BMGT 499B
Next Generation Financial Cyberinfrastructure

Data Science for Finance

Louiqa Raschid

April 2 2013

April 2 Agenda

- 50 minutes:

  Financial markets and data management. (Lahaie)

- 50 minutes today and will continue on April 9:

  Overview of the next generation financial cyberinfrastructure and data science for finance. (Raschid)

- 20 minutes today and will continue on April 9:

  Review of future topics.

- 10 minutes

  Logistics for the class
  - Class schedule.
  - Preparation for the class.

  Grading
  - Participation
  - Project

OFR Data Mandate

- Financial instrument reference data: Information on the legal and contractual structure of financial instruments such as prospectuses or master agreements, including data about the issuing legal entity and its adjustments based on corporate actions.
- Legal entity reference data: Identifying and descriptive information such as legal names and charter types, for financial entities that participate in financial transactions, or that are otherwise referenced in financial instruments.
- Positions and transactions data: Terms and conditions for new contracts (transactions) and the accumulated financial exposure on an entity's books (positions).
- Prices and related data: Transaction prices and data used in the valuation of positions, development of models and scenarios, and the measurement of micro-prudential and macro-prudential exposures.

Presentation by Chuck LaHaie

Next Gen Fin Cyberinfrastructure

- Underlying legislation and policies.
- OFR mandate: Focus on the data.
- Examples of financial supply chains.
- Topics
  - Data definitions and protocols and messaging standards.
  - Types of financial instruments; properties; valuations; levels of risk; taxonomies.
  - Ontologies and FIBO and Semantic technologies.
  - LEI (CICI).

Data Science for Finance

- Operational Risk.
- Visual and Network Analytics.
- De-identification and Privacy.
- Knowledge Extraction from Financial Document Collections.
- Semantic search.
- What is not going to be covered.

Background

- Underlying legislation (policy).
  - 1929-1933 Glass,Glass+Steagall Act; 1999 repeal.
  - http://en.wikipedia.org/wiki/Glass%E2%80%93Steagall_Act
  - Separation of commercial and investment banking.
  - Federal Reserve system and FDIC.
  - Stability of markets but also negative consequences.
  - MANY reporting requirements; no teeth since no standard(s).
- Commodity Futures Modernization Act, 2000
  - wiki entry; deregulation of OTC derivatives.
- July 2010 Dodd Frank Act and the OFR mandate
  - OFR: Focus on the data and computational tools.
  - Lots of standards, e.g., LEI (CICI). Enforcable?
  - CFPB (Consumers); Volcker rule.

Without the NextGen FCI …

- Regulators, analysts, and the financial press are denied an understanding of capital market operations sufficient to forge knowledgeable and prudent financial policy.
- Business analysts are uncertain of the quality of their internal risk and accounting numbers. [Operational risk.]
- The academic community is lacking the information required to examine and analyze market operations and behavior. [New field: Data science for finance.]
- The public lacks the information to make sound financial decisions reducing confidence in our financial systems and increasing the risk of bank runs and systemic crisis.
- Another way to think about it … being better prepared with up-to-date data to make critical decisions during crisis situations like the Great Recession of 2008, the ongoing EURO crisis, etc.

Financial supply chains

- Examples of financial supply chains
  - Share capital/common stock (IPO; Corporate reporting to SEC, FDIC; Trading in a market to establish prices.)
  - Mortgages (financial contract; land title) and MBS.
  - Futures contracts; derivatives.
  - Commercial paper.

- What is the quality (of data) along these financial supply chains?
  Operational risk.
  - How did sub-prime mortgages get bundled into the ABACUS 2007-AC1 R(esidential)MBS rated Aaa by Moody's and AAA by S&P?

According to the SEC's complaint, the deal closed on April 26, 2007, and Paulson & Co. paid Goldman Sachs approximately $15 million for structuring and marketing ABACUS. By Oct. 24,2007, 83 percent of the RMBS in the ABACUS portfolio had been downgraded and 17 percent were on negative watch. By Jan. 29, 2008, 99 percent of the portfolio had been downgraded. Investors in the liabilities of ABACUS are alleged to have lost more than $1 billion.

  - AIG and the slippery slope of underwriting financial contracts using exotic CDSes (credit default swaps).

Modeling and monitoring the flow of data across a financial supply chain

- U.S. residential mortgage supply chain [Mark Zandi, Chief economist, Moody's Analytics].
  - 2004-2007: Loans to individuals with credit score < 660 (700 average).
  - 2006: 50% of the loans had no documentation; robo-signing scandal.
  - 2006: 600 billion of subprime and alt-A loans; 75% originated by the private sector; equivalent to total U.S. credit card debt.
- A mortgage is both a financial contract and legal title over an asset (property);
  - MERS is a repository and electronic transaction system designed to manage the legal title registration process.
  - The mortgage companies set up MERS and MERS effectively owns all the titles.
  - MERS was poorly designed and has no audit trail.

- How could better data management practices have helped to monitor the residential mortgage supply chain and send out early warning signals?
- Nancy Wallace, Haas School, UC Berkeley.

Modeling and monitoring the flow of data across a financial supply chain

- Supply chain of exotic financial products.
  - Commodity Futures Modernization Act, 2000.
  - Derivatives are not futures; they are not securities; they are not insurance products; ergo … they are not regulated!
  - Based on the belief that investors, markets and management could self-regulate without causing any problems to the global financial supply chain.
  - Credit Default Swaps (CDS)
  - Greece defaulted on $100 billion in 2012 triggering a CDS payment of $3 trillion.
  - AIG went down the slippery slope of underwriting CDSes.
  - CDSes largely accounted for the insolvency of Bear Sterns and Lehman.
- DATA DRIVEN models and DATA SCIENCE for finance.
- ISDA
  - If a CDS is a tradable asset then it must be settled at some market valuation.
  - If a CDS is an insurance product then it requires capital reserves.
  - What if it is neither?
  - Importance of SEMANTICS and REASONING to complement POLICY and ENFORCEMENT.

OFR Data Mandate

- Financial instrument reference data: Information on the legal and contractual structure of financial instruments such as prospectuses or master agreements, including data about the issuing legal entity and its adjustments based on corporate actions.
- Legal entity reference data: Identifying and descriptive information such as legal names and charter types, for financial entities that participate in financial transactions, or that are otherwise referenced in financial instruments.
- Positions and transactions data: Terms and conditions for new contracts (transactions) and the accumulated financial exposure on an entity's books (positions).
- Prices and related data: Transaction prices and data used in the valuation of positions, development of models and scenarios, and the measurement of micro-prudential and macro-prudential exposures.

Challenge questions? (revisit)

- Would you accept an information system / database solution with no PK/FK integrity; incomplete information; no attempt to reduce data inconsistency?
- Would you accept an information system such as MERS with no audit trail?
- How could better data management practices have helped to monitor the residential mortgage supply chain and send out early warning signals? What is the correct level (disaggregate level) of reporting on residential mortgages?

Next Gen Fin Cyberinfrastructure

- Topics ….
- What is not going to be covered?
  - Financial models (financial engineering).
  - Investment strategies.
  - High frequency (algorithmic) trading (HFT).

Background

- Dodd Frank Act of July 2010 establishes the Office of Financial Research.
- National Science Foundation Workshop on Knowledge Representation for Financial Information Management – NSF FIW 2010 (minutes after Act was signed by President Obama)
- Workshop on Next Generation Financial Cyberinfrastructure in July 2012 sponsored by the NSF and the CRA/CCC.
- Proposal to the NSF on DATA SCIENCE for financial research and financial cyberinfrastructure.

Data definitions; protocols; messaging standards

- The Financial Information eXchange (FIX) Protocol is a messaging standard developed specifically for the real-time electronic exchange of securities transactions. FIX is a public-domain specification.

http://fixwiki.fixprotocol.org/fixwiki/FIXwiki

- FpML® (Financial products Markup Language) is the business information exchange standard for electronic dealing and processing of financial derivative instruments. It establishes a new protocol for sharing information on, and dealing in swaps, derivatives and structured products. It is based on XML. All categories of OTC (over-the-counter derivatives) will eventually be incorporated into the standard.

http://www.fpml.org/about/ataglance.html

- SWIFT carries payment instructions exchanged between banks and market infrastructures operated by Central Banks or for retail payment, e.g., an Automated Clearing House. In clearing and settlement, SWIFT provides secure and reliable messaging and common message standards.

http://www.swift.com/solutions/messaging/index.page

Challenges/Limitations?

- These are essentially message/transport standards to describe individual financial transactions.
- Semantics are not included, i.e., there may be an XML tag "name" but there is no specification (if it is the name of an organization or an individual or a financial instrument). See FIBO for details.
- PK/FK/Referential integrity/Consistency.
- There is no GLOBAL END-to-END consistency across the supply chain, e.g., a payment be made against an entity that has been removed from the supply chain.

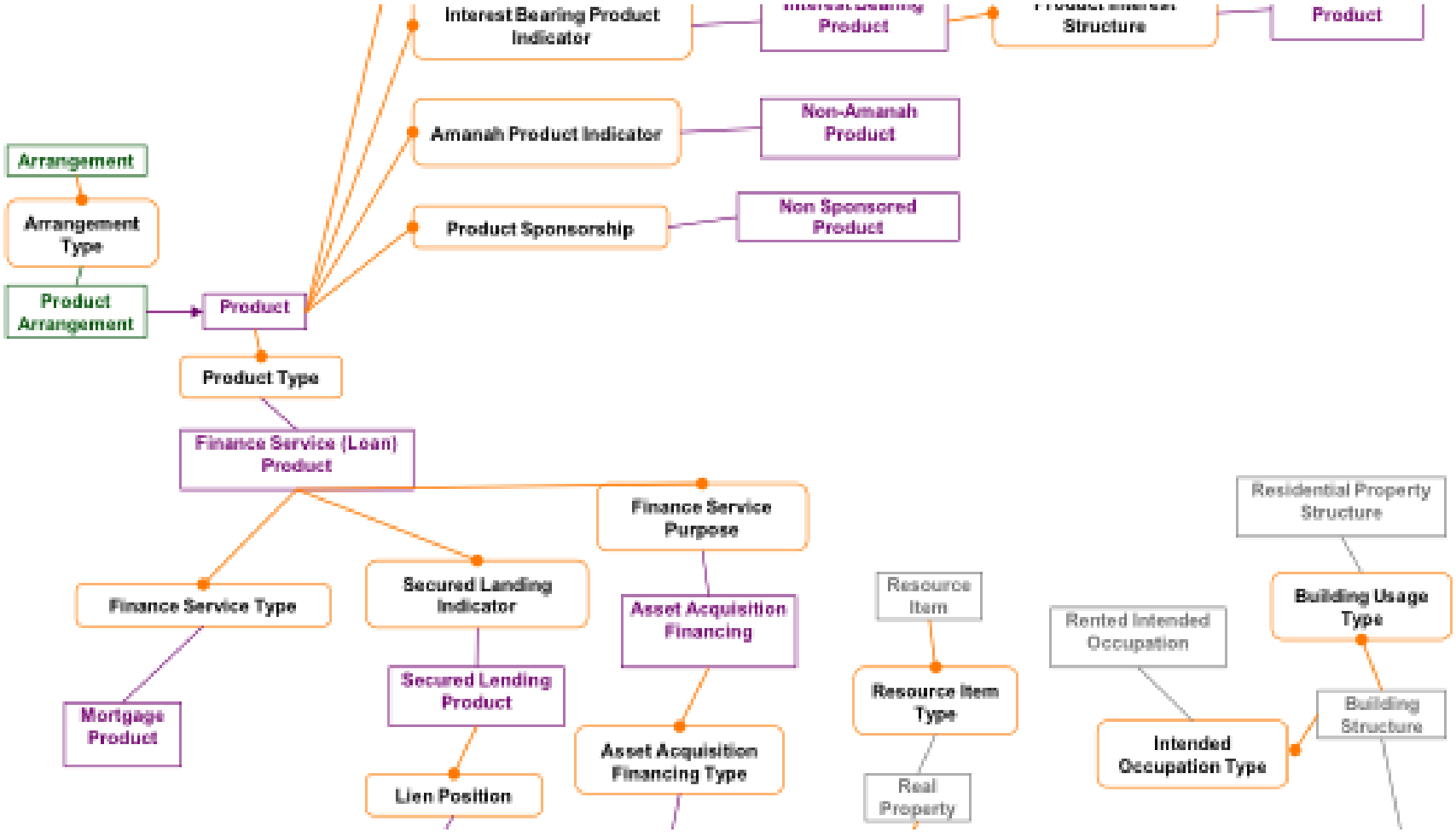
Challenges/Limitations?

- What are the benefits of the Legal Entity Identifier (LEI)?
  - Fully identify organizations and their relationships.
  - Complete network of counterparties to each contract.
  - (LEI and CICI will be discussed later.)
- Financial transactions must be overlaid over the BIG NETWORK representing all the financial entities (persons, organizations, instruments, contracts, …) to provide a comprehensive picture of the “transactional supply chain”.
- One and two sided reporting and impact on data quality and operational risk.
  - (Discussed later.)

Types of financial instruments; properties; valuations; taxonomies

- Example of fixed deposit instrument (fixed term; fixed rate); aka certificates of deposit in US or fixed term bonds in the UK.
- CDs are a low risk investment in the U.S.
- In many countries these can be risky investments with unregulated agencies and potential of a high payoff.
- Taxonomy elements:
  - Frequency of payout [monthly, annual, maturity, ..]
  - Depositor [can withdraw deposit; loan against asset]
  - Debtor [can change rate; can make early repayment]
- Next week's discussion.

Taxonomy of residential mortgage products



Ontologies and FIBO and semantic technologies

- Ontologies and FIBO and Semantic technologies.
- Models
  - Relational model/SQL; ER model; XML;
  - RDF/OWL
- “Retrieve all instruments that are a subclass of a Swap and have the property YY”.
  - Relational: Must know complete schema; hard coded queries.
  - RDF/OWL: Inference
-

Formal Collaboration to Define Financial Industry Business Ontology

ENTERPRISE DATA MANAGEMENT EDM COUNCIL → **Financial Industry Business Ontology** ← OMG WE SET THE STANDARD™

FIBO

- Non Profit industry association founded to facilitate enterprise-wide control over data content. Activities include standards, identification, semantics, data quality.
- Members include 80 major financial institutions worldwide (BOA, Barclays, Citigroup, Deutsche, Fannie Mae, Federal Reserve Bank NY, HSBC, Goldman Sachs, JP Morgan Chase, Morgan Stanley, RBC, RBS, UBS, Wells Fargo, etc…

- Object Management Group is an international, open membership, not-for-profit computer industry consortium that focuses on standards development work.
- Members include over 500 organizations worldwide including government agencies, small and large IT users, vendors and research institutions

OTC Derivatives POC Mission

Demonstrate to the regulatory community and the financial services industry how semantic technology and the Financial Industry Business Ontology (FIBO) model can be effectively utilized to fulfill regulatory governance, data standardization, risk management and data analytics requirements specified by the Dodd-Frank legislation

Goals and Objectives of OTC Derivatives Semantic POC

Provide Standard Financial Data Terminology

Provide standardized and unambiguous terminology that can electronically describe the content of an OTC Swap using the FIBO conceptual model and ontology.

Provide Mapping between FpML and FIBO

Provide mapping between messaging protocols e.g. FpML and common domain models e.g. FIBO to enable interoperability and to avoid imposing changes and alterations to existing messaging protocols

Classify Swaps into Asset Classes

Classify swap instruments into asset classes based upon rules that evaluate the attributes and structure of the actual swap data, rather than relying on descriptive codes.

Provide Query Capabilities

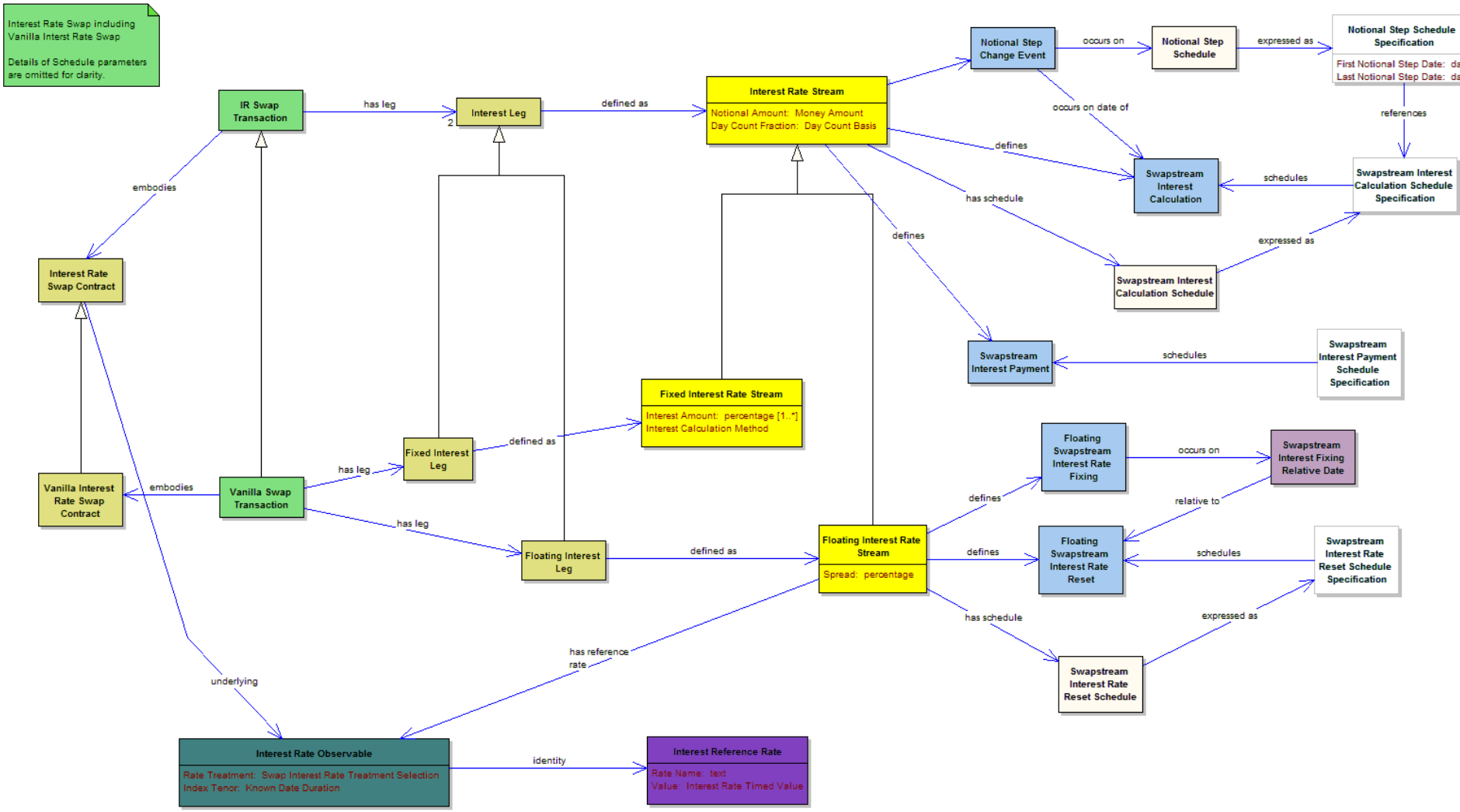
Provide query capabilities that report legal entity hierarchies, levels of asset class taxonomies, exposures of risk across ownership groups and asset classes (tbd), as well as extract data to be used for risk analytics

Electronically Reflect Contractual Provisions

Demonstrate how provisions of master agreements (e.g. calculations of net exposures, timing of transfers of collateral, forms of eligible collateral etc.), can be electronically realized (future phase?).

FIBO Business Semantic Conceptual Model

~~*Captures Knowledge* about IR Swaps~~



Potential Beneficiaries of Semantic Technology for Financial Data Standards

Financial Service Organizations

Trade investments efficiently, securely, customized to need

Manage risk for financial transactions and ensure regulatory compliance

Markets/Exchanges

Maintain orderly, liquid markets

Achieve transparent pricing and risk

Business Data Reporters and Reference/Market Data Vendors

Structure market data for publishing

Provide notification and analysis for investment managers

Regulators

Maintain market fairness, detect and address abusive trading practices, systemic risk

Standards Organizations

Define consensus standards facilitating interoperation of market participants and regulators

Industry Associations

Represent the interests of the market participants

Technology Providers

Provide technology to support market participants, reporters and regulators

Why Semantic Technology?

- Major step towards reducing data chaos
- *Ontology* = Knowledge Representation
  - Specification of a Conceptualization
- Based upon Description Logic
  - A mathematically verifiable symbolic logic that allows *reasoning* about *entities* and the many *properties* that describe entity *relationships*
- Describes entities in terms of:
  - Concepts (classes)
  - Relationships (properties)
  - Individuals (instances)
- Makes inferencing possible
  - Infers relationships and memberships in classes per axioms via a "*Reasoner*"

Jackson Pollock "*Convergence*"

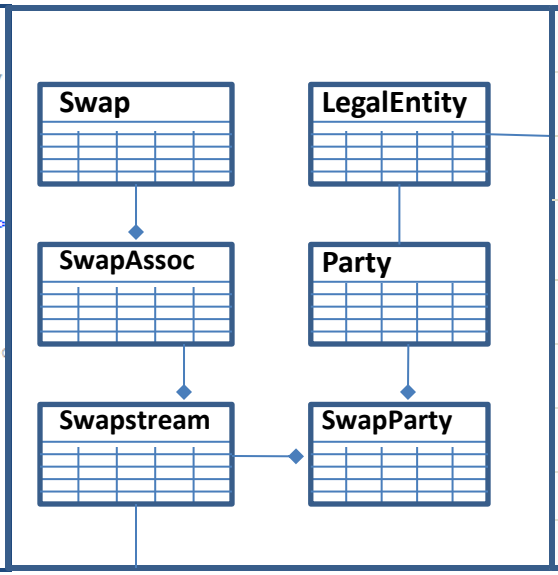
From Data Chaos ***To Data Order***

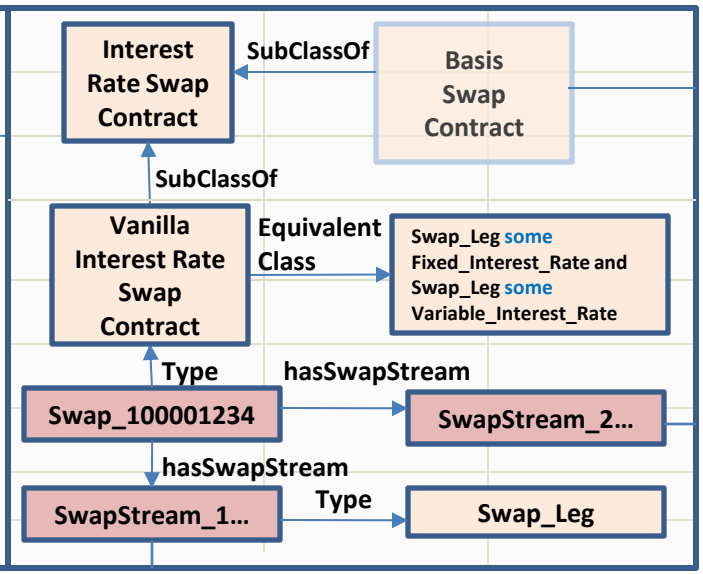
Aligns linguistically with how we think and speak!

Subject *(domain)*	Predicate *(property)*	Object *(range)*

RDF Triples/ Statements

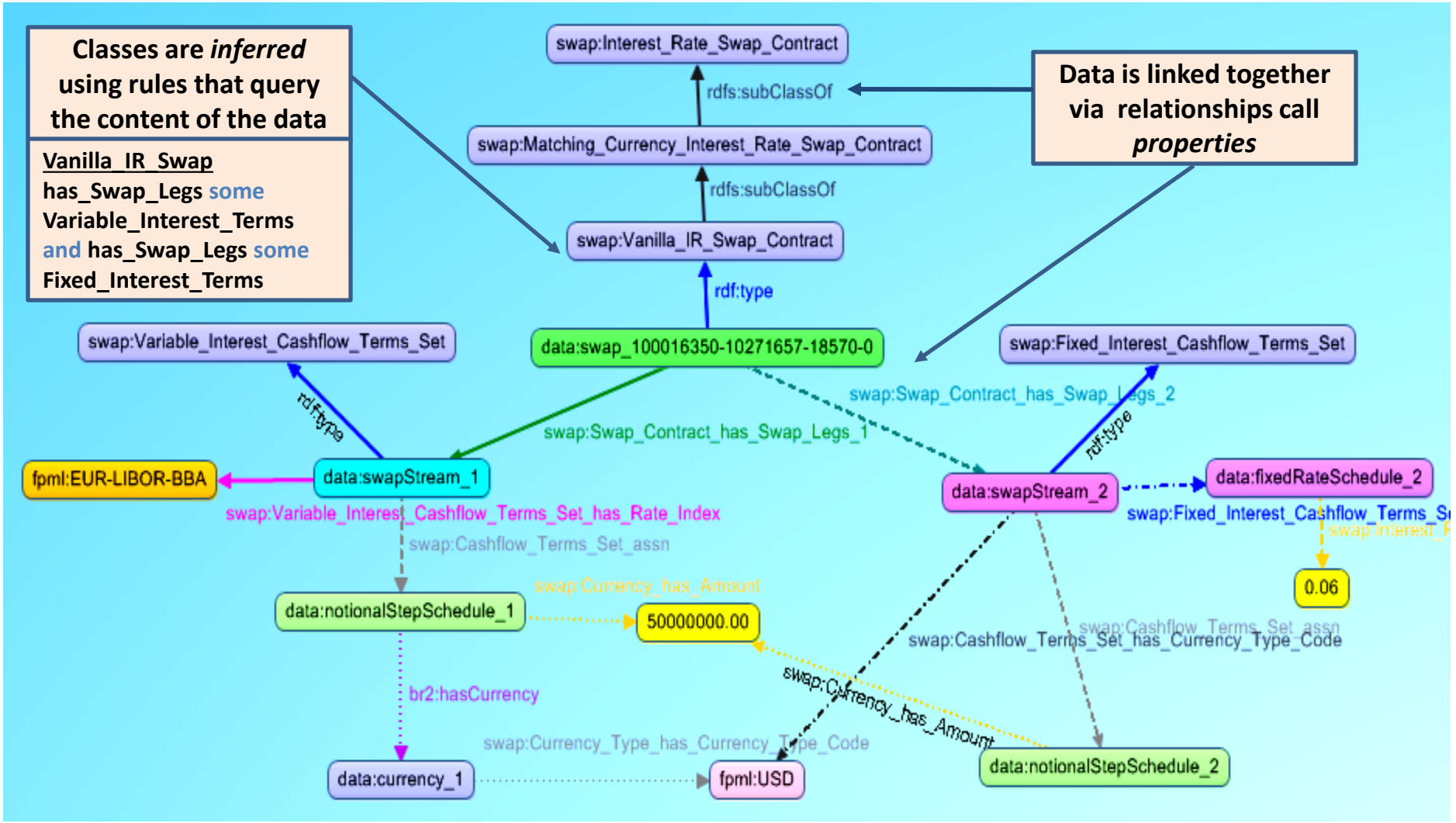
Semantics Offers Differentiating Value to Meet Regulatory Requirements





XML	Relational	Semantics
• Lingua franca of web service messaging payloads following W3C standards • Used to tag data elements with standard labels that conform to a predefined schema • Forms structured data hierarchies • Document hierarchy can be queried	• Dominant database implementation • Highly mature software and tools • Data is physically organized within tables and accessed by matching related columns in different tables that fulfill various conditions	• Emerging form of knowledge representation offers highly intelligent form of data organization • Conceptually describes the meaning of data and its relationships in a way that both people and computers can understand • Supports classification, reasoning and agility
• While tags associate data to labels, the *meaning* of the labels is *not* understood by the computer thereby limiting the extent to which the data can be used and referenced	• Knowledge within application logic • Design, construction, access, schema change and management are brittle, labor, time, resource intensive	• Limited, but growing, set of software, tools • Can supplement XML and relational database • Can begin with knowledge representation and evolve towards operational implementations

Operationalizes Knowledge into Data Structures Reflecting the Actual Data Relationships



* Gruff 3.0 courtesy of Franz, Inc.

Logistics

- Presentations are an INTEGRAL component of the class. Students MUST review material prior to the class.
- Class schedule - note that we meet in VMH 1336 on April 9 and 23.http://www.umiacs.umd.edu/~louiqa/2013/BMGT499B/Readings.html
- Participation
  - Questions before class.
  - Comments after class.
  - Use your name to receive credit.
- Grading
  - Participation and presentation 40-60 %
  - Term paper 40-60 %
- Presentation and paper
  - Pick topic by April 16.
  - 10 minute presentation on April 30.
  - Individual 10 page (or more) report by May 17.

What we did and did not cover?

- What we covered
  - Taxonomies
  - FIBO and taxonomies (Bennett)
  - Semantic Technologies (Newman)
  - SIFI Perspective (Langsam)
  - Regulator perspective (Flood)
- What we did not cover
  - Legal Entity Identifier (Nichols)
  - Mortgage supply chain (Wallace)
- Examples
  - Using a messaging protocol or data standard
  - Financial contracts and markup using the FIBO
  - Financial contract representation and inference - OWL/RDF