study: EU-SILC 2010

Data Files



```
:dataFile
  a disco:Datafile ;
  dcterms:identifier
    "ARG1900-P-H.dat" ;
  dcterms:description
    "Person records"@en ;
  disco:caseQuantity 2667714 ;
  dcterms:format "ascii" ;
  dcterms:provenance
    "Minnesota Population Center"@en ;
  owl:versionInfo
    "Version 1.0, IPUMS sample"@en .
```


:dataFile

a disco:Datafile ;

dcterms:spatial [

a dcterms:Location ;

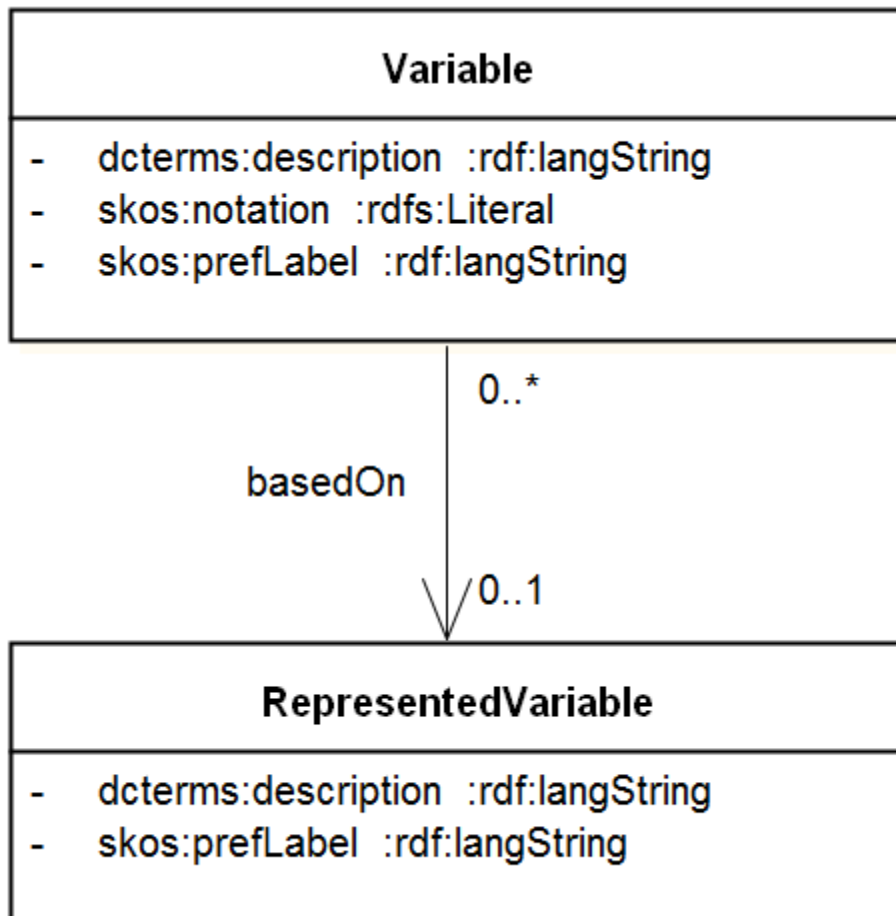
rdfs:label

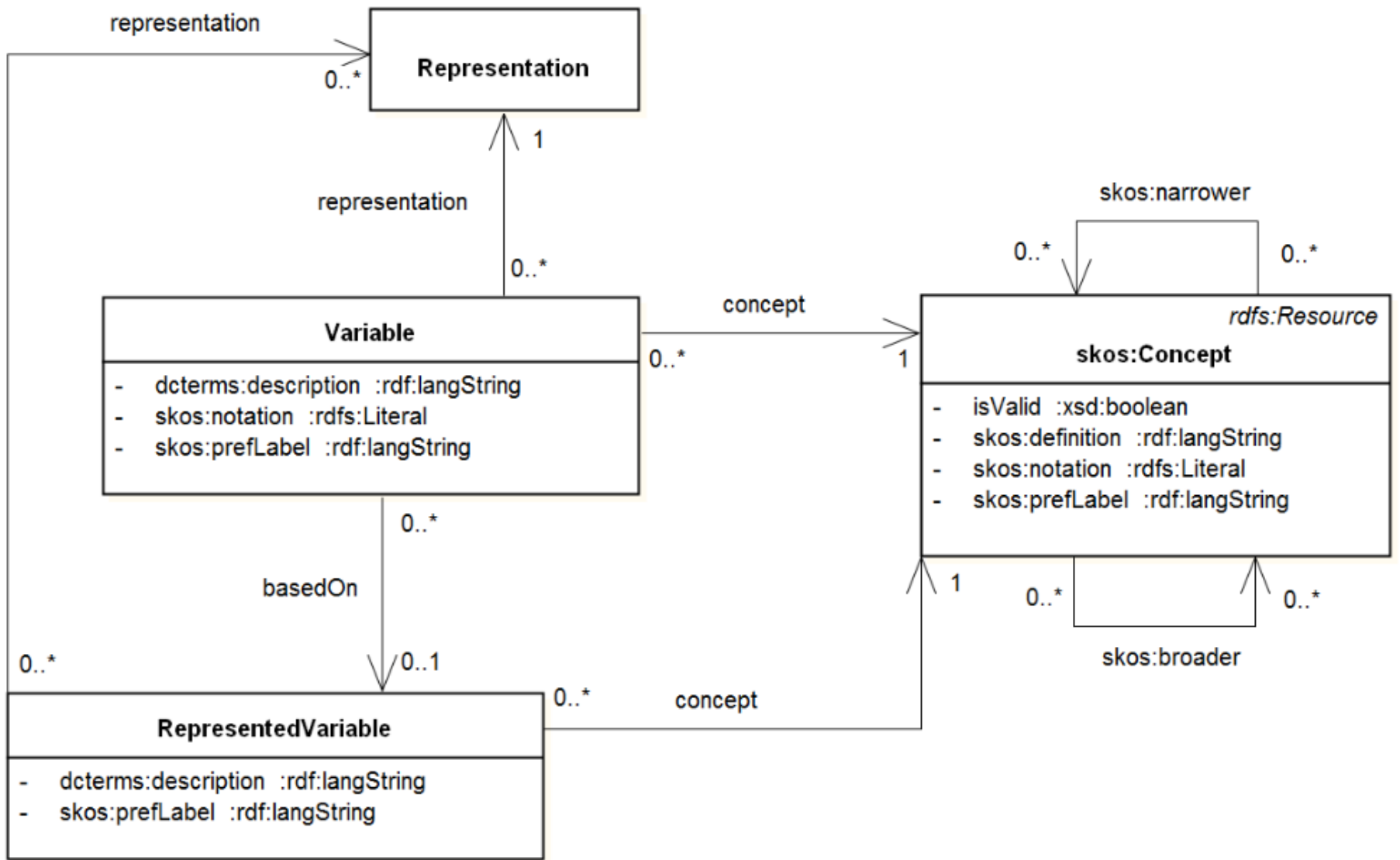
"Argentina, national coverage"@en];

dcterms:temporal :periodOfTime .

Controlled Vocabularies

Variables





ddi:AR80A401

a **disco:Variable** ;

skos:notation "AR80A401" ;

skos:prefLabel "Sex"@en ;

disco:basedOn ddi:SexVD ;

disco:question ddi:QuestionGender .

ddi:SexVD

a **disco:RepresentedVariable** ;

disco:universe ddi:UniversePerson ;

disco:representation ddi:SexRepr ;

disco:concept ddi:IpumsC1 ;

skos:prefLabel "Sex"@en ;

dcterms:description

"Sex data element"@en.

missy:PB100

a **disco:Variable** ;

skos:notation "PB100" ;

skos:prefLabel

"Quarter of the personal
interview"@en ;

skos:concept :concept ;

disco:question :question .



Variables (Names + Labels)

Data Set:

EI-SILC 2010 cross-sec p-data

Data Set (ID):

<urn:ddi:de.gesis:logicalDataSet_EU-SILC-
2010-cross-sec-p-data-2010_rev3:0.1>



Variable Concept

Data Set:

 EI-SILC 2010 cross-sec p-data

Data Set (ID):

<urn:ddi:de.gesis:logicalDataSet_EU-SILC-
2010-cross-sec-p-data-2010_rev3:0.1>

Variable Name:

 PB010



Study

Data Set

Variable label

Variable name: 'AGE'



Topical Coverage

Variable name:

B21

Study:

"Structure of Earnings Survey - 2006"@en



Variables having no concepts

Study title:

EU-SILC 2006

Variable Representation

Study

Variables

Change variable

Pos.	Name
1	PB010
2	PB020
3	PB030
4	PB040
6	PB060
8	PB100
10	PB110
12	PB120
14	PB130
16	PB140
18	PB150
20	PB160
22	PB170
24	PB180
26	PB190
28	PB200
30	PB210
32	PB220A
34	PE010

Variable Details

Values / Frequencies

EU-SILC / 2009 / PB100



AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HU	IE	IS	IT	LT
LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK					

ALL

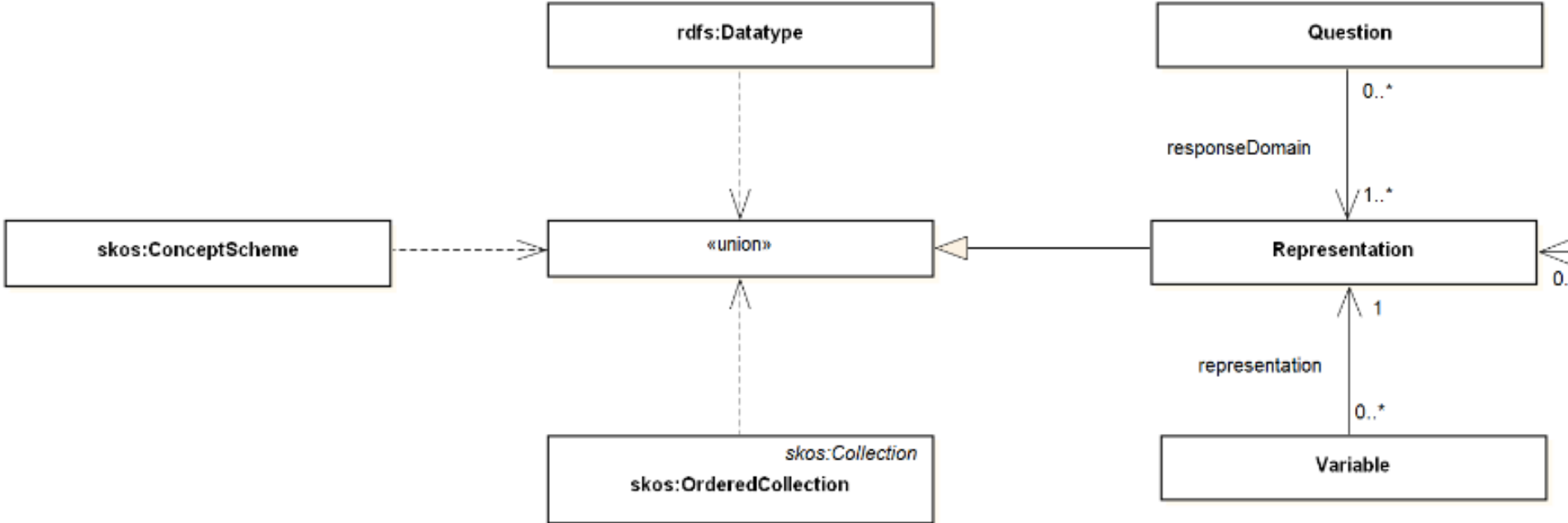
PB100: Quarter of the personal interview

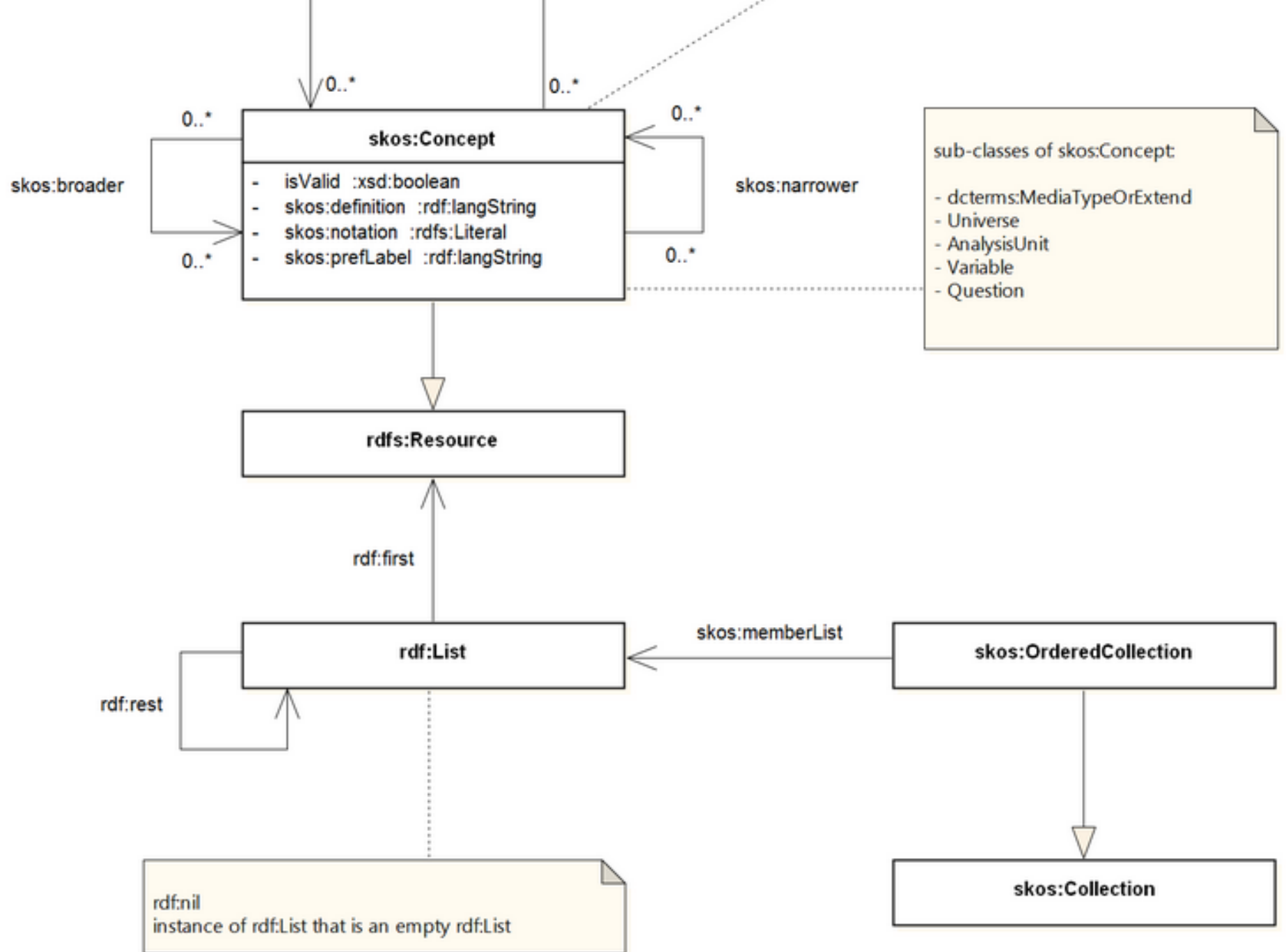
Current variable Frequencies

Statistics

Total (Missing) 470950 (4195)

Value	Value Label	Frequency	Total %	Valid %
1	January,February,March	102710	21.6	21.8
2	April, May, June	243708	51.3	51.7
3	July, August, September	66927	14.1	14.2
4	October, November, December	57605	12.1	12.2
(M)	Missing	4195	0.9	
	Valid Total	470950	99.1	100
	Total	475145	100	





Valid Codes and Categories

```
missy:1
```

```
  a skos:Concept ;
```

```
  skos:notation "1" ;
```

```
  skos:prefLabel
```

```
    "January, February, March" ;
```

```
  disco:isValid true .
```

Invalid Codes and Categories

```
missy:Missing  
  a skos:Concept ;  
  skos:notation "M" ;  
  skos:prefLabel "Missing" ;  
disco:isValid false .
```

Variable - Variable Representation

```
missy:PB100  
  a disco:Variable ,  
  missy:Variable ;  
  skos:notation 'PB100' ;  
  disco:representation  
    :representationPB100 .
```

Variable Representation

```
:representationPB100
  a disco:Representation ,
    skos:OrderedCollection ;
skos:memberList (
  missy:1
  missy:2
  missy:3
  missy:4
  missy:Missing ) .
```



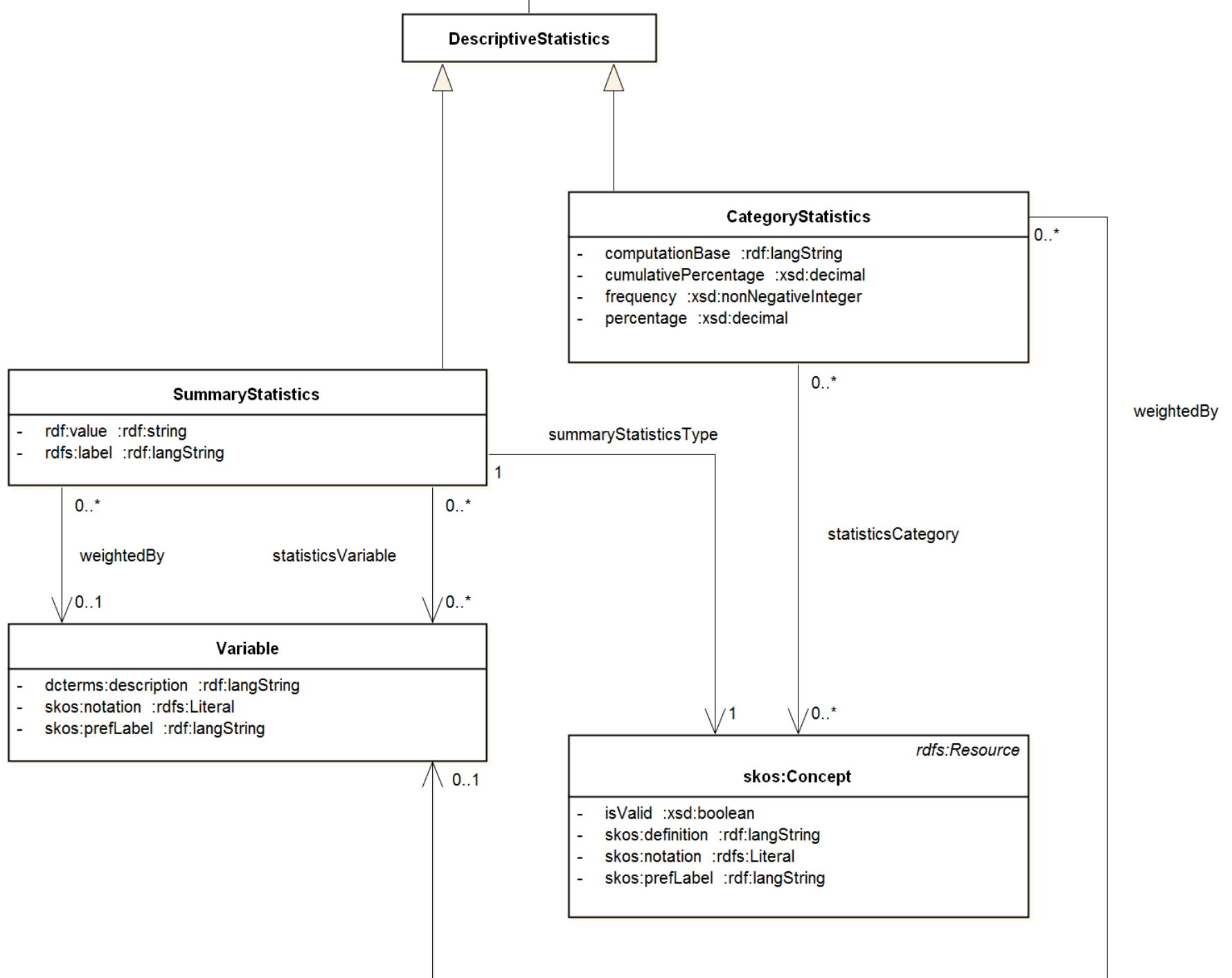
Variable Representation

Codes and Categories

Variable:

missy:PB100

Descriptive Statistics



Summary Statistics

Study

Variables

Change variable

Pos.	Name
1	PB010
2	PB020
3	PB030
4	PB040
6	PB060
8	PB100
10	PB110
12	PB120
14	PB130
16	PB140
18	PB150
20	PB160
22	PB170
24	PB180
26	PB190
28	PB200
30	PB210
32	PB220A
34	PE010

Variable Details

Values / Frequencies

EU-SILC / 2009 / PB100



AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HU	IE	IS	IT	LT
LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK					

ALL

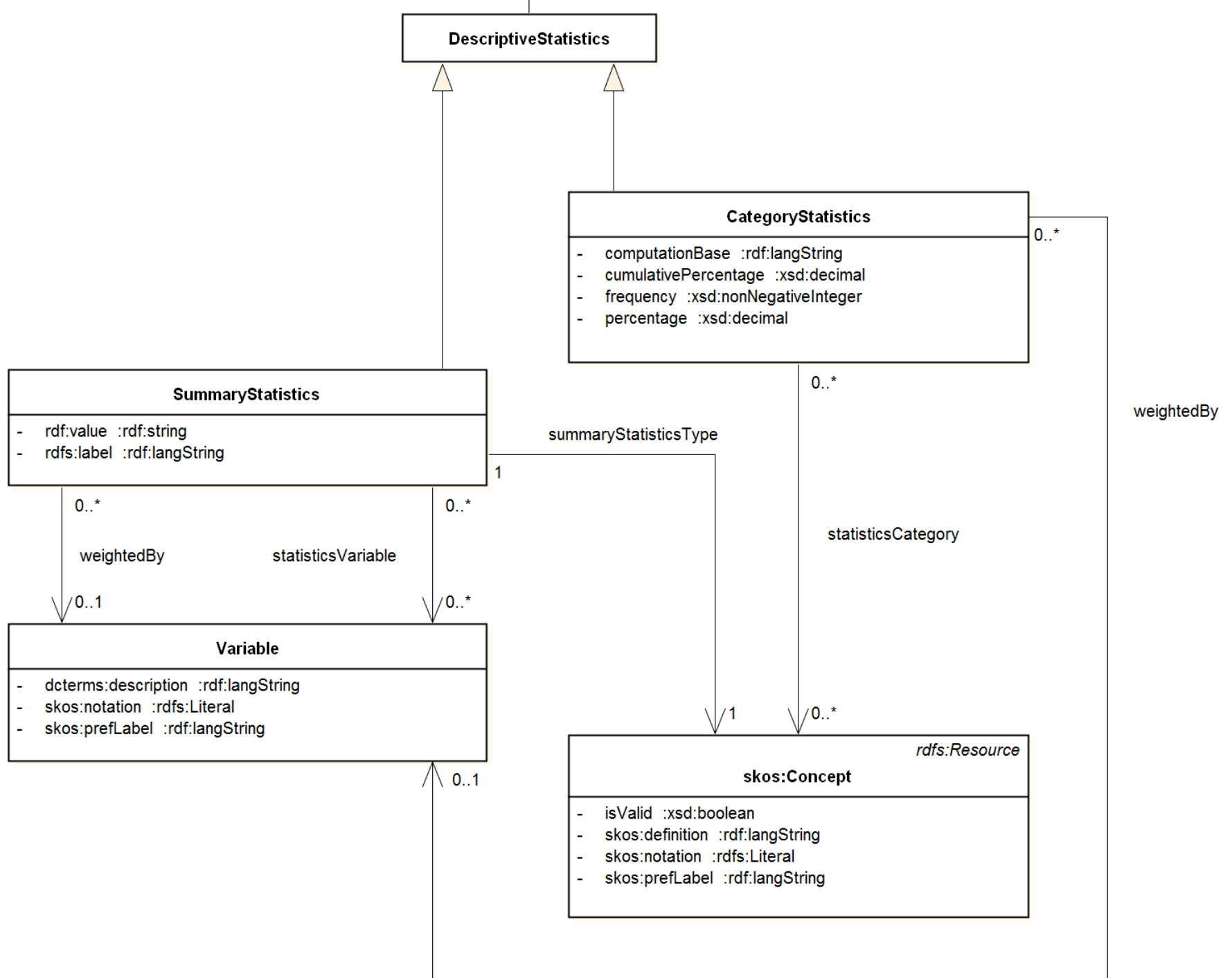
PB100: Quarter of the personal interview

Current variable Frequencies

Statistics

Total (Missing) 470950 (4195)

Value	Value Label	Frequency	Total %	Valid %
1	January,February,March	102710	21.6	21.8
2	April, May, June	243708	51.3	51.7
3	July, August, September	66927	14.1	14.2
4	October, November, December	57605	12.1	12.2
(M)	Missing	4195	0.9	
	Valid Total	470950	99.1	100
	Total	475145	100	



missy:**Minimum**

a disco:SummaryStatistics ,

missy:SummaryStatistics;

disco:statisticsVariable

missy:PB100 ;

missy:country

:AllCountriesOfStudy ;

disco:summaryStatisticType

ddicv-sumstats:Minimum ;

rdf:value "1".

Spatial Coverage of Study

```
:AllCountriesOfStudy  
  a dcterms:Location ,  
    missy:Country;  
  rdfs:label  
    "all countries of study";  
  missy:code "" .
```

Change variable

Pos.	Name
1	PB010
2	PB020
3	PB030
4	PB040
6	PB060
8	PB100
10	PB110
12	PB120
14	PB130
16	PB140
18	PB150
20	PB160
22	PB170
24	PB180
26	PB190
28	PB200
30	PB210
32	PB220A
34	PE010

EU-SILC / 2009 / PB100



AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HU	IE	IS	IT	LT
LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK					
ALL																

PB100: Quarter of the personal interview

Current variable

Statistics

Total (Missing)	23832 (0)
Min/Max (valid)	2 / 3
Mean	2.64
Standard Deviation	0.4795

PB100_F: Flag

Current variable

Statistics

Total (Missing)	23832 (0)
-----------------	-----------

Value	Value Label	Frequency	Total %	Valid %
1	filled	23832	100.0	100.0
	Valid Total	23832	100.0	100
	Total	23832	100	

missy:**Minimum**

a disco:SummaryStatistics ,
missy:SummaryStatistics;
disco:statisticsVariable
missy:PB100 ;

missy:country

<<http://sws.geonames.org/2921044>> ;

disco:summaryStatisticType
ddicv-sumstats:Minimum ;
rdf:value "1".

missy:**Maximum**

a disco:SummaryStatistics ,

missy:SummaryStatistics;

disco:statisticsVariable

missy:PB100 ;

missy:country

:AllCountriesOfStudy ;

disco:summaryStatisticType

ddicv-sumstats:Maximum ;

rdf:value "4".

missy:**Mean**

a disco:SummaryStatistics ,

missy:SummaryStatistics;

disco:statisticsVariable

missy:PB100 ;

missy:country

:AllCountriesOfStudy ;

disco:summaryStatisticType

ddicv-sumstats:ArithmeticMean ;

rdf:value "2.17".

missy:**StandardDeviation**

a disco:SummaryStatistics ,

missy:SummaryStatistics;

disco:statisticsVariable

missy:PB100 ;

missy:country

:AllCountriesOfStudy ;

disco:summaryStatisticType

ddicv-sumstats:StandardDeviation ;

rdf:value "0.9061".

```
missy:ValidCases  
  a disco:SummaryStatistics ,  
    missy:SummaryStatistics;  
disco:statisticsVariable  
  missy:PB100 ;  
missy:country  
  :AllCountriesOfStudy ;  
disco:summaryStatisticType  
  ddicv-sumstats:ValidCases ;  
rdf:value "470950".
```

missy:**PercentOfValidCases**

a disco:SummaryStatistics ,

missy:SummaryStatistics;

disco:statisticsVariable

missy:PB100 ;

missy:country

:AllCountriesOfStudy ;

disco:summaryStatisticType

ddicv-sumstats:PercentOfValidCases ;

rdf:value "99.1".

missy:**InvalidCases**

 a disco:SummaryStatistics ,

 missy:SummaryStatistics;

disco:statisticsVariable

 missy:PB100 ;

missy:country

 :AllCountriesOfStudy ;

disco:summaryStatisticType

 ddicv-sumstats:InvalidCases ;

rdf:value "4195".

```
missy:PercentOfInvalidCases
  a disco:SummaryStatistics ,
    missy:SummaryStatistics;
disco:statisticsVariable
  missy:PB100 ;
missy:country
  :AllCountriesOfStudy ;
disco:summaryStatisticType
  ddicv-sumstats:ValidCases ;
rdf:value "0.9".
```

Variable Pointing to Summary Statistics

```
missy:PB100
  a disco:variable ,
    missy:Variable ;
missy:summaryStatistics (
  missy:Minimum
  missy:Minimum_DE
  missy:Maximum
  missy:Maximum_DE
  ... )
```



Summary Statistics: Minimum

Variable:

missy:PB100

Spatial Coverage:

all countries of study



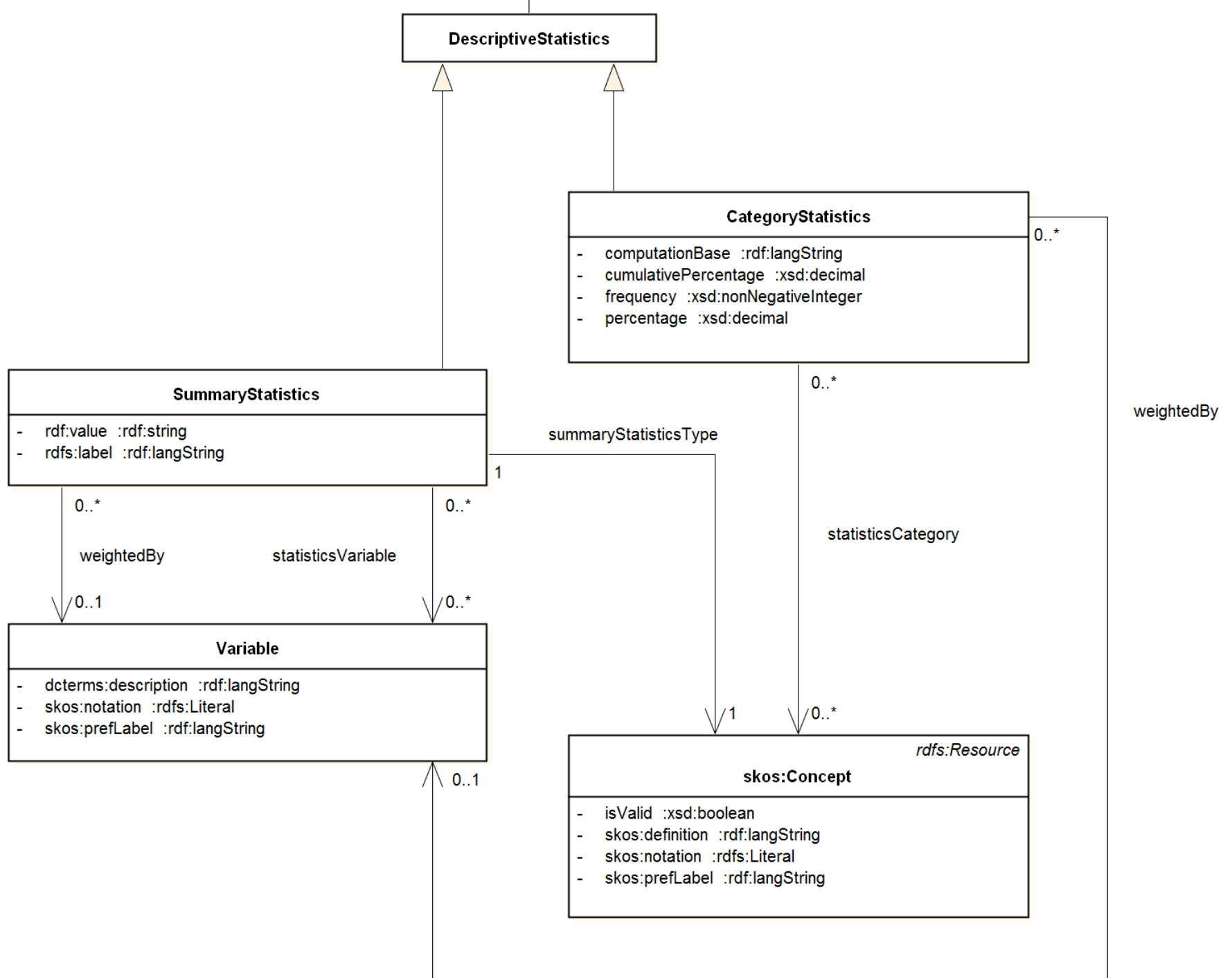
Summary Statistics: Valid Cases

Variable:

missy:PB100

Spatial Coverage: DE

Category Statistics



Valid Codes and Categories

```
missy:2
  a skos:Concept , missy:Concept ;
  skos:notation "2" ;
  skos:prefLabel
    "April, May, June" ;
  disco:isValid true ;
missy:categoryStatistics (
  missy:CS_2_AllCountries
  missy:CS_2_DE ) .
```

Invalid Codes and Categories

```
missy:Missing  
  a skos:Concept , missy:Concept ;  
  skos:notation "M" ;  
  skos:prefLabel "Missing" ;  
  disco:isValid false ;  
  missy:categoryStatistics (  
    missy:CS_M_AllCountries ) .
```

Valid Cases

```
missy:CS_2_AllCountries  
  a disco:CategoryStatistics ,  
    missy:CategoryStatistics ;  
disco:statisticsCategory missy:2 ;  
missy:country  
  :AllCountriesOfStudy ;  
disco:frequency 243708 ;  
disco:percentage 51.3 ;  
disco:cumulativePercentage 51.7 ;  
disco:computationBase "valid" .
```

Invalid Cases

```
missy:CS_M_AllCountries
  a disco:CategoryStatistics ,
    missy:CategoryStatistics ;
disco:statisticsCategory
  missy:Missing ;
missy:country
  :AllCountriesOfStudy ;
disco:frequency 4195 ;
disco:percentage 0.9 ;
disco:computationBase "invalid" .
```



Category Statistics:

Frequency (Invalid Cases)

Variable: missy:PB100

Spatial Coverage: All Countries of Study

Code: missy:Missing



Category Statistics:

Cumulative Percentage (Valid Cases)

Variable: missy:PB100

Spatial Coverage: DE

[<http://sws.geonames.org/2921044>](http://sws.geonames.org/2921044)

Code: missy:2

Category Label: 'April, May, June'

Study

Variables

Change variable

Pos.	Name
1	PB010
2	PB020
3	PB030
4	PB040
6	PB060
8	PB100
10	PB110
12	PB120
14	PB130
16	PB140
18	PB150
20	PB160
22	PB170
24	PB180
26	PB190
28	PB200

Variable Details

Values / Frequencies

EU-SILC / 2009 / PB100

AT

BE

BG

CY

CZ

DE

DK

EE

EL

ES

FI

FR

HU

IE

IS

IT

LT

LU

LV

MT

NL

NO

PL

PT

RO

SE

SI

SK

UK

ALL

PB100: Quarter of the personal interview

Current variable

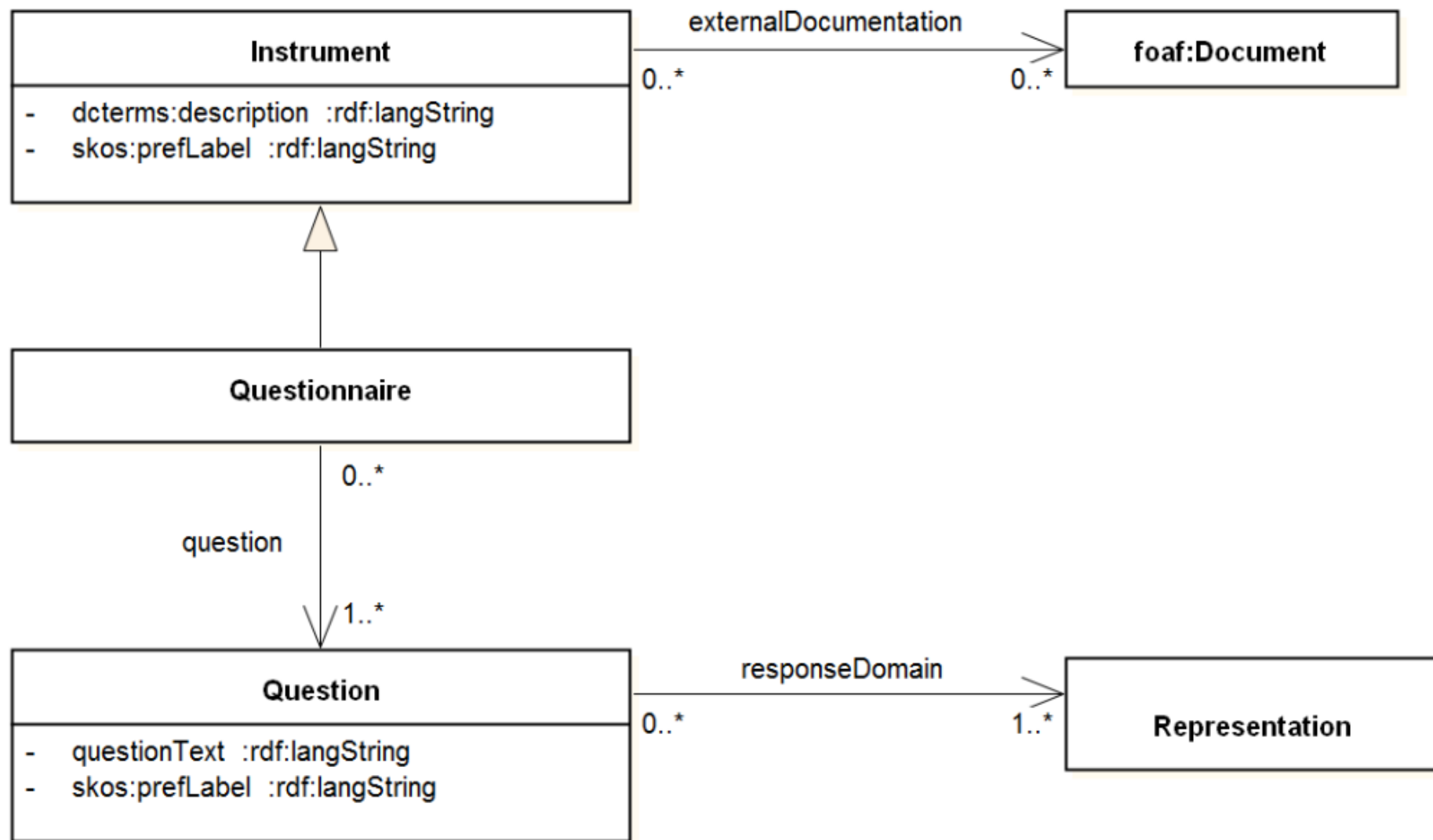
Frequencies

Statistics

Total (Missing) 23832 (0)

Value	Value Label	Frequency	Total %	Valid %
2	April, May, June	8539	35.8	35.8
3	July, August, September	15293	64.2	64.2
	Valid Total	23832	100.0	100
	Total	23832	100	

Data Collection



```
:variableYearOfBirth  
  a disco:Variable  
    skos:notation "RB080" ;  
    skos:prefLabel "Year of birth"@en ;  
    dcterms:subject :concept ;  
    disco:question :questionYearOfBirth.
```

```
:questionYearOfBirth  
  a disco:Question ;  
  disco:questionText  
    "What is your date of birth?"@en .
```



Variable Question

Data Set:

 EI-SILC 2010 cross-sec p-data

Data Set (ID):

<urn:ddi:de.gesis:logicalDataSet_EU-SILC-
2010-cross-sec-p-data-2010_rev3:0.1>

Variable Name:

 PB010



Variables

Series:

EU-SILC 2005

Question text:

'What is your date of birth?'@en

Relationships to other Vocabularies

PHDD

Mapping DDI-XML to Disco

DDI 4

DDI 4

- **Model**-driven further development of DDI
- Model → generate multiple **representations** (OWL, XSD, Java, RDB, ...)
- **Functional views** are published in a step by step manner

Disco + DDI 4

Do Not Wait for DDI 4!

- Own **functional view** for disco
- **Mapping**: disco $\rightarrow$ DDI 4 (OWL representation)
- Easy **migration**

**Let's Disco
Now!**



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.

- [1st 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
- [2nd 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