Ontology Summit 2016

Finance and Retail Track
May 9, 2016

Co-Champions:
Mike Bennett
Andrea Westerinen
Finance and Retail Track: Goals and Mission

• Explore challenges and solutions in the finance and retail domains related to:
  – Integration and interoperability/reuse of ontologies and their data
  – Compliance and risk management

• Understand finance and retail reference ontologies, especially:
  – FIBO
  – GoodRelations and schema.org
Presentations (Session 1)

• Michael Uschold (Semantic Arts) and Lynn Calahan (Wells Fargo)
  – Referencing the Home Mortgage Disclosure Act in Building a Loans Ontology

• Mike Bennett (Hypercube) and Juan Sequeda (Capsenta)
  – Creating a Virtual Knowledge Base for Financial Risk and Reporting

• David Saul (State Street Corporation)
  – Do you know where your data is? How FIBO Makes Data Smarter and More Governable
Presentations (Session 2)

- Cory Casanave (Model Driven Solutions and OMG)
  - Conceptual Modeling with SIMF
- Elisa Kendall (Thematix Partners LLC)
  - Using Business Architecture and Semantics to Drive Data Quality Improvement in Banking
- Mirek Sopek (Makolab)
  - The Quest for Meaning and Trust on the Web – schema.org in Finance and Retail
- Rebecca Tauber and Andrea Westerinen (Nine Points Solutions, LLC)
  - Semantic Interoperability and Knowledge Engineering: Use of GoodRelations in a Climbing Gear Retail Ontology
Interoperability Challenges

• Development of comprehensive vocabularies and ontologies
• Ability to integrate both semantics/concepts and data
  – And in finance, support compliance reporting and risk management
  – Also related to tooling
• Knowledgeable “practitioners”
  – Who know what, where and how to use and extend the vocabularies and ontologies
Challenges – “Fragmented Data”

- Office of Financial Research’s (OFR) Financial Research Advisory Committee (FRAC)
- In today’s fragmented data environment, ... the data is sourced from many places and reported to a variety of independent data repositories without a common data standard. The lack of a common standard for data meaning results in the use of common words that mean different things and the expression of common concepts using a variety of words. In order to reconcile the problems created by the lack of a common set of concepts and associated terms, industry participants are engaged in a continual process of data reconciliation. Many of these processes are manual and prone to error. Also, as these reconciliation efforts are multiplied across a variety of interdependent processes, the result is a mismatch of underlying data, divergence in calculation, and lack of comparability.
Current Initiatives - Finance

• Tame Risk with Smart Data, Not Big Data (AmericanBanker.com)

• “... need to create useful data rather than just lots of data comes as large global institutions face expenditures ranging from $150 million to $350 million each to comply with new post-credit crisis regulatory requirements in the United States, Europe and elsewhere. That is "significantly larger" than the level of expenditures required previously ... The challenge ... will be for firms to effectively pull together the correct "smart" data from across many systems across many divisions or subsidiaries in many geographies ... The lack of integration has to be overcome ...”

• “The aim is to allow data to be shared and reused across applications, different parts of a company and different organizations, as needed. For global financial institutions, a key initiative for creating the smart, reusable data is the creation of the Financial Industry Business Ontology (FIBO) ...”
FIBO

- Reference ontology providing common concepts and language for financial instruments, processes and legal entities
- Designed with broad input from subject-matter experts to address the multiple needs of banks, financial services firms and regulators
- Logical choice for mapping and integrating financial data
Current Initiatives - Retail

- No clear solution for modeling and ontologies
- For online retail, goal is for customers to find products and services via online search
  - Online sales in the U.S. forecast to grow from $349.20 bn in 2015 by 14.4% to reach $399.48 bn in 2016 (Centre for Retail Research)
  - Requires that search engines "understand" the details of the businesses' offerings
  - One of the main sources for providing such information is rich snippets and schema.org
  - Over a third (36.6%) of search results analyzed by Searchmetrics included at least one snippet with information derived from Schema.org within search results pages
- In 2012, schema.org integrated GoodRelations e-commerce ontology
Schema.org and GoodRelations

• Searchmetrics analyzed the most frequently appearing snippets (or integrations), derived from information presented by websites using Schema
  – “Movie” integrations came out on top (accounting for 27% of integrations)
  – Second (accounting for 21% of the total) were “Offers” which accompany search results related to online retail and ecommerce sites
  – Joint third were “TV series” and “Product” snippets

• Comparable results found by SimilarTech in their Technologies Market Share for document standards
Other Retail Solutions

• ARTS from the National Retail Foundation
  – Operational Data Model
  – Data Warehouse Model
  – XML Schemas for interoperability (build on the operational data model and provide defined structure for applications “to speak the same language”)

• REA (Resources-Events-Agents) models of business process
Track Findings (I)

• FIBO and GoodRelations “not intuitive” but require learning and experimentation
  – Knowledgeable practitioners needed
  – Documentation available but sometimes difficult to understand what to use, where to link and what to extend
  – Issues also tied to correctly modeling the domain - For example:
    • What needs to be represented and when?
    • A bank as a business entity, versus a lender, versus a lender on my mortgage
    • Schema.org Product versus Offer (an offer “lives” within a product)
Track Findings (II)

• Differences between data in the real world versus data needed for compliance reporting
  – Need support for both
  – Ideally as close as possible
  – Reporting mandates understanding the domain and then mapping to requirements

• Possible to integrate data across repositories by mapping to “reference ontologies” (such as FIBO)
  – Two approaches to data integration - ETL and federation
Track Findings (III)

• Few tools aid in development and integration of concepts or data
  – For terminology development/vocabulary management
    • Subject matter experts “were comfortable with Excel”
    • Extraction tools created own taxonomies without reference to standards, or defined relevance values without hierarchies
    • Require support for collaborative definition, annotation, provenance editing and review process
  – For logical data schemas, can validate with ontologies
    • SPARQL queries
    • “Reverse – from ontology to schema, proved impractical”
Track Findings (IV)

- Reuse of concepts and semantics -> interoperability
  - For example, schema.org is reuses (is extended by) both the GoodRelations and FIBO ontologies
  - Simplicity of vocabulary with backing semantics of ontologies
- Factors limiting reuse
  - Many ontologies are too large and complex, not sufficiently granular and may not be sufficiently documented
  - Semantic languages such as OWL do not support partial import of an ontology and force the inclusion of many more concepts than may be deemed appropriate or (worse) correct
- Modular design and patterns valuable to integration and reuse
  - Financial modules in FIBO
  - Retail patterns such as brands versus products versus collections specific to purpose versus offers
Track Findings (V)

- Recommend creating ontologies from granular, cohesive and self-contained modules and design
- Smaller modules and patterns combined together and extended to form larger ontologies
- Could extend use of patterns to “integration” and mapping patterns
  - Addressing:
    - Valid concept/semantics but terminology is confusing to a domain
    - When, why and how to extend versus creating a new, domain-specific concept
    - Literals versus general concepts with ability to subtype
BACKUP
References

- http://ontologforum.org/index.php/ConferenceCall_2016_03_24
- http://nrf.com/resources/retail-technology-standards-0
- http://reatechnology.com/what-is-rea.html