eswc17
14th ESWC 2017
28th May - 1st June 2017
Portoroz, Slovenia

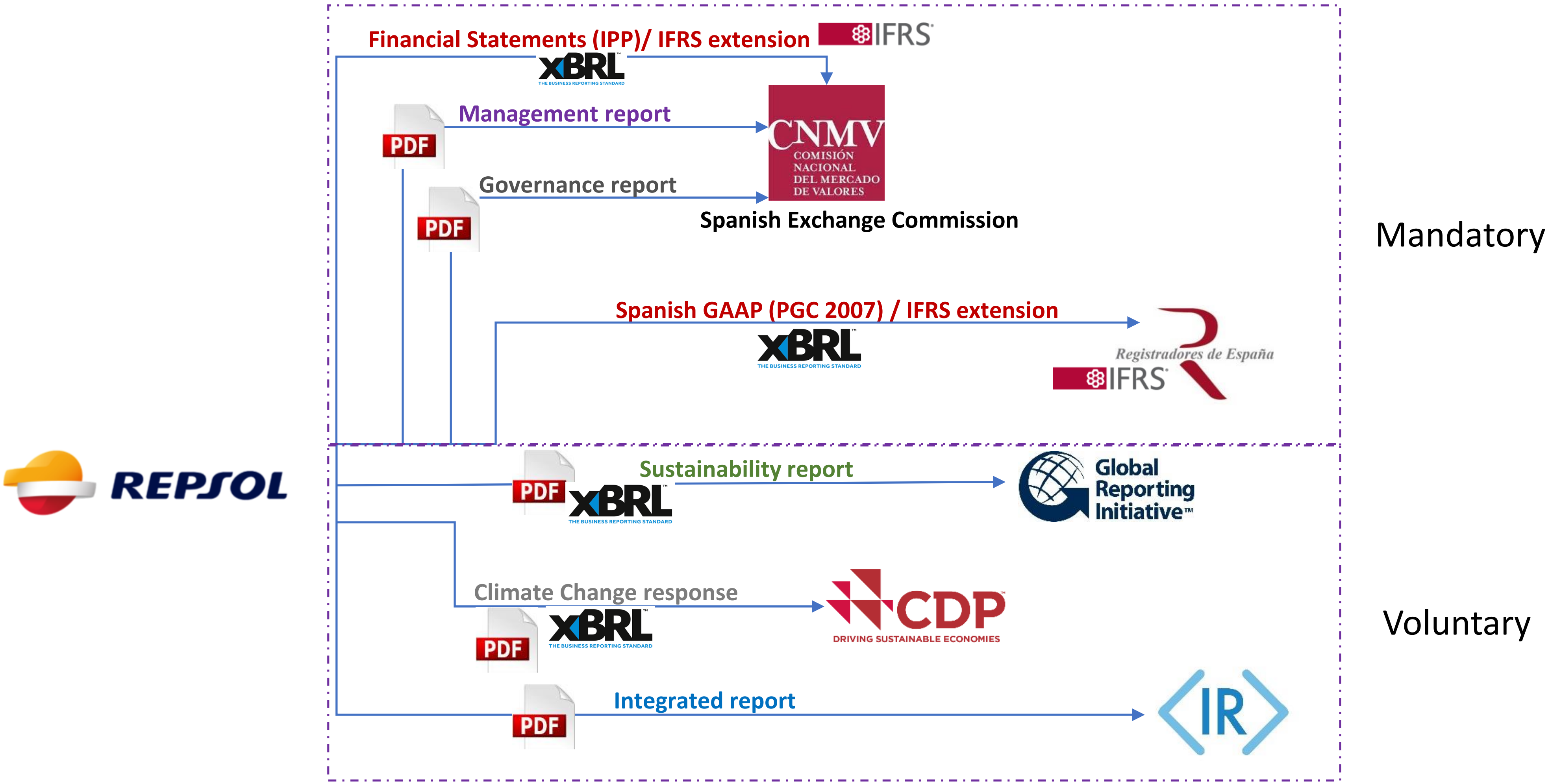
University of BRISTOL

CDP
DRIVING SUSTAINABLE ECONOMIES

Adopting Semantic Technology for Effective Corporate Transparency

Maria Mora-Rodriguez,
University of Bristol and Technical Manager at CDP
maria.mora@bristol.ac.uk
maria.mora@cdp.net
Ghislain Auguste-Atemezing (Mondeca Lab)
Chris Preist (University of Bristol)

Our problem


REPSOL
xBRL
PDF
IFRS
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
Registradores de España
Global Reporting Initiative™
CDP
DRIVING SUSTAINABLE ECONOMIES
IR

Financial Statements (IPP)/ IFRS extension

Management report

Governance report

Spanish Exchange Commission

Spanish GAAP (PGC 2007) / IFRS extension

Mandatory

Sustainability report

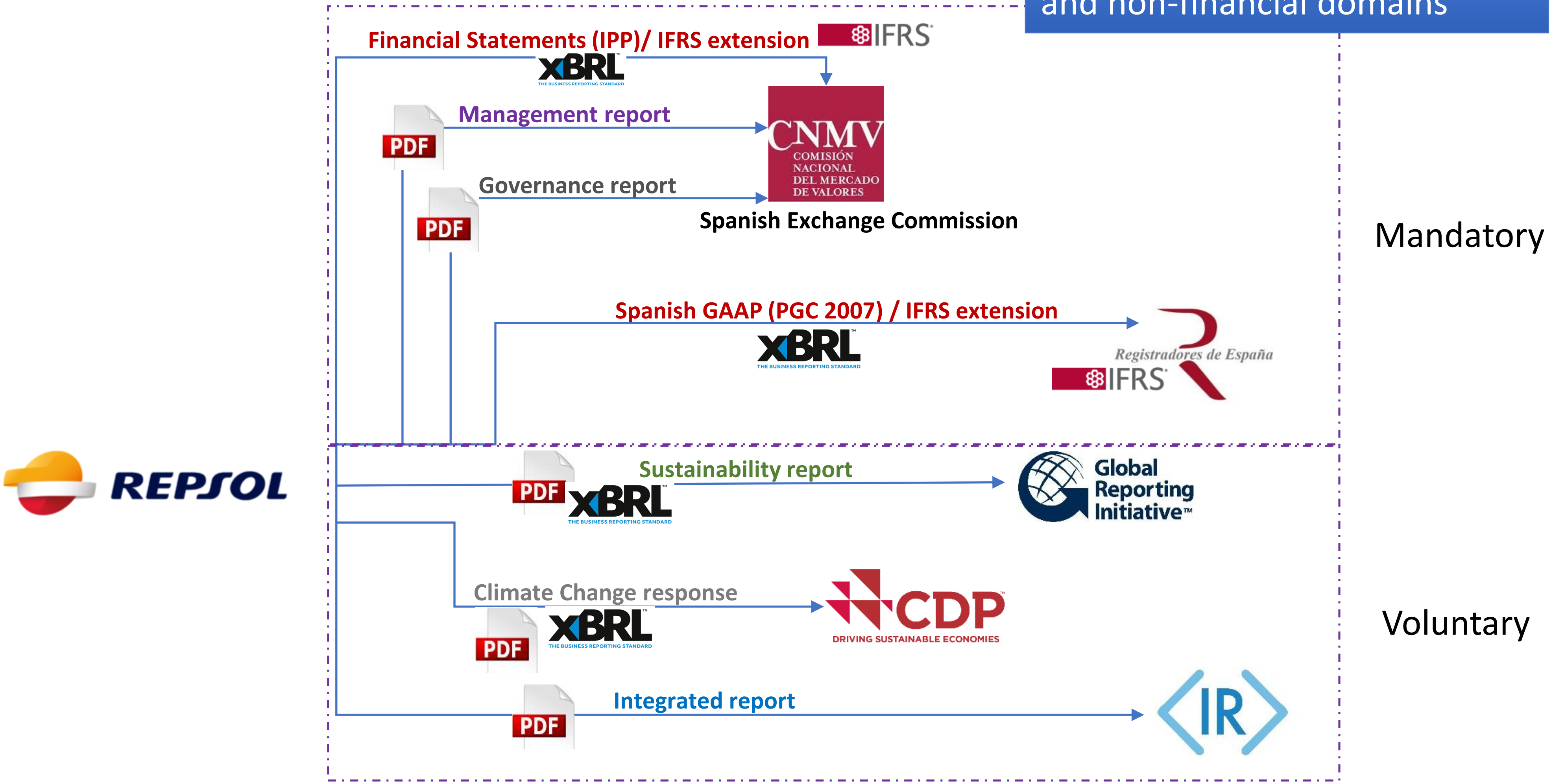
Climate Change response

Integrated report

Voluntary

Our problem

A. Data connectivity - financial and non-financial domains

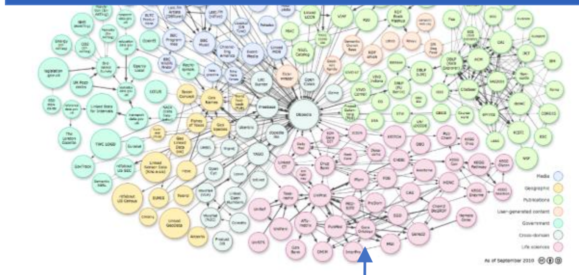


Financial Statements (IPP)/ IFRS extension
IFRS
XBRL
Management report
PDF
Governance report
PDF
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
Spanish Exchange Commission
Mandatory
Spanish GAAP (PGC 2007) / IFRS extension
XBRL
Registradores de España
IFRS
REPSOL
Sustainability report
PDF
XBRL
Global Reporting Initiative
Climate Change response
PDF
XBRL
CDP
DRIVING SUSTAINABLE ECONOMIES
Voluntary
Integrated report
PDF
IR

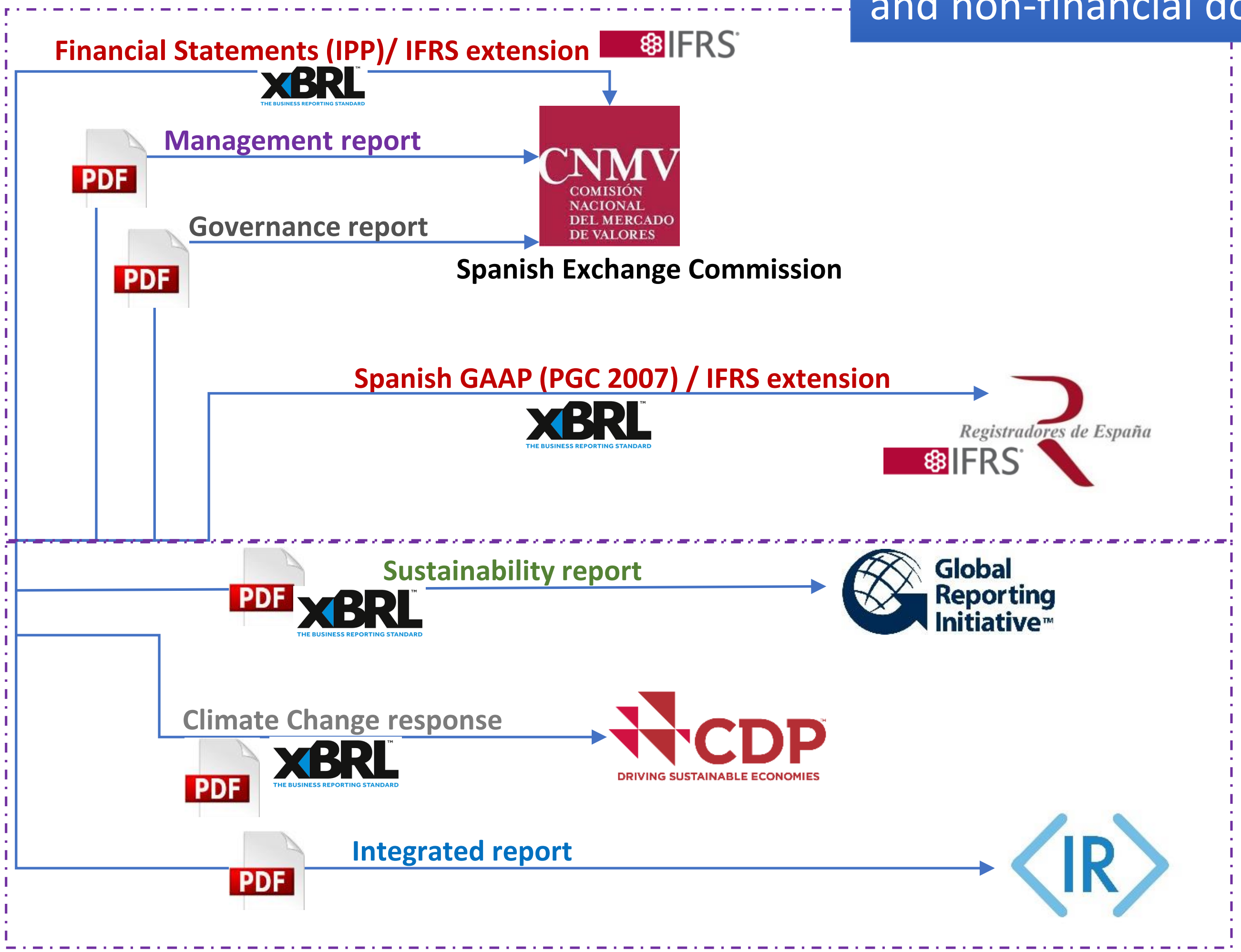
Our problem

A. Data connectivity - financial and non-financial domains

B. Connectivity with existing dataset in the LOD space



REPSOL



Financial Statements (IPP)/ IFRS extension

Management report

Governance report

Spanish Exchange Commission

Spanish GAAP (PGC 2007) / IFRS extension

Sustainability report

Climate Change response

Integrated report

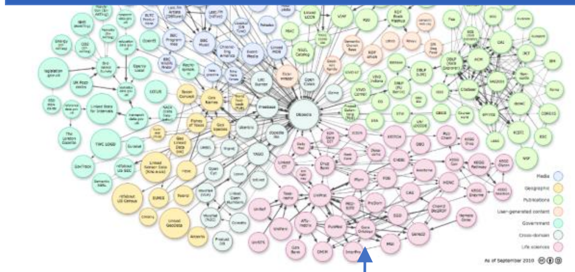
Mandatory

Voluntary

IFRS
XBRL
THE BUSINESS REPORTING STANDARD
PDF
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
Registradores de España
Global Reporting Initiative™
CDP
DRIVING SUSTAINABLE ECONOMIES
IR

Our problem

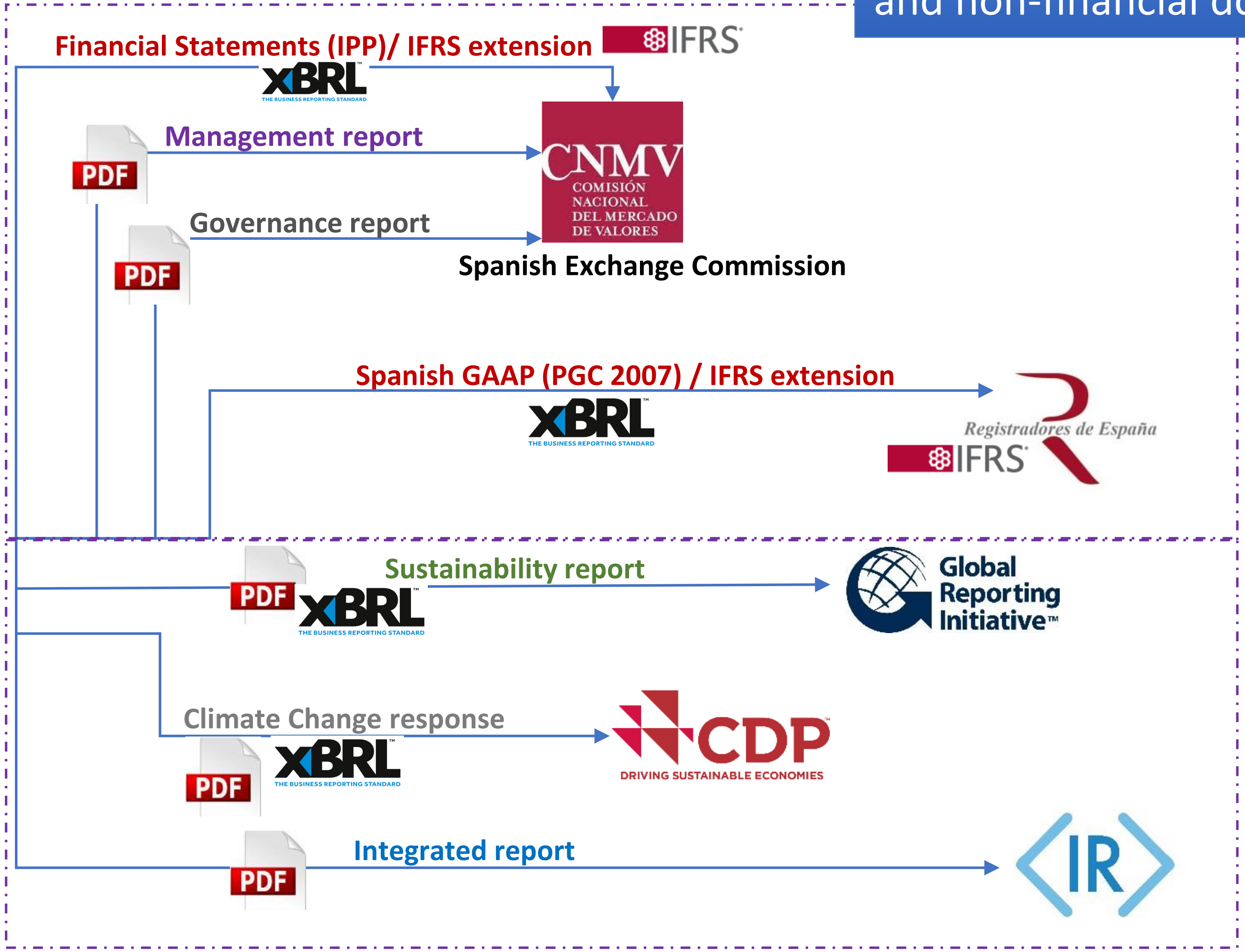
B. Connectivity with existing dataset in the LOD space



REPSOL

1. Better data contextualisation of company data

A. Data connectivity - financial and non-financial domains



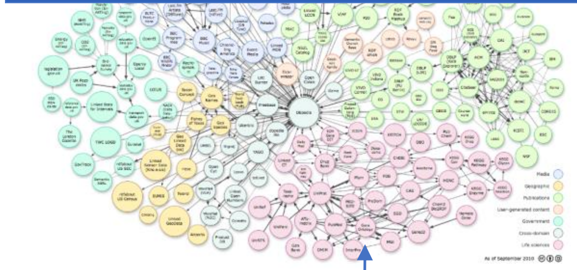
Financial Statements (IPP)/ IFRS extension
XBRL
IFRS
Management report
PDF
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
Governance report
PDF
Spanish Exchange Commission
Spanish GAAP (PGC 2007) / IFRS extension
XBRL
Registradores de España
IFRS
Sustainability report
PDF
XBRL
Global Reporting Initiative
Climate Change response
PDF
XBRL
CDP
DRIVING SUSTAINABLE ECONOMIES
Integrated report
PDF
IR

Mandatory

Voluntary

Our problem

B. Connectivity with existing dataset in the LOD space

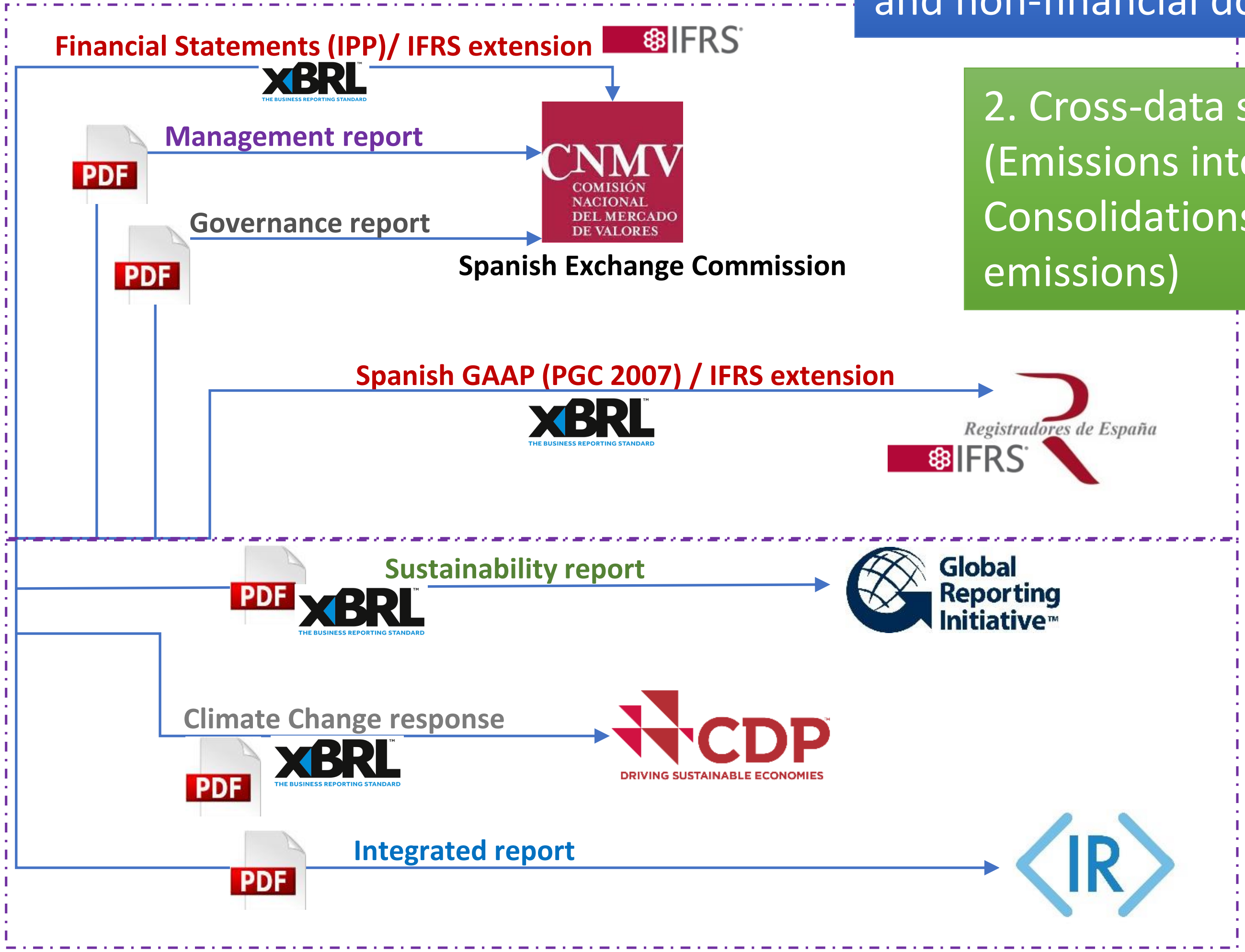


REPSOL

1. Better data contextualisation of company data

A. Data connectivity - financial and non-financial domains

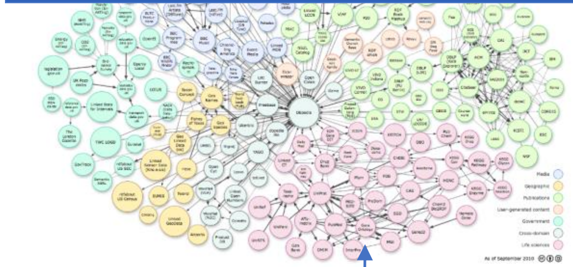
2. Cross-data source analysis (Emissions intensity figure: Consolidations sales / CO2 emissions)


Financial Statements (IPP)/ IFRS extension
XBRL
IFRS
Management report
PDF
Governance report
PDF
CNMV COMISIÓN NACIONAL DEL MERCADO DE VALORES
Spanish Exchange Commission
Spanish GAAP (PGC 2007) / IFRS extension
XBRL
Registradores de España
IFRS
Sustainability report
PDF
XBRL
Global Reporting Initiative
Climate Change response
PDF
XBRL
CDP DRIVING SUSTAINABLE ECONOMIES
Integrated report
PDF
IR

Voluntary

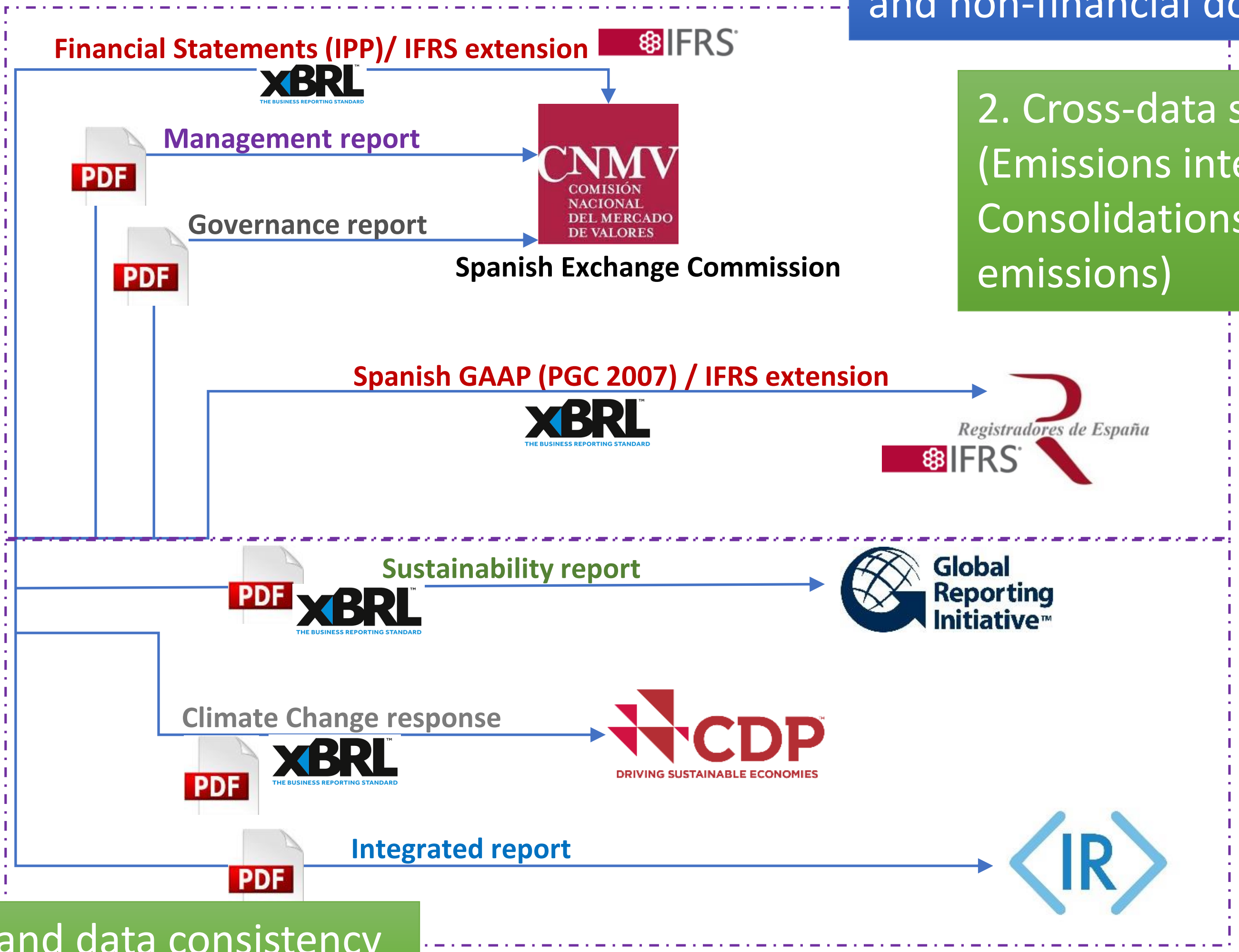
Our problem

B. Connectivity with existing dataset in the LOD space



A. Data connectivity - financial and non-financial domains

2. Cross-data source analysis (Emissions intensity figure: Consolidations sales / CO2 emissions)



Financial Statements (IPP)/ IFRS extension
XBRL
IFRS
Management report
PDF
Governance report
PDF
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
Spanish Exchange Commission
Spanish GAAP (PGC 2007) / IFRS extension
XBRL
Registradores de España
IFRS
Sustainability report
PDF
XBRL
Global Reporting Initiative™
Climate Change response
PDF
XBRL
CDP
DRIVING SUSTAINABLE ECONOMIES
Integrated report
PDF
IR

REPSOL

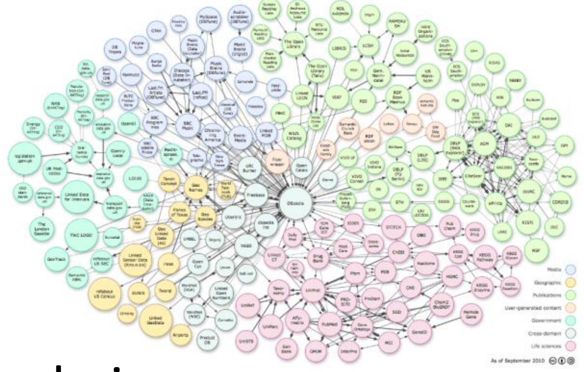
Voluntary

1. Better data contextualisation of company data

3. Data reliability and data consistency Cross-checking

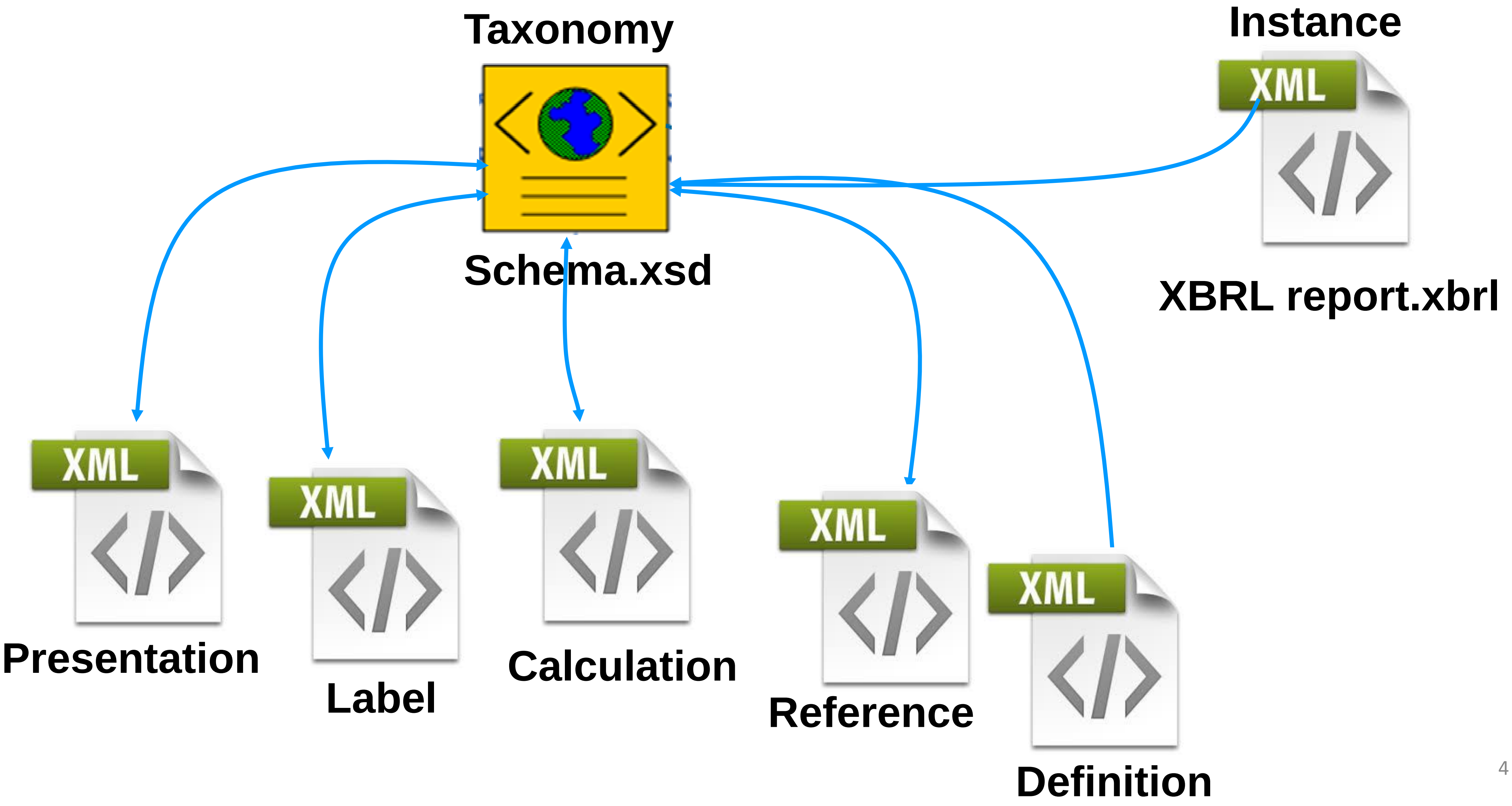
Our attention

- Current availability and adoption of XBRL data
  - XBRL is already being used by over **10 million companies**, 100 regulators and 60 governments worldwide.
  - XBRL is becoming in the common denominator between financial and non-financial company disclosure.


Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2010

- Opportunities that Linked data can offer
  - Converting independent silos of XBRL data into interconnected pieces.

Understanding XBRL



Taxonomy

Schema.xsd

Instance

XBRL report.xbrl

Presentation

Label

Calculation

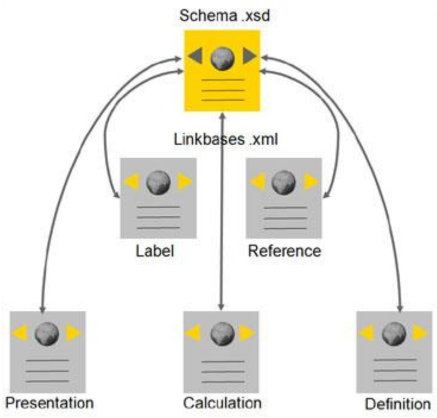
Reference

Definition

XML
XML
XML
XML
XML
XML

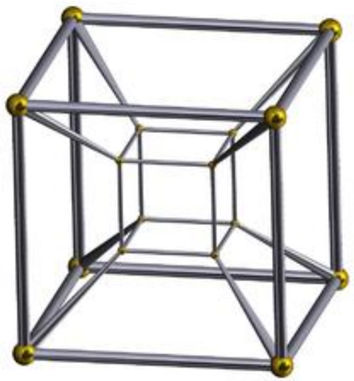
About XBRL - evolution

2003


Schema .xsd
Linkbases .xml
Label
Reference
Presentation
Calculation
Definition

XBRL 2.1

2005

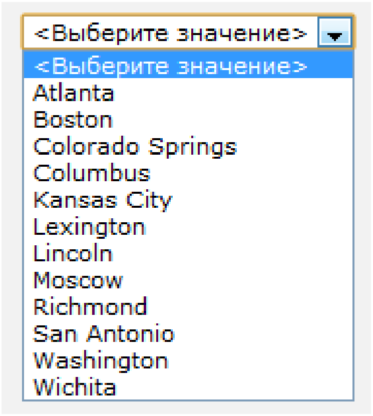


Dimensions 1.0 Specification

2009

Formula 1.0 Specification

2014


<Выберите значение>
<Выберите значение>
Atlanta
Boston
Colorado Springs
Columbus
Kansas City
Lexington
Lincoln
Moscow
Richmond
San Antonio
Washington
Wichita

Extensible Enumerations 1.0

Our work

lod live

What is the context of the company Repsol?

What was the emission intensity of Repsol in 2015?

How reliable is the equity figure presented in DBpedia?

Applicability

Visualisation LodLive

Data contextualisation

Cross-data analysis

Data accuracy

Accessibility

Data publication

Apache Jena Fuseki ->SPARQL

semantic web framework Jena

Interlinking

Linking to the Web of Data

LIMES->DBPedia : LOD cloud

LIMES Link Discovery Framework for Metric Spaces

Ontology

XBRL to Linked data

XBRL Lightweight vocabulary
- Mapping with well-known vocabularies
- Dereferenceable URIs
- XBRL Data from the (CNMV) [XBRL 2.1] and the CDP [XBRL 2.1 + Dimensions 1.0 + Enumerations 1.0].

CNMV COMISIÓN NACIONAL DEL MERCADO DE VALORES

XBRL THE BUSINESS REPORTING STANDARD

CDP DRIVING SUSTAINABLE ECONOMIES

Related work

How to translate XBRL Financial data into XBRL

- Transforming XBRL taxonomies from well-known **open government data initiatives** (SECs EDGAR, CNMV) into RDF (Garcia and Gil, 2009)
- **RDF Cube vocabulary** -(Kampgen *et al*.,2014)
- Experimental Initiative called Edgar **Linked Wrapper**, which provides access to XBRL filing from the SEC as Linked Data. Each new US-GAAP taxonomy means a new semantic vocabulary.

http://edgarwrap.ontologycentral.com/

How to translate XBRL Sustainability data into RDF

- **Sustainability data** -> RDF: GRI taxonomy into RDF (Madlberger *et al*., 2013)

Architecture for a better financial data integration

- Architecture for evolving information systems enabling better **financial data integration**. XBRL financial data + DBPedia +Yahoo!!Finance (Goto *et al.*, 2013)

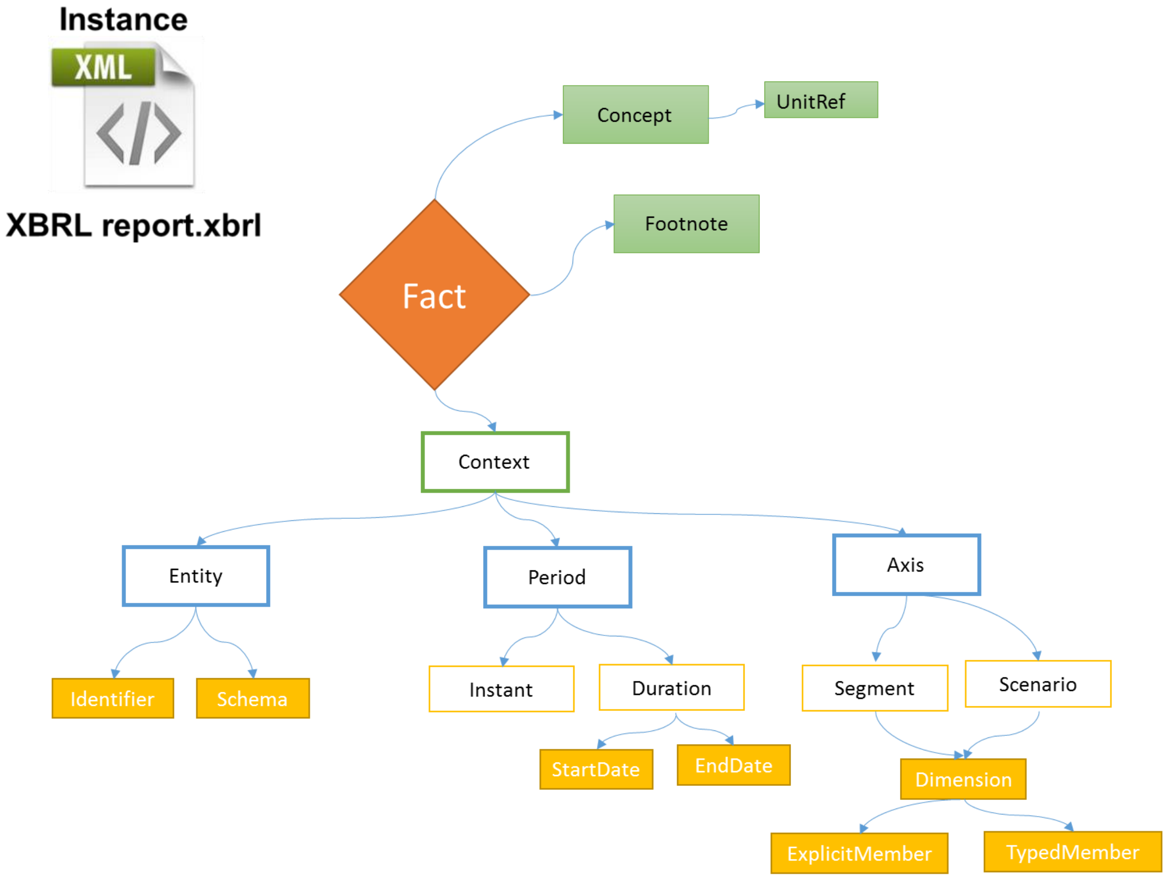
Related work - difficulties

- Representing XBRL data in RDF graphs and as Linked data:
  - Lack of formal semantics and inference mechanisms
  - Difficulties to find correspondences with well-known vocabularies (SKOS, FOAF, etc..)
  - Lack of general solutions to transform any XBRL filings into RDF format.
    - covering the full XBRL Specification: XBRL 2.1, Dimensions 1.0, Enumerations….

I HAVE A SURPRISE FOR YOU...

Let’s talk about technical bits….

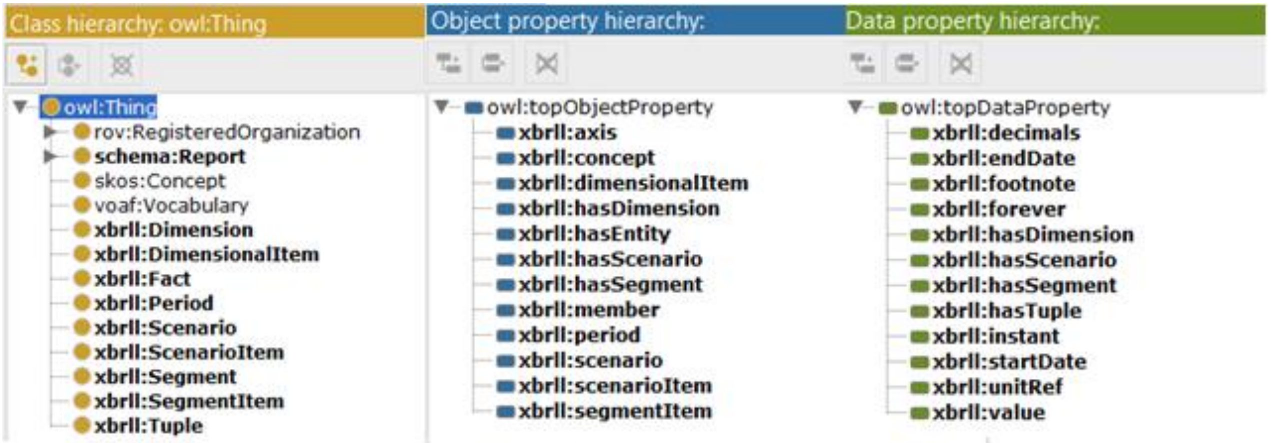
XBRL into Linked data → Interlinking → Data publication → Applicability



Instance
XML
XBRL report.xbrl
Fact
Concept
UnitRef
Footnote
Context
Entity
Period
Axis
Identifier
Schema
Instant
Duration
StartDate
EndDate
Segment
Scenario
Dimension
ExplicitMember
TypedMember

How XBRL data is represented

SKOS SCHEMA.ORG FOAF ROV



Class hierarchy: owl:Thing

- owl:Thing
  - rov:RegisteredOrganization
  - schema:Report
  - skos:Concept
  - voaf:Vocabulary
  - xbrll:Dimension
  - xbrll:DimensionalItem
  - xbrll:Fact
  - xbrll:Period
  - xbrll:Scenario
  - xbrll:ScenarioItem
  - xbrll:Segment
  - xbrll:SegmentItem
  - xbrll:Tuple

Object property hierarchy:

- owl:topObjectProperty
  - xbrll:axis
  - xbrll:concept
  - xbrll:dimensionalItem
  - xbrll:hasDimension
  - xbrll:hasEntity
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:member
  - xbrll:period
  - xbrll:scenario
  - xbrll:scenarioItem
  - xbrll:segmentItem

Data property hierarchy:

- owl:topDataProperty
  - xbrll:decimals
  - xbrll:endDate
  - xbrll:footnote
  - xbrll:forever
  - xbrll:hasDimension
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:hasTuple
  - xbrll:instant
  - xbrll:startDate
  - xbrll:unitRef
  - xbrll:value

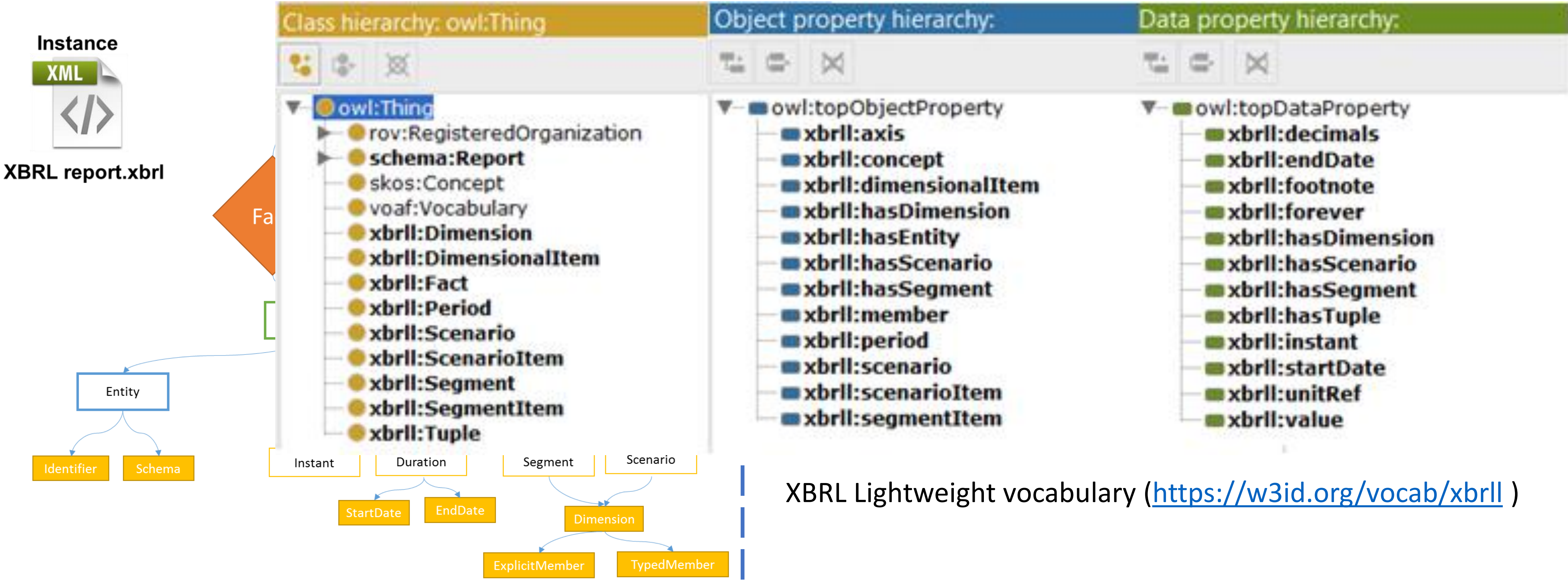
XBRL Lightweight vocabulary (https://w3id.org/vocab/xbrll)

XBRL into Linked data

Interlinking

Data publication

Applicability


Instance
XML
XBRL report.xbrl
Fa
Entity
Identifier
Schema
Instant
Duration
Segment
Scenario
StartDate
EndDate
Dimension
ExplicitMember
TypedMember

Class hierarchy: owl:Thing

- owl:Thing
  - rov:RegisteredOrganization
  - schema:Report
  - skos:Concept
  - voaf:Vocabulary
  - xbrll:Dimension
  - xbrll:DimensionalItem
  - xbrll:Fact
  - xbrll:Period
  - xbrll:Scenario
  - xbrll:ScenarioItem
  - xbrll:Segment
  - xbrll:SegmentItem
  - xbrll:Tuple

Object property hierarchy:

- owl:topObjectProperty
  - xbrll:axis
  - xbrll:concept
  - xbrll:dimensionalItem
  - xbrll:hasDimension
  - xbrll:hasEntity
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:member
  - xbrll:period
  - xbrll:scenario
  - xbrll:scenarioItem
  - xbrll:segmentItem

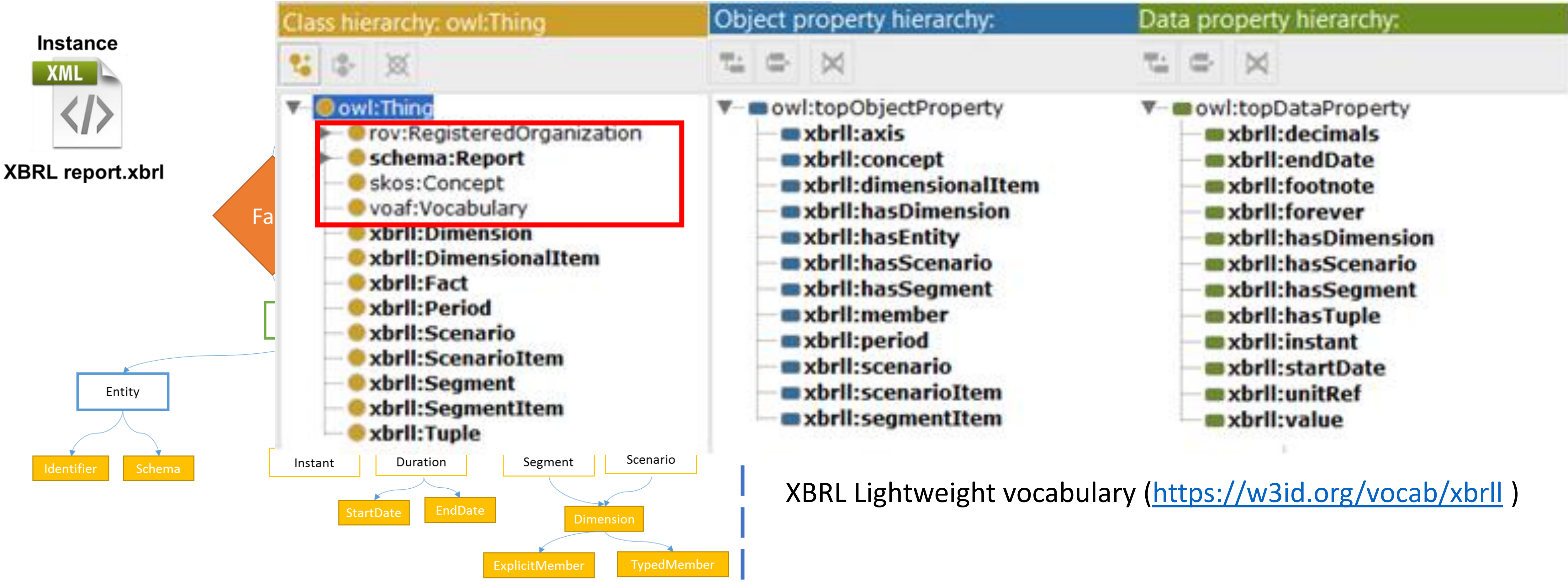
Data property hierarchy:

- owl:topDataProperty
  - xbrll:decimals
  - xbrll:endDate
  - xbrll:footnote
  - xbrll:forever
  - xbrll:hasDimension
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:hasTuple
  - xbrll:instant
  - xbrll:startDate
  - xbrll:unitRef
  - xbrll:value

XBRL Lightweight vocabulary (https://w3id.org/vocab/xbrll)

How XBRL data is represented

XBRL into Linked data → Interlinking → Data publication → Applicability


Instance
XML
XBRL report.xbrl
Fa
Entity
Identifier
Schema
Instant
Duration
StartDate
EndDate
Segment
Scenario
Dimension
ExplicitMember
TypedMember

Class hierarchy: owl:Thing

- owl:Thing
  - rov:RegisteredOrganization
  - schema:Report
  - skos:Concept
  - voaf:Vocabulary
  - xbrll:Dimension
  - xbrll:DimensionalItem
  - xbrll:Fact
  - xbrll:Period
  - xbrll:Scenario
  - xbrll:ScenarioItem
  - xbrll:Segment
  - xbrll:SegmentItem
  - xbrll:Tuple

Object property hierarchy:

- owl:topObjectProperty
  - xbrll:axis
  - xbrll:concept
  - xbrll:dimensionalItem
  - xbrll:hasDimension
  - xbrll:hasEntity
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:member
  - xbrll:period
  - xbrll:scenario
  - xbrll:scenarioItem
  - xbrll:segmentItem

Data property hierarchy:

- owl:topDataProperty
  - xbrll:decimals
  - xbrll:endDate
  - xbrll:footnote
  - xbrll:forever
  - xbrll:hasDimension
  - xbrll:hasScenario
  - xbrll:hasSegment
  - xbrll:hasTuple
  - xbrll:instant
  - xbrll:startDate
  - xbrll:unitRef
  - xbrll:value

XBRL Lightweight vocabulary (https://w3id.org/vocab/xbrll)

How XBRL data is represented

XBRL into Linked data

Interlinking

Data publication

Applicability

Dereferenceable URIs (We need Unique Identifiers!!)

```
<http://data.mondeca.com/id/fact/f001> a
ns1:Fact ;
    ns1:hasEntity
<http://data.mondeca.com/id/entity/000084
7838> ;
    ns1:concept cdp:IntroductionCompany;
    ns1:period [ ns1:startPeriod "2015-
01-01"^^xsd:date ;
            ns1:endPeriod "2015-12-
31"^^xsd:date] ;
    ns1:value “Repsol is international
integrated Oil and Gas Company whose main
activity consists of the upstream and
downstream business.” ;
```

```
<xbrli:context id="ctx_00">
    <xbrli:entity>
      <xbrli:identifier
scheme="http://www.cdp.net/CIK">0000847838</xbrli:identifie
r>
    </xbrli:entity>
    <xbrli:period>
      <xbrli:startDate>2015-01-01</xbrli:startDate>
      <xbrli:endDate>2015-12-31</xbrli:endDate>
    </xbrli:period>
  </xbrli:context>
<cdp:IntroductionCompany contextRef="ctx_00">Repsol is
international integrated Oil and Gas Company whose main
activity consists of the upstream and downstream
business.<cdp:IntroductionCompany>
```

RDF representation of a simple XBRL fact (CDP)

XBRL into Linked data

Interlinking

Data publication

Applicability

```
<http://data.mondeca.com/id/fact/f0004> a ns1:Fact ;
   ns1:concept ifrs-gp:IntangibleAssetsNet;
   ns1:entity <http://data.mondeca.com/id/entity/A7837472>;
   ns1:period [ ns1:instant "2015-12-31"^^xsd:date] ;
   ns1:hasTuple <http://data.mondeca.com/id/tuple/t89493> ;
   ns1:unitRef < http://dbpedia.org/resource/EUR> ;
   ns1:value 94588000.0 ;
   ns1:decimals 0 .
   ns1:hasScenario
<http://data.mondeca.com/id/scenario/s0001> .
 <http://data.mondeca.com/id/fact/f0004> a ns1:Fact ;
   ns1:concept ifrs-gp:GoodwillNet ;
   ns1:entity <http://data.mondeca.com/id/entity/A7837472>;
   ns1:period [ ns1:instant "2015-12-31"^^xsd:date] ;
   ns1:hasTuple <http://data.mondeca.com/id/tuple/t89493> ;
   ns1:unitRef < http://dbpedia.org/resource/EUR>;
   ns1:value 350.0 ;
   ns1:decimals 0 .
   ns1:hasScenario
<http://data.mondeca.com/id/scenario/s0001> .
 <http://data.mondeca.com/id/scenario/s0001> a ns1:Scenario
   ns1:scenarioItem [ns1:concept ipp:Modelo; ns1:value “GE”]
   ns1:scenarioItem [ns1:concept ipp:Apartado; ns1:value
“Individual”]

 <http://data.mondeca.com/id/tuple/t89493> a ns1:Tuple ;
    ns1:concept ipp-gen:BalanceIndividual;
    ns1:hasTuple <http://data.mondeca.com/id/tuple/t89494>
```

```
<xbrli:context id="S22015_A-78374725_ici">
  <xbrli:entity>
       <xbrli:identifier
scheme="http://www.cnmv.es/xbrl/ipp/A-
78374725">REPSOL, S.A</xbrli:identifier>
  </xbrli:entity>
  <xbrli:period>
     <xbrli:instant>2015-12-31</xbrli:instant>
  </xbrli:period>
  <xbrli:scenario>
      <ipp:Modelo>GE</ipp:Modelo>
      <ipp:Apartado>Individual</ipp:Apartado>
   </xbrli:scenario>
</xbrli:context>
 <xbrli:unit id="euro">
   <xbrli:measure>iso4217:EUR</xbrli:measure>
</xbrli:unit>
 <ipp-gen:InformacionFinancieraSeleccionada>
     <ipp-gen:BalanceIndividual>
           <ifrs-gp:IntangibleAssetsNet
decimals="0" contextRef="S22015_A-78374725_ici"
unitRef="euro">94588000</ifrs-
gp:IntangibleAssetsNet>
           <ifrs-gp:GoodwillNet decimals="0"
contextRef="S22015_A-78374725_ici"
unitRef="euro">350</ifrs-gp:GoodwillNet>
      </ipp-gen:BalanceIndividual>
</ipp-gen:InformacionFinancieraSeleccionada>
```

RDF representation of a tuple XBRL fact (CNMV)

XBRL into Linked data → Interlinking → Data publication → Applicability

```
<http://data.mondeca.com/id/fact/f0005> a
ns1:Fact ;
  ns1:concept  cdp:EmissionValueGrossCO2e
  ns1:entity
<http://data.mondeca.com/id/entity/00008478
38> ;
  ns1:period [ ns1:startPeriod "2015-01-
01"^^xsd:date; ns1:endPeriod "2015-12-
31"^^xsd:date] ;
  ns1:hasDimension
<http://data.mondeca.com/id/dimension/d0001
> ;
   ns1:unitRef
<http://dbpedia.org/page/Carbon_dioxide_equ
ivalent>;
   ns1:value 21068516;
   ns1:decimals 0 .
 .
<http://data.mondeca.com/id/dimension/d0001
> a ns1:Dimension ;
   ns1:dimensionItem [ns1:axis
TotalEmissionDataAxis; ns1:value “id01”]
```

```
 <xbrli:context id="ctx_8_2_2015_id01">
    <xbrli:entity>
      <xbrli:identifier
scheme="http://www.cdp.net/CIK">0000847838</xbrli:identifier>
      <xbrli:segment>
        <xbrldi:typedMember
dimension="cdp:TotalEmissionDataAxis">
<cdp:GreenhouseInventoryBoundariesID>id01</cdp:GreenhouseInvento
ryBoundariesID>
        </xbrldi:typedMember>
      </xbrli:segment>
    </xbrli:entity>
    <xbrli:period>
      <xbrli:startDate>2015-01-01</xbrli:startDate>
      <xbrli:endDate>2015-12-31</xbrli:endDate>
    </xbrli:period>
</xbrli:context>
 <xbrli:unit id="CO2e">
   <xbrli:measure>cdp:CO2e</xbrli:measure>
</xbrli:unit>
  <cdp:EmissionValueGrossCO2e decimals="0"
contextRef="ctx_8_2_2015_id01"
unitRef="CO2e">21068516</cdp:EmissionValueGrossCO2e>
<cdp:Scope contextRef="ctx_8_2_2015_id01">cdp-
enum:Scope1</cdp:Scope>
```

RDF representation of a dimensional XBRL fact (CDP)

XBRL into Linked data → Interlinking → Data publication → Applicability

- LIMES allows detecting similar Linked datasets (Link Discovery).
- LIMES works specifying the searching criteria and the target endpoint.
    - (find references in other places)

- **Our searching criteria** is the company name contained in the RDF generated.
- **Endpoint target:** DBpedia
- **Trigram metric** (legalName, sch:Organization)
- Results are included as SameAs relationships

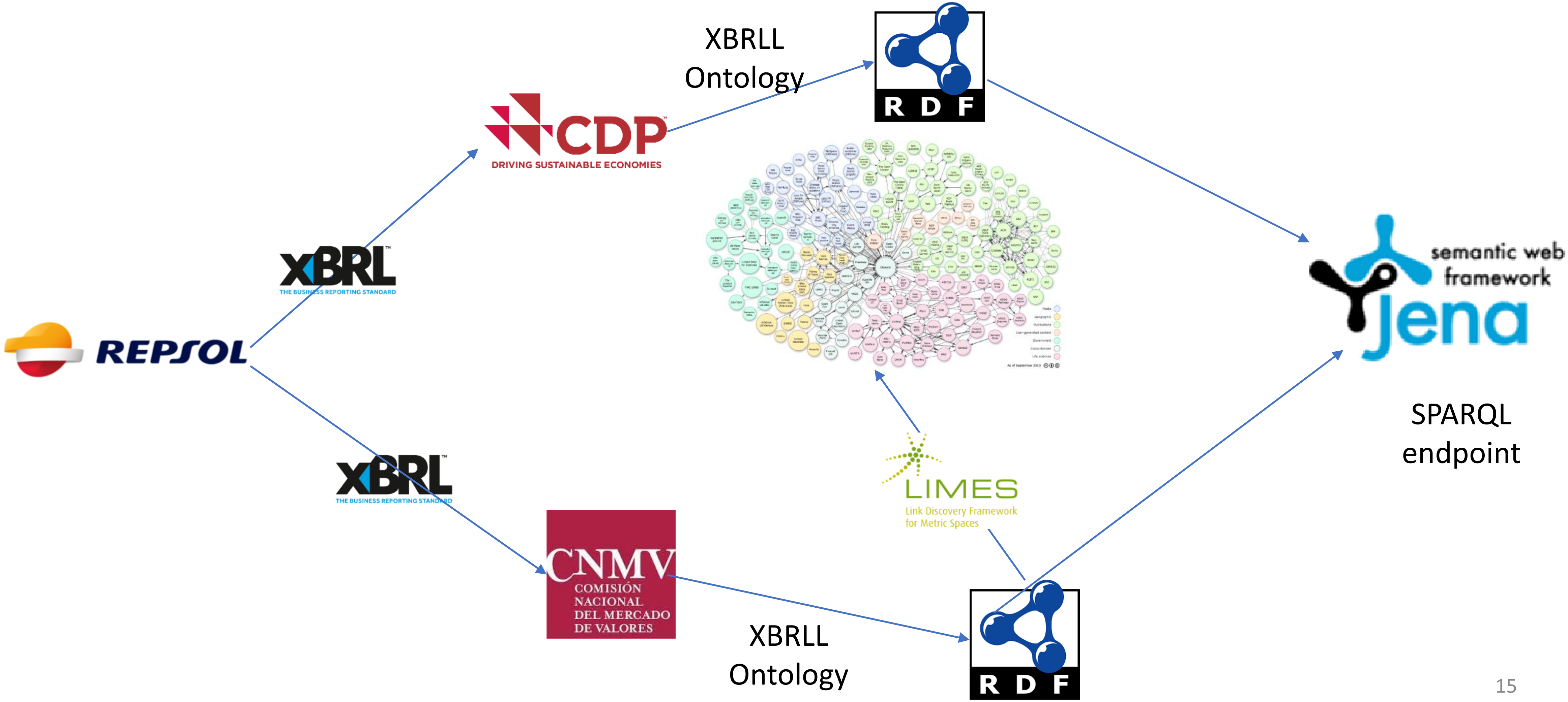
LIMES
Link Discovery Framework
for Metric Spaces

XBRL into Linked data

Interlinking

Data publication

Applicability


REPSOL
XBRL
THE BUSINESS REPORTING STANDARD
CDP
DRIVING SUSTAINABLE ECONOMIES
XBRL
THE BUSINESS REPORTING STANDARD
CNMV
COMISIÓN NACIONAL DEL MERCADO DE VALORES
RDF
LIMES
Link Discovery Framework for Metric Spaces
RDF
semantic web framework
Jena

XBRLL Ontology

XBRLL Ontology

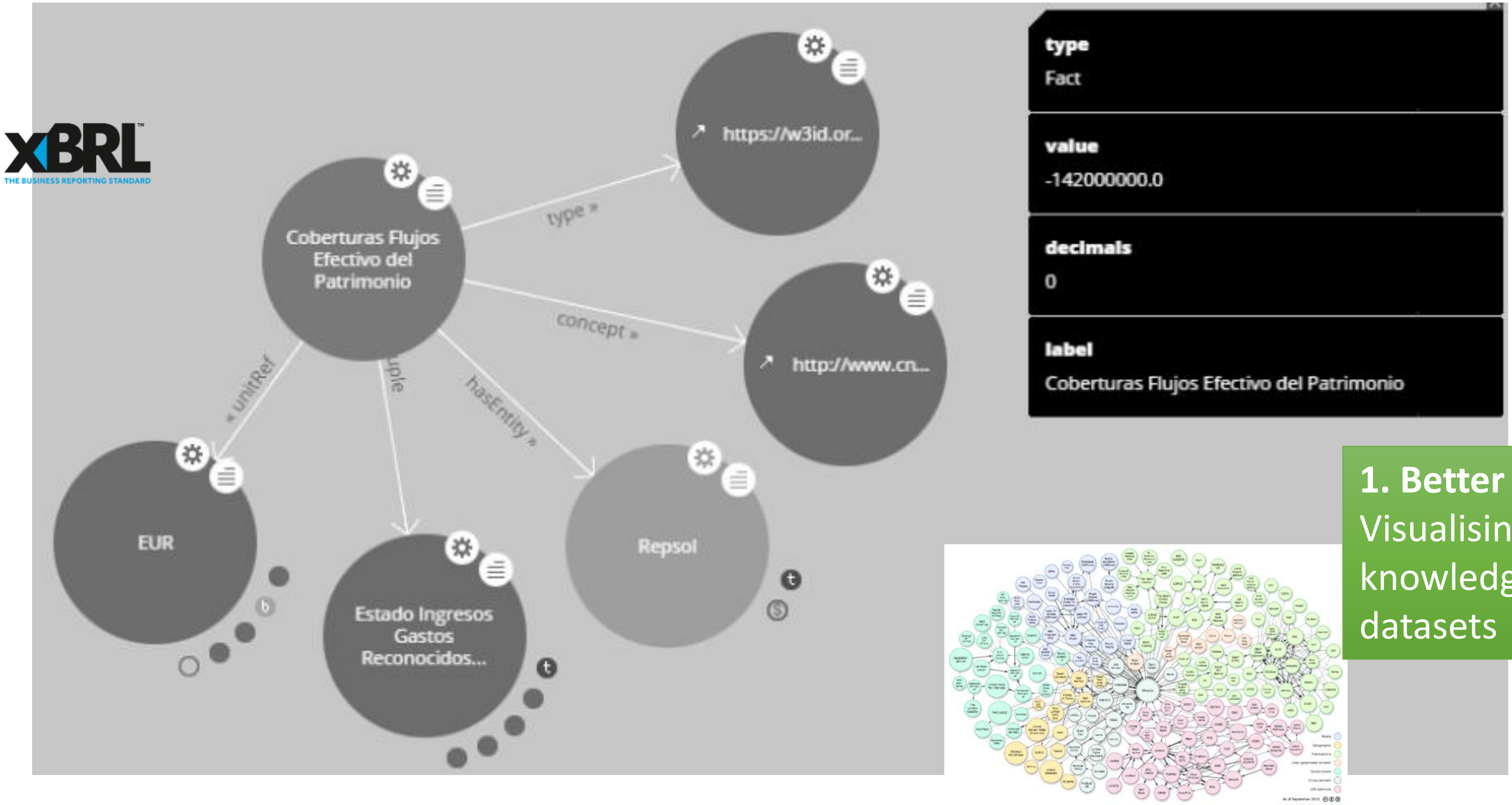
SPARQL endpoint

XBRL into Linked data

Interlinking

Data publication

Applicability


Coberturas Flujos Efectivo del Patrimonio
type »
https://w3id.or...
concept »
http://www.cn...
« unitRef
hasEntity »
EUR
Estado Ingresos Gastos Reconocidos...
Repsol

type
Fact

value
-142000000.0

decimals
0

label
Coberturas Flujos Efectivo del Patrimonio

XBRL THE BUSINESS REPORTING STANDARD

lodlive browsing the web of data

semantic web framework Jena

1. Better data contextualisation
Visualising XBRL data as Graphical knowledge connected to other datasets

XBRL into Linked data

Interlinking

Data publication

Applicability

Goal 1. Data Coverage Using DBpedia

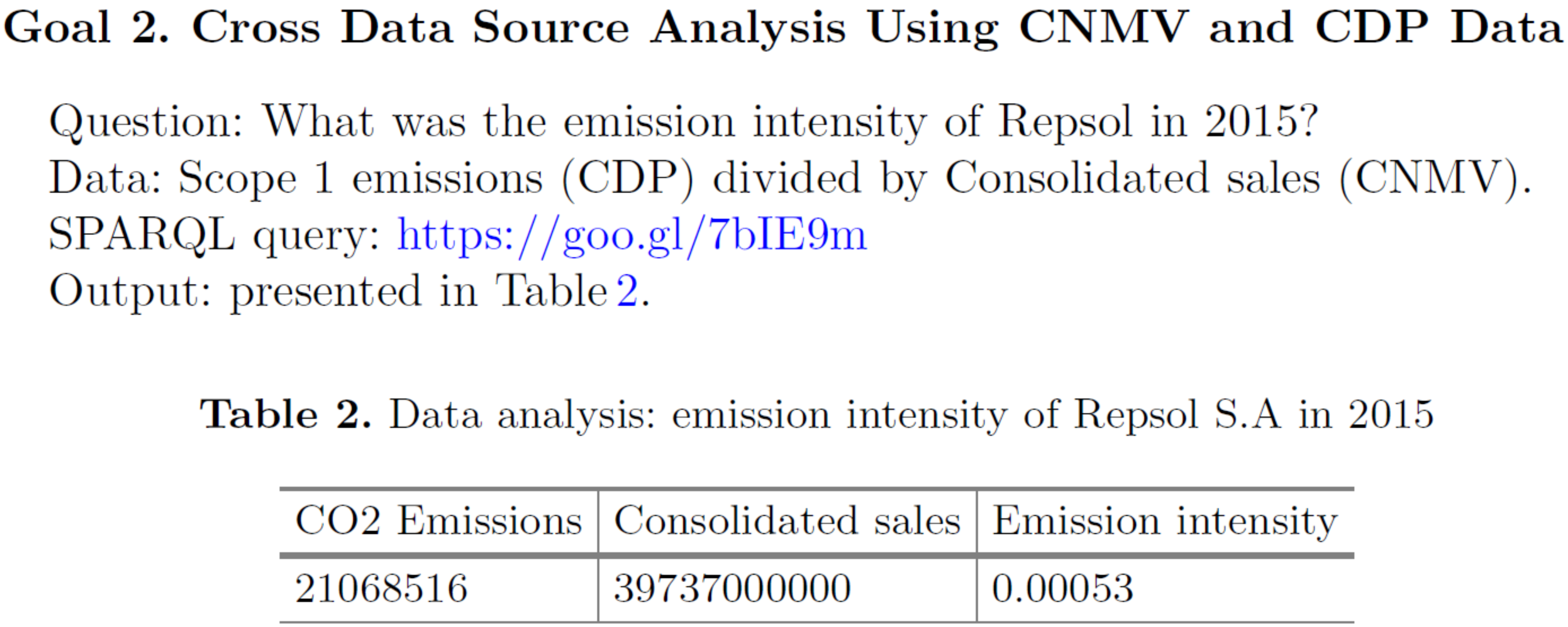
Question: What is the context of the company Repsol?
Data: Abstract, subsidiary and industry.
SPARQL query: https://goo.gl/if8ydG
Output: presented in Table 1.

1. Better data contextualisation
Accessing to more data about a company

Table 1. Data coverage: information about the context of Repsol S.A

URL	Abstract	Subsidiary	Industry
http://dbpedia.org/resource/Repsol	Repsol S.A. is an integrated global energy company based in Madrid, Spain. It carries out upstream and downstream activities throughout the entire world. It has more than 24,000 employees worldwide	http://dbpedia.org/resource/Petronor	http://dbpedia.org/resource/Petroleum_industry

XBRL into Linked data → Interlinking → Data publication → **Applicability**



Goal 2. Cross Data Source Analysis Using CNMV and CDP Data

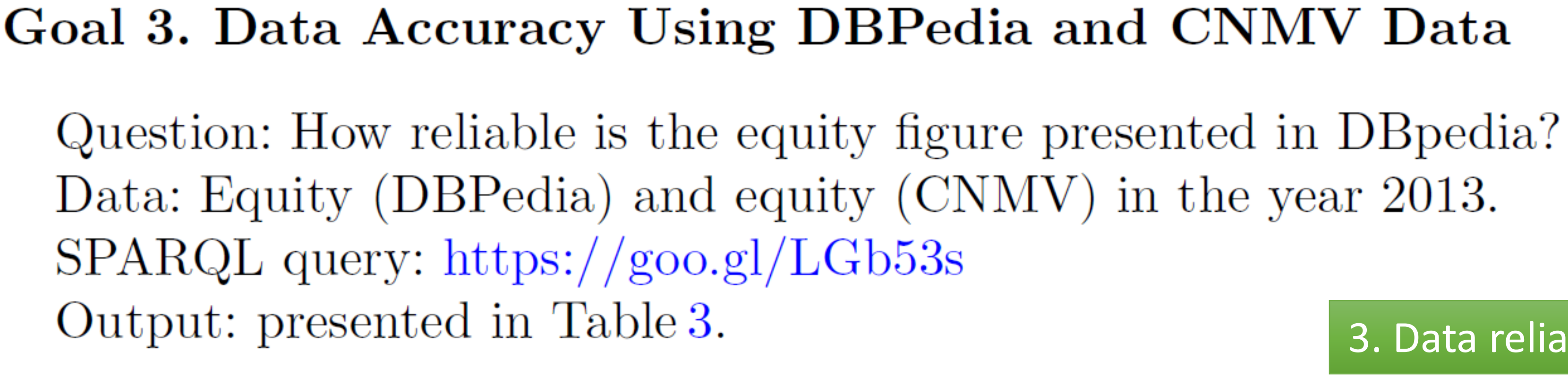
Question: What was the emission intensity of Repsol in 2015?
Data: Scope 1 emissions (CDP) divided by Consolidated sales (CNMV).
SPARQL query: https://goo.gl/7bIE9m
Output: presented in Table 2.

Table 2. Data analysis: emission intensity of Repsol S.A in 2015

CO2 Emissions	Consolidated sales	Emission intensity
21068516	39737000000	0.00053

2. Cross-data source analysis
(Emissions intensity figure:
Consolidations sales / CO2 emissions)

XBRL into Linked data
Interlinking
Data publication
Applicability



Goal 3. Data Accuracy Using DBPedia and CNMV Data

Question: How reliable is the equity figure presented in DBpedia?
Data: Equity (DBPedia) and equity (CNMV) in the year 2013.
SPARQL query: https://goo.gl/LGb53s
Output: presented in Table 3.

3. Data reliability and data consistency

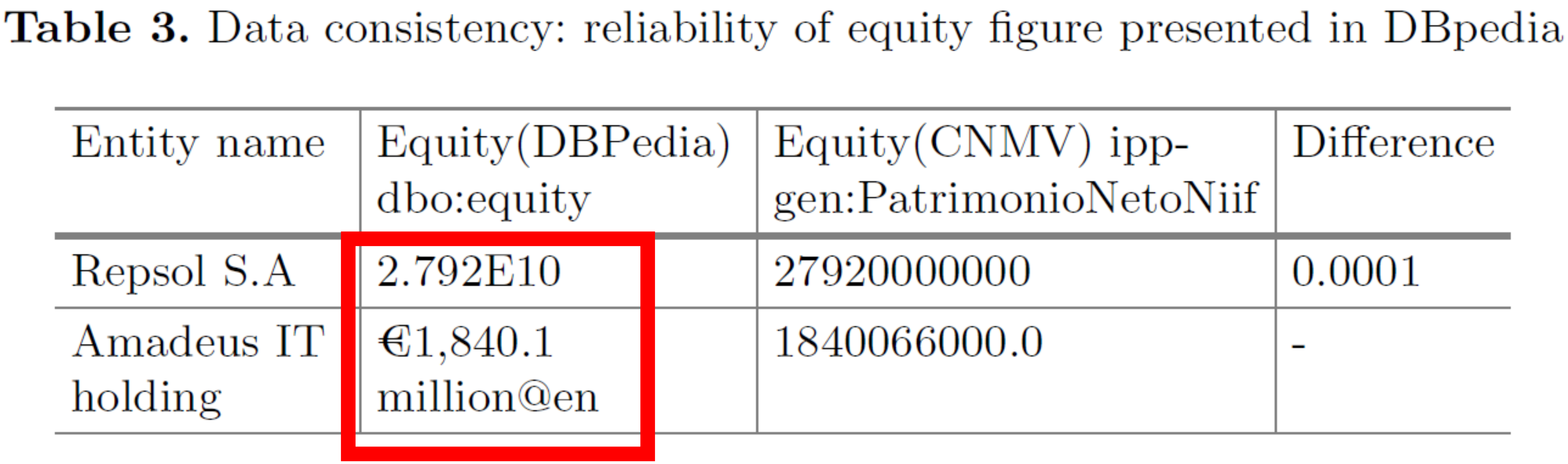


Table 3. Data consistency: reliability of equity figure presented in DBpedia

Entity name	Equity(DBPedia) dbo:equity	Equity(CNMV) ipp-gen:PatrimonioNetoNiif	Difference
Repsol S.A	2.792E10	27920000000	0.0001
Amadeus IT holding	€1,840.1 million@en	1840066000.0	-

Conclusions

- **Corporate Transparency needs** technologies to enable connectivity between existing company data from different domains (financial and non-financial) and formats.

- **Linked data principles** can encourage better corporate data publication and therefore data analysis.

- **XBRL** enables a standard and accurate representation of corporate data with advanced validation rules.

- **XBRL and Linked data can complement to each other.**

- **Our work is focused on** (1) applying Linked Data practices and tools on existing Financial and non-financial XBRL datasets and (2) to show applicable results.

- **Our potential academic contribution:** a generic Ontology to translate any XBRL filing into Linked data
  - Covering the full XBRL specification
  - Keeping correspondences with well-known vocabularies (schema.org, Registered Organization Vocabulary, FOAF)

- **Our potential industrial contribution:** our work brings value and enables use from existing company data in a connected way: financial + non-financial data + Linked Open Datasets.

Thank you.

maria.mora@bristol.ac.uk
maria.mora@cdp.net