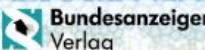


Integration and Exploration of Financial Data using Semantics and Ontologies

Roberto García

HCI and Data Integration Research Group
Universitat de Lleida, Spain

mit Unterstützung von
with the support of



EUROPEAN
BANKING
AUTHORITY



XBRL | EUROPE



EUROFILING XBRL WEEK IN FRANKFURT 6-7-**8**-9 JUNE 2017

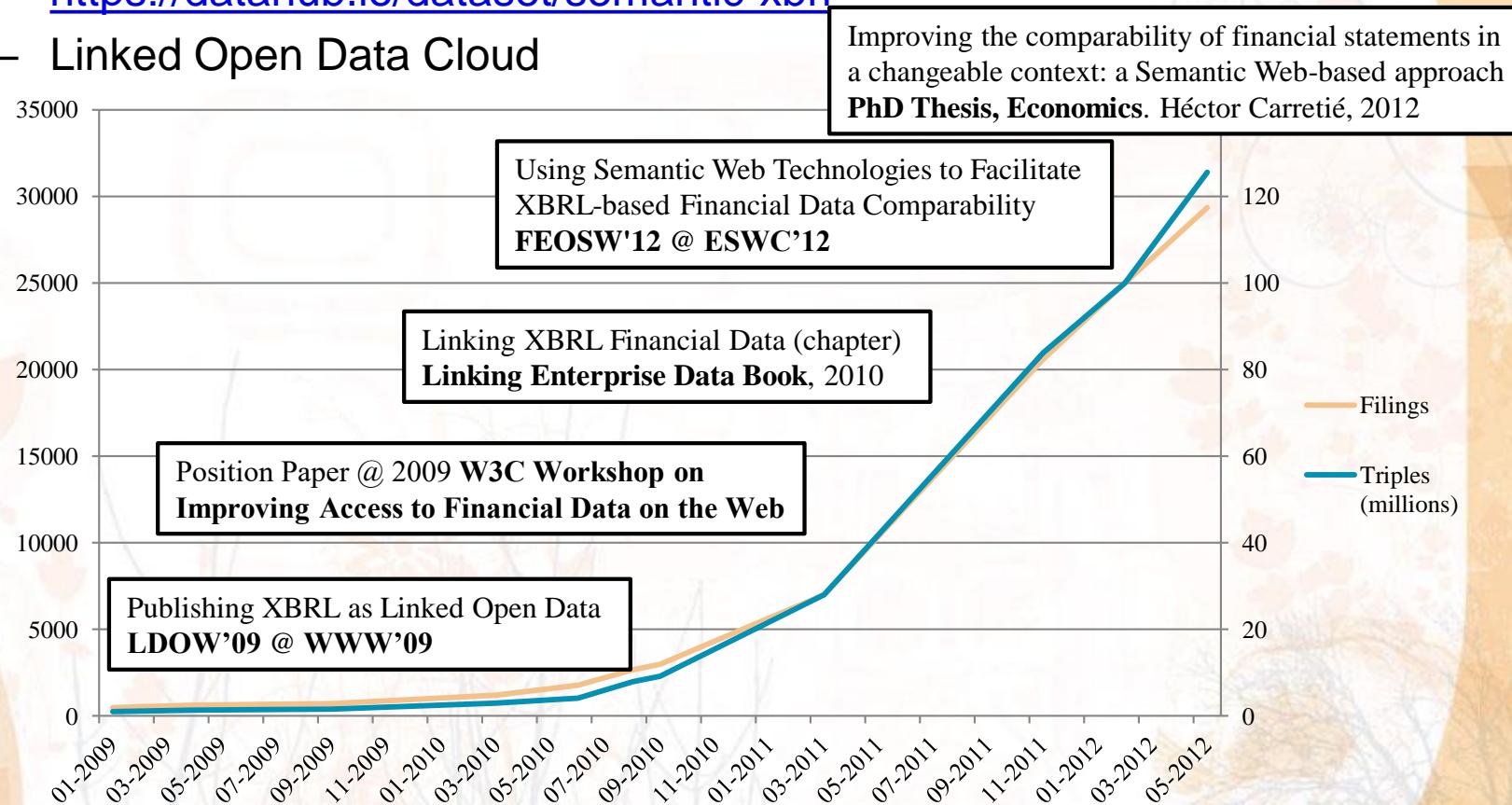
19th XBRL Europe day | Eurofiling 23rd workshop | Tutorials | Academic Track

Introduction

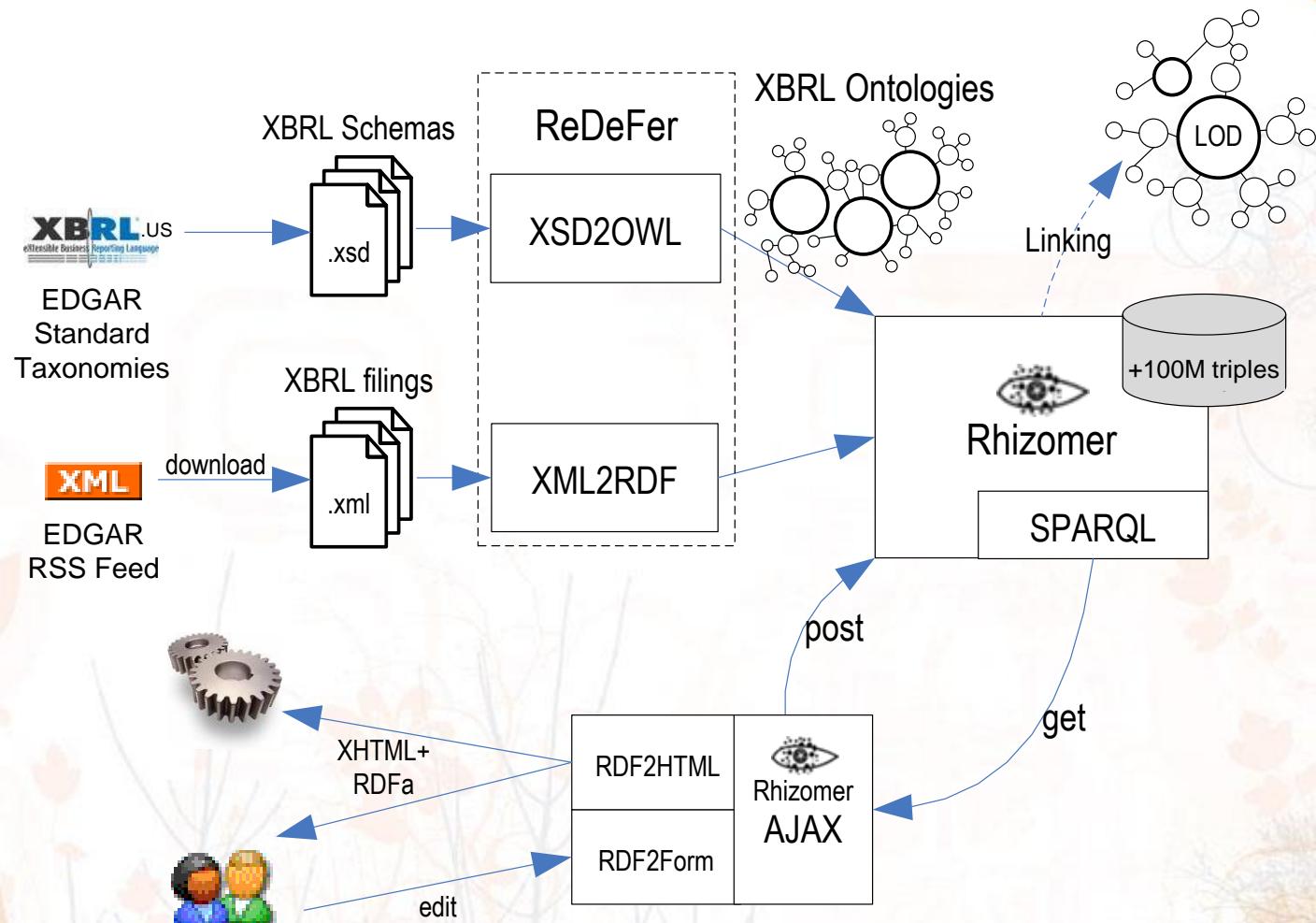
- **Report experience** first attempt to map “significant” amount of **XBRL to RDF**
 - 2008 +100M triples... and as far as we know
- **Test** Semantic Web and Linked Data for financial data **integration and exploration**
- **Convert it, map** XBRL to RDF
 - Apply generic mapping: [ReDeFer](#)
 - XSD2OWL, XML2RDF
- Once as RDF+OWL, improve **comparability** using semantic links

Semantic XBRL

- Dataset size (from US SEC)
 - May 2012: **125,55 million triples** from **29342 XBRL filings**
 - <https://datahub.io/dataset/semantic-xbrl>
 - Linked Open Data Cloud

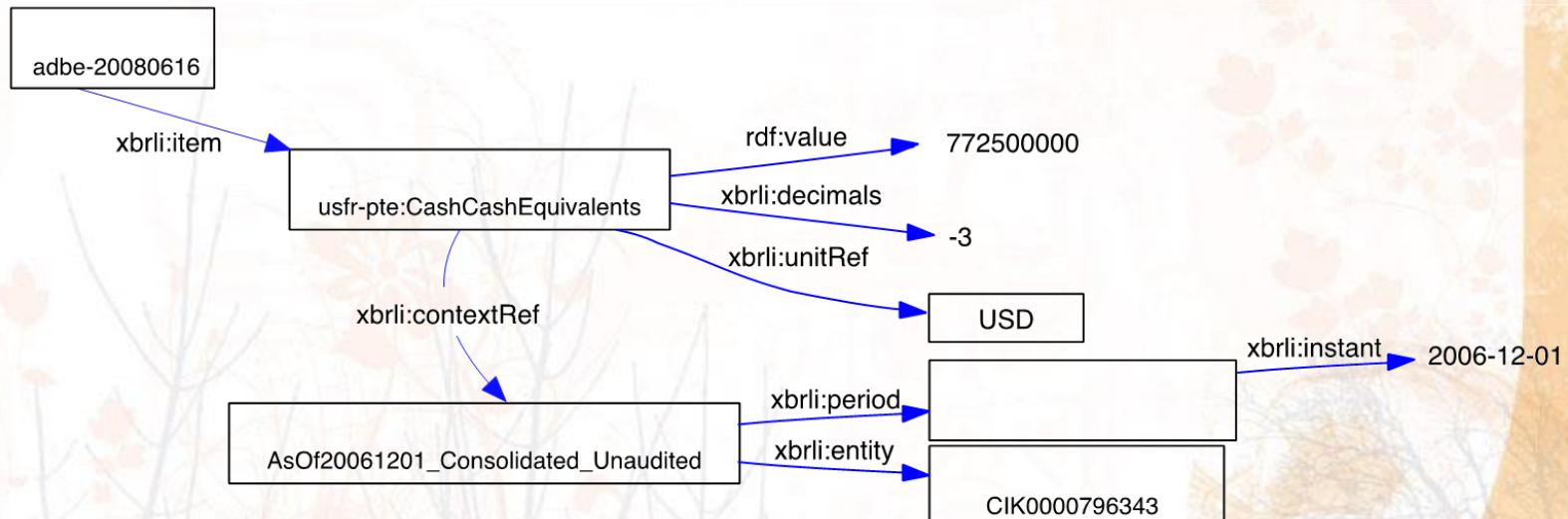


Approach



XBRL XML to RDF

- ReDeFer [XML2RDF](#),
model XML **tree** using **triples**
 - xsd:element and xsd:attribute → rdf:Property
 - xbrli:id and xbrli:identifier → rdf:Resource ID
 - Other resources, anonymous

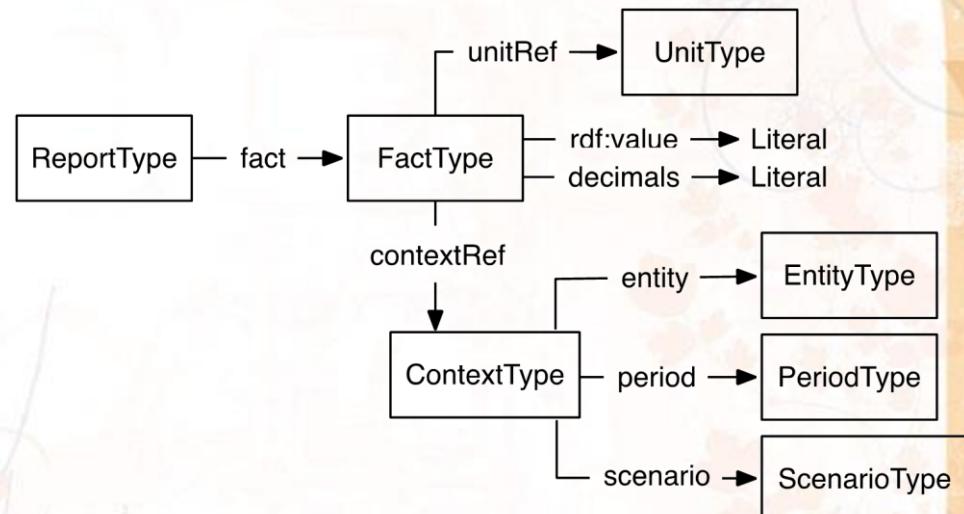


XBRL XSD to OWL

- XBRL Schemas: XBRL 2.1, US GAAP¹, Spanish PGC...
- ReDeFer [XSD2OWL](#)

XSD2OWL mappings

XML Schema	OWL
element attribute	rdf:Property owl:DatatypeProperty owl:ObjectProperty
element@substitutionGroup	rdfs:subPropertyOf
element@type	rdfs:range
complexType	owl:Class
complexType//element	owl:Restriction
extension@base restriction@base	rdfs:subClassOf
@maxOccurs, @minOccurs	owl:maxCardinality, owl:minCardinality
sequence, choice	owl:intersectionOf, owl:unionOf

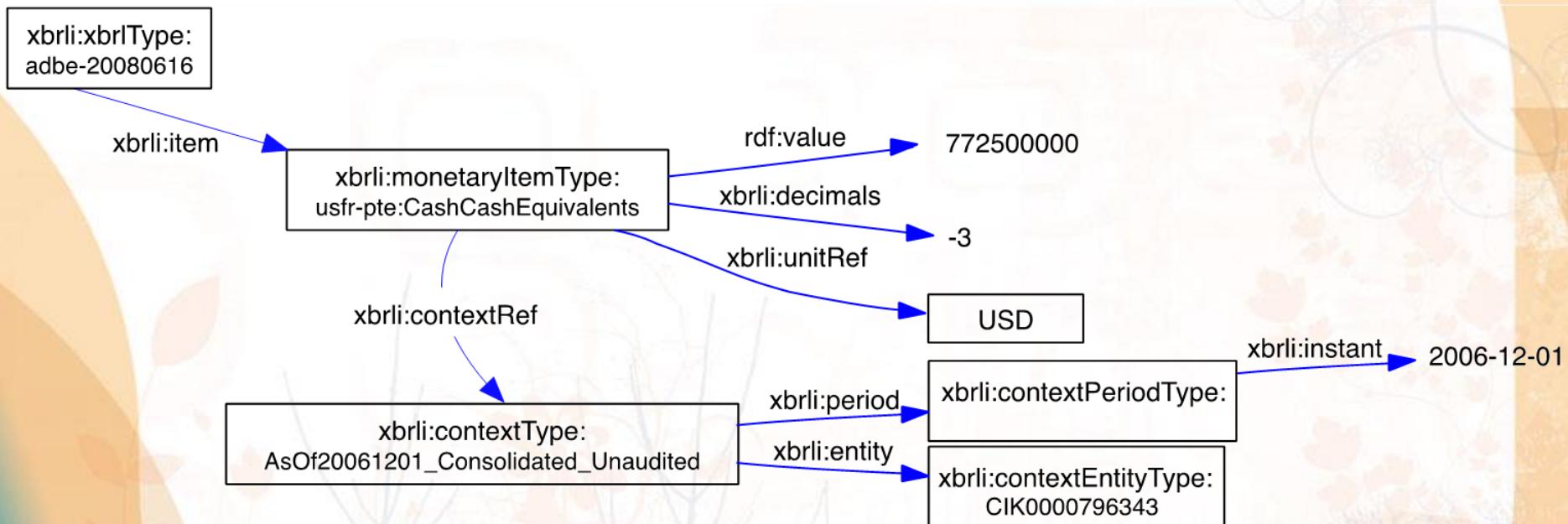


Core classes and properties for XBRL Instance

¹ Ontologies available from <http://rhizomik.net/ontologies/bizontos>

XML2RDF plus XSD2OWL

- Enrich RDF with links to **classes** for corresponding XSD ***complexType***:



Comparability

- Generate **semantic links** across accounting standards (XBRL taxonomies)
 - Automatic ontology alignment tools
 - Experts input
 - Numeric analysis
- Experiment:
 - Telefonica S.A. Balance Sheet 2008 and 2009
 - XBRL for Spanish CNMV
 - XBRL for US SEC

Semantic Links US SEC - Spanish CNMV

Telefónica S.A. Balance sheet filled to **US SEC** (th. of €)

ASSETS	2009
A) NON-CURRENT ASSETS	84,311
Intangible assets	15,846
Goodwill	19,566
Property, plant and equipment	31,999
Investment properties	5
Investments in associates	4,936
Non-current financial assets	5,988
Deferred tax assets	5,971
B) CURRENT ASSETS	23,83
Inventories	934
Trade and other receivables	10,622
Current financial assets	1,906
Tax receivables	1,246
Cash and cash equivalents	9,113
Non-current assets held for sale	9
TOTAL ASSETS (A + B)	108,141

Telefónica S.A. Balance sheet filled before **Spanish CNMV** (th. of €)

ACTIVOS	2009	Diff. 2009
A) ACTIVO NO CORRIENTE	84.311	
1. Inmovilizado intangible:	35.412	
a) Fondo de comercio	19.566	
b) Otro inmovilizado intangible	15.846	
2. Inmovilizado material	31.999	
3. Inversiones inmobiliarias	5	
4. Inversiones en empr. grupo y asoc. L/P	4.936	
5. Inversiones financieras a largo plazo	5.499	5.988 €
6. Activos por impuesto diferido	5.971	
7. Otros activos no corrientes	489	
B) ACTIVO CORRIENTE	23.830	
1. Activos no corrientes mantenidos para la venta	9	
2. Existencias	934	
3. Deudores comerciales y otras cuentas a cobrar:	9.718	
a) Clientes por ventas y prestaciones de servicios	8.288	10.622 €
b) Otros deudores	2.334	
c) Activos por impuesto corriente	-	903
4. Otros activos financieros corrientes	1.906	
5. Otros activos corrientes	2.150	
6. Efectivo y otros activos líquidos equivalentes	9.113	
TOTAL ACTIVO (A + B)	108.141	

Spanish CNMV (PGC taxonomies)	US SEC (IFRS taxonomies)	Semantic Mappings
ipp-gen: ActivoNoCorrienteNiif 84.311 €	ifrs: NoncurrentAssets 84.311 €	ipp-gen:ActivoNoCorrienteNiif owl:equivalentClass ifrs:NoncurrentAssets
ifrs-gp: TradeAndOtherReceivablesNet Current = ipp-gen: ClientesVentasPrestaciones Servicios + ipp-gen: OtrosDeudores 8.288€ + 2.334€	ifrs: TradeAndOtherCurrentReceivables 10.622€	<p>ifrs-gp:TradeAndOtherReceivablesNetCurrent owl:equivalentClass ifrs:TradeAndOtherCurrentReceivables</p> <pre> CONSTRUCT { [] a ifrs-gp: TradeAndOtherReceivablesNetCurrent; xbrli:contextRef ?context; xbrli:unitRef ?unit; xbrli:decimals ?decimals; rdf:value ?value. } WHERE { ?cvps a ipp-gen: ClientesVentasPrestacionesServicios; xbrli:contextRef ?context; xbrli:unitRef ?unit; xbrli:decimals ?decimals; rdf:value ?cvps-value. ?od a ipp-gen:OtrosDeudores; xbrli:contextRef ?context; xbrli:unitRef ?unit; xbrli:decimals ?decimals; rdf:value ?od-value. } BIND(?cvps-value+?od-value AS ?value) }</pre>

PGC Fact and Value (thousands of €)	IFRS Equivalent Query
Assets	
ipp-gen:ActivoNoCorrienteNiif = 84.311 €	Get ifrs:NoncurrentAssets
ifrs-gp:IntangibleAssetsNet = 35.412 €	No equivalent
ifrs-gp:GoodwillNet	
ifrs-gp:IdentifiedIntangibleAssets	
ifrs-gp:PropertyPlantEquipment	
ifrs-gp:InvestmentProperties	
ifrs-gp:EquityMethodInvestments	
ifrs-gp:OtherFinancialAssets	
+ ifrs-gp:OtherAssets	
ifrs-gp:DeferredTaxAssets	
ipp-gen:ActivoNoCorriente	
ifrs-gp:NonCurrentFinancialAssets	
ifrs-gp:Inventories	
ipp-gen:ClientesVendidos	
ipp-gen:OtrosDeudores	
ifrs-gp:TradeAndOtherReceivables	
ifrs-gp:OtherFinancialAssetsCurrent	No equivalent
ipp-gen:OtrosActivosCorrientes (2.150€) - ifrs-gp:CurrentTaxReceivables (903€) = 1.246€	Get ifrs:CurrentTaxAssets
ifrs-gp:CashAndCashEquivalents = 9.113€	Get ifrs:CashAndCashEquivalents
ipp-gen:TotalActivoNiif	No equivalent

DESCRIBE ?r
 WHERE {
 ?r a ifrs:NoncurrentAssets }

DESCRIBE ?r
 WHERE {
 ?r a
 ifrs:OtherNoncurrentFinancialAssets }

Conclusions

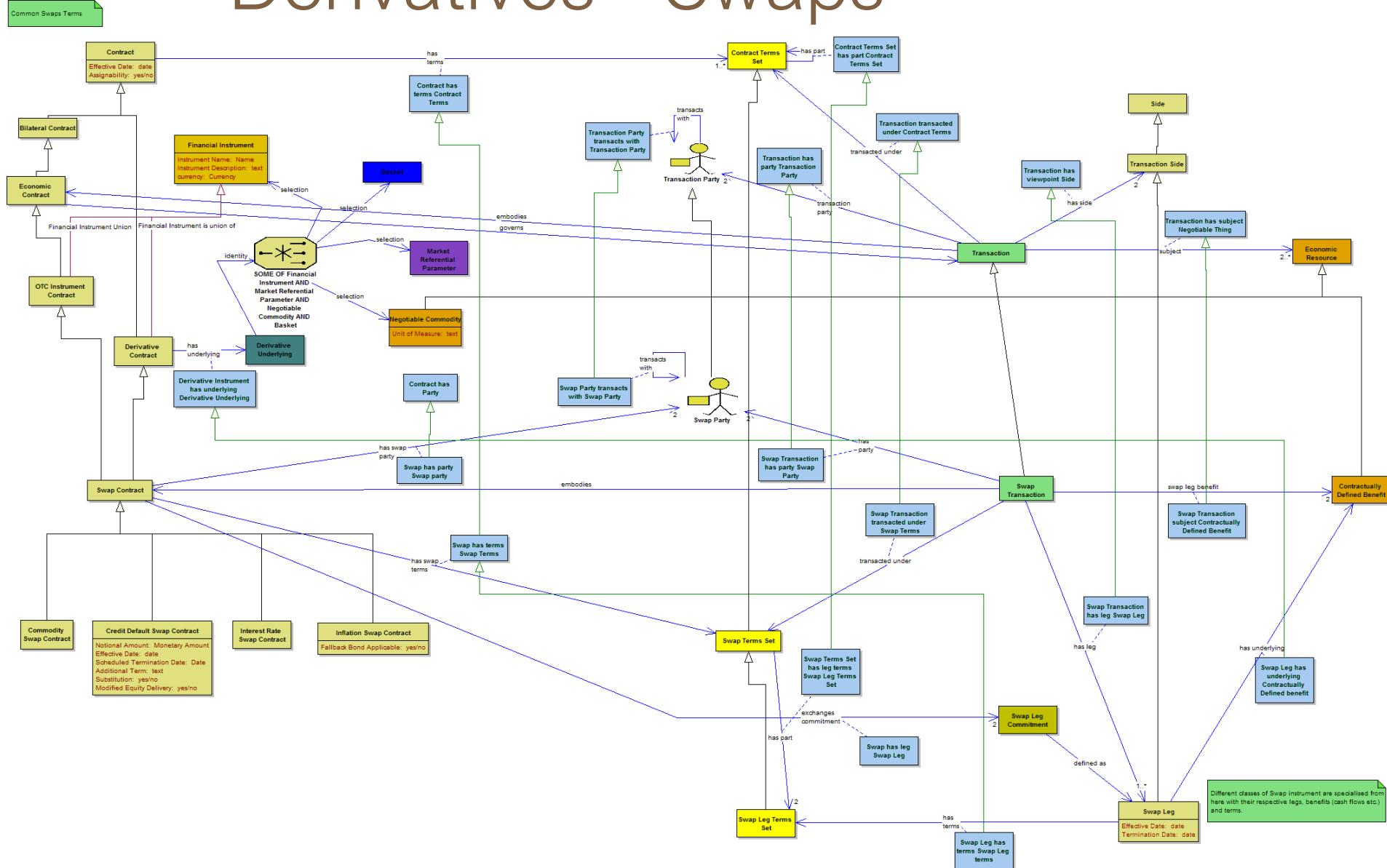
- **XBRL tools:** for financial information edition and maintenance (formulae,...)
XML2RDF  XSD2OWL
- **Semantic Web tools:** data integration, cross querying,... **COMPARABILITY**
 - Across filings, companies, accounting principles,...
- However, **lack of impact** of Semantic XBRL
 - Project discontinued 2013
 - Too early?
 - Too much semantics... or too little?

Financial Industry Business Ontology

- FIBO™ open industry standard for defining financial contracts terms, facts and relationships
- 11 core domains, 49 modules and 418 ontology files:
 - Foundations
 - Business Entities
 - Indices and Indicators
 - Securities Reference Terms
 - Derivatives
 - Loans
 - Market Data
 - Collective Investment Vehicles
 - Corporate Actions
 - Payments
 - Portfolios and Holdings
 - ...

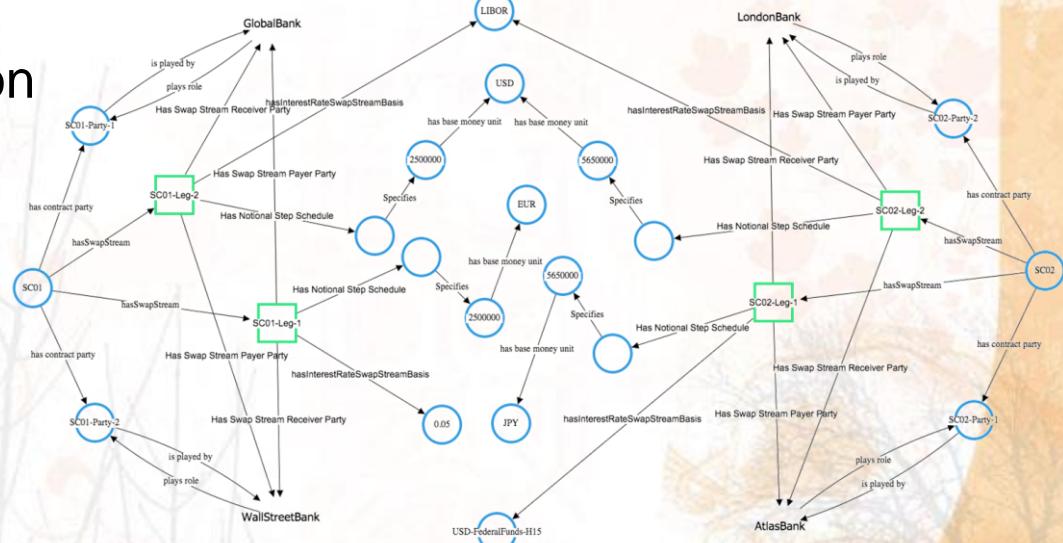


Derivatives - Swaps



Experience with US Banks

- Consulting for Cambridge Semantics
 - <http://www.cambridgesemantics.com>
- Banks interest in **heavyweight semantics**, data integration for regulators reports automation
- Proof of Concept with Interest Rate Swaps
 - Modelling
 - Exploration
 - Automatic classification (reasoning)
 - Cross-currency swaps
 - Single-currency swaps
 - Float-Fixed swaps
 - ...



Experience with US Banks

ANZO ON THE WEB

CAMBRIDGE SEMANTICS

The image displays two screenshots of the ANZO platform interface, which integrates semantic web technologies from Cambridge Semantics.

Top Screenshot: This screenshot shows a bar chart titled "Swaps Network" with three bars. The Y-axis is labeled "hasNotionalStepSchedule/s" and ranges from 5.000.000 to 9.000.000. The X-axis categories are not explicitly labeled but correspond to the three bars. The chart is part of a dashboard titled "Swaps Network". On the left, there are filters for "Linked Data Sets", "Data Types", "Fixed Rates", and "Floating Rate". Under "Floating Rate", "USD-FederalFunds-H15" and "LIBOR" are selected. Below these are filters for "Amount" (Min: 1.000.000, Max: 8.531.000) and "Currency" (USD, JPY, EUR). A sidebar on the right contains a search bar and a "Properties" panel.

Bottom Screenshot: This screenshot shows a detailed view of a swap stream entity named "SC03-Leg-1". The "Properties" panel shows tabs for "Info" and "Filters". The "Info" tab displays the ID as "SC03-Leg-1" and the label as "SC03-Leg-1". The "Filters" tab lists various predicates and their values, such as "swap contract", "Has Swap Stream", "Has Swap Stream Payer Party", "Has Swap Stream Receiver Party", "Has Notional Step Schedule", and "Has Interest Rate Swap Stream Basis". To the right of the properties panel is a semantic graph visualization. The graph features nodes represented by circles and squares, and edges represented by arrows. A central node labeled "SC03-Leg-1" is highlighted with a green border. Arrows point from this node to other nodes, labeled with predicates like "Has Swap Stream Payer Party", "Has Swap Stream Receiver Party", "Has Notional Step Schedule", "Has Interest Rate Swap Stream Basis", and "Has Swap Stream". Other nodes include a square node, a circle node labeled "Specifies", and a circle node labeled "has contract party". A toolbar at the top of the graph interface includes buttons for "run", "undo", "redo", "open", "save", "print", "export", "find", "layout", "options", and various node and edge manipulation tools.

Thank you for your attention Questions?

Roberto García

@rogargon

roberto.garcia@udl.cat

<http://rhizomik.net/~roberto/>



HCI and Data Integration Research Group
Universitat de Lleida, Spain

