Creating a Virtual Knowledge Base for Financial Risk and Reporting

Juan Sequeda, Capsenta Inc.
Mike Bennett, Hypercube Ltd.

Ontology Summit 2016
24 March 2016
Risk reporting – New regulatory requirements

• The Basel Committee for Banking Supervision
• Set of 14 principles, across 4 broad themes
• Identifies a number of Systemically Important Banks
  • G-SIB or G-SIFI
  • “Too Big to Fail”
  • Domestic SIFIs to fc
• Mandates basic data integration and re
The Search for a Conceptual Something

• Data dictionary?
• Business Glossary?
• Business dictionary
• Terminology?
• Thesaurus?
• Or what?...

• All these words about words
Creating a Reference Ontology

• Requires an understanding of cognitive and terminological principles
  • Semiotics
  • Classification theory (taxonomy)
  • Abstraction and encapsulation of concepts
  • Core concepts grounded in the reality of contracts, accounting etc.

• Use an Upper Ontology

• Needs to be a common language across the business

• The Two Ontological Questions:
  • What kind of Thing is this?
  • What distinguishes it from other Things?

• Needs to be framed in formal logic
  • Unambiguous across business and machines
# FIBO: Scope and Content

<table>
<thead>
<tr>
<th>Upper Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIBO Foundations: High level abstractions</td>
</tr>
<tr>
<td>FIBO Business Entities</td>
</tr>
<tr>
<td><strong>FIBO Contract Ontologies</strong></td>
</tr>
<tr>
<td>Securities (Common, Equities)</td>
</tr>
<tr>
<td>Derivatives</td>
</tr>
<tr>
<td>Funds</td>
</tr>
<tr>
<td><strong>FIBO Pricing and Analytics (time-sensitive concepts)</strong></td>
</tr>
<tr>
<td>Pricing, Yields, Analytics per instrument class</td>
</tr>
<tr>
<td><strong>FIBO Process</strong></td>
</tr>
<tr>
<td>Corporate Actions, Securities Issuance and Securitization</td>
</tr>
<tr>
<td><strong>Future FIBO: Portfolios, Positions etc.</strong></td>
</tr>
<tr>
<td>Concepts relating to individual institutions, reporting requirements etc.</td>
</tr>
</tbody>
</table>
Semantic Data Virtualization technologies

- Introducing Capsenta R2RML-compliant “wrappers”
- Maps conventional databases to RDF
- ETL (Extract Transform and Load) vs. NoETL (Federated)
- Example investigation: Wealth management integration issues
Mapping Relational Databases to RDF and OWL

<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
<th>AGE</th>
<th>CID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alice</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>Bob</td>
<td>NULL</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CID</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Austin</td>
</tr>
<tr>
<td>200</td>
<td>Madrid</td>
</tr>
</tbody>
</table>
Ultrawrap: Semantic Data Virtualization

Semantic Queries

R2RML Mapping

Reference Ontology

Internal Data
- Wealth Management
  - Debt
  - Bond
  - Equity
- Ordinary Share
- Preference Share

External Data
- BCBS 329
- Golden Source
- Asset Control

Legacy Data Sources and Systems

Oracle
SQL Server
Postgres
Unstructured
Structured
HIVE
Impala, etc.

Risk, Compliance
Search
API
Dashboard
Reports
Ultrawrap: Semantic Data Virtualization

1) Create Mappings
   • Create mappings for each Source databases to the Target Ontology
   • Mappings are represented in a declarative language: R2RML

2) Use Mappings
   • Use the mappings to integrate the data
   • ETL transforms the source data into data in terms of the Target Ontology
   • NoETL executes queries written in terms of the Target Ontology over each of the sources
<TriplesMap1>
a rr:TriplesMap;
  rr:logicalTable [ rr:tableName "Person" ];
  rr:subjectMap [ rr:template "http://www.ex.com/Person/ [ID]";
      rr:class foaf:Person ];
  rr:predicateObjectMap [ rr:predicate foaf:based_near ;
      rr:objectMap [ rr:parentTripelMap <TripleMap2>; rr:joinCondition [ rr:child "CID"; rr:parent "CID";]
  ]].
</TriplesMap1>
Ultrawrap ETL

RDBMS

Ultrawrap ETL

R2RML Mapping

SPARQL

Result

SPARQL

RDF Graph Triplestore
Ultrawrap NoETL

- SQL
- SPARQL
- R2RML Mapping
- SPARQL Result
- SQL Results
- RDBMS
- Semantics

Hypercube

Capsenta
The Smart Data Company™
NoETL Architecture
Hybrid NoETL and ETL Architecture
Financial Institution with Golden Copy

Reference Ontology

Data Lineage - Metadata management

Golden Copy DB

Source DB 1

Source DB N

Market Data Feed 1..n

Other DB N

Market Data Feed

Query

Response

Reporting
Data Strategy Considerations

• Real-time data:
  • Unlikely to want to stand up in triple store
  • Usually apply market data feed (pricing, yield) directly to position data

• Critical reference data
  • Provenance etc. determined upstream of Golden Copy
  • Any benefit to replicating golden copy data in triple store?
    • Possible re-use of GC data in risk, compliance

• Data Migration Strategy
  • Implementation of semantics provides an opportunity to review data requirements,
  • Also consider third party data licensing / rights management
Virtualization versus Triple Store

• When to stand up the Triple Store?
  • Only when that is already in place or a new triple-based application

• When to query through to source DBs directly?
  • Whenever possible

• Depends on:
  • Data provenance / DQ measures
  • Temporality

• Virtual Federation
  • Not have to create the “tower” of a triple store, but you get what looks like one.
  • SPARQL as fast as SQL
Integration: Mapping Considerations

- Generally not one to one
  - Develop “Mapping Patterns”

- One data element may represent different ontological things
  - Defined in different contexts
  - How is the context (in this case the presence of other data elements) dealt with?

- One target ontological Thing or property may be represented by some unique combination of database elements
  - Example: use of “type” enumerations on a class (this is just one design choice in the DB design)
Mapping Considerations (1)

Reference Ontology

Source DB

Loan Borrower

Loan Borrower Name
Loan Borrower SSID
Loan Borrower Address Line 1
Loan Borrower ZIP Code

All of this (graph)
Maps to...
this.

Loan

hasBorrower

Borrower

Party In Role

Is a

Autonomous Agent

hasIdentity

Address

hasAddress

hasAddressElement

PostalCode

Hypercube
Capsenta
The Smart Data Company™
Mapping Considerations (2)

Capsenta has a specific set of transformations that deal with enumerations; Enum entries are transformed to OWL classes.

This class

Maps to...

This set of values:

Security

<table>
<thead>
<tr>
<th>Security Name:</th>
<th>string</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issuer:</td>
<td>string</td>
</tr>
<tr>
<td>Identifier:</td>
<td>Identifier</td>
</tr>
<tr>
<td>Security Type:</td>
<td>SecurityType</td>
</tr>
</tbody>
</table>

SecurityType (enum)

<table>
<thead>
<tr>
<th>Ordinary Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond</td>
</tr>
<tr>
<td>MediumTermNote</td>
</tr>
<tr>
<td>MortgageBackedSecurity</td>
</tr>
</tbody>
</table>

This class

Reference Ontology

Source DB

Ordinary Share

Preference Share

Equity

Debt

Preference Share

Medium Term Note

Security

Security Type:
Semantic Query-based Applications

- Risk and Compliance
  - Compliance and what-if reports
  - Internal reports
  - Risk reports
- Generate and update using semantic queries
  - Understood and signed off by business stakeholders
Reporting

• Reports:
  • Timely, accurate, complete
  • Able to generate new kinds of report at short notice
  • Increased frequency during times of crisis

• Reports are created using semantic queries
  • Tools exist for graphical creation of semantic queries
  • Business analysts can create these queries

• Starting point: replicate existing reports using semantic queries
  • No additional IT effort required to update reports
  • New reports generated as needed provided the concepts are in the Reference Ontology
Summary

• Semantic queries are used to recreate reports and dashboards
  • Provide the means to extend and vary these quickly
  • Minimal application development.

• Non-disruptive:
  • Existing bank systems of record are used
  • No need for centralized data storage (relational or graph database).
  • Existing data quality measures remain unaffected

• Probabilistic mapping methods to semi-automatically create the mappings
Thank you! Questions?

• Juan Sequeda, PhD
  • Founder – SVP Technical Sales and Research, Capsenta Inc.
  • www.Capsenta.com
  • Twitter: @JuanSequeda

• Mike Bennett
  • Director, Hypercube Ltd.
  • www.hypercube.co.uk
  • Twitter: @MikeHypercube