Semantic Modeling of Mortgage Backed Securities: Case Study

Mike Bennett, Enterprise Data Management Council
Yefim Zhuk, Sallie Mae
Overview

• EDM Council Semantics Repository
• Regulatory challenges
• Project to demonstrate how semantic tagging of securities data would assist systemic risk monitoring
• Extended Semantics Repository for loans terms
• Subject Matter Expert (SME) reviews experience and pointers
Proof of Concept Project

- **Structure**: Industry collaboration headed up by EDM Council and IBM Research initiated on request by the European Central Bank

- **Goal**: Demonstrate the viability of tagging financial contracts to regulators using standard semantics and identifiers in support of risk analytics

- **Test Case**: Mortgage backed securities (end-to-end linkages of facts about instruments, pools, mortgages, lending information, etc. that are currently held in disparate databases)
Aims of the Project

• Demonstrate benefits of semantic tagging
• Example chosen: Mortgage Backed Securities
  – Extend EDM Council Semantics Repository
  – Create logical model for database design
  – Populate this with sample data (not done in the end)
• Demonstrate what can be done with this for systemic risk oversight
Semantics Repository

- Formal and factual representation of the facts about, definitions of, and relationships associated with all financial instruments
- Goal is industry consensus (and adoption of) a common language for the financial industry
- Result of industry collaboration, implemented with semantic web techniques and presented in business-readable formats (spreadsheets and non-technical diagrams)
- Formal nature of the model means it can be used to specify or derive data models and common message formats
Semantics Repository

- Business Conceptual Model of financial securities terms and definitions
- Uses OWL notation
  - Optimized for business readability
  - Presented in a UML tool framework using the OMG’s Ontology Definition Metamodel (ODM) standard with some extensions and changes for readability and tool support
- Includes debt instruments such as asset backed and mortgage backed securities (MBS)
- Also includes process flow models for securities issuance
The Problem

• Mortgage Backed Securities:
  – Tranched
  – Tranches designed for specific cashflow behaviors
  – Aggregated loan terms
  – Terms for individual loans are not included in the terms for MBS prospectus or terms sheets

• The data for loans, pools, tranches, securities
  – live in different silos
  – have different data models

• End to end linkage not easily attainable
The Project

• Participants
  – Enterprise Data Management Council
  – IBM Research
  – European Central bank
  – Banks, risk tool vendors etc.

• Initial deliverables from IBM Research
• Ongoing project
Project Overview

• Used EDM Council Semantics Repository as basis
• Extended with further terms for MBS
  – Extensive business subject matter expert review
  – Looked at issuance and pool creation process flows
  – Agency and non agency (private label) mortgage backed securities
• Adds terms for individual loans, borrower facts
Activities

• IBM Research created a “Semantic Data Model” (SDM) from the SR content
• Reviewed and extended the Semantics Repository for agency and non agency MBS terms
• Solicited data model elements from a number of leading risk analysis vendors and users
• Added these to the SDM
• Reverse engineered into the Semantics Repository
• Some of these were loan specific, so added new Loans section
• Subjected this to business SME review
MBS Semantic Model
Semantic versus Data Models

• Set theory classes not OO classes
• Relationships are unidirectional
  – Pair of relationship + inverse = one OO relationship
• Open World Assumption
  “Absence of evidence is not evidence of absence”
  – Every fact which defines a thing is included even if data would never be
    available or is not needed
• Taxonomy
  – Multiple inheritance
  – Supports real world multiple classifications
• Data model enumerations
  – Often contain mixed semantics in reference data models
  – Render entries as separate semantic classes where possible
Transformation

• Identify which SR constructs correspond to which SDM constructs
• SR: OWL constructs
• SDM: UML Class Model (Logical Data Model) constructs
• Example:
  – Ontology: Pair of functional object properties, where one is the inverse of the other
  Transforms to
  – Data model: Association, with the two ends being the two object properties
## Transformation

<table>
<thead>
<tr>
<th>OWL construct</th>
<th>UML data model construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWL Class</td>
<td>Class</td>
</tr>
<tr>
<td>Generalization</td>
<td>Generalization</td>
</tr>
<tr>
<td>Object Property</td>
<td>Association end</td>
</tr>
<tr>
<td>Property Constraints</td>
<td>Multiplicity</td>
</tr>
<tr>
<td>Datatype Property</td>
<td>Attribute</td>
</tr>
<tr>
<td>Sub property</td>
<td>not mapped</td>
</tr>
<tr>
<td>Disjoint</td>
<td>not mapped</td>
</tr>
</tbody>
</table>
Semantic Data Model
Reverse Engineering

• Some terms were extension or enhancement to existing MBS terms
  – Added to the existing body of the model

• Some terms were specific to loans
  – Added new section on “Loans” to the SR
Experience

• The new terms from the risk applications were very physical
  – Lots of mapping between these
  – Semantics not always clearly understood, leading to terms being combined that actually had different meanings

• Reverse engineering is very inefficient
  – Requires second guessing of intended semantics
  – Reverse engineering logical SDM to Semantics also inefficient
    • “Kludges” to reuse field semantics
    • Example: taking natural person terms and populating with corporate borrower facts (names, domicile)

• But it was what we had to work with
• In the long run, should be able to drive content and changes from the conceptual view
Observations

• The “Semantic Data Model” became less semantic as time went on
  – Addition of local and foreign keys to each data table
  – Design for reuse of similar terms with different semantics

• This is all as it should be for a “Logical Data Model”
  – But it’s not semantic

• In future, should segregate SDM, LDM
Recommendations

• Semantic Model (ontology):
  – Use OWL or similar FOL constructs (with presentation for business SMEs)

• Semantic Data Model:
  – Direct copy of Semantic Model in ERM or UML Class notation
  – Identify transformations

• Logical Data Model
  – Embodies “design” thinking as required for population of a database (keys etc.)

• Transformations
  – SDM to LDM is a design activity
  – Going from SR to SDM could potentially be automated
Loans SME Reviews

• Many of the reverse-engineered terms in the PoC SDM were very context specific:
  – Loan type = mortgage loan
  – Borrower type = natural person

• A complete semantic model needs to be applicable to all loan and borrower types
  – Specialization for mortgages, natural persons etc.

• Recruited business domain experts from loans (retail lending, agencies etc.)

• Should align with MISMO (loans industry data standard) when possible, but had to start with what we had
Collaborative Work Experience

Several SMEs from Sallie Mae (SLM) participated in the collaborative work on Loan Ontology

Well organized and very productive work by SMEs from SLM, Wells Fargo, National Australia Bank, Bank of America and more (thanks to M.B. for his leadership!)

The complexity and scale of the industry model brings many challenges to transparent knowledge exchange
Modular and Mapping approaches in the company-to-industry frameworks

Lessons learned:

- There is no single truth – no single correct model
- Modular approach is necessity
- Multiple models bring concept flavors/duplications
- If we use references or try to integrate, we’ll have to change one or both models
Modular and Mapping approaches in the company-to-industry frameworks

Another standard approach is to provide a map. This allows us to keep the models untouched and have the A-B Map-ontology resolve conceptual flavors or duplications. The Map-Ontology can include a set of rules to resolve some possible conflicts (Future work).
Ontology of Ontologies
Upper and Pluggable Ontologies

We are getting there while preparing for financial industry ontology standard (work in progress by EDMC/OMG).

Sharing some ideas:

Everything is an ontology: upper, low level, maps…

Upper – references and meta-data to lower level ontologies

Maps – resolve conflicts and object flavors/duplications
Thank you!
Questions?

References

• [www.hypercube.co.uk/edmcouncil/docs/bennett_mike.ppt](http://www.hypercube.co.uk/edmcouncil/docs/bennett_mike.ppt), Mike Bennett, The Enterprise Data Management Council Semantics Repository Case Study (Semantics Technology conference June 2010)
• [http://portal.acm.org/citation.cfm?id=1937131](http://portal.acm.org/citation.cfm?id=1937131), Mike Bennett, EDM council semantics repository next steps: exploring a consensus semantics framework (ODISE 2010)
• Open Ontology Repository (OOR) Initiative, OOR Initiative is to the promote the global use and sharing of ontologies, [http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository#nid17YN](http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository#nid17YN)
• Knowledge Engineering and Common Sense Reasoning, [http://cyc.com](http://cyc.com)
• Distributed Active Knowledge and Process, Yefim (Jeff) Zhuk, US Patent, Collaboration of knowledge & services
• Rules Collector System and Method, Yefim Zhuk, Transformation of “tribal knowledge” into rule-based applications and working systems

Mike Bennett, mbennett@edmcouncil.org | Yefim (Jeff) Zhuk, yefim.zhuk@salliemae.com