Evaluation of Application of Ontology and Semantic Technology for Improving Data Transparency and Regulatory Compliance in the Global Financial Industry

By

Jinchun Chen

B.E. Computer Science & Technology
Guangdong University of Foreign Studies, China, 2010

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MANAGEMENT STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2015

Signature of Author: Signature redacted
MIT Sloan School of Management
May 8, 2015

Certified by: Signature redacted
Stuart Madnick
John Norris Maguire Professor of Information Technology, MIT Sloan School of Management &
Professor of Engineering Systems, MIT School of Engineering
Thesis Supervisor

Accepted by: Signature redacted
Michael A. Cusumano
SMR Distinguished Professor of Management
Program Director, M.S. in Management Studies Program
MIT Sloan School of Management

[Page intentionally left blank]

Evaluation of Application of Ontology and Semantic Technology for Improving Data Transparency and Regulatory Compliance in the Global Financial Industry

By

Jinchun Chen

Submitted to MIT Sloan School of Management
on May 8, 2015 in Partial Fulfillment of the
requirements for the Degree of Master of Science in
Management Studies.

ABSTRACT

In the global financial industry, there are increasing motivations for financial data standardization. The financial crisis in 2008 revealed the risk management issues, including risk data aggregation and risk exposure reporting, at many banks and financial institutions. After the crisis, the Dodd-Frank Act required transaction data of derivatives trades to be reported to Swap Data Repositories (SDRs). In addition, the Basel Committee on Banking Supervision (the Basel Committee) issued the *Principles for effective risk data aggregation and risk reporting* (BCBS 239) in January 2013. These new regulatory requirements aim to enhance financial institutions' data aggregation capabilities and risk management practices.

Using ontology and semantic technology would be a plausible way to improve data transparency and meet regulatory compliance. The Office of Financial Research (OFR) has considered a project recommended by Financial Research Advisory Committee (FRAC) to explore the viability of a comprehensive ontology for solving existing data challenges, such as the Financial Industry Business Ontology (FIBO). FIBO, which could be a credible solution, is an abstract ontology for data that is intended to allow firms to explain the semantics of their data in a standard way, which could permit the automated translation of data from one local standard to another.

This thesis studies the new regulatory requirements, analyzes the challenges of implementing these regulations, proposes a possible solution, and evaluates the application of semantic technology and FIBO with a use case. The thesis tries to explain how semantic technology and FIBO could be implemented and how they could be beneficial to risk data management in the financial industry.

Thesis Supervisor: Stuart Madnick
Title: John Norris Maguire Professor of Information Technology, MIT Sloan School of Management & Professor of Engineering Systems, MIT School of Engineering

[Page intentionally left blank]

Acknowledgements

Firstly, I would like to thank my thesis advisor, Professor Stuart Madnick, for providing me with an opportunity to work on this interesting topic. He accepted me in his research group and allowed me to choose the thesis topic that I felt most excited about. His guidance, assistance, patience, and feedback drove my work towards completion. Also, I would like to thank Allen Moulton for giving me insight and expertise in the domain of my thesis. His valuable feedback and continuous encouragement are indispensable for my work. Meanwhile, Robert Stowsky also provided useful comments and corrections for the thesis. I am so grateful for having their help throughout the research process.

I would also like to thank Marty Loughlin and Phi Nguyen from Cambridge Semantics for offering significant help in my thesis.

Further, I would like to thank Maura Herson and Chanh Phan, who are leading the MSMS program that brought me to MIT. I sincerely appreciate their support and effort in building this big warm family. I also want to thank all my classmates in my program for providing constant support and encouragement. I have really enjoyed my time in the MIT Sloan Community, which features inspiring forums, lectures, events, and other kinds of activities. I will always cherish the moments and memories I have experienced at MIT Sloan.

My special thanks also go to Wenying Tan, who encouraged me, ignited my passion, and offered assistance to my studies at MIT.

The thesis is dedicated to my mother, Lianfang Cen. All the above magic moments could not have happened without her sacrifices, her selfless contributions, and her continual support. She has provided me with unconditional love, guidance, patience, encouragement, and mental and financial support throughout my life. She could not be better as a mother.

[Page intentionally left blank]

Table of Contents

Acronyms

BCBS - Basel Committee on Banking Supervision

CFTC - Commodity Futures Trading Commission

CUSIP - Committee on Uniform Securities Identification Procedures

EDM - Enterprise Data Management

FCIC - Financial Crisis Inquiry Commission

FIBO - Financial Industry Business Ontology

FIX - Financial Information eXchange

FpML - Financial Products Markup Language

FRAC - Financial Research Advisory Committee

FSB - Financial Stability Board

LEI - Legal Entity Identifier

M&A - Mergers and Acquisitions

OFR - Office of Financial Research

OMG - Object Management Group

OTC - Over-the-counter

OWL - Web Ontology Language

POC - Proof of Concept

RDF - Resource Definition Framework

SDR - Swap Data Repository

SEC - Securities and Exchange Commission

SIB - Systemically Important Bank

SIFI - Systemically Important Financial Institutions

SIG - Supervision and Implementation Group

WGSS - Working Group on SIB Supervision

XML - eXtensible Markup Language

List of Figures

1 Introduction

1.1 Global Financial Crisis

The financial crisis that happened from 2007 to 2008, also known as the Global Financial Crisis or the Subprime Mortgage Crisis, was considered one of the greatest recessions since the 1930s. People witnessed the collapse of the stock market. Some large financial institutions, including Lehman Brothers, went bankrupt, while many other surviving banks had to accept bailouts from the government. Personal savings were wiped out and many people lost their jobs.

It has been several years since this financial crisis. Regulators and financial institutions have been making an effort to find out the root cause of this crisis and to implement solutions to prevent future crises. Lots of reports were published and numerous meetings were held for discussion. Root causes and lessons have been identified from different perspectives.

1.1.1 Problems Revealed in Financial Crisis

A variety of problems emerged during the recession. Although it is generally believed that this financial crisis was triggered by the bursting of the U.S. housing bubble, experts hold different points of view on how much weight to give this explanation of a very complex situation. The U.S. Senate Permanent Subcommittee on Investigation released a report on April 13, 2011, stating that this economic collapse was "the result of high risk, complex financial products; undisclosed conflicts of interest; and the failure of regulators, the credit rating agencies, and the market itself to rein in the excesses of Wall Street" (Levin, Bean, & Henderson, 2011).

In other words, practices of risk management were not strong enough to support the financial system. Transaction data was not transparent throughout the financial market, especially for the small players and individuals. Risk exposure was hard to estimate in firms. Regulators could not gain a holistic view of systemic risks in the financial market in time. Poor regulations and supervisions facilitated the growth of bad practices, pulling the market into collapse.

1.1.2 Lessons from Financial Crisis

The Financial Crisis Inquiry Commission (FCIC), which was created to investigate the causes of the financial crisis and explain them to the public, submitted a final report in January 2011. The report pointed out several conclusions of the investigation (Financial Crisis Inquiry Commission, 2011).

1. **This financial crisis was avoidable:** People tended to ignore some warning signals before the crisis and failed to manage risk among financial systems. This crisis could have been avoided if better management practices had been in place.
2. **Failures of regulators:** Insufficient regulation and supervision devastated the stability of the financial market. Deregulation and a high reliance on self-regulation in financial institutions disabled the function of monitoring and controlling by regulators. It showed that even though regulators should have had enough power to control and prevent the crisis, they failed to do so.
3. **Failures of financial institutions:** Inadequate corporate governance and risk management were key causes of this crisis. Systemically Important Financial Institutions (SIFIs) took too much risk on short-term funding, which lead to great losses afterwards.
4. **Uncontrolled over-the-counter (OTC) derivatives encouraged the crisis.** There was no reporting and disclosure requirement needed for participants in the market and no government agencies were responsible for monitoring them. OTC derivatives grew to $673 trillion in notional amount at the end of June 2008 in the market. However, they were unregulated, and this lack of transparency turned into the center of the storm in this crisis.

The government's weak preparation and inconsistent response to the collapse and credit rating agencies' improper measures were also responsible for this crisis.

1.2 Solutions Proposed to the Crisis

Lessons were learned from this crisis, and actions should be taken to prevent another one. Regulations and supervision need to be strengthened in case a similar crisis emerges in the future. Meanwhile, on the industry's side, corporate governance—especially IT governance—and risk management practices need to be enhanced.

1.2.1 New Regulations

Numerous regulations were introduced by governments and regulators after the recession. The Dodd-Frank Wall Street Reform and Consumer Protection Act (the "Dodd-Frank Act"), which was signed by President Barack Obama, came into effect on July 21, 2010. The Dodd-Frank Act brought about changes in regulatory requirements, which impacted the whole financial environment.

In 2010, the Basel Committee on Banking Supervision (the "Basel Committee" or "BCBS") introduced Basel III (or the "Third Basel Accord"), targeting capital, liquidity, leverage, and procyclicality problems, in response to the financial crisis. It aimed to strengthen global capital and liquidity rules, improve risk management of the firms, and bring out data transparency and disclosure of transactions (Basel Committee on Banking Supervision, 2011).

In addition, the Basel Committee issued *the Principles for effective risk data aggregation and risk reporting* (the "Principles" or "BCBS 239" in short) in January 2013. These principles were created to "strengthen risk data aggregation and risk reporting practices at banks to improve risk management practices" (Basel Committee on Banking Supervision, 2013b).

1.2.2 Requirements on Regulators Side

Requirements of regulators were clarified by these new regulations. For example, the Dodd-Frank Act required 243 rulemakings, 67 studies, and 22 periodic reports, according to one estimation (Davis Polk Wardwell LLP, 2010). Accordingly, several new regulatory agencies were established, including the Financial Stability Oversight Council (the "Council"), the Office of Financial Research (the "OFR"), and the Bureau of Consumer Financial Protection (the "Consumer Bureau"). New regulatory power and responsibilities were either created or transferred among new agencies and some existing agencies (Dodd, 2010).

Among these new entities, OFR takes an active role in collecting and providing information and financial data to investors, policymakers and regulators, including the Financial Stability Oversight Council (the "FSOC"), for anticipating emerging risk, identifying vulnerabilities in the whole market, and improving data quality and regulations (OFR, 2010). Data extraction, aggregation, processing and analyzing are the most concerns of OFR.

1.2.3 Requirements on Industry Side

Requirements of the financial industry were published as well. In order to bring data transparency to participants in the derivatives market, the Dodd-Frank Act required that transaction data of derivatives trades should be reported to centralized data warehouses, which are named Swap Data Repositories (SDRs) (OFR Financial Research Advisory Committee, 2014).

Basel III works along with Basel I and II. It focused on operational risks of banks and required these institutions to meet a range of quantitative ratios of various assets. For example, Basel III required the minimum of 4.5% of common equity and 6% of Tier I capital. Basel III covered capital, risk coverage, containing leverage, risk management and supervision, and market discipline, as well as global liquidity standards and supervisory monitoring (Basel Committee on Banking Supervision, 2013b).

Meanwhile, the Basel Committee required Global Systemically Important Banks (G-SIBs), which were designated by the Financial Stability Board (the "FSB"), to meet the Principles by January 2016. It was also suggested that a national supervisor apply the Principles to Domestic Systemically Important Banks (D-SIBs) to spread the recommended practices (Basel Committee on Banking Supervision, 2013a).

It is expected that fulfilling these requirements will lead to better risk management, risk data aggregation, risk reporting, and data transparency in financial institutions.

1.3 Challenges for Meeting the Requirements

A large number of regulations point toward improving risk management, especially risk data management practices. Although requirements have been clarified regarding these regulations, challenges and difficulties have emerged throughout the implementation process.

1.3.1 Overall Industrial Challenges in Risk Data Management

Financial institutions face communication problems among systems internally due to legacy systems and new mergers and acquisitions (M&A). Databases were sometimes designed quite differently using various structures or vocabularies. It would be hard for firms to prepare consolidated reports without significant manual intervention, which impairs timeliness, accuracy and flexibility. How to use a common data model to automatically extract data from disparate sources, and how to aggregate all data into reports in the face of changeable requirements are the major problems to be solved.

Through vast external communication among financial institutions and reports delivery to regulators, it is possible that these firms use same words for different things or different words for same things. Unstandardized and inconsistent definitions act as "foreign languages" to different data sources. Therefore, data aggregation cannot always be automatically processed by machines.

1.3.2 Overall Challenges in Regulatory Reports Analysis

For regulators, there is no dominating data reporting standard. FpML and FIX are major standards, but they are not fully compatible. Lack of consistent data standards hinders regulators from building a holistic picture of systemic risk in the whole market. Meanwhile, as data requirements could be changed from time to time due to different situations, a flexible and extensible data model remains to be developed.

Although the Dodd-Frank Act ordered all OTC derivatives transactions to be reported to SDRs, the OFR Financial Research Advisory Committee (the "FRAC") revealed that the data collected in SDRs were "fragmented and inconsistent" (OFR Financial Research Advisory Committee, 2014). Different formats and standards of data are reported and stored in SDRs, increasing the difficulties of aggregation and analysis.

1.3.3 Actions Needed to Overcome Challenges

Based on the previous descriptions and analysis, in order to meet regulatory requirements and overcome challenges, it is essential to find best practices for managing risk data. Clear and standardized definitions of financial terms, which could cover existing transactions types, also need to be set up. Meanwhile, data formats and standards need to be extensible to preserve new financial products and requirements introduced in the future. This requires the cooperation of the OFR with financial institutions, SDRs, and regulators, including the Commodity Futures Trading Commission (the "CFTC"), to develop a viable approach for solving existing data challenges. As SDRs are required to submit annual compliance reports to the CFTC (CFTC, 2015), better data management practices are also critical for solving issues of data aggregation from multiple SDRs and reporting to regulators.

In its annual report for 2014, the OFR mentioned that they had been seeking practical ways to create a financial instrument reference database "with the greatest benefit and lowest cost to the public and private sectors" (Office of Financial Research, 2014). Benefits of using ontology were mentioned for creating the financial instrument reference database. However, an evaluation of ontology's feasibility and performance remains to be accomplished.

1.4 Thesis Structure

This thesis studies challenges for meeting requirements on the industry side and proposes solutions to these challenges. Including the background information and introduction provided in Chapter 1, there are 7 chapters in this thesis. Chapter 2 provides a review of literature and defines key terms, including systemic risk, semantic technologies, ontology, FIBO, and other data standards. Chapter 3 provides an analysis of implementing BCBS 239. Chapter 4 and 5 lay out a use case serving as a POC experiment for implementing semantic technology and part of FIBO on Anzo. Chapter 6 presents the conclusion of this study, and Chapter 7 identifies several directions for future research to address.

2 Literature Review

Given that the final goal of this study is to assess and control risks, particularly systemic risk in the whole financial market, it is important to understand the current state of research addressing systemic risk.

2.1 Systemic Risk

2.1.1 Definitions of Systemic Risk

Systemic risk in financial markets is still a relatively new research domain, and there is no formal and agreed upon definition of systemic risk. In the *Handbook on Systemic Risk*, Jean-Pierre Fouque et al. pointed out a temporary working definition in use, saying that systemic risk is "the risk of a disruption of the market's ability to facilitate the flows of capital that results in the reduction in the growth of the global GDP." Eric Hughes et al. also stated that systemic risk is "a threat to a large, single component of the financial systems poses to the system as a whole" (Fouque & Langsam, 2013).

In other words, systemic risk is the risk of impairing the function of the whole system, which would in turn have a huge impact on economic growth. Meanwhile, because of globalization and the interconnection between world economic entities, the "system" here does not mean any specific local market, but the global market. Lessons have shown that a recession from a country or region could have a chain reaction, causing a global crisis.

2.1.2 Approaches for Assessing Systemic Risk

Eric Hughes et al. proposed several computational approaches to help agencies, such as OFR, collect financial data for analyzing and producing reports (Fouque & Langsam, 2013). Different approaches are based on corresponding database models, including relational databases, parallel databases, NoSQL databases, and semantic databases. Financial organizations need to consider both the advantages and the disadvantages of each approach and choose the most suitable one.

Relational databases. Relational databases have become the industry's norm as a mature and robust approach for a long time. Data warehouses can be built to store data in tables. Incoming data will be transformed into specific data models of targeted data warehouses via Extract, Transform and Load (ETL) process. After collection, data can be aggregated into data marts, which are databases designed for data analysis. Structured Query Language (SQL) is used for operating in relational databases without specifying details for data fetching methodology, which could be taken over by database systems. The processes of relational databases are depicted by Figure 1.

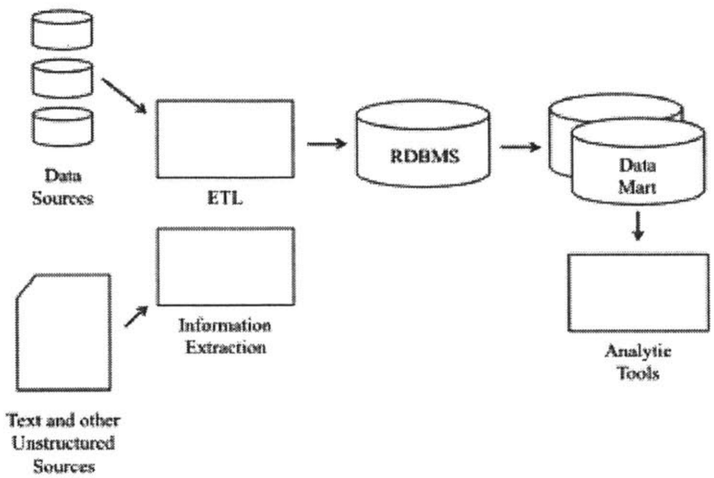


Figure 1: Relational Databases Process (Fouque & Langsam, 2013)

However, because of the relatively fixed design of data structure, it is difficult to promote a single data structure or standard to various systems in different organizations. Communications can be difficult among multiple databases. When data sizes become larger and larger, relational databases could be inefficient. Meanwhile, unstructured sources such as text, emails, and articles cannot be processed by relational databases directly. This means that knowledge and information cannot be automatically generated without manual intervention.

Parallel databases. Compared to relational databases, massively parallel processing (MPP) could be used to manage large data sets. MPP databases store data across a cluster of distributed servers, which have independent storage and computing capacities. Consequently, as data size increases, users simply need to add more servers in response. At the same time, parallel means that parallelism instead of indexes was adopted during querying in MPP databases. Parallel databases also load and process data in parallel processes, which increases both processing efficiency and flexibility. Therefore, data analytics can be performed directly on corresponding data warehouses in parallel.

Nonetheless, parallel databases could be regarded as an advanced version of relational databases because unstructured sources still cannot be processed directly.

NoSQL databases. NoSQL databases have relatively loose requirements and restrictions on data compared to traditional databases. They can improve the capacity of data aggregation and analytics of heterogeneous data sets. NoSQL is often used in systems that have steady flow of incoming data and no need of modification. In this case, ordinary SQL queries are usually abandoned and only the searching function works.

Compared to traditional databases, NoSQL databases have the advantage of supporting multiple data models, including semi-structured data. Because new columns can be inserted at any time, the flexibility of database design is significantly improved. Data in NoSQL databases is also stored and shared in a network of distributed servers, enjoying similar benefits to MPP databases. Meanwhile, ETL is not required for data ingesting, enabling quick analysis along with the risk of low data quality.

Nevertheless, as current development of NoSQL databases is still at early stage, there are concerns about security and access control, which are sensitive issues in financial firms.

Semantic databases. Semantic is a promising technology for automatic processing. Unlike traditional databases, semantic databases represent the data using subject, relationship, and object, which are called triples. One of the advantages of using triples is enabling automatic logical reasoning, which is valuable to the financial industry. Considering it is still a new technology, more research needs to be conducted in this field.

2.2 Semantic Technologies

2.2.1 Introduction of Semantic Technologies

Semantic technologies use the explicit representation of meaning to enable seamless interoperability and communication among different data sources, without the need of changing physical data structures, vocabularies, and other core opponents. Compared with some data exchange standards, such as XML, which focus on the syntactic or symbol-processing level, semantic technologies are created for the level of meaning of information.

People are more interested in knowledge rather than pure data or numbers. However, there are challenges for machines to process data and develop comprehensive knowledge automatically without understanding the exact meanings of each data element. Semantic technologies are able to process information based on the meaning, or semantics, of data. By using logical languages, semantic technologies can attach explicit structures and meaning to corresponding data, providing the possibility of automated processing through machines. While semantic technologies provide shared meaning of data, a linked virtual data structure will be created for searching, querying, and reasoning the relevant data across different data sources and systems (Rittgen, 2008).

Anne M. Cregan provided a strategy overview of semantic technologies in the Handbook of Ontologies for Business Interaction, as illustrated in Figure 2 (Rittgen, 2008).

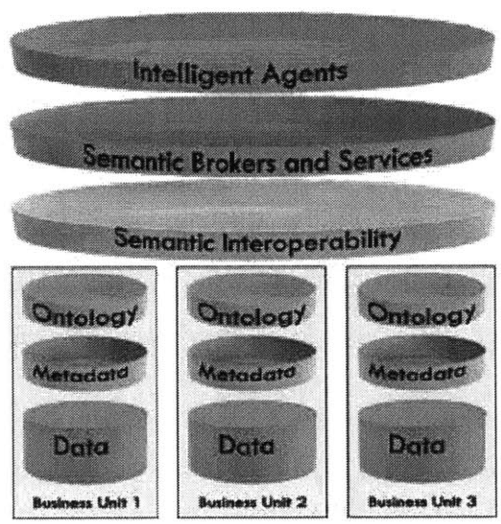


Figure 2: Strategy Overview of Semantic Technologies (Rittgen, 2008)

Data. Data can be in any structured or unstructured format from disparate data sources, including traditional databases, Excel spreadsheets, and documents such as PDF files.

Metadata. Metadata is "data about data," describing all the information about physical data. It acts as a bridge, linking physical data to ontology models without reconstructing the existing structures of databases.

Ontology. Ontologies are logical conceptual structures, organizing metadata and describing individuals, concepts, and the relationships among them.

Semantic Interoperability. Ontology models form a semantic network and a virtual view across multiple databases, providing semantic interoperability. All data looks as if it is stored in a single database that is constructed by these ontology models, although physical data actually comes from disparate databases.

Semantic Brokers and Services. Semantic interoperability furnishes a platform for semantic operations, including intelligent search and reasoning, across the interlinked data.

Intelligent Agents. Intelligent agents are user interfaces to solve specific tasks, providing the functionalities of semantic operations to process interlinked data across disparate sources. All the previous levels are fundamental for intelligent agents.

2.2.2 Applications of Semantic Technologies

Semantic technologies aim for two goals. The first is to enable distributed data sources, which could have various designs, to communicate and interoperate with each other. The other one is to offer tools for semantic operations, including searching, querying and reasoning. Meanwhile, although some of the data is stored in structure formats such as XML and traditional databases, there are still lots of data presented in unstructured formats such as RDF documents (Rittgen, 2008).

Semantic technologies, or specific Semantic Web applications, typically use files to express "triples," namely the relationships of "Subject-Verb-Object." These relationships are defined in files called Resource Definition Framework (RDF), which can be referred to as ontologies as defined in Web Ontology Language (OWL) files (Fouque & Langsam, 2013).

Semantic technology applications furnish semantic functionalities. Reasoning engines can draw inferences based on the "Subject-Verb-Object" relationship. Implicit knowledge can be explicitly expressed by the reasoning results. Semantic querying enables semantic searches of data with the query language called SPARQL. The query is run on the precondition of an unambiguous definition of the context and terms. Automatic classification works with the reasoning engine to automatically classify information. In the financial industry, automatic classification is valuable for classifying financial products and performing scenario-based analysis.

2.2.3 Comparison between Conventional Data Model and Semantic Model

A report of the Object Management Group (OMG) pointed out that there were some factors that distinguished semantic models from conventional logical data models (Object Management Group, 2014b).

1. **Design.** Logical data models represent "the results of some design effort," although they are independent from any database platform (Object Management Group, 2014b). Data model designers work from a range of knowledge and solutions. Hence these data models normally do not generate new knowledge. Instead, they present data within a predefined knowledge scope. For semantic models, they do not represent any design or specific solution. They are designed for representing and describing items and relationships in problem domains. As the logical framework is not predefined, semantic models usually can provide new knowledge and explicitly express implicit information derived from semantic reasoning and automatic classification.
2. **Representation format of items.** As logical data models represent the logical design of databases, these models normally contain traditional structures, such as tables, keys, and so forth. In contrast, semantic models do not represent logical design. They explicitly represent entities as kind of "thing" in the semantic model by using set theory concepts rather than specific tables.
3. **Reuse data structures.** Logical data models usually reuse common physical data structures for efficiency without considering actual meaning of data elements. In contrast, semantic models have separate definitions of each "thing" and only reuse data structures if they express general concepts.
4. **Inheritance.** Logical data models are normally inherited from a single hierarchy of class, while semantic models use multiple inheritances. Entities in semantic models can be classified into different taxonomies according to different properties.
5. **Optionality.** In the logical data model, a practical approach is to have some optional properties or fields that may not be applied to any given data elements. These optional properties can be utilized in different ways according to specific needs. In comparison, semantic models only include applicable factors to a given type according to semantic meaning.

6. **Enumerations.** It is common practice in logical data models to create a single data element for enumeration to distinguish among different kinds of "thing." However, for semantic models, each "thing" is an independent class in the enumeration, distinguishing itself from the others by definitions.
7. **Assumption.** The logical data model is built on Closed World Assumption (CWA), which means "absence of evidence is evidence of absence." In contrast, the semantic model is built on Open World Assumption (OWA), which means "absence of evidence is not evidence of absence" (Object Management Group, 2014b). In practice, facts can be represented in the semantic model without actual data. This enables semantic models to define different concepts.

2.3 Ontology

The increasing prevalence of Semantic Web and Web 2.0 raised the interest of ontology technology, which sits in the heart of semantics (Shadbolt, Hall, & Berners-Lee, 2006). Ontology can be implemented in different areas, and can be especially powerful in the business domain.

The term "ontology," borrowed from philosophy, often refers to "a formalization of conceptualization" (Fouque & Langsam, 2013). It is a formal and logically consistent model that can be read by both human and machines.

Mike Atkin and Mike Bennett pointed out that there were practical differences between Business Ontology (also called Conceptual Ontology) and Operational Ontology (Fouque & Langsam, 2013). In the financial industry, business ontology is a conceptual model that provides concepts and relationships in business. It is based on business people's minds, and it is not optimized for practical applications. An OMG specification defined the business conceptual model as "a model which represents and only represents business subject matter," which means information and data about "things" in the business domain, "without reference to the design of any solution or data model representation" (Object Management Group, 2014a). Business ontology focuses more on defining and describing all the "things" in the business domain without considering technical constraints.

On the contrary, operational ontology is designed for semantic applications. It is defined as "an ontology which is intended for use within some application" (Object Management Group, 2014a).

There are some limitations and principles applied to this ontology in order to make it operable for machines. For example, a conceptual ontology can be exhaustive in order to perfectly capture and describe all aspects of and relationships in the universe. But its implementation will be impractical if it takes one year for a machine to run a simple semantic query. Therefore, usually only a sub-set of the corresponding business ontology will be involved to create an operational ontology model.

2.4 Financial Industry Data Standards

As each financial institution has a variety of business requirements, proposing a single agreed upon data standard seems impossible. Yet there are two widely accepted standards in the market.

2.4.1 FpML

The Financial products Markup Language (FpML) is an open industry-standard protocol, based on eXtensible Markup Language (XML). Since its initialization in 1997 as a research project, FpML has been developed for processing complex financial products, especially OTC derivative transactions (Fouque & Langsam, 2013).

FpML was initially used in the swaps affirmation service, confirmation, and trade warehouse service of DTCC (the Depository Trust & Clearing Corporation) for credit default swaps. Later, with support of the SWIFT network, FpML was also used in describing the trade life cycle events of various products.

With the growing numbers of trade data encoded, currently FpML is largely used by several entities, including brokers and dealers in the credit market, exchanges and clearing institutions, and financial institutions, for various purposes (Fouque & Langsam, 2013).

2.4.2 FIX

According to the FIX Trading Community, the Financial Information eXchange (FIX) protocol is another prevailing data messaging standard being used by buy-side and sell-side firms, trading platforms and regulators to exchange trading information. It is widely applied in equity trading from pre-trade to post-trade spaces, supporting Straight-Through Processing (STP) via a machine readable format. It also covers fixed income, foreign exchange, and derivative markets. In addition, FIX is also utilized for regulatory reporting by clearing organizations such as the CME

Group for OTC derivatives. Currently FIX is owned by members of the FIX Trading Community (FIX Trading Community, 2015).

2.5 FIBO

As demand increases for solving the data management and integration problem, the Financial Industry Business Ontology (FIBO) was initiated in 2008 by the Enterprise Data Management Council (the EDM Council) to draft a "Semantic Repository" (Semantics Repository, 2015). The EDM Council is a non-profit trade association founded by the financial industry, aiming to improve practices of data management. Currently FIBO acts as a formal standard of the OMG and is managed by the EDM Council membership. (EDM Council, 2015).

2.5.1 Definition of FIBO

According to the definition introduced by the EDM Council, FIBO is a business conceptual standard that provides a set of formal models that define and describe the structure of financial industry concepts, especially "contractual obligations of financial instruments, legal entities, market data and financial processes" (EDM Council, 2015). The Committee draft of ISO 20022, "Financial Industry Business Information Model" (FIBIM), provided a set of logical data model terms, and ISO 10962, "Classification of Financial Instruments," provided a basis for categorizing securities. Later, draft models of derivatives, market data, funds and loan terms were also developed (Bennett, 2013).

Given that FIBO is positioned as business ontology rather than as an operational ontology, the focus on developing FIBO is to describe and express the financial world comprehensively, without taking technical constraints into account.

Nonetheless, the final objective of developing FIBO is not making a beautiful model on paper just to put it aside. Hence one of today's emerging issues is to transfer FIBO into operational ontology models that can be processed by machines in real life.

2.5.2 Elements of FIBO

On the Semantics Repository webpage, the EDM Council points out that there are some key elements of FIBO ontology, including Business Conceptual Ontology, Adaptive Web Presentation Facility, FIBO OMG Specifications, and FIBO Operational Ontologies. The structure is presented in Figure 3 (Semantics Repository, 2015).

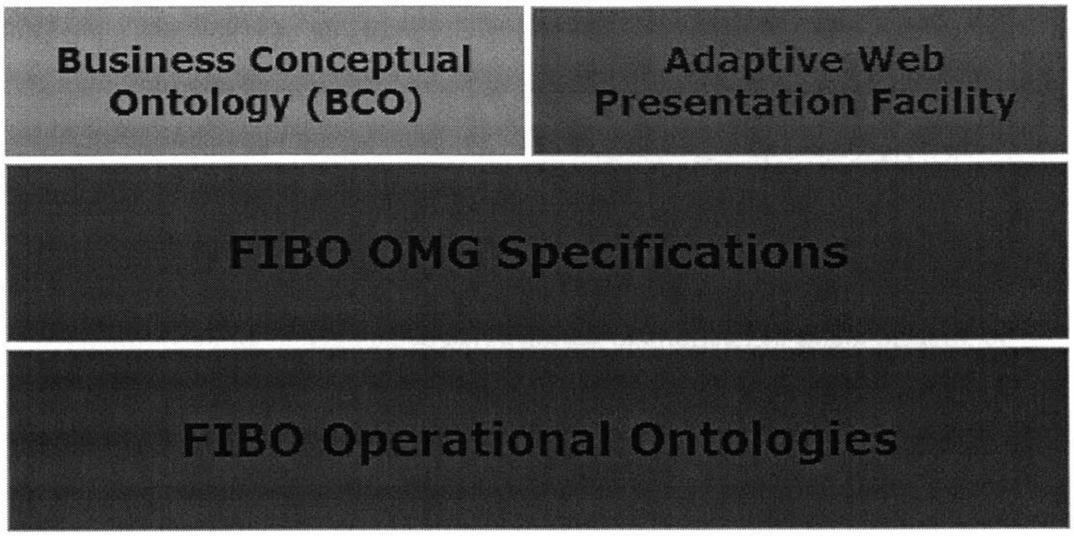


Figure 3: Element of FIBO (Semantics Repository, 2015)

1. **Business Conceptual Ontology (BCO).** FIBO contains a range of models grounded in legal and financial realities, framed within Basic Business Ontology (BBO). Models are created independently from any specific application or data structure, covering business entities, securities, derivatives, and other types of financial instruments. Basic Business Ontology, or Global Terms, provides a fundamental framework at a more general level, supporting each detailed financial model. These global terms are not limited to the financial industry, acting as archetypes for describing the simplest kind of "thing."
2. **Adaptive Web Presentation Facility.** FIBO specifications should be presented to and validated by experts on the business side. To illustrate the models, several presenting methodologies are needed, including spreadsheets with tables of terms, diagrams with boxes and lines, and other types of visualization channels.
3. **FIBO OMG Specifications.** FIBO specifications are a set of Request for Comments (RFC) documents under the OMG standards process. These proposed specifications are required to cover all possible use cases in business domains.
4. **FIBO Operational Ontologies.** Operational ontologies are based on RDF/OWL and are created for specific purposes. Compared to conceptual ontology, which defines general meaning, operational ontology should take instance data into account and focus on specific use case. Although FIBO conceptual ontologies provide a full range of terms, relationships and semantic functions such as multiple inheritances, operational ontologies

should consider technical constraints and take only a sub-set of full models, enabling the efficient delivery of results.

2.5.3 Scope of FIBO Application

The primary purpose of developing FIBO is to provide a common and unambiguous shared meaning across data sources. It could be used to automate the process of data extraction and aggregation, helping financial institutions to perform risk data analysis. Mike Bennett also pointed out that there are at least two benefits that could be delivered by FIBO, including provision of common semantics and new ways of analyzing large data sets (Bennett, 2013).

To monitor systemic risk, the OFR needs to capture all relevant factors, including the financial instruments' issuer, counterparties, and other details of trade at the granularity of transaction level. As these factors can be modeled in the "Subject-Verb-Object" structure, FIBO is able to capture and process this data via semantic technology. FIBO can also be applied in a broader way to fully capture all the factors relative to systemic risk, including institutional facts (such as positions, assets, and holdings) and systemic facts (such as currencies and financial sectors) (Fouque & Langsam, 2013). As FIBO is still under development and can be easy extended, it is expected that FIBO can cover more domains in the future.

2.6 LEI Problem

A universal and unique identifier is useful for identifying and matching entities across databases, including semantic databases. Although there are several existing schemes for identifying entities, currently there is no such industry-wide identifier in use. Lessons learned during the financial crisis showed that financial firms had difficulties identifying and aggregating their risk exposure to counterparties, as many relevant entities and subsidiaries were involved.

2.6.1 Definition of LEI

According to the OFR's report, a Legal Entity Identifier (LEI) is "a number that identifies a legally separable business entity that engages in financial transactions" (OFR, 2010). In another report, LEI was defined as "a reference code to uniquely identify a legally distinct entity that engages in a financial transaction" (Lei & Brothers, 2013).

The Trade Associations Global LEI Proposal (the Proposal) stated that a LEI solution should, among other characteristics, be persistent, neutral, singular, unique, extensible (scalable), structural fixed, reliable (quality), and interoperable. It also declared that LEI should cover transacting entities, issuing entities, reference entities, reporting entities, ultimate parent entities, and other participants in financial transactions (The Trade Associations Global LEI Proposal, 2011).

In other words, LEI information can be used in financial transactions to identify parties involved, including counterparties or other participants in the market, enabling logical reasoning and risk exposure analysis.

2.6.2 Benefits of Adopting LEI

Several benefits were identified for adopting LEI, including reducing operational errors and risk, saving costs, and providing a consistent framework for collecting transactional data (Fouque & Langsam, 2013).

The Proposal also claimed that LEI can help "organizations to more effectively measure and manage counterparty exposure, while providing substantial operational efficiencies and customer service improvements" (The Trade Associations Global LEI Proposal, 2011).

OFR's annual report pointed out that LEI can increase the credibility of risk reports for both supervisors and financial firms. With the development of a holistic view of risk exposure, regulators can enhance their capability to monitor and analyze systemic risk. At the same time, risk managers can evaluate corresponding internal risk more efficiently. For regulators, LEI can improve their capacities of risk analysis, supervision, and regulation when the affiliation of each entity can be clarified. From a cost saving perspective, a universal LEI can save costs related to risk data management, including data collecting, cleaning, aggregation, and reporting to regulators. Meanwhile, for financial organizations, operational risk and internal risk management can be improved by adopting LEI. In such cases, they will be able to roll up their risk exposure to specific firms. In general, the industry's market discipline can also be improved due to complete reflection of risk (Lei & Brothers, 2013).

2.6.3 Challenges for Establishing LEI

Although the adoption of a global LEI has plenty of benefits, it is still hard to establish this global identifying system. There are several reasons why this system is not ready yet.

The most concern is lack of incentives. Because benefits of adopting LEI rely on large scale of application, any individual firm does not have sufficient incentive to change its system and apply the new standard. It will take a long time to get enough firms involved to enjoy network effect.

Another problem is coordination. Because of the sheer number of participants in the global market, it is difficult to agree upon and implement a single global solution. Moreover, the application scope and granularity of LEI are not clear. It could involve longtime political negotiations and technical development.

To overcome these challenges, stakeholders, including regulators, industry, standard setters, should be involved and collaborated. Regulatory pressure could act as an external incentive to force firms to adopt LEI. Successful development and implementation of LEI also require stakeholders' support and coordination.

3 Case Analysis: Implementation of BCBS 239

Assessing and controlling systemic risk is a huge, complicated project, requiring corresponding support from policies, IT infrastructure, and advanced risk management practices. The financial crisis has revealed that none of these three factors were sufficient enough for a robust financial system. Given these facts, guidelines and mandates were published in order to improve regulators' performance, financial firms' IT infrastructure, and industrial risk management practices, especially in the domain of risk data management.

The Principles for effective risk data aggregation and risk reporting (the Principles), also known as BCBS 239, was released in January 2013. The Principles aim to improve the risk data aggregation capabilities of financial institutions and regulatory reporting process. The final goal is to strengthen risk management in financial markets and decision-making processes and resolvability in financial firms. It is an important step for both firms and regulators to assess and control systemic risk. These principles should be applied not just at the group level, but to all lines of business and legal entities of the group (Basel Committee on Banking Supervision, 2013a).

3.1 Background Introduction

BCBS 239 revealed that the lack of sufficient information technology and data architecture to support risk management was one of the most noticeable lessons from the financial crisis. Financial institutions' ability to aggregate risk data internally and perform a corresponding analysis of this data was called into doubt. In this crisis, financial institutions failed to estimate and handle the systemic risk with timeliness and accuracy, damaging financial stability in the whole market. Consequently, industrial risk management practices turned out to be the most critical issue to be addressed.

After the crisis, several new regulations and guidelines were proposed, including BCBS 239. The financial industry was kept under pressure about data because all participants, including regulators and financial institutions, have to deal with new regulatory requirements aimed at improving risk management and largely involving best practices of data management. The importance of leveraging information technology in financial institutions, especially developing better practices of data management, is stressed.

3.2 Concerns of Financial Institutions

BCBS 239 forces financial institutions, especially SIBs, to improve their data management practices, not just for internal decision processes and risk management, but also for external regulatory reporting. The requirements of meeting these principles could be regarded as both opportunities and as challenges for firms.

As opportunities, financial institutions could take this chance to develop a better practice of leveraging internal data and information flows, which could generate a much more comprehensive and holistic view on their data. Given increasing competition in the market, better data management practices based on innovative information technology would create competitive advantages for these institutions. Considering risk management capabilities also reflects the reputation and credibility of the organizations, so these financial institutions can take this chance to increase brand impact and the value of the companies. Improving data management can also improve strategic decision-making for investments, which would in turn result in higher profitability. In addition, once firms are able to better estimate their risk exposure to each asset, they can avoid taking too many risks. Therefore their operational risk and cost can be reduced.

On the other hand, as challenges, financial institutions need to take these problems seriously and find proper solutions. Challenges include difficulties in changing internal processes and significant costs of improving IT infrastructure.

To some extent, financial institutions are publishers of financial knowledge. They have plenty of internal and external data in stock. However, leveraging this data is possibly hindered by their habits of working in silos and legacy applications. Changing internal processes could be costly and difficult.

Meanwhile, firms need to facilitate scenario-based analysis and report to regulators for cross-industry analysis. However, self-assessments revealed that they have a shortage in IT governance and infrastructure, data aggregation, and reporting processes (Basel Committee on Banking Supervision, 2013b). Considering risk management, which is largely relied on in risk data management, is essential to keeping financial institutions running healthily. Therefore, financial institutions need to spend resources to improve risk data management practices to prevent future crises.

3.3 Concerns of Regulators

When Lehman Brothers collapsed, the financial industry did not have effective risk data aligned with all entities under its name, making it difficult for regulators to estimate the impact of its collapse. Taking this lesson into account, for regulators, the major domains required for improvement were risk reporting and monitoring.

According to the definition provided in BCBS 239, the term "risk data aggregation" means "defining, gathering, and processing risk data" and "sorting, merging or breaking down sets of data" according to the requirements of regulatory reporting (Basel Committee on Banking Supervision, 2013b). In addition, "risk data" includes all data which is critical to managing financial institutions' risk. Risk data and reports should be complete and comprehensive to provide sufficient information about firms.

In other words, financial firms' capacity for risk data aggregation should be strong enough in response to requirements of regulatory reports. As regulators have great concerns about the quality of regulatory reports, they need to ensure that firms have adequate data aggregation capacities.

The principles listed in BCBS 239 were intended to be implemented to all G-SIBs and D-SIBs if necessary. A Timetable for implementation was provided in BCBS 239 as well. G-SIBs, which were designated before November 2012, must meet the Principles by January 2016, while other G-SIBs need to meet the Principles within three years of designation. The process will be monitored and assessed by national supervisors and the Basel Committee.

In summary, regulators and supervisors should provide clear guidance and necessary assistance to financial firms in order to ensure that firms can build up a sufficient capacity of risk data aggregation and reporting on time.

3.4 Principles in BCBS 239

There are 14 principles covering four domains in BCBS 239, including governance and infrastructure, risk data aggregation, risk reporting, and supervisory guidance (Basel Committee on Banking Supervision, 2013a).

3.4.1 Overarching governance and infrastructure

A strong governance framework, risk data architecture and IT infrastructure are preconditions and foundations of other principles. Also, financial firms need to ensure that both internal and external communication and reporting are clear and consistent. Principle 1 requires strong governance arrangements that are in compliance with policies created by the Basel Committee. In an organization's overall risk management framework, the board and senior management should ensure the involvement of "identification, assessment and management of data quality risks" (Basel Committee on Banking Supervision, 2013a). Meanwhile, standards and policies should be made to ensure the confidentiality, integrity and availability of data.

Principle 1 also claimed that risk data aggregation and risk reporting practices should be fully documented. Considering M&As are common in financial industry, the same practices should be applied to any new initiatives and validated independently, ensuring all processes are functioning appropriately and as intended. These practices should be independent from the business structure as well.

Along with governance, Principle 2 requires robust data architecture and IT infrastructure that fully support risk data aggregation and reporting practices. Both business and IT functions should cooperate closely and take the roles and responsibilities of managing risk data and information.

3.4.2 Risk data aggregation capabilities

Based on strong and robust governance, data architecture and IT infrastructure, financial institutions should develop and maintain risk data aggregation capabilities to provide comprehensive reports for reflecting and monitoring risks. Risk data qualities could show how well a firm can provide the foundation for reporting and decision making. Principles 3 to 6 stress different characteristics of data quality required, including accuracy and integrity (Principle 3), completeness (Principle 4), timeliness (Principle 5), and adaptability (Principle 6).

3.4.3 Risk reporting practices

Outstanding quality of data is necessary and fundamental, but it is not sufficient and guaranteed for generating desired reports and making effective decisions. Based on proper governance, data architecture and IT infrastructure mentioned in the first domain, and high quality data required by the second one, firms should be able to generate risk reports that can provide the right information to the right people at the right time. As risk reports are essential both to senior managers in firms, who use them to make decisions, and to regulators, who use them to gain a holistic view of systemic risk, Principles 7 to 11 raise several important qualities that are required for risk reports. They are accuracy and preciseness (Principle 7), comprehensiveness (Principle 8), clarity and usefulness (Principle 9), frequency (Principle 10), and distribution (Principle 11).

3.4.4 Supervisor review, tools and cooperation

The last part of BCBS 239 identifies national banking supervisors' roles in monitoring and guiding financial institutions in implementing the above principles. All previous domains work under this framework. Principle 12 claims that supervisors have the responsibility to review their bank's compliance with all above principles. Supervisors need to conduct this review regularly to ensure that these principles result in the expected outcomes. If anything goes wrong, Principle 13 stresses that supervisors should require appropriate remedial actions be taken by financial institutions and that they use a set of tools to address material deficiencies in the data aggregation and risk reporting process of these firms. Furthermore, Principle 14 states that supervisors should cooperate effectively and share information with relevant supervisors when financial institutions operate in different jurisdictions.

In summary, Figure 4 provides brief descriptions of each domain and the relationships among them.

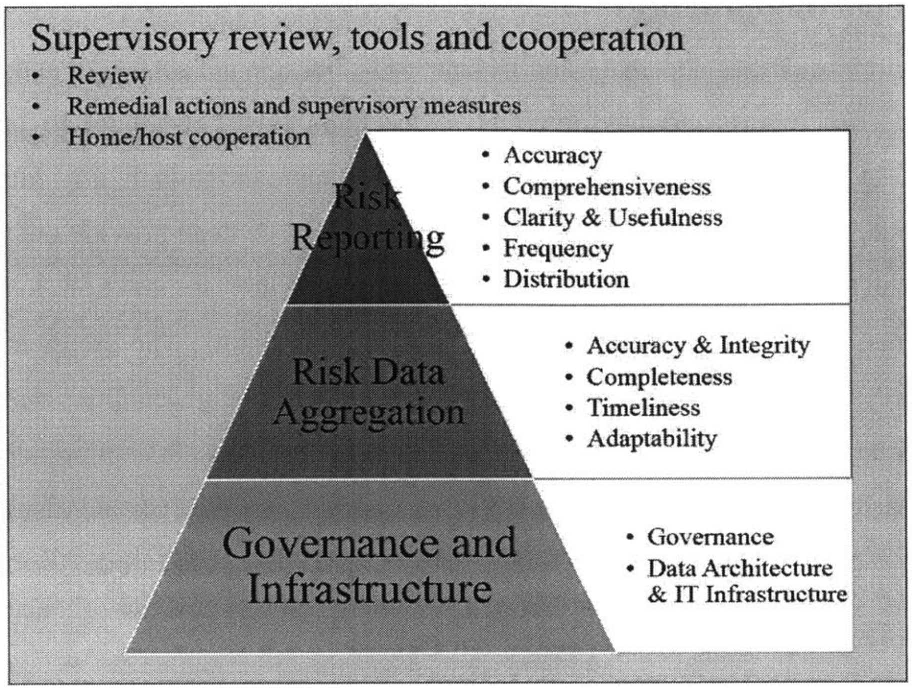
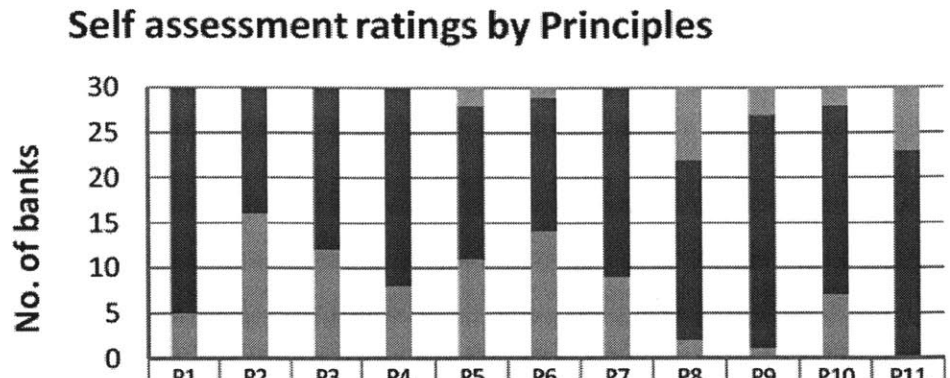


Figure 4: Principles in BCBS 239

3.5 Challenges to Be Solved

Besides Principles 12 to 14, which are designed for supervisors, SIBs need to implement necessary changes and meet other principles as scheduled. The Basel Committee's Supervision and Implementation Group (the "SIG") is responsible for monitoring and assessing the process of implementing BCBS 239 by G-SIBs. SIG developed a "stocktaking" self-assessment questionnaire to understand firms' capabilities. The questionnaire contained 87 questions for 11 principles and was analyzed by the Basel Committee's Working Group on SIB Supervision (the "WGSS"). Results showed several key weakness existing in G-SIBs, as illustrated in Figure 5 (Basel Committee on Banking Supervision, 2013b).



	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Fully compliant	0	0	0	0	2	1	0	8	3	2	7
Largely compliant	25	14	18	22	17	15	21	20	26	21	23
Materially non-compliant	5	16	12	8	11	14	9	2	1	7	0
Non-compliant	0	0	0	0	0	0	0	0	0	0	0
Average rating	2.8	2.5	2.6	2.7	2.7	2.6	2.7	3.2	3.1	2.8	3.2
	Governance/ infrastructure		Risk data aggregation capabilities				Risk reporting practices				

Figure 5: Self-Assessment Rating (Basel Committee on Banking Supervision, 2013b)

This 2013 survey revealed that there were several outstanding challenges for implementing all 11 principles. In the above table, P1 means Principle 1 and so on. The WGSS believed that the results broadly reflected participating banks' current state of implementation, even though some potential weaknesses could be missing (Basel Committee on Banking Supervision, 2013b). Although BCBS 239 was designed to apply to each level and entity at each firm, it turned out that many firms only focused on the group level and that risk reports were only available at the senior management and board level. Besides, their risk reports did not cover all types of risk, which could not meet principles of comprehensiveness and completeness. One of the most critical issues was that instead of automated data aggregation, organizations relied heavily on manual process, which hindered the accuracy and timeliness of data aggregation, especially when under pressure during a crisis.

3.5.1 Challenges in Governance and Infrastructure

Strong IT system infrastructure and corporate governance within firms are critical for compliance with BCBS 239. According to the results, although 83% of participating banks rated themselves as largely compliant with Principle 1 (Governance), Principle 2 (Data architecture and IT infrastructure) received the lowest rating among all principles at 2.5 (full score 4). These results suggested that data architecture and IT infrastructure would be a bottleneck for SIBs.

To meet Principle 1 and Principle 2, organizations need to upgrade legacy IT systems and establish relative standards and policies. Meanwhile, formal, comprehensive, and fully documented frameworks should be made to support data aggregation and reporting as preconditions. All these activities require awareness, high engagement, and support from senior management. Moreover, cooperation between business and IT functions and clear ownership of data are also critical factors to ensure effective and efficient implementation.

However, improving IT infrastructure and data management practices means a large investment of financial and human resources. As the benefits of implementation and the danger of ignoring risk management may not be revealed in the short-term, it is possible that financial institutions are reluctant to expend resources and attention in these fields. In large organizations, working in silos also hinders effective cooperation between the business and IT teams. This means that they need to establish a sustainable model for improving IT infrastructure and practices.

3.5.2 Challenges in Risk data aggregation

Risk data accuracy, completeness, timeliness, and adaptability are emphasized in BCBS 239. High quality of data is the precondition of high quality of reports. However, some legacy practices in use by financial institutions are obstacles to reaching these goals. The 2013 survey also reflected a low score in this section, as graded by banks themselves. Results showed that ratings of principles related to risk data aggregation were also relatively low, especially for Principle 3 (Accuracy and Integrity) and Principle 6 (Adaptability).

As financial institutions still rely on manual process of data aggregation and extraction, accuracy and timeliness could be impaired due to unavoidable human mistakes. Although manual interventions are necessary for some cases, banks should take a trade-off between manual processes and automated processes.

Even though data quality problems could be solved by several tools and software, there are still multiple challenges for managing data, especially in a specific context such as analyzing or integrating financial instruments and transaction records.

Inconsistent taxonomies, metadata, identifiers and dictionaries also hinder the data aggregation process, especially during data integrating and comparing among repositories. In the past, financial institutions used to manage data vertically following business processes, without too much horizontal communication. However, today, collaboration and information sharing across different teams are more stressed. The nature of financial processes and activities is built on interconnections with different entities through all financial instruments and transactions. Information and data need to be better shared, both internally among departments and subsidiaries and externally with different participants in the market, including regulators.

Efficient data communication depends on a mutual understanding of data transferred. Nevertheless, because of the complexity of the financial industry, there is no single common "dictionary" for each data attribute. Each organization could have its own definitions and method of processing different key terms, lowering the possibility of harmonizing these terms into common meanings even within an organization. Meanwhile, the data structure alone cannot provide information about each data element. When a system tries to communicate with another system, whether internally or externally, it must understand the intended meaning of each data element in order to map data. In such cases, a common language is needed to ensure the correctness of mapping to corresponding elements during interaction of systems (Bennett, 2013).

In large financial institutions, new mergers and acquisitions deepen the problem of communication among systems because of different database designs. Meanwhile, newly developed systems with extended functionalities may encounter difficulties when talking to legacy systems, which were developed a long time ago and are difficult to upgrade today.

Data completeness and adaptability require banks to not only to consider all the risks, but to also be flexible enough to support on-demand, ad-hoc reporting in different scenarios. This requires flexibility and an extensive data framework for data aggregation, especially when regulators have unclear reporting requirements. Nonetheless, due to a large number of manual processes and inflexible data models, banks currently have difficulties adapting and customizing their methods in the face of changing requirements.

3.5.3 Challenges in Risk reporting practices

In order to provide comprehensive reports to regulators, financial institutions need to extract internal data from multiple databases. Meanwhile, institutions need to externally deliver the reports to regulators and SDRs. Outstanding risk reporting practices are based on robust IT infrastructure and reliable data aggregation practices. Although relevant ratings looked the highest among all principles, according to the analysis of WGSS, firms tended to rate themselves higher than their actual level of compliance. Considering the logic of the questionnaire, it was unlikely that firms could do an excellent job in reporting without robust infrastructure and reliable data aggregation practices. During interactions among systems, there are some data challenges that remain to be solved.

The accuracy of risk reports is based on the accuracy of risk data aggregated. Organizations need to reconcile and process all this data within a consistent framework. However, when performing integration from multiple data sources, there is no single data model that could be used because of different designs of data models from different data sources. As there is no common "vocabulary" to define each term, financial institutions would use different words for the same things and the same words for different things (Bennett, 2013). This increases the difficulties of integrating and consolidating data on the regulators' side without a manual intervention. Again, manual processes can increase the risk of human mistakes. Data accuracy, timeliness and the frequency of reports generated would be harmed, hindering the development of a holistic view of the systemic risk in the market.

At the same time, in order to improve the comprehensiveness and usefulness of reports, organizations not only need to align all these risk data to semantic concepts, but they also need to keep meaning consistent or use a common dictionary across the groups. Inconsistent definitions of financial terms in reports hindered firms' ability to perform scenario-based analysis and make decisions at different levels. The risk monitoring capacity of regulators could also be impaired due to unclear reports.

3.6 System Dynamics View

Figure 6 depicts a brief system dynamics model that aims to explain interactions among major items. Although not all involving factors and relationships are shown in the model, several causal loops demonstrate how these important factors link with others.

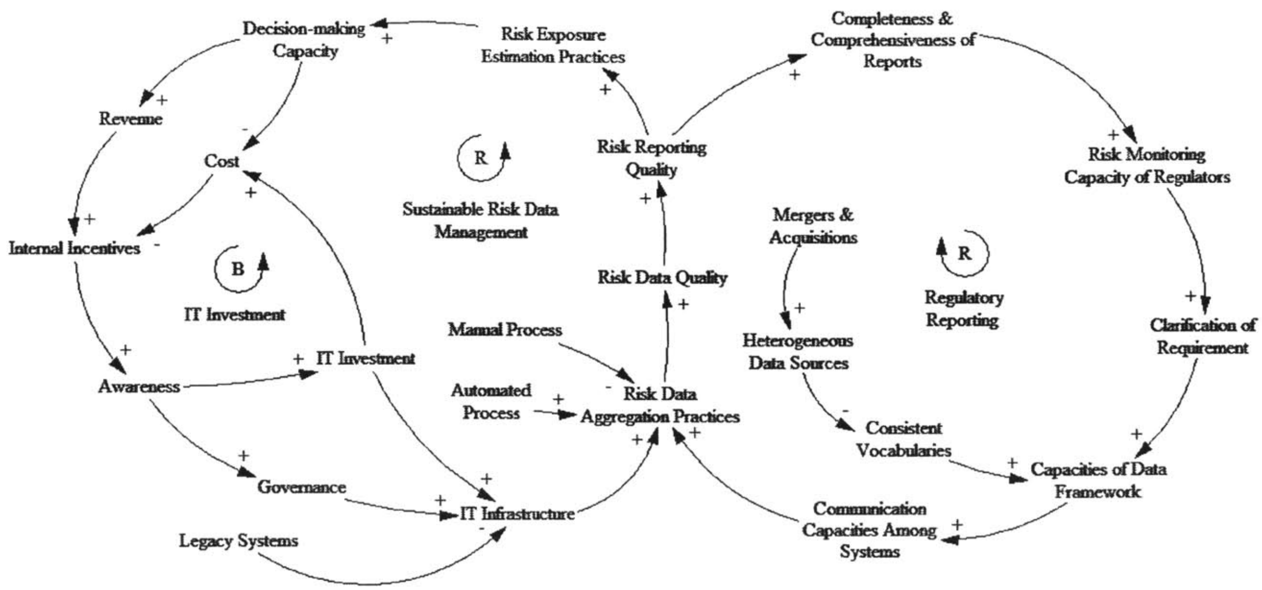


Figure 6: System Dynamics View

Reinforcing loops:

1. Sustainable Risk Data Management. Strong governance practice is fundamental to IT infrastructure and risk management practices. In return, strong risk management practices help financial institutions better estimate risk exposure and enhance their decision-making capacity, which can lead to an increase in revenue and a decrease in unnecessary costs. The incentives of gaining higher profit will further raise awareness of investing in IT and implementing better governance practices. A virtuous circle and a sustainable data management practices model can be established.
2. Regulatory Reporting. Except for building a sustainable internal model, better risk management practices can also lead to robust regulatory reporting. High quality risk reporting practices generate complete and comprehensive reports, which enhance the risk monitoring capacity of regulators. Hence they are able to provide clear and clarified requirements of data. Clear requirements are beneficial to financial institutions for building a robust data framework to capture all required data across different systems. Conversely, poor reports can impair the risk monitoring capacity of regulators, leading to unclear and unsound requirements. In this scenario, unstable data frameworks damage a firm's capacity to aggregate and report data accurately.

3.7 Possible Solution

Before proposing solutions, it is necessary to identify critical issues. Based on the previous analysis, there are several key points to be addressed.

First of all, senior management and boards of these financial institutions should have enough motivation and determination to further engage with, invest in and support the implementation of sound practices. As they are sensitive to profit, the solutions need to be inexpensive, easy to deploy, and clearly beneficial to the business's profitability and operation processes.

Moreover, on the data management side, a common, consistent, extensible data model should be established for minimizing unnecessary manual intervention during data extraction, aggregation, and communication. Keep in mind that the final goal is to monitor systemic risk in near real time. This requires linking all data that can be provided by all banks with its real meaning to prepare comprehensive and complete risk reports with both timeliness and accuracy.

Therefore, for generalization, the data model should be able to be independent from platforms, systems, and databases. For customization, the data model should be able to meet requirements from different business levels and perform scenario-based analysis without significant changes. Meanwhile, the data model should also be acceptable by the industry, covering all participants, to gain seamless communication and scale.

Although many consulting firms have proposed general solutions to meeting BCBS 239, there has been little consensus on detailed solutions designed to improve risk data management practices by utilizing innovative technologies. Taking the above considerations into account, ontology technologies can be a suitable solution to existing data issues. In the previous chapters, solutions about semantic databases and FIBO have been introduced broadly. Specifically, FIBO can be a reliable solution to overcoming the challenges of implementing BCBS 239.

FIBO is designed to provide a common and unambiguous shared meaning across data sources. It could be used to automate the processes of data extraction and aggregation, helping financial institutions to perform risk data analysis. For providing regulatory reports, regulators just need to identify information requested and map it to FIBO, which is flexible and extensible. In such cases, financial institutions could get the clear requirements from regulators, making the reporting process more efficient and effective.

As a common conceptual data model, FIBO can be utilized to align data to meaning according to a common "dictionary," which provides a detailed definition of each term and establishes links and relationships among each entity, allowing organizations to make connections. It helps both humans and machines understand the exact meaning of data across different databases with different designs and structures. Because it is independent from platforms, FIBO also fits different data sources through matching fields, providing convenience in maintenance and extension. Therefore, fewer manual interventions are necessary during data extraction and aggregation.

Compared to other data standards (including FpML, FIX), FIBO has several significant advantages. As a conceptual model, FIBO is easy to adopt and has no need to change the fundamental physical design of databases. It can also be compatible with existing data structures. Detailed definitions of each term could form a universal language, providing a tool for systems to talk with each other efficiently. Apart from generating regulatory reports, FIBO could also be used to customize reports for different usage levels, largely improving decision-making processes within organizations.

3.8 Implementation Plan for Meeting BCBS 239

This thesis proposes an implementation plan for financial firms to meet requirements of BCBS 239 by adopting FIBO. Both general strategic solutions and specific solutions to data management are provided. Figure 7 depicts each step of the implementation plan in general.

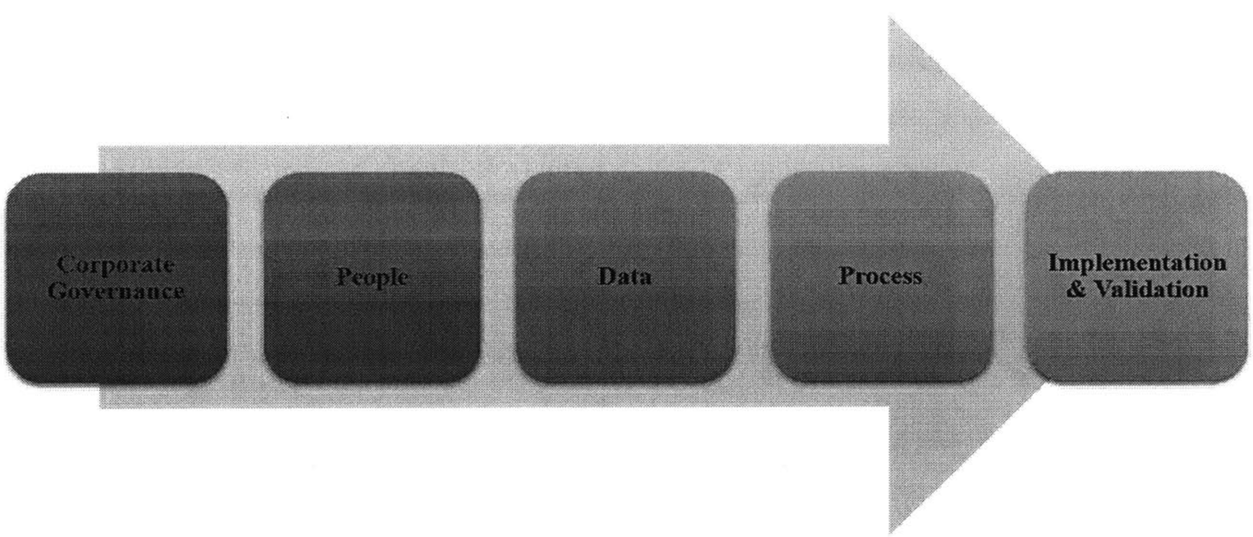


Figure 7: General Implementation Plan

3.8.1 General Implementation Plan

In response to Principles 1 to 11 of BCBS 239, this thesis proposes a general implementation plan which aims to solve issues revealed in each domain. It is assumed that firms have taken some kind of self-assessment to assess their current capacities and identify gaps.

1. **Corporate Governance.** The first step of the implementation plan is to get stakeholders engaged in raising awareness of risk data management and in publishing relevant policies. As stated above, strong governance and support from senior management in financial firms are fundamental to successfully meeting the requirements of BCBS 239. They need to establish solid risk control frameworks and identify their risk appetites. Furthermore, firms need to formulate data policies, including a measurement of data quality. Senior management also needs to review their IT strategy to ensure that sufficient resources are available to develop and maintain their IT infrastructure. To summarize, at this stage firms need to facilitate an adequate framework and implement policies for the following steps.
2. **People.** Grounded by strong corporate governance, the second step would be to clarify the roles and responsibilities of each party involved. As risk data is considered as a kind of strategic asset, risk data management should be one of firms' core businesses. Therefore, risk data should be owned by both the business side and by IT teams. Clear roles and responsibilities during the whole process should be clearly identified. Both sides should be highly engaged and work together closely.
3. **Data.** Once foundational factors are guaranteed, firms should take the third step, which is to understand their own data and requirements from the regulator side and to establish their data framework accordingly. The data framework should be applied across the organization in each level and subsidiary to enable seamless and consistent communication. As mentioned above, FIBO is a promising solution, providing data alignment with semantic meaning and flexibility in response to changing regulatory requirements.
4. **Process.** When data issues are solved, firms can take next step to develop processes of risk data aggregation, risk reporting, risk analysis, and other relevant activities in the predefined risk control frameworks. Firms need to find the most suitable data processing approaches to meet different business needs and changeable regulatory requirements. By

adopting semantic technologies, firms can take advantage of some semantic applications, which provide powerful data processing functionalities to firms.

5. **Implementation and Validation.** The final step is to transition from preparation to implementation. Firms need to propose a detailed schedule and roadmap for this project for monitoring and tracking. Meanwhile, validation is required to make sure that the implementation process is always on track. Although there is a deadline for meeting the requirements of BCBS 239, improving data management practices can be an ongoing project. Feedback from validation is important for reviewing and adjusting previous steps.

There are challenges for the implementation plan. Getting stakeholders engaged and raising awareness sometimes are not easy. In order to create sufficient incentives of processing the implementation plan, it is critical to clearly identify both benefit and cost. It is also expected to encounter some difficulties and resistance when adopting new process. Therefore, it should be an ongoing task, which needs feedback to correct and optimize itself continuously. Meanwhile, the success of implementation requires fully support of senior management.

3.8.2 Implementation Plan of FIBO

Apart from a strategic implementation plan, this thesis also proposes another implementation plan of FIBO in solving issues in Step 3 and Step 4 as mentioned above. It cannot be denied that meeting the requirements of FIBO is a holistic and complicated project that requires much more than implementing a new technology.

Adopting new technology will definitely not be sufficient to guarantee success in meeting regulatory requirements. However, technology can play a significant role in helping firms overcome critical challenges, which emphasizes the importance of implementing ontology technology and FIBO. Figure 8 depicts each step of the FIBO implementation plan.

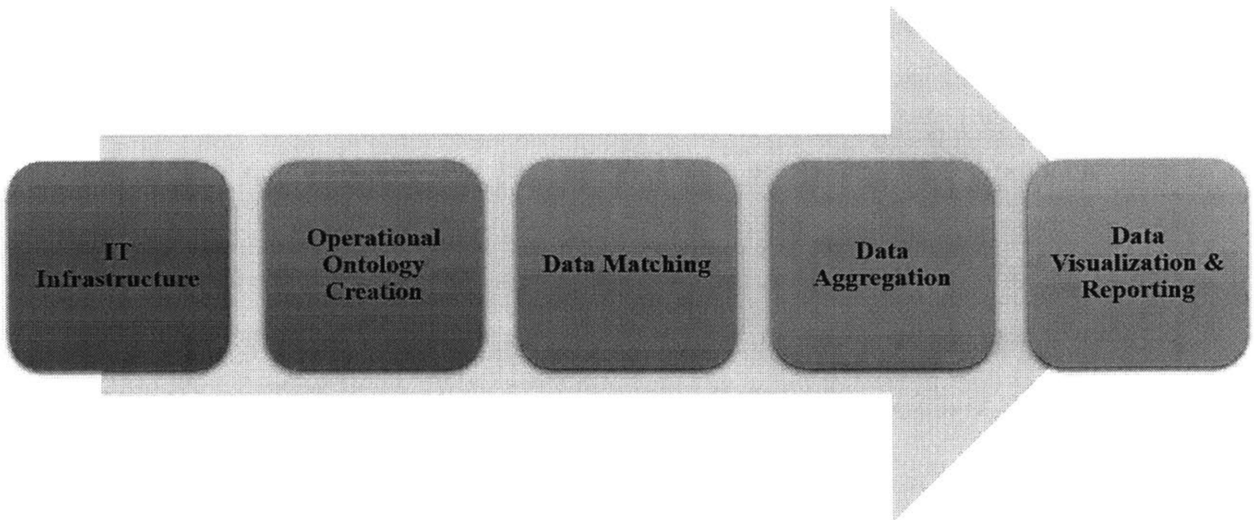


Figure 8: Implementation Plan of FIBO

1. **IT Infrastructure.** As stressed above, strong IT infrastructure is a precondition for downstream activities, including data aggregation and reporting. In such cases, senior management should ensure FIBO can be effectively implemented across different business lines and entities without technical constraint. Some fundamental infrastructure, such as messaging and network distribution, should be sufficient to support data sharing and communication. Both software and hardware satisfy the needs of semantic technologies and FIBO.
2. **Operational Ontology Creation.** The second step is creating operational ontology models. As FIBO is a conceptual ontology, it is not optimized for practical application. The business side needs to work with the IT team to decide how to transfer FIBO into operational ontology models, which could actually be applied in real application. They need to take several considerations into account, including regulatory requirements, business needs, and the technical constraints of their systems.
3. **Data Matching.** After FIBO operational ontology models are created, the next step is to perform matching between actual data with these models. Financial firms need to fully understand the exact semantic meanings of each data element before they begin matching. Businesses should ensure that the data matching is accurate and that it exactly reflects the actual meaning of these instruments, as alignment to actual meaning is precondition for generating comprehensive risk reports. All the matching rules need to be well documented in order to be referred in the future. In this step, a universal identifier system,

such as LEI, is important for identifying each legal entity in transactions. It is indispensable for building the whole network.

4. **Data Aggregation.** When data matching rules are well defined, firms can take the fourth step for data aggregation, including data extraction and quality checking. After extraction, all data from disparate data sources looks as if it is in a single database for operating. Meanwhile, after data extraction, organizations also need to monitor the data quality with relevant standards to ensure that aggregated data is fit-for-purpose and able to provide adequate data for risk reporting. Quality checking also provides confidence that the risk reports are all based on reliable and comprehensive data.
5. **Data Visualization & Reporting.** The final step is to customize data visualization and generate reports. Based on the needs of different business lines and purposes, firms can customize data visualization or dashboards and perform semantic operations. A report is also a kind of data visualization. Firms can generate reports for internal risk analysis and regulatory reporting. While generating reports, all data in ontology models can be easily reconciled and organized according to various requirements and scenarios. As long as requirements are clear, dashboards or report formats can be designed regardless of what the data structures look like.

There are also challenges of implementing FIBO. The benefit of improving IT infrastructure could be gained in long term, but the cost could be immediate. Therefore, firms need to balance long-term interest with short-term interest. Creating operational ontology models and matching data with these models can also be complicated, given a large amount of data sources and systems exist within each financial groups. The whole process could be time-consuming and require close collaboration between business and IT function. In addition, firms need to clearly understand requirements of different business level and regulatory reporting to customize data visualization and perform analysis.

4 Use Case of Insider Trading

The advantages of adopting semantic technology to solve data management problems have been discussed in previous chapters. FIBO could be a potential approach to help firms overcome the critical challenges of complying with BCBS 239, which involves strong data management practices, including data extraction, integration, and processing across disparate data sources. The only concern now is whether semantic technology, or FIBO, is feasible and operable in real life. However, only a few Proof of Concept (POC) experiments have been performed in this field.

In order to evaluate the feasibility of semantic and ontology technology in data aggregation, this thesis uses a use case of insider trading surveillance as a POC. This POC experiment is based on a demo provided by Cambridge Semantics (Cambridge Semantics, 2015).

4.1 Background Introduction

Risks of trading exist not only in flawed financial system mechanisms with unintended poor data management practices, but also in flawed surveillance practices with illegal human behavior.

According to the definition provided by the U.S. Securities and Exchange Commission (SEC), the term "insider trading" consists of both legal and illegal conduct. This use case is more interested in surveillance of illegal part, which means "buying or selling a security, in breach of a fiduciary duty or other relationship of trust and confidence, while in possession of material, nonpublic information about the security" (SEC, 2015b). A typical case is that an employee trades the securities of a specific company after acknowledging the company's significant and confidential information. To be consistent, the term "insider trading" will always take the definition of illegal conduct in the following sessions.

As regulatory requirements have increased, existing surveillance processes and technology are neither effective nor efficient enough to detect and monitor suspected trading. There are several critical challenges that are difficult to solve with traditional approaches.

1. Data is derived from various systems and databases that have different designs. As insider trading surveillance involves many different internal and external data, each data source could have very different structures and definitions of each data element, increasing the difficulty of extracting and analyzing data in a single platform.

2. Developing and maintaining surveillance platforms is expensive. Given that M&A and the development of new products are common in financial institutions, it is costly and inflexible to expand the surveillance functionalities to new systems with a pre-determined model. Meanwhile, fixed data models have to be modified every time in response to new changes in regulations requirements.
3. Traditional databases are not good at handling unstructured data formats. Apart from structured data, such as transaction records in relational databases or spreadsheets, other information from unstructured data sources, such as E-mails and articles, are also important data that are worthy of attention. Sensitive information can be revealed in the chat history of messaging applications, which are costly when monitored by humans only.

4.2 Experiment Design

This thesis built an experiment to explore the application of semantic technology for solving existing challenges.

4.2.1 Objective

The success of the experiment depends on several aspects:

1. Each data element is unambiguously defined in the ontology models, which are readable to both humans and machines.
2. Data is extracted automatically from diverse data sources without manual intervention.
3. The results of data aggregation and mapping are accurate.
4. Visualization of data output can be customized with different designs.

4.2.2 Data Sources

This experiment uses several Excel spreadsheets provided by Cambridge Semantics (Cambridge Semantics, 2015) and unstructured data fetched online for a simple illustration. In reality, data sources can be from structured formats, including different databases or Excel sheets, and from unstructured formats, including E-mails, online articles or PDFs documents.

4.2.3 Tools

The experiment is based on a series of components from the Anzo application, which was developed by Cambridge Semantics, including:

1. Anzo Connect: Live SQL connectivity with different types of databases.
2. Anzo Unstructured: Unstructured text monitoring via pipeline processing.
3. Anzo for Microsoft Excel: Live Excel connectivity, interacting with Excel spreadsheets.
4. Anzo on the Web: Customized visualization of reports and analysis.

4.2.4 Implementation Steps

There are several steps for implementing ontology models in the experiments:

1. Develop semantic definitions for all relevant objects and relationships in the problem domain. In practical application, the definitions and entities involved should be decided by the business side, in accordance with regulatory requirements and internal policies. In this experiment, definitions are only roughly established for simplification.
2. Develop ontology models referring to corresponding semantic definitions. The models should cover entities and relationships between them. These models are used in data matching and extraction not only from structured databases, but also from unstructured ones. In this chapter, the POC experiment only focuses on evaluating feasibility of semantic technology. Therefore, for simplification, ontology models were built as purpose-based, reflecting the physical data structure in some way. In practical application, ontology models should be built independently from any specific structure of databases, only referring to semantic definitions. An exploration of application of independently built ontology models, such as FIBO, is discussed in the next chapter.
3. Match data elements from disparate databases to ontology models that fully understanding their meanings. Because it is likely that data sources use the same language (or field name) for different meanings and use different languages for the same meaning, it is important to know the data and document semantic meanings of each data element before actual matching.
4. Extract and aggregate data across multiple databases via semantic application. After data matching for each relevant database, the semantic application (Anzo in this experiment) aggregates all the data into corresponding ontology models in the online server. In this experiment, Anzo for Excel can perform data extraction from Excel files, while Anzo Unstructured can perform data extraction from online articles and E-mails via pipeline processing.

5. Customize data visualization in the dashboard. In reality, different levels of business could have different requirements for focuses and security. Business can create their own dashboards with different authority levels for different usages. The results shown in dashboards could be exported as Excel files. In this experiment, Anzo on the Web can provide dashboards for customized visualization.

In practical application, it is also important to perform relevant semantic operation for trading surveillance, including reasoning, querying, and classifying. Because the final purpose is to detect and monitor suspect transactions, it is necessary to perform relative operations and prepare the reports, taking advantage of semantic technology. Meanwhile, regulators or management can perform scenario-based analysis to investigate the transactions data in different perspective. In this experiment, the semantic operation is temporarily not explored because the experiment focused on data matching and aggregation.

4.3 Semantic Definitions

To create this use case, it was important to understand the context at the beginning and elaborate on the exact semantic meanings of each concept. It required deliberate consideration of the exact meaning of several critical concepts and entities involved in this case. Because the use case is about employees who engage in security trading based on confidential information, the accurate semantic meanings and concepts such as "security," "duty," "employee," and "confidential information" need to be well defined.

In reality, semantic definitions should be carefully discussed and reviewed to ensure the unambiguous and clear meaning of each entity and relationship. Because this experiment's sole purpose is to explore the application of semantic technology, definitions of relevant terms are not strictly developed.

4.4 Ontology Models

Two kinds of ontology models are built for the experiment, including business ontology (or conceptual ontology) and operational ontology.

4.4.1 Conceptual Ontology

A conceptual ontology is used to illustrate the business concept of entities and relationships, providing information to humans rather than machines. Businesses can use a conceptual ontology model to clarify their thoughts, provide intuitive understanding of the issue, and guide the creation of operational ontology models. The conceptual ontology model is built on the basis of semantic meanings of each entity involved.

In this experiment, a brief conceptual model diagram is built for the exploration. The main goal of building this conceptual ontology is to express and describe each "thing" in surveillance of insider trading. Several entities and relationships are identified in the model in Figure 9.

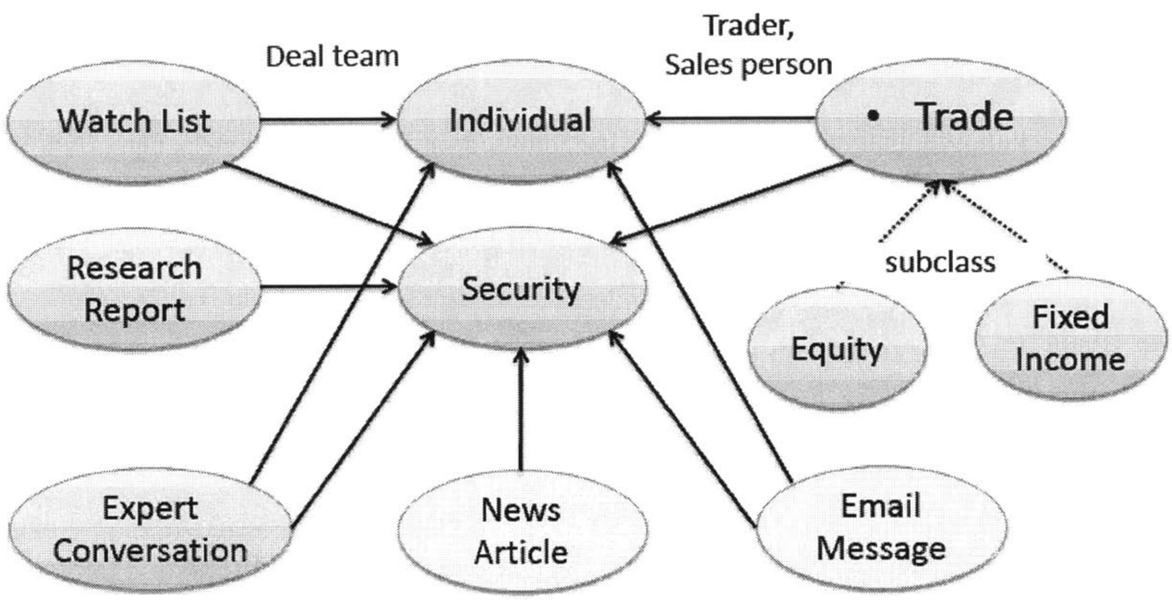


Figure 9: POC Ontology Diagram (Cambridge Semantics, 2015)

The diagram presents an accessible view of this use case. For instance, Equity and Fixed Income are subclasses of Trade, while Trade is initiated by Individual. Security has different relationships with multiples entities, including Trade, Watch List, Research Report, Expert Conversation, News Article, and Email Message. The detail is not shown on the diagram for simplicity.

In addition to diagram form, conceptual ontology also can be presented in table form. Figure 10 shows examples of two different taxonomies, referring to relative definitions in FIBO (Semantics Repository, 2015). They provide context for understanding this use case.

In reality, the conceptual model can be much more complex and formal than this brief experiment. All relevant properties, or at least all critical properties, should be included.

Taxonomy	Term	Semantic Definition
Security	CUSIP 6	Identifier of the issuer.
	Issuer	The issuer of the Security; that is the party which is the holder's counterparty to the rights and obligations defined in the security.
	Issuer location	Jurisdiction (country, county, state, province, city) in which the security is legally recorded for regulatory and/or tax purposes.
	Symbol	Synonym or short name of issuer.
Trade	Buy/Sell	The role of transaction performed.
	CUSIP	The identifier of the security which is being traded.
	Date	Date on the trade occurred.
	Description	A user defined short description often used to uniquely identify a financial instrument. This usually follows a recognized structure and can be used to describe the principal unique features of the instrument.
	Price	The price of the instrument.
	Sales person	The person who promote the sales of the security.
	Security	The security which is being traded.
	Trade Volume	The share or the size of security to be traded.
	Trader	The person who directly conduct the trade of security.

Figure 10: Semantic Meaning of Part of POC Ontology

4.4.2 Operational Ontology

Compared to conceptual ontology, operational ontology focuses more on implementation, and it is optimized for semantic application. The objective of operational ontology is to build a common structure that can be implemented across disparate data sources. On one hand, operational ontology is a reflection of corresponding conceptual ontology so that it is still a common dictionary for elaborating each concept in the problem domain. On the other hand, it is necessary to consider physical and technical constraints to create an operational ontology. In this use case, given that relevant physical data exists in conventional databases already, the corresponding operational ontology is built taking both conceptual ontology and conventional data structure into account.

In this experiment, the operational ontology is created via Anzo Ontology Editor and stored in OWL files. A graph view of the operational ontology model is shown in Figure 11.

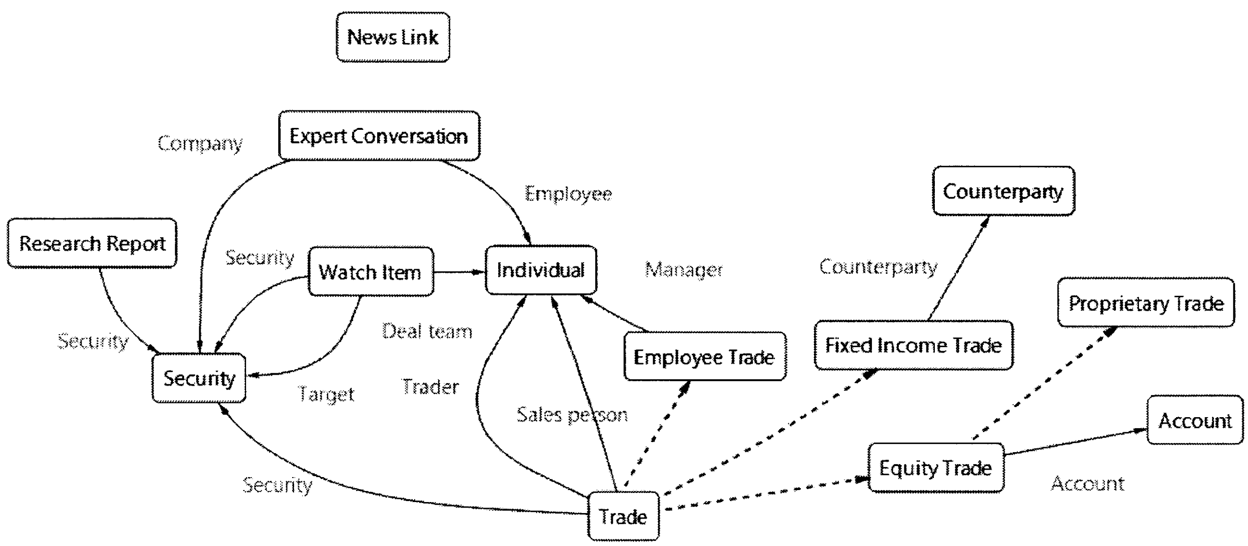


Figure 11: POC Operational Ontology Graph

In the graph, boxes indicate each entity. Green dotted lines the indicate relationship as "has sub class," while solid lines represent types of relationships among entities. For example, Trade has subclasses Fixed Income Trade and Equity Trade. In this example, Employee Trade is also regarded as subclass of Trade in order to inherit its properties. Meanwhile, Trade has Traders and Sales Person, which belong to individuals. However, detailed properties are not shown in the graph either.

In reality, operational ontology should be independent from platform and specific logical data to ensure this ontology can be implemented across different data sources. The representation of "things" should also be more formal, using the "Subject-Verb-Object" structure.

4.5 Data Matching and Extraction

When the ontology model is complete, it is time to implement the model across different data sources and aggregate the data. In this experiment, data matching and extraction can be done in Anzo for Excel at the same time.

The table in Figure 12 illustrates the exact matching rules between actual data in different Excel files and operational ontology models. Fields or column names of different files are matched with their corresponding properties of different classes. Only a part of the rules are listed for discussion.

Files	**Field Names**	**POC Class**	**POC Property**
Employee Blotter	Watch List Item	Security	Issuer
	Watch List Item	Trade	Description
	CUSIP	Trade	CUSIP
	CUSIP	Security	CUSIP6
	Ticker	Security	Symbol
	Buy / Sell	Trade	Buy/Sell
	Trade Position	Trade	Trade Volume
Equity Trade Blotter	CUSIP	Trade	CUSIP
	CUSIP	Security	CUSIP6
	Ticker	Security	Symbol
	Description	Trade	Description
	Description	Security	Issuer
	Price	Trade	Price
	Buy / Sell	Trade	Buy/Sell
	Trade Size	Trade	Trade Volume
Fixed Income Trade Blotter	Issuer	Security	Issuer
	CUSIP	Trade	CUSIP
	CUSIP	Security	CUSIP6
	Buy / Sell	Trade	Buy/Sell
	Trace Price	Trade	Price
	Description	Trade	Description
Proprietary Trade Blotter	CUSIP	Trade	CUSIP

	CUSIP	Security	CUSIP6
	Ticker	Security	Symbol
	Description	Trade	Description
	Description	Security	Issuer
	Price	Trade	Price
	Buy / Sell	Trade	Buy/Sell
	Trade Size	Trade	Trade Volume
	Volume	Equity Trade	Volume
Research Reports	Company	Security	Issuer
	Company	Research Report	Company
	CUSIP	Research Report	CUSIP
	CUSIP	Security	CUSIP6
Watch List	Client / Issuer	Security	Issuer
	CUSIP	Security	CUSIP6
	Ticker	Security	Symbol
	Issuer Location	Security	Issuer Location
	Target / Buyer	Security	Issuer
	Target CUSIP	Security	CUSIP6
	Target Ticker	Security	Symbol
	Target Location	Security	Issuer Location

Figure 12: Matching Rules for POC

There are several points that are worthy of attention.

1. Anzo for Excel provides the function of extracting a field with a certain formula. For example, there is no unique LEI used in this use case. Instead, a part of CUSIP is used as a simplified identifier. CUSIP stands for Committee on Uniform Securities Identification Procedures, which is used to identify securities. A typical CUSIP number consists of nine characters, uniquely defining a type of securities of an issuer (SEC, 2015a). Normally the first 6 digits can be used as an identifier for a company or issuer, although it is not always unique. In this experiment, the first 6 digit (named CUSIP6) is extracted using function "LEFT ()".
2. It is possible, or even common, to have different names in different data sources for the same concept. For example, in Employee Blotter file, the field "Trade Position" actually

means "trade volume," while in the Equity Trade Blotter file, the same concept is called "Trade Size."

In a separate system, it is not unacceptable to use arbitrary field names in conventional databases as long as users understand what each field means. However, when performing data aggregation across these data sources, it is important to have a common "dictionary" for describing and defining each meaning. In real life, businesses should work closely with IT to identity the semantic meaning of each element and perform accurate matching.

3. It is also possible to have the same names in different files to describe different concepts. For example, in the Equity Trade Blotter file, the "description" field stores information about the security issuer and trade description, while the "description" field in the Fixed Income Trade Blotter file only represents trade description. It is important to understand what exact information is stored in the field and match it with accurate elements in ontology model.

Figure 13 is a screen shot of the data matching process. It shows how to use Anzo for Excel to match each data field in Excel files with corresponding elements in ontology model.

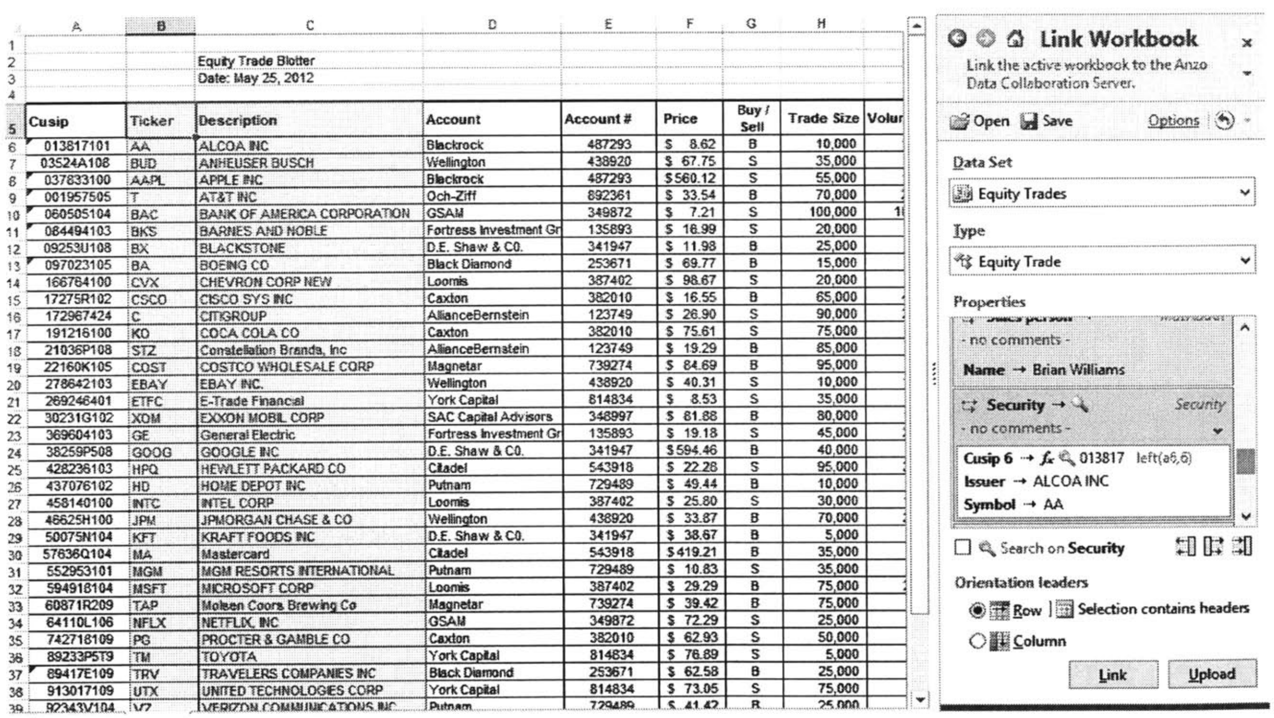


Figure 13: Screen Shot of Matching POC Ontology

To summarize, in a real application the business should fully understand the meaning of all data elements and thoroughly document them. It is possible that the name of a property is quite different from, or even opposite to, the field name of the corresponding data element, while that property name is exactly the same as an irrelevant field name. Therefore it is necessary to keep the matching rules documented to avoid confusion.

After matching, Anzo for Excel will automatically upload all the data into the models in the online server. If automatic linkages have been set in advance, any future modification in the spreadsheet files will be synchronized with server's data.

4.6 Customized Data Visualization

Once data is available in the server, different designs of visualization, or dashboards, can be built according to each specific need. Data displayed in the dashboards are the unambiguous results of data aggregation from disparate data sources (in the case of this experiment, different Excel files).

In this experiment, some simple dashboards are built for the exploration of customized visualization function. Filters, or "Lenses" as they are called in Anzo for Web, can be used in scenario-based analysis, based on different criteria. Figure 14 illustrates an example of customized dashboard, which contains filters on the left side and queried results on the right side.

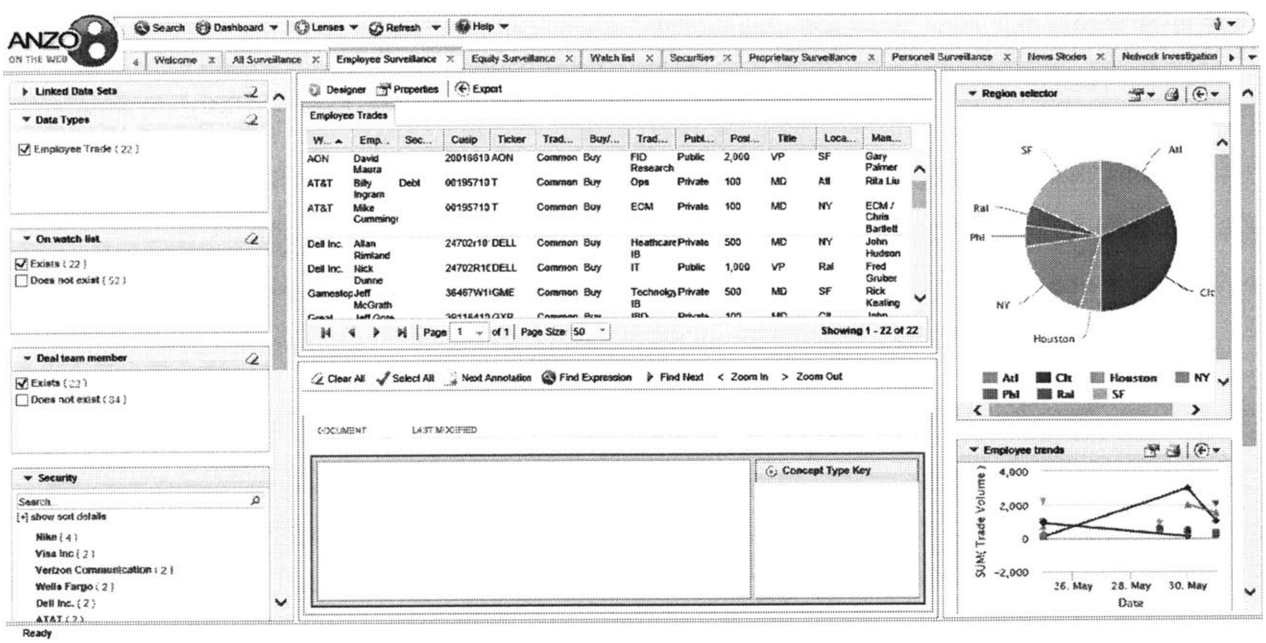


Figure 14: Customized Dashboard

To perform insider trading investigation, user is able to filter the data in the left column and get results showing in tables and charts. Figure 15 shows an example of how user can review data and evidence, including unstructured text, of suspected trades. In this example, a news is captured by pipeline, revealing that there is a significant change of senior management in Molsen Coors Brewing Company. As this company is on the watch list, all relevant trades are reported and traders' information are shown in the graph for further investigation.

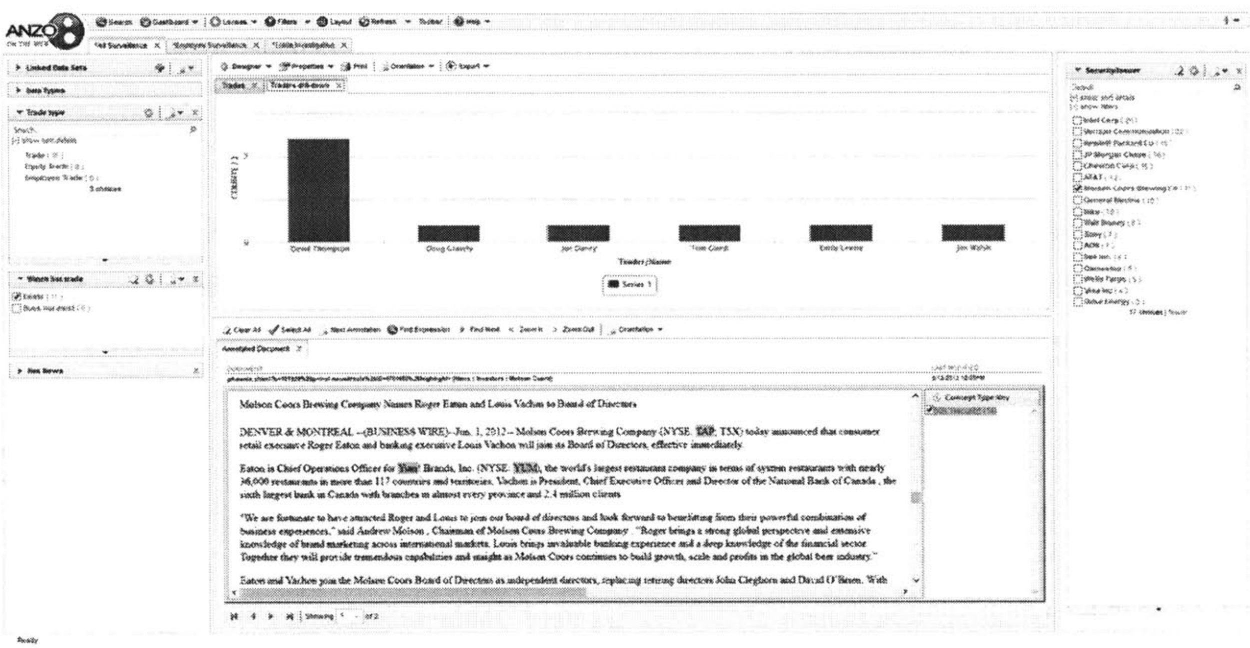


Figure 15: Insider Trading Investigation (Cambridge Semantics, 2015)

In reality, dashboards are created more deliberately. Businesses can set different security levels and provide access to different user groups to ensure privacy as well. Based on customized dashboards and filters, the same data can be used in different levels for different purposes without modifying the actual data model, reducing development and maintenance costs. Once regulatory requirements change, financial firms need to simply change the visualization of data to prepare updated reports. The results shown in dashboards are able to be exported via Excel file format.

5 Exploration of FIBO Application in the Use Case

The previous POC experiment has shown the capability of semantic technology for data extraction across different data sources. Semantic technology enables data aggregation using a single common model that can be used with disparate databases designed with different structures and vocabularies. Compared to directly building standardized data warehouse using ETL tools, semantic technology provides the context, avoiding confusion among fields in each source by providing a readable format to both humans and machines. The only manual intervention during data extraction is the matching part, which is one-off process. Once the rules are set, any update in these files will be synchronized in the server accordingly.

Considering the main purpose of the POC experiment is to evaluate the application of semantic and ontology technology for solving data aggregation issues, the ontology model was created as a purpose-based ontology, which means that the ontology model is implemented through a bottom-up approach. In comparison, FIBO is a general conceptual model that is created independently from any specific data structures and implemented through a top-down approach. To explore the implementation of FIBO, this chapter uses part of FIBO concepts to replace corresponding parts in the previous ontology model and evaluate its feasibility.

5.1 Legal Entities

According to the definitions in the OMG, the concept and relationships of legal entities are defined in FIBO for Business Entities, which is a formal and factual representation of the structure of relevant business entities. The actual definition of term "Legal Entity" is based on ISO Legal Entity Identifier effort (Object Management Group, 2014a).

5.1.1 Component of Legal Entities

Legal entities ontologies define the basis of legal personhood and business entities. They describe the concepts and relationships between legal entities, along with other information about organization, membership, classification, address, etc. (Object Management Group, 2014a).

The module of legal entities contains several separate ontology models, including:

1. **Legal Persons.** A legal person is either a natural person as an adult or an organization that is constituted by some legal instruments and has the ability to accrue liability on its

own. The semantic meaning of "Legal Person" is "any entity which can incur legal obligation and can be sued at law" (Object Management Group, 2014a).

2. **Formal Business Organizations.** This ontology model covers concepts and properties related to formal business organizations, including membership, classification, and address information. A formal business organization can be either incorporated or non-incorporated. The semantic meaning of a "Formal Business Organization" is an organization "with some formal contractual agreement among its principles" (Object Management Group, 2014a).
3. **Corporate Bodies.** This ontology covers corporate bodies, which are incorporated by equity, by guarantee, and by agreement. A corporate body is a formal organization, artificially created with legal personhood.
4. **LEI Related Entities.** In this ontology, contractually capable business entities are well defined. The semantic meaning of "Contractually Capable Entity" is "a unique entity that is legally or financially responsible for the performance of financial transactions, or has the legal right in its jurisdiction to enter independently into legal contracts, regardless of whether it is incorporated or constituted in some other way." Legal entity identifier is also defined in this ontology, meaning "a code which uniquely identifies a formal business organization and is allocated to it as a result of it being recognized as being a contractually capable entity" (Object Management Group, 2014a).

5.1.2 Replacement of Legal Entities

In the previous POC experiment, data fields of "employee," "manager," "trader," "sales person," and "deal team" in different files can be regarded as legal persons according to the definition. In the previous operational ontology model, they are matched with the class "Individual" and property "Name." When replacing these previous matching rules with FIBO concepts, both FIBO Legal Persons Ontology and its related ontology called People Ontology, which is designed in FIBO Foundations, will be utilized.

Figure 15 is a view of the entities and relationships of these two FIBO ontology models from the Anzo ontology editor. Only the part of Legal Persons Ontology that is related to natural personhood is shown in the graph.

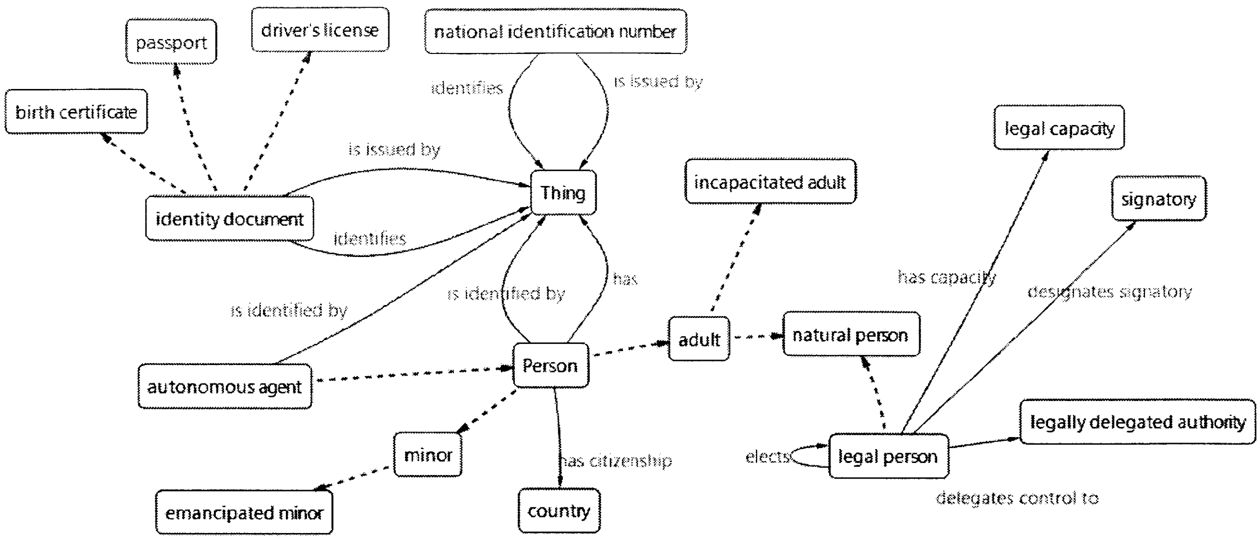


Figure 16: FIBO Legal Persons Ontology and People Ontology

In the diagram, it shows that "natural person" is a subclass of both "legal person" and "adult" classes. Meanwhile, "adult" is also a subclass of "Person." It means that "natural person" can inherit properties of its parent classes, or even parent's parent, namely "Person."

Furthermore, Figure 16 is a table format presenting new mapping rules using FIBO concepts. Only those elements which are related to natural person are listed in the table.

Files	**Field Names**	**POC Class**	**POC Property**	**FIBO ontology**	**FIBO Class**	**FIBO property**
Employee Blotter	Employee	Individual	Name	Legal Persons	Natural Person	has person name
	Group / Manager	Individual	Name	Legal Persons	Natural Person	has person name
Equity Trade Blotter	Trader	Individual	Name	Legal Persons	Natural Person	has person name
	Sales Person	Individual	Name	Legal Persons	Natural Person	has person name
Fixed Income Trade Blotter	Trader	Individual	Name	Legal Persons	Natural Person	has person name
	Sales Person	Individual	Name	Legal Persons	Natural Person	has person name
Watch List	Deal Team	Individual	Name	Legal Persons	Natural Person	has person name

Figure 17: Matching Rules for Natural Person

In addition, "issuer," "counter party," "company," and other related concepts in those files represent legal persons, formal business organizations, and contractually capable business entities at the same time. It is a typical example of multiple inheritances, which is an important feature in semantic technology.

To simplify the exploration of FIBO, the experiment here only treats these entities as legal persons, or specifically as judicial persons. Hence both entities of human and organization can use the same Legal Persons Ontology model. When replacing these previous matching rules with FIBO concepts, both FIBO Legal Persons Ontology and another related ontology called Formal Organizations Ontology, which is a part of FIBO Foundations, will be utilized. Moreover, in order to identify entities, LEI Related Entities Ontology is also involved.

Figure 17 is a view of the entities and relationships of these three FIBO ontology models from the Anzo ontology editor. Only the part of Legal Persons Ontology that is related to judicial persons is shown in the graph. Also, in LEI Related Entities Ontology, only the Legal Entity Identifier class is shown.

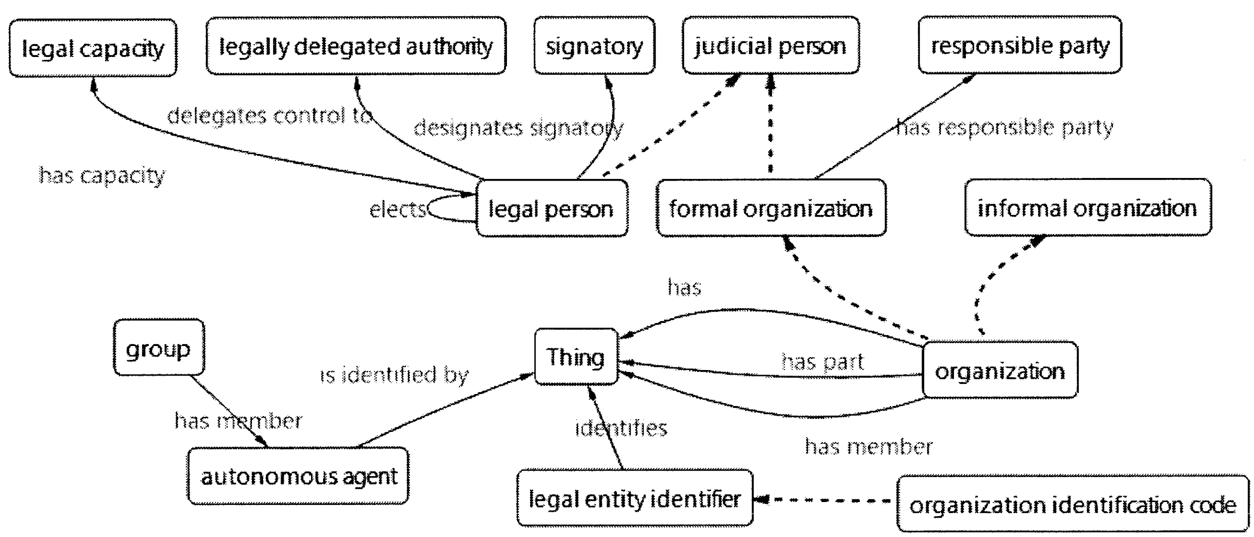


Figure 18: Ontology of Legal Persons, Formal Organizations, and LEI Related Entities

The diagram shows that "judicial person" is a subclass of both "legal person" and "formal organization" classes. Meanwhile, "formal organization" is also a subclass of "organization." It means that "judicial person" can inherit properties of its parent classes, or even its parent's parent, namely "organization."

Figure 18 is another table format presenting new mapping rules using FIBO concepts. Only those elements which are related to judicial persons are listed in the table.

Files	Fields Names	POC Class	POC Property	FIBO Ontology	FIBO Class	FIBO Property
Employee Blotter	Watch List Item	Security	Issuer	Legal Persons	Judicial Person	has name
	CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
Equity Trade Blotter	CUSIP	security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
	Description	Security	Issuer	Legal Persons	Judicial Person	has name
	Account	Account	Account name	Legal Persons	Judicial Person	has name
	Account #	Account	Account number	LEI Related Entities	Legal Entity Identifier	has unique identifier
Fixed Income Trade Blotter	Issuer	Security	Issuer	Legal Persons	Judicial Person	has name
	CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
	Counterparty	Counter party	Name	Legal Persons	Judicial Person	has name
	Counterparty #	Counter party	Counterparty	LEI Related Entities	Legal Entity Identifier	has unique identifier
Proprietary Trade Blotter	CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
	Description	Security	Issuer	Legal Persons	Judicial Person	has name
Research Reports	Company	Security	Issuer	Legal Persons	Judicial Person	has name
		Research Report	Company	Legal Persons	Judicial Person	has name
	CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
Watch List	Client / Issuer	Security	Issuer	Legal	Judicial	has name

				Persons	Person	
	CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier
	Target / Buyer	Security	Issuer	Legal Persons	Judicial Person	has name
	Target CUSIP	Security	CUSIP6	LEI Related Entities	Legal Entity Identifier	has unique identifier

Figure 19: Matching Rules for Judicial Person and LEI

In the table, "CUSIP 6" means the first 6 digits of CUSIP, acting as a temporary identifier for the specific company or issuer. However, "CUSIP 6" is not the single identifier in this experiment. Account number and counterparty number are also used as another kind of identifier in this case.

In summary, this illustrates that FIBO concepts have sufficient capacity to express basic financial terms and can replace corresponding items in a purpose-based ontology. After replacement, confusion about field names is eliminated.

5.1.3 FIBO Operational Ontology Model

The previous section sets up some new rules for matching data elements with FIBO concepts. However, they involved four ontology models, which consist of tens of classes and relationships. Not all the classes, relationships, or properties are necessary and useful in this experiment.

Meanwhile, there are two fields featuring people's names in the Employee Blotter file. The person in the "Group/Manager" field is the manager of the person in the "Employee" field. To distinguish these two elements, another relationship called "is managed by" and the inverse version, "manages," should also be included in the ontology model. Therefore, it is important to create an operational ontology model based on subsets of FIBO.

In such cases, an operational ontology model is built based on a subset of FIBO concepts, removing unnecessary entities and introducing new properties. Meanwhile, inheritance networks are also simplified. The main idea is to create an operational ontology model that conforms to FIBO specifications and that is optimized for application without many redundancies.

Figure 19 is a diagram of the operational ontology created for this experiment.

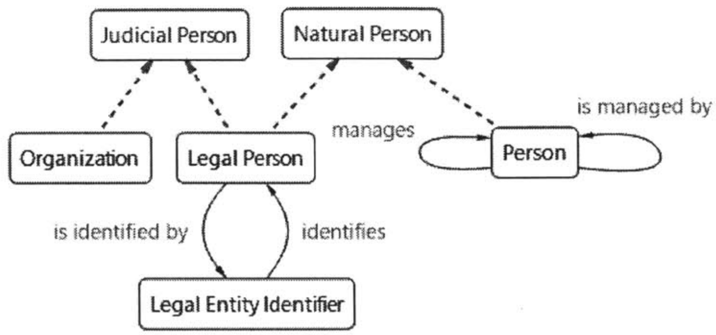


Figure 20: Operational Ontology for Legal Persons

5.1.4 Implementing Operational Ontology

After creating an operational ontology, as the rules are already set, the next step is matching existing data sources with this new ontology model. Because the purpose of this part of the experiment is to explore FIBO implementation rather than to solve the case, the natural person and judicial person are still separate. There are no relationships to connect these two entities yet.

For instance, Figure 20 shows the matching process for judicial persons, namely companies. Formula "LEFT ()" is also used to extract the first six digits of CUSIP.

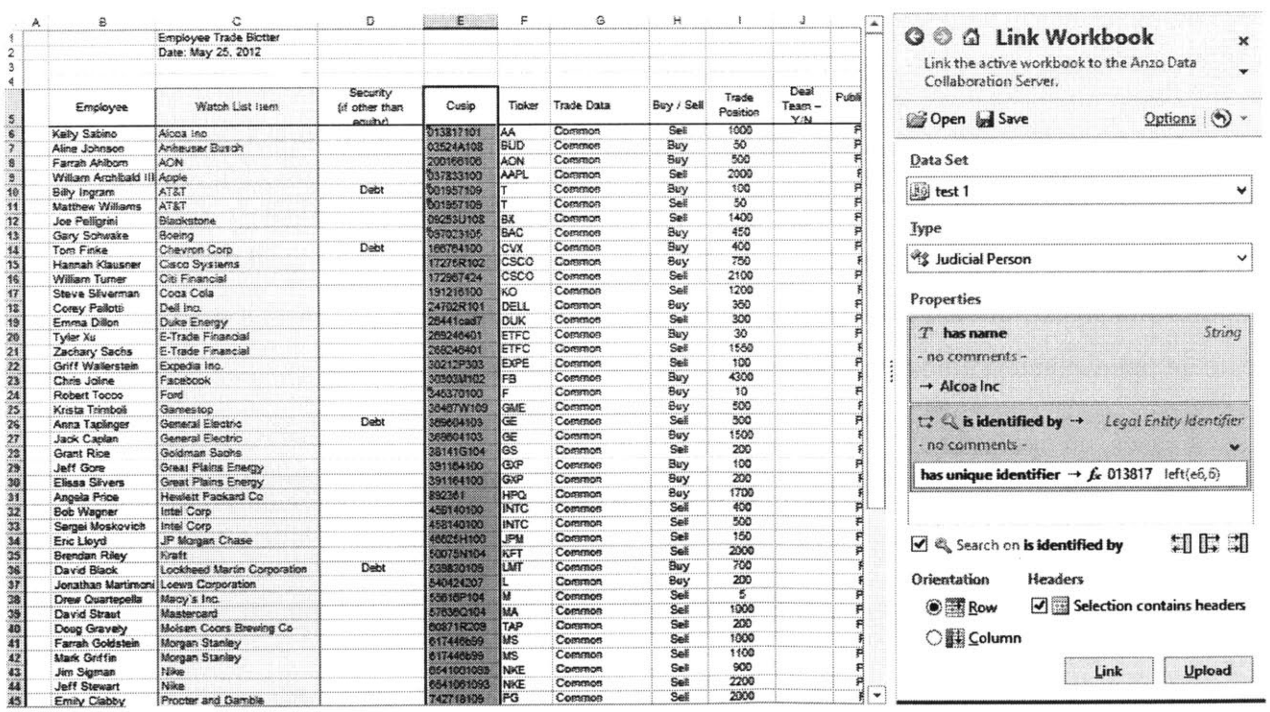


Figure 21: Screen Shot of Matching Legal Persons

The outcome of the data aggregation can be viewed in the dashboards created in Anzo for Web as in Figure 21. During extraction, the LEI, or first six digits of CUSIP, is set as a key to mandate that only one LEI can exist in the aggregated data.

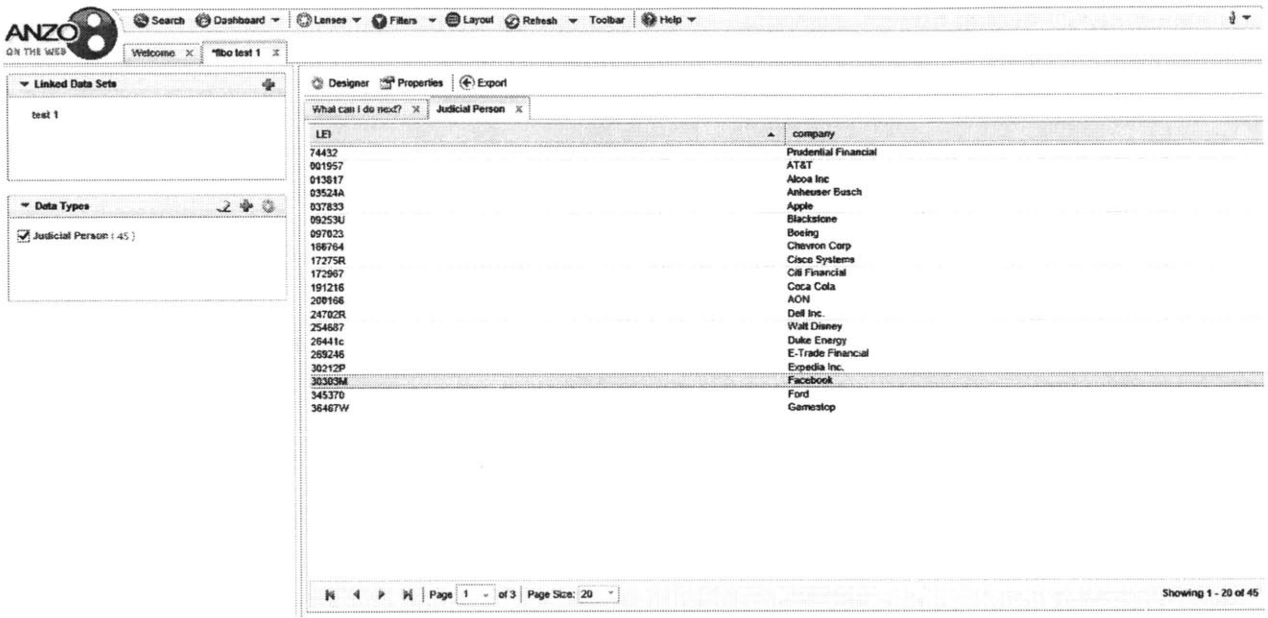


Figure 22: Data Aggregation Visualization for Legal Persons

In real life, because there is no universal LEI, an error can occur if same string of numbers represents different entities, or if different strings of numbers represent the same entities.

5.2 Involvement of Other FIBO Concepts

The previous experiment shows the feasibility of implementing FIBO concepts on a small scale. Identifying legal entities is a basic function of the whole FIBO semantic application. Until now, these entities were separated without relationships among them. To further explore the application of FIBO, some other FIBO concepts need to be involved.

5.2.1 Securities

In this section, concept of security in FIBO is involved in building connections among these entities. Compared to FIBO Foundations and FIBO Business Entities, Securities Reference Terms are not finalized and remained to be refactored (Semantics Repository, 2015). However, some of their structures and elements are still useful for constructing an operational ontology in this case.

Figure 22 is diagram of the operational ontology created for this experiment.

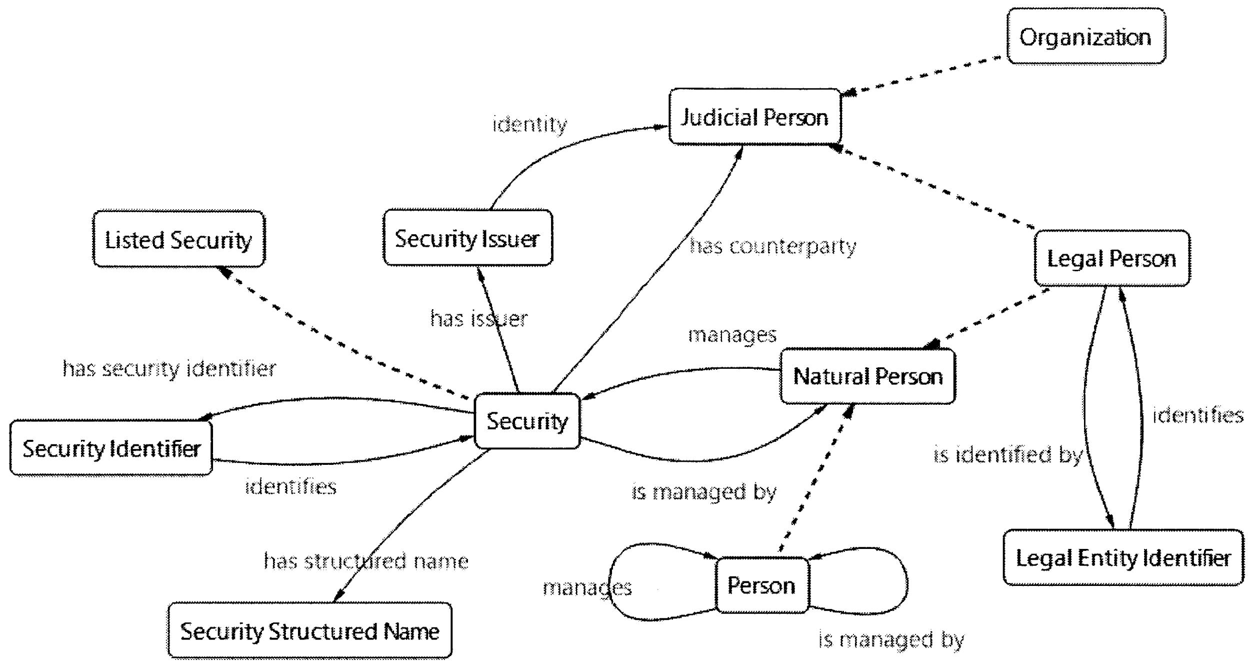


Figure 23: Consolidated Operational Ontology Model

The graph shows that a company or organization can act as security issuer or counterparty of a security. This security is identified by a unique security identifier, specifically the full CUSIP in this case. A natural person can take a role, such as trader, to manage a certain security.

Figure 23 is the table format presenting the matching rules of the operational ontology model. Compared to the diagram, which focuses on elaborating the relationships, the table format focuses on representing what “thing” each data element is.

Considering that this ontology model was created to show how FIBO connects entities, some properties, including prices, dates, and other simple facts, are not involved in the model. Meanwhile, inheritances are simplified and shortened for this specific use case. In real life, even the FIBO operational ontology should be deliberately created and should conform to the FIBO conceptual model.

Files	Field Names	POC Class	POC Property	FIBO Class	FIBO property
Employee Blotter	Employee	Individual	Name	Natural Person	has person name
	Watch List Item	Security	Issuer	Judicial Person	has name
	CUSIP	Trade	CUSIP	Security Identifier	has unique identifier
	CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Ticker	Security	Symbol	Security Structured Name	has short name
	Group / Manager	Individual	Name	Natural Person	has person name
Equity Trade Blotter	CUSIP	Trade	CUSIP	Security Identifier	has unique identifier
	CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Ticker	Security	Symbol	Security Structured Name	has short name
	Description	Security	Issuer	Judicial Person	has name
	Trader	Individual	Name	Natural Person	has person name
	Sales Person	Individual	Name	Natural Person	has person name
Fixed Income Trade Blotter	Issuer	Security	Issuer	Judicial Person	has name
	CUSIP	Trade	CUSIP	Security Identifier	has unique identifier
	CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Counterparty	Counter Party	Name	Judicial Person	has name
	Counterparty #	Counter Party	Counterparty	Legal Entity Identifier	has unique identifier
	Trader	Individual	Name	Natural Person	has person name
	Sales Person	Individual	Name	Natural Person	has person name
Proprietary Trade Blotter	CUSIP	Trade	CUSIP	Security Identifier	has unique identifier
	CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Ticker	Security	Symbol	Security Structured Name	has short name
	Description	Security	Issuer	Judicial Person	has name
Research Reports	Company	Security	Issuer	Judicial Person	has name
	CUSIP	Research Report	CUSIP	Security Identifier	has unique identifier
	CUSIP	Security	CUSIP6	Legal Entity	has unique

				Identifier	identifier
Watch List	Client / Issuer	Security	Issuer	Judicial Person	has name
	CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Ticker	Security	Symbol	Security Structured Name	has short name
	Deal Team	Individual	Name	Natural Person	has person name
	Target / Buyer	Security	Issuer	Judicial Person	has name
	Target CUSIP	Security	CUSIP6	Legal Entity Identifier	has unique identifier
	Target Ticker	Security	Symbol	Security Structured Name	has short name

Figure 24: Consolidated Matching Rules

5.3 Conclusion

In conclusion, the experiment for FIBO exploration in this chapter, accompanied with the previous POC for semantic technology, indicate that FIBO is capable of solving data issues in financial industries.

5.3.1 Feasibility of Implementing FIBO

Although the POC experiment and FIBO exploration are, to some extent, simplified versions, they not only show the feasibility of implementing FIBO, but also provide a model for doing so.

1. Semantic technology enables common structure models among different data sources using common vocabularies. Even data field names that are different for the same concept or the same for different concepts among the files can be accurately matched to corresponding elements of operational ontology with a full understanding their actual semantic meanings. The POC showed how operational ontology can be created and how to set up the rules for matching.
2. Data aggregation can be conducted across disparate data sources without changing each physical data structure. Once the matching is done, relevant data will be extracted and uploaded to a single data repository. No manual intervention is needed during extraction and aggregation. The experiment showed how the data can be extracted using Anzo for Excel.

3. As data is stored in "triple" structures, it is possible to use semantic reasoning, querying, and auto classification. Semantic technology enables customized visualizations for different purposes and scenarios to generate new knowledge. It is really hard to do so in conventional databases due to a relatively fixed data entry format.
4. As a general conceptual ontology, FIBO can be transformed into operational ontology, which can be utilized in a real case. The exploration experiment showed how to use FIBO Foundations, FIBO Business Entities, and other FIBO concepts for this use case. The results showed that FIBO can be operable in a real case.
5. The ontology models are readable to both humans and machines. Concepts of entities and relationships among them are clearly presented in the diagrams and tables. FIBO provides detailed explanations and definitions for each concept of financial entities and relationships. It would be convenient for a developer to create corresponding operational ontology models.

5.3.2 Challenges of Implementation

However, there are still some challenges to be solved while implementing a FIBO operational ontology.

1. In the application of FIBO, a universal unique LEI is quite important. As semantic model relies on a definition of each "thing," a unique identifier that is critical to distinguish among different elements belonging to the same "thing." The lack of a unique and universal identifier may cause problems when this same "thing" is stored multiple times in a data repository. Meanwhile, LEI is also important for performing semantic operations such as reasoning. In a real application, firms may find it difficult to implement this model when there is no universal identifier for them to calculate total risk exposure.
2. Considering that POC and FIBO exploration are dealing with a relatively small scale of data fields and files, it is still practical to examine the semantic meanings of each item in files and match them in the right place in ontology model. In real life, there could be thousands of fields in all the data sources with confusing names and meanings. Although the matching process can be regarded as one-off work, total workload is still huge and may cause some misunderstanding.

3. In a real application, some fields in the files are created only for the specific local cases. As FIBO is a general business ontology, it is important to know how to balance descriptive functions and operative functions to create the operational ontology. There are also some rules to follow when creating a FIBO-conformed model. The ontology model may be extremely complicated to the point of incomprehensibility, especially once the system contains enormous fields and files. It is difficult to estimate the effort that would be involved in creating this operational ontology and performing matching with these data sources.

6 Conclusion

To conclude, this thesis discussed several topics, ranging from the root cause of improving data management practices to an experiment that provides a possible solution for implementation. It covered an analysis of the importance of better data management practices, suggested a possible approach for implementing improvements, provided proof of the solution's feasibility, and discussed challenges for its implementation. Although there is still plenty of future work and development to be accomplished, through the POC experiments, the semantic and FIBO models nevertheless show that they are capable of solving data aggregation issues, which provide critical challenges to monitoring systemic risk and stabilizing financial markets.

6.1 Motivation for Better Data Management Practices

Participants in the financial industry, especially financial institutions, are motivated to implement better data management practices for multiple reasons.

Internal Factors. Internal factors that raise financial firms' interest in improving data management include leveraging existing data and information to generate new knowledge, enabling robust models for decision making, doing a better job of monitoring and controlling risk exposure, and enhancing firms' reputation and credibility. Due to increased competition in the financial market, firms can use better data management as competitive advantage and be a leader in IT innovation. Better risk data management could also increase revenue and decrease operational risks and costs through better risk control.

External Factors. Meeting regulatory compliance is the most significant external factor that urges financial firms to improve their data management practices. The Dodd-Frank Act and BCBS 239 are recent important regulations that each firm needs to follow. The self-assessment held by the SIG revealed that these firms were poor in IT infrastructure and data aggregation capability. To meet the regulatory requirement, financial firms have to find approaches to improve their data aggregation practices along with their existing infrastructure.

Although there are some data standards that can be used for data aggregation, they have different kinds of restrictions and disadvantages. At the same time, semantic technology not only shows advantages compared to other approaches, but also has the ability to improve data management practices.

6.2 Semantic technology, Ontology and FIBO

There are several significant advantages of semantic applications. The most critical one is that semantic technology provides a common, unambiguous dictionary for describing and defining entities and relationships in the problem domain. Moreover, conceptual ontology models are built independently from any specific data structures so that they can be placed on top of existing data structures in various data sources. Semantic technology enables communication among disparate data sources, even if they have different designs, without changing each fundamental physical data structure. Among emerging ontology technologies, FIBO is an important conceptual model that is created specifically for financial industry.

Given that semantic and ontology applications have these advantages, the next concern is how to create a conceptual ontology, transform it into an operational ontology, and implement this operational ontology into a real application. As FIBO is created and maintained by the EDM Council and the OMG, the rest of the task would involve transforming the FIBO conceptual model into an operational model and then implementing it into real case.

The thesis first used a use case to study the capability of semantic technology for data aggregation across different data sources. A POC experiment is created to evaluate the feasibility of semantic technology. The results show that semantic applications can actually work for solving data issues.

Given that the ontology model is a purpose-based model, this thesis also conducted another experiment to replace several items with FIBO items rather than using real FIBO concepts in order to show the feasibility of FIBO. The results still show positive outcomes, although there will be some challenges implementing this model in real life.

6.3 Challenges

Semantic technology in general and FIBO in particular are far from perfect at their current stage. Implementing semantic technology and FIBO will pose some challenges.

The first challenge is creating an operational ontology from its corresponding conceptual ontology model. There is no “right” rule to create an operational ontology other than some guidance. Because an operational ontology is the one that is actually used in a real applications, technical constraints need to be taken into account when it gets transformed from a conceptual ontology model. Business and IT functions need to work closely to find a balance point between the conceptual model and the real application. They also need to figure out technical constraints and a more suitable and decidable subset of the FIBO family.

Another challenge is performing the actual matching of physical data and elements in the ontology model. Fully understanding the exact semantic meaning of each field in each file is a prerequisite of accurate matching. It is acceptable when there are only a few files and a few fields. However, in real life there can be hundreds of fields in thousands of files in the problem domain. Although setting the rules of matching is mostly a one-off process, the total processing time is still huge.

At the same time, lack of a unique LEI is also a challenge. Given that the ontology model uses a "Subject-Verb-Object" format to describe a "thing," a unique identifier for distinguishing each entity is a fundamental part of the model. However, in the financial industry, there is still no universal identifier for legal entities, which can sometimes cause problems.

7 Future work

Although the thesis has shown that semantic and ontology technologies, especially FIBO, are feasible and can be implemented in real life, in order to implement them on a large scale, there is still a lot of work to do in the near future.

7.1 Further Development of FIBO

FIBO is not yet a mature product. Currently, only some fundamental parts are at the finalized stage. Many parts, especially those for financial instruments, are still under review or have not even been started. It is expected that it will take a long time to fully develop FIBO. More effort should be done to establish the detailed specifications of each type of financial instrument and expand the application scope.

7.2 POCs to Be Conducted

More POCs experiments should be conducted along with the development of FIBO. Currently, there are only a few POCs for FIBO. The POC in this thesis also only explores part of the FIBO family. To prove that FIBO can be implemented in a larger scale, it requires more research and further experiments.

Meanwhile, a sound method for creating a more formal operational ontology model has yet to be developed. In this thesis, a simplified version of operational ontology is created by taking a subset of FIBO. In a real application, a solid and formal model that rigorously conforms to FIBO should be created.

7.3 Compatibility with Other Data Standard

Because there are some existing data standards, one interesting topic of exploration would be how, if at all, FIBO can be a mediator for regulatory reporting, or even for other applications, with the input of some data standards such as FIX. It will also be interesting to discover whether the data and information organized by FIBO can be utilized by other tools such as ACTUS. In these cases, FIBO acts as a mediator, taking input from or sending output to other data standards. How, if at all, FIBO works with other data standards remains to be investigated.

7.4 LEI Issues

One of the most critical unsolved problems is LEI. Due to lack of incentives and coordination, a universal LEI system is not ready yet. Considering the importance of LEI in FIBO, it is worthwhile to research ways to generate and implement LEI, with a special focus on how to make it sustainable and acceptable to mainstream of financial institutions.

In summary, semantic and ontology technology is a growing and promising trend that cannot be neglected. It is not a problem of "whether" but "how" to use FIBO to bring data transparency to the financial industry. Therefore, future work should be focused on further development, better implementation, application scope expansion, and peripheral technologies' support.

8 Bibliography

Basel Committee on Banking Supervision. (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems* (Vol. 2010).

Basel Committee on Banking Supervision. (2013a). Principles for effective risk data aggregation and risk reporting, (January).

Basel Committee on Banking Supervision. (2013b). Progress in adopting the principles for effective risk data aggregation and risk reporting. Retrieved from http://www.bis.org/publ/bcbs268.htm

Bennett, M. (2013). The Financial Industry Business Ontology: Best Practice for Big Data. *Journal of Banking Regulation*, *14*(3-4), 255–268. Retrieved from http://www.palgrave-journals.com/jbr/archive/index.html

Cambridge Semantics. (2015). Compliance Surveillance | Cambridge Semantics. Retrieved May 5, 2015, from http://www.cambridgesemantics.com/solutions/compliance-surveillance

CFTC. (2015). Data Repositories - CFTC. Retrieved from http://www.cftc.gov/industryoversight/datarepositories/index.htm

Davis Polk Wardwell LLP. (2010). Summary of the Dodd-Frank Wall Street Reform and Consumer Protection Act , Enacted into Law on July 21 , 2010.

Dodd, C. (2010). Summary : Restoring American Financial Stability. *Senate Committee on Banking, Housing, and Urban Affairs*, 1–11.

EDM Council. (2015). Financial Industry Business Ontology. Retrieved from http://www.edmcouncil.org/financialbusiness

Financial Crisis Inquiry Commission. (2011). *The financial crisis inquiry report.*

FIX Trading Community. (2015). About FIX Trading Community. Retrieved from http://www.fixtradingcommunity.org/

Fouque, J.-P., & Langsam, J. A. (2013). *Handbook on systemic risk.* Cambridge ; New York : Cambridge University Press, 2013.

Lei, T., & Brothers, W. L. (2013). Frequently Asked Questions : Global Legal Entity Identifier (LEI) February 2013, (August), 1–6.

Levin, C., Bean, E. J., & Henderson, A. C. (2011). WALL STREET AND THE FINANCIAL CRISIS : Anatomy of a Financial Collapse Staff Director to the Minority.

Object Management Group. (2014a). Financial Industry Business Ontology - Business Entities, (May).

Object Management Group. (2014b). Financial Industry Business Ontology Foundations, (January).

Office of Financial Research. (2014). 2014 Annual Report.

OFR. (2010). Office of Financial Research Created Under the Dodd - Frank Wall Street Reform and Consumer Protection Act : Frequently Asked Questions Q : What is the Office of Financial Research (OFR)?

OFR Financial Research Advisory Committee. (2014). Discussion Topic on Swap Data Repository and Data Standards, 1–3.

Rittgen, P. (2008). *Handbook of ontologies for business interaction.* Hershey, PA : Information Science Reference, c2008.

SEC. (2015a). SEC.gov | CUSIP Number. Retrieved March 31, 2015, from http://www.sec.gov/answers/cusip.htm

SEC. (2015b). SEC.gov | Insider Trading. Retrieved March 30, 2015, from http://www.sec.gov/answers/insider.htm

Semantics Repository. (2015). Semantics Repository. Retrieved from http://www.edmcouncil.org/semanticsrepository/index.html

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic Web revisited. *IEEE Intelligent Systems*, *21*(3), 96–101. Retrieved from DOI: 10.1109/MIS.2006.62.

The Trade Associations Global LEI Proposal. (2011). Requirements for a Global Legal Entity Identifier (LEI) Solution, (May).