

# Ontological Challenges for Financial Information: Lessons from the Rhizomik Initiative

Workshop on Improving Access to Financial Data on the Web  
5-6 October 2009 - Arlington, Virginia USA

Roberto García

[rgarcia@diei.udl.cat](mailto:rgarcia@diei.udl.cat)

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



HCI & Data Integration Research Group  
Universitat de Lleida, Spain



# Introduction

- **Explore** Semantic Web technologies for financial information
- Most public **financial data** available as XBRL
- **Reuse** this data, **map** XBRL to RDF
  - Apply generic mapping: [ReDeFer](#)
    - XSD2OWL, XML2RDF
- Publish as Linked Data using [Rhizomer](#)

# Approach

- Direct, non-intrusive
- Reuse XBRL data and schemas
- Facilitate roundtrip:
  - XBRL → SW → XBRL
- Showcase Semantic Web benefits
- Test data: SEC's EDGAR program

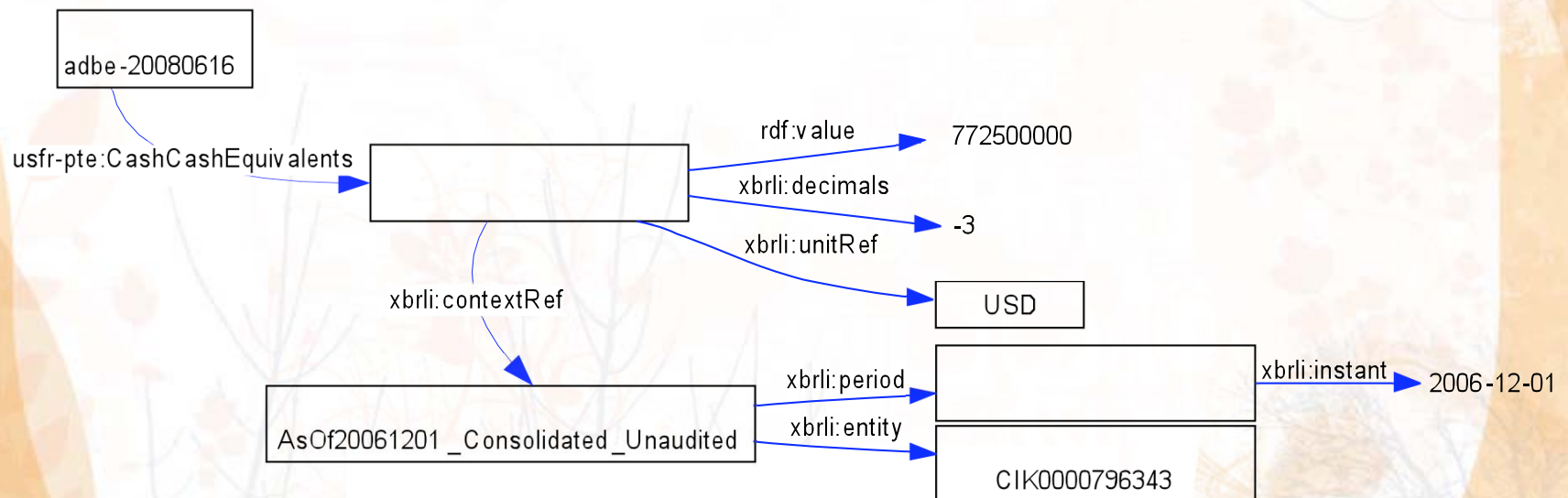
# ReDeFer

- Part of the Rhizomik initiative
- ReDeFer (in and out Semantic Web)
  - RDF2HTML+RDFa
  - RDF2SVG
  - **XML2RDF**
  - **XSD2OWL**

Use cases: MPEG-7, MPEG-21, ODRL,...

# 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



Rendered by [RDF2SVG](#)

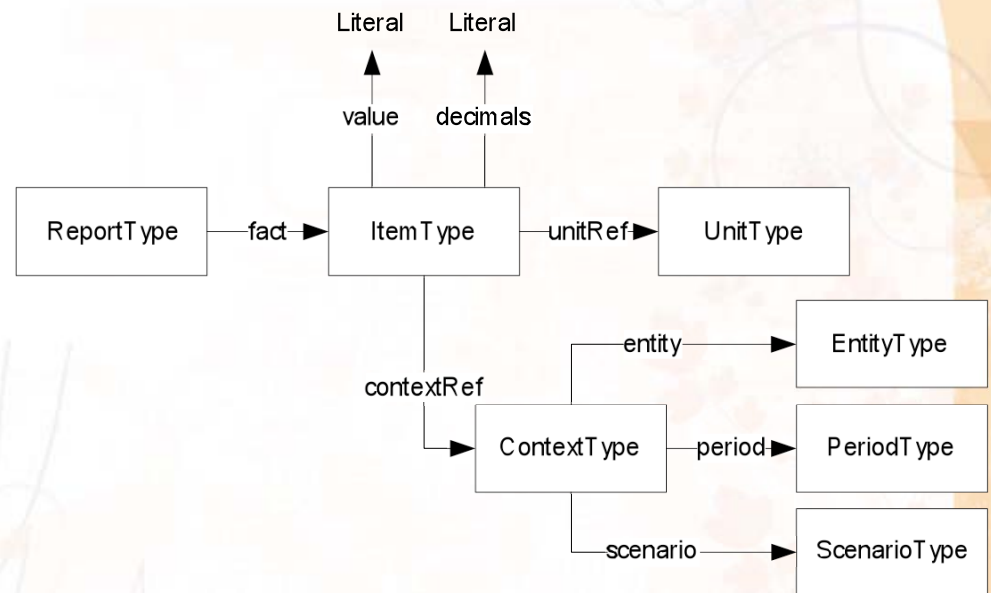


# XBRL XSDs to OWL

- XBRL Schemas: XBRL 2.1, US GAAP<sup>1</sup>,...
- 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

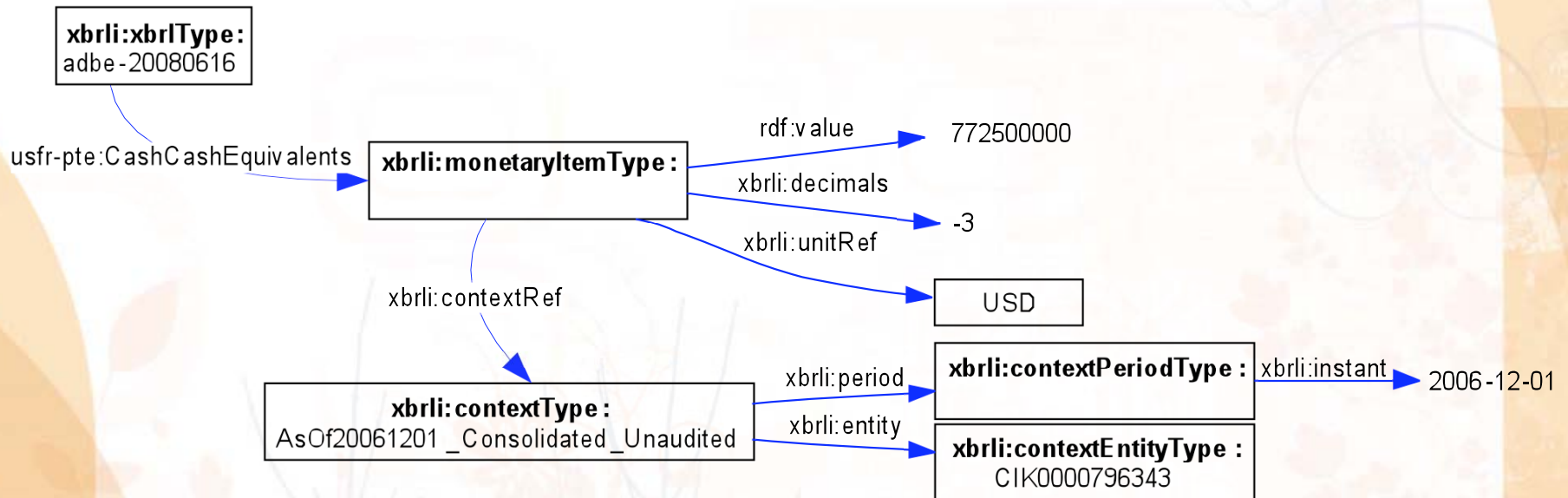


Core classes and properties for XBRL Instance

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

# XML2RDF plus XSD2OWL

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



# XBRL Sources

- SEC's EDGAR filings data files:

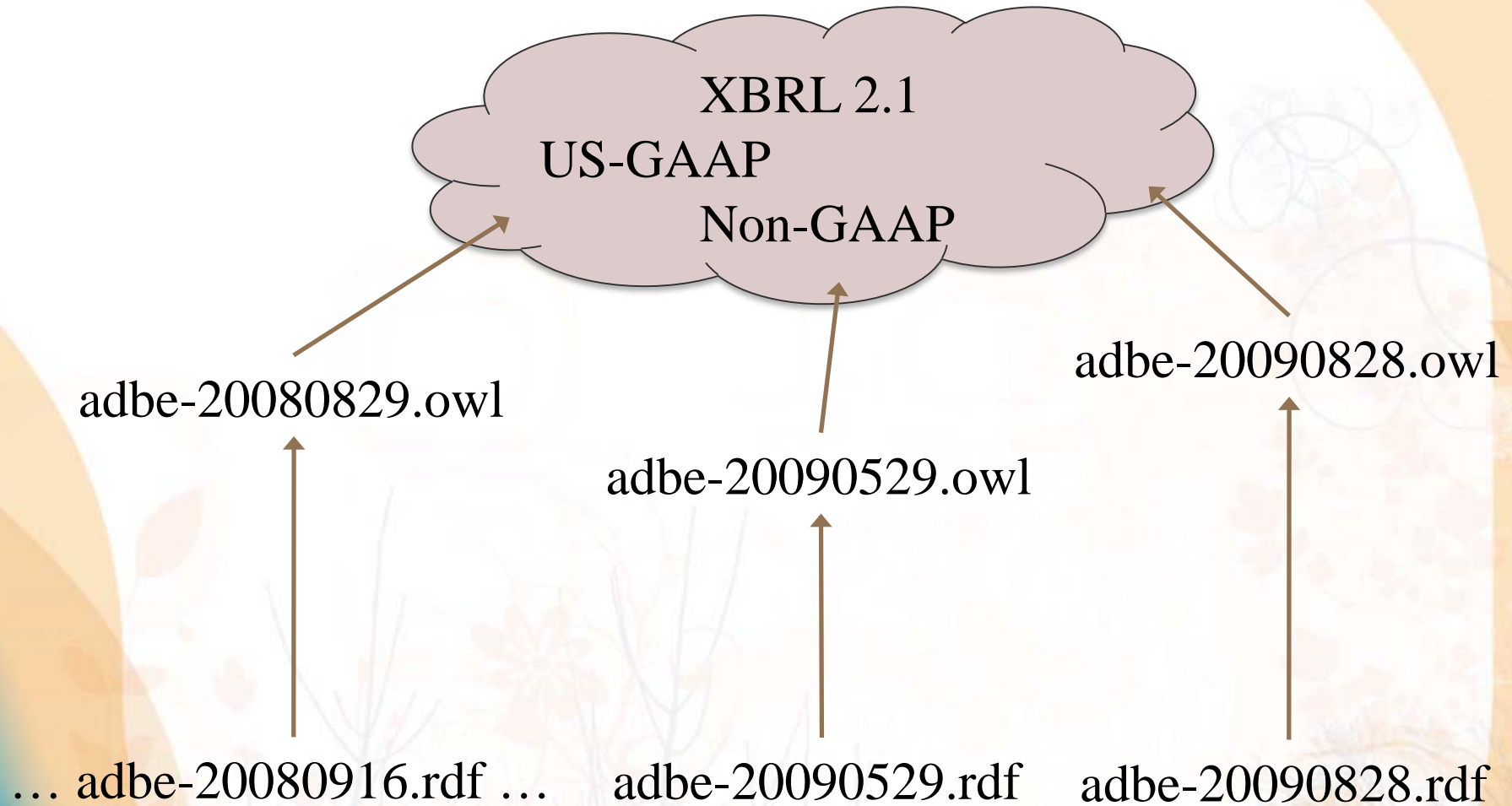
Description	Document	Type	Size	
XBRL INSTANCE DOCUMENT	<a href="#">adbe-20090828.xml</a>	EX-101.INS	1349030	⇒ XML2RDF
XBRL TAXONOMY EXTENSION SCHEMA DOCUMENT	<a href="#">adbe-20090828.xsd</a>	EX-101.SCH	88847	⇒ XSD2OWL
XBRL TAXONOMY EXTENSION CALCULATION LINKBASE DOCUMENT	<a href="#">adbe-20090828_cal.xml</a>	EX-101.CAL	108182	⇒ Calculations
XBRL TAXONOMY EXTENSION LABELS LINKBASE DOCUMENT	<a href="#">adbe-20090828_lab.xml</a>	EX-101.LAB	688129	⇒ Labels
XBRL TAXONOMY EXTENSION PRESENTATION LINKBASE DOCUMENT	<a href="#">adbe-20090828_pre.xml</a>	EX-101.PRE	343710	⇒ Presentation
XBRL TAXONOMY EXTENSION DEFINITION LINKBASE DOCUMENT	<a href="#">adbe-20090828_def.xml</a>	EX-101.DEF	75107	⇒ Definition



<http://www.sec.gov/Archives/edgar/xbrlrss.xml>  
<http://www.sec.gov/Archives/edgar/usgaap.rss.xml>



# XBRL Ontologies



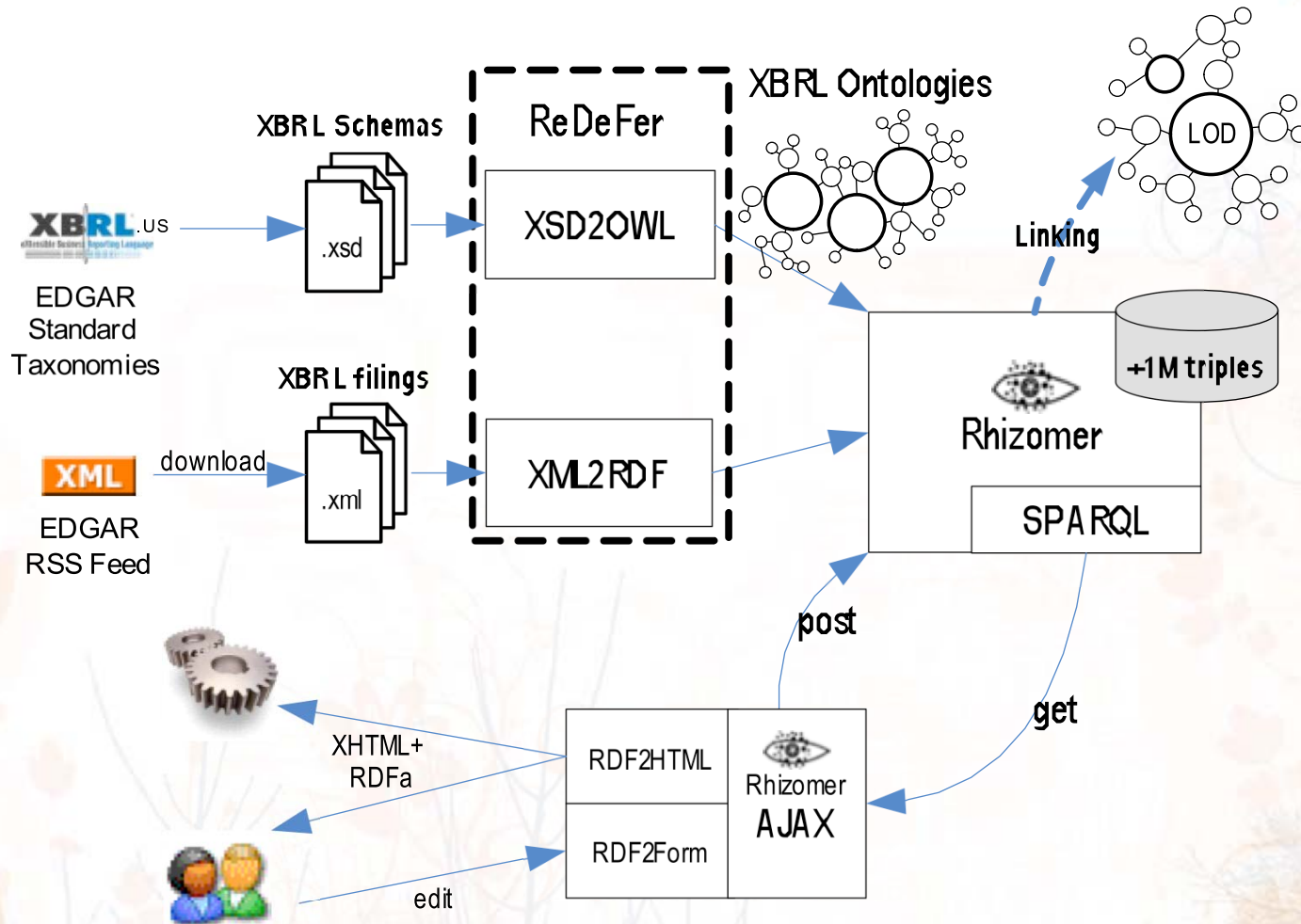
# Ontology Alignment

- Integrate ontologies for different filings
  - E.g.:  
*adbe-20080530.owl:InvestmentLeaseReceivable = adbe-20080829.owl:Investment**In**LeaseReceivable*
  - Ontology alignment tools (edit distance)
  - Perform queries across filings

# Semantic XBRL

- Dataset size
  - October 2009: **1,54 million triples** from **716 XBRL filings**.
- Linked Data:
  - Entities: companies in DBPedia, use name or Central Index Key (CIK)
  - Units: e.g. USD → [http://dbpedia.org/resource/United\\_States\\_dollar](http://dbpedia.org/resource/United_States_dollar)

# Architecture

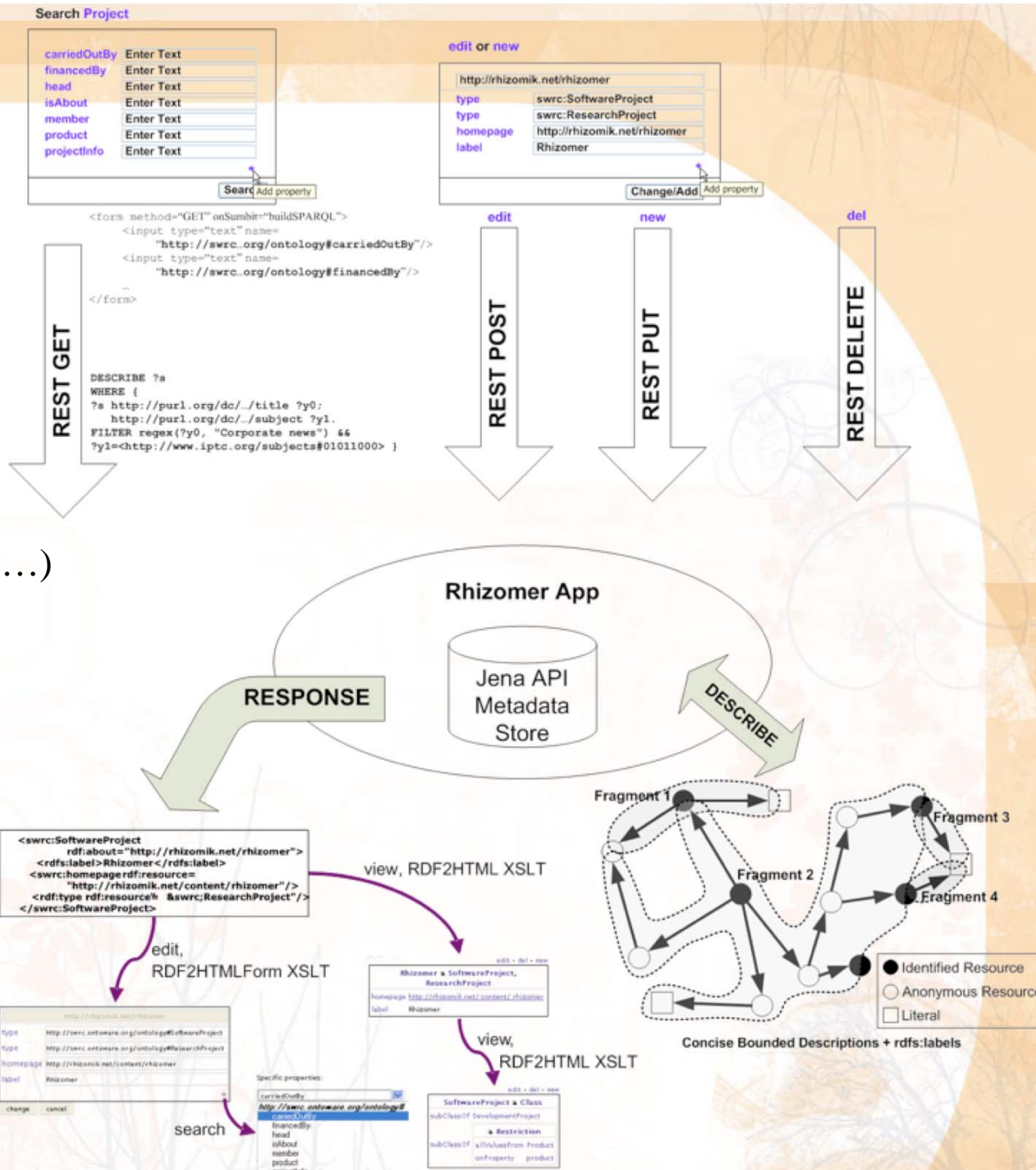


# Prototype Demo

## Rhizomer:

- publish
- query
- browse
- edit
- mashup (map, timeline,...)
- etc.

...Linked Data





# Get a context description from its URL

[http://rhizomik.net/semanticxbrl/adbe-20081216/From20071201-To20081128\\_None\\_None\\_StatementOperatingActivitiesSegmentAxis\\_KnowledgeWorkerSolutions](http://rhizomik.net/semanticxbrl/adbe-20081216/From20071201-To20081128_None_None_StatementOperatingActivitiesSegmentAxis_KnowledgeWorkerSolutions)



The image shows the top navigation bar of the Rhizomik website. On the left is the Rhizomik logo, which consists of the word "Rhizomik" in a serif font next to a stylized eye icon with a globe inside. To the right of the logo, the text "rhizomer | logout" is displayed. Below this, a horizontal menu contains the items "Rhizomik - Semantic XBRL - XBRL Ontologies - Contact". The background of the navigation bar features a faint, abstract graphic of a globe and network connections.

HTML

[edit | delete | history]

There isn't HTML content associated with this page.

back - go to... - forward

edit - new - del

**[From20071201-To20081128\\_None\\_None\\_StatementOperatingActivitiesSegmentAxis\\_KnowledgeWorkerSolutions](#) a contextType**

entity CIK0000796343

**a contextPeriodType**

period endDate 2008-11-28

startDate 2007-12-01

Referrers



Powered by Rhizomik



# Reuse Wikipedia data (DBpedia)

[http://rhizomik.net/semanticxbrl/?query=DESCRIBE<http://dbpedia.org/resource/Adobe\\_Systems>](http://rhizomik.net/semanticxbrl/?query=DESCRIBE<http://dbpedia.org/resource/Adobe_Systems>)

<b>Adobe Systems a ComputerCompaniesOfTheUnitedStates, CompaniesEstablishedIn1982, Company, CompaniesBasedInSiliconValley, Company108058098, Organisation, Resource, SoftwareCompaniesOfTheUnitedStates, CompaniesListedOnNASDAQ</b>	
abstract	<p>Adobe Systems Incorporated is an American computer software company headquartered in San Jose, California, USA. The company has historically focused upon the creation of multimedia and creativity software products, with a more-recent foray towards rich Internet application software development. Adobe was founded in December 1982 by John Warnock and Charles Geschke, who established the company after leaving Xerox PARC in order to develop and sell the PostScript page description language. In 1985, Apple Computer licensed PostScript for use in its LaserWriter printers, which helped spark the desktop publishing revolution. The company name Adobe comes from Adobe Creek, which ran behind the house of one of the company's founders. Adobe acquired its former competitor, Macromedia, in December 2005, which added newer software products and platforms, such as Adobe Flash and Adobe Flex, to its product portfolio.As of February 2009, Adobe Systems has 7,173 employees,Cite error: Invalid &amp;lt;ref&amp;gt; tag; refs with no name must have content about 40% of whom work in San Jose. Adobe also has major development operations in Orlando, Florida; Seattle, Washington; San Francisco, California; Ottawa, Ontario; Minneapolis, Minnesota; Newton, Massachusetts; San Luis Obispo, California; Hamburg, Germany; Noida, India; Bangalore, India; Bucharest, Romania; Beijing, China. Since 1995, Fortune has ranked Adobe as an outstanding place to work. Adobe was rated the fifth-best U.S. company to work for in 2003, sixth in 2004, 31st in 2007, 40th in 2008, and eleventh in 2009. Cite error: Invalid &amp;lt;ref&amp;gt; tag; refs with no name must have contentIn 2007 Adobe ranked 9th on the list of largest software companies in the world.In May 2008, Adobe Systems India was ranked 19th in great place to work in India. In October 2008, Adobe Systems Canada Inc. was named one of "Canada's Top 100 Employers" by Mediacorp Canada Inc., and was featured in Maclean's newsmagazine.</p>
comment	Adobe Systems Incorporated is an American computer software company headquartered in San Jose, California, USA.
companyLogo	File:AdobeSystems.svg
companyName	Adobe Systems Incorporated
companySlogan	Better by Adobe
companyType	Public company
foundation	Mountain View
foundation	California
foundationplace	Mountain View
foundationplace	California
hasPhotoCollection	Adobe Systems
homepage	www.adobe.com
homepage	www.adobe.com
industry	Software industry
industry	Software industry
keyPeople	Charles Geschke
keyPeople	Shantanu Narayen

# Get all *Investment[In]LeaseReceivable*

<http://rhizomik.net/semanticxbrl/?query=DESCRIBE ?o WHERE { ?s ?p ?o . ?p <http://www.w3.org/2000/01/rdf-schema#subPropertyOf> <http://rhizomik.net/ontologies/bizontos/xbrl-edgar/adbe-20080530.owl#InvestmentLeaseReceivable> . }>

edit - new - del

## adbe-20080616

	<b>a monetaryItem</b>
InvestmentLeaseReceivable	contextRef AsOf20080530 Consolidated Unaudited
	decimals -3
	unitRef USD
	value 207239000
InvestmentLeaseReceivable	contextRef AsOf20071130 Consolidated Unaudited
	decimals -3
	unitRef USD
	value 207239000

Referrers

edit - new - del

## adbe-20080916

	<b>a monetaryItem</b>
InvestmentInLeaseReceivable	contextRef AsOf20071130
	decimals -3
	unitRef USD
	value 207239000
InvestmentInLeaseReceivable	contextRef AsOf20080829
	decimals -3
	unitRef USD
	value 207239000

Referrers

edit - new - del

## adbe-20090227

	contextRef BalanceAsOf 28Nov2008 Unaudited
--	--



# Integrate different CIK variants

## SPARQL Construct Query

Example 2: make all URIs for Adobe CIK equivalent.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xbrli: <http://rhizomik.net/ontologies/2007/11/xbrli-instance-2003-12-31.owl#>

CONSTRUCT { ?e owl:sameAs <http://rhizomik.net/semanticxbrl/CIK0000796343> }
WHERE {
  ?s xbrli:entity ?e .
  FILTER (REGEX(STR(?e), ".*796343.*"))
}
```

Construct

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xbrli="http://rhizomik.net/ontologies/2007/11/xbrli-instance-2003-12-31.owl#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  <rdf:Description rdf:about="http://rhizomik.net/semanticxbrl/adbe-20080616/CIK0000796343">
    <owl:sameAs rdf:resource="http://rhizomik.net/semanticxbrl/CIK0000796343"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://rhizomik.net/semanticxbrl/adbe-20071217/CIK0000796343">
    <owl:sameAs rdf:resource="http://rhizomik.net/semanticxbrl/CIK0000796343"/>
```

Load



# Conclusions

- **XBRL tools:** for financial information edition and maintenance (formulae,...)

XML2RDF

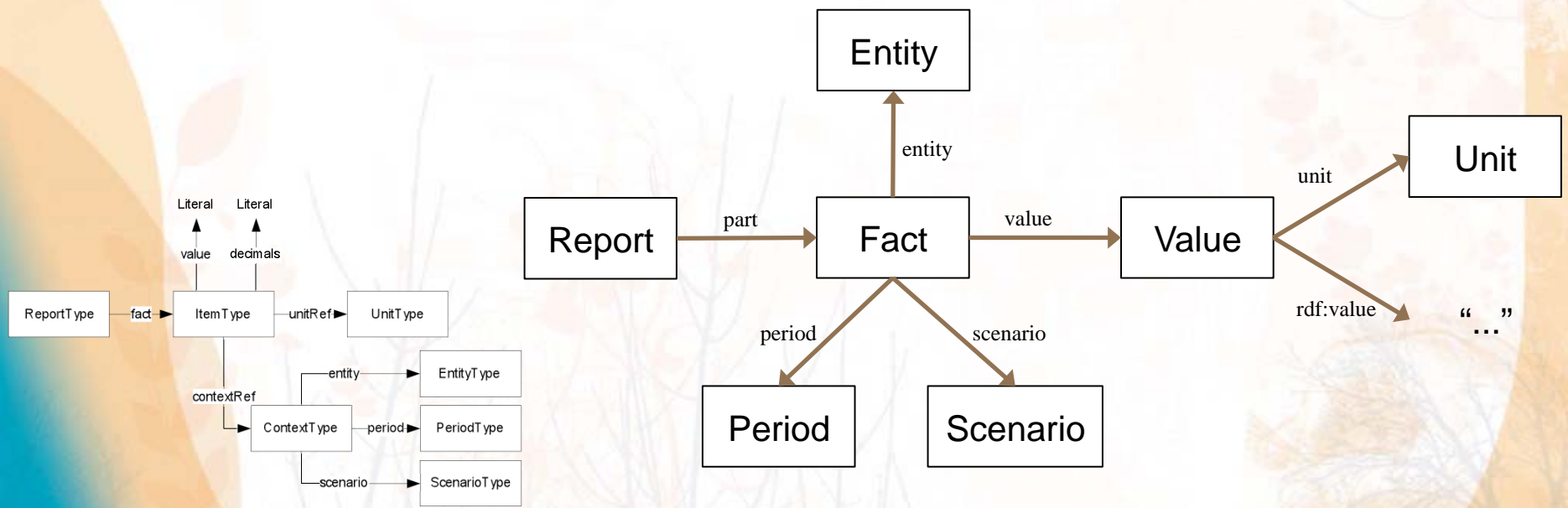


XSD2OWL

- **Semantic Web tools:** data integration (linked data), cross querying,...
  - Across filings, companies, accounting principles,...

# Future Work

- Resulting RDF is too “XML-styled”
- Rules or CONSTRUCT to tailor to needs
  - E.g. **facts as resources** instead of properties, facilitate querying and browsing



# Future Work

- Systematic **mappings** across filings for the **same company**
- Explore mappings:
  - Across companies
  - Across accounting principles
  - ...

# Future Work

- Calculations: math ops among values
  - Difficult to model using OWL. Rules?
- Labels:
  - rdfs:label
- Presentation: report layout
  - Layout ontology?
- Definition: semantic rels (is-a, whole-part...)
  - Reuse upper ontologies

# Thank you for your attention

Roberto García

[rgarcia@diei.udl.cat](mailto:rgarcia@diei.udl.cat)

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



HCI & Data Integration Research Group  
Universitat de Lleida, Spain

