

Financial Industry Ontologies for Risk and Regulation Data (FIORD) – A Position Paper

Justin McCarthy¹, Laurentiu Vasiliu², Allan D. Grody³, Cal Muckley⁴, Donald Lawrence⁵, Fokion Zervoudakis⁵, Said Tabet⁶, Jeroen van Grondelle⁷, Thanassis Bouras⁸, Kiran Fernandes⁹, Theodoros Krintas¹⁰, Panos Xidonas¹⁰, Kyriakos Chourdakis¹¹ and Adamantios Koumpis¹²

¹GRCTC University College Cork, Ireland, ²Peracton Ltd., Dublin, Ireland, ³Financial InterGroup, London, UK

⁴University College Dublin, Ireland, ⁵University College London, UK, ⁶EMC Corporation, Belgium, ⁷Be Informed, Apeldoorn, The Netherlands, ⁸Ubitech Athens, Greece, ⁹Durham University Business School, Durham, UK, ¹⁰Attica Wealth Management, Athens, Greece, ¹¹Nomura Holdings, London, UK, ¹²DERI, National University of Ireland, Galway, Ireland.

Abstract. This paper presents a proposed approach to address risk and regulation management within the highly active and volatile financial domain by employing semantic based technologies within a collaborative networks environment. Firstly the problems and motivation are introduced, with accent on big data and high frequency trading issues that are creating major problems to the current software systems. Secondly the state of the art on Big Data, Regulation and Risk Management are presented. Next the FIORD platform architecture is detailed and the envisioned approach explained. Finally conclusions are presented where benefits for real time monitoring are emphasized so high frequency trading irregularities are detected in real time for the benefit of involved financial institutions.

Keywords: Big Data, High Frequency Trading, Collaborative Networks, Semantic Web, Ontologies, Reasoning, Risk, Regulation, Financial Industry

1 Introduction

Gathering, querying and analysis of various types of data is an area that has developed in recent years as the volume and complexity of data has grown. This has led in turn to the concept of "Big Data", that is strongly anchored within collaborative networks concept: without collaborative networks, Big Data would not have developed explosively in almost all domains of science, industry and life. In this respect, according to IBM, "2.5 quintillion bytes of data are created daily" [1]. The Financial Services sector is amongst the most data driven of industries. However, the industry has relied on older technologies to handle this ever-increasing data and analytics burden. The regulatory environment for this industry requires an understanding of multiple types of data including: order and trade execution data, market data for bids and offers, last sales and volume information, order book information, news and

economic data, post trade data such as trade allocation information and payment and settlement instructions; corporate event notifications and data on financial products and financial market participants. This results in large volumes of data - billions of market related messages, industry related economic information and data on individual company and contract markets distributed globally in real time. As example, just the market data feeds expressing bids and offers distributed on just the below listed equity and options exchanges in the US on a most recent and typical day April 25, 2013 peaked at 5.51 million transactions a second. Historically, the peak rate was 6.8 million per second achieved on December 21, 2012 [2].

The changes in regulation, the increased interconnectedness of markets globally and the concomitant increases in financial data have resulted in a number of problems for financial participants and their regulators:

- The growing volume of order, cancellation and trade data
- The growing number of trading markets to be monitored,
- Non standardized data structures and overloaded middle ware techniques;
- Legacy technology solutions – both hardware and software;
- Lagging people and organisation skills in the technology, business and scientific communities necessary to understand Big Data, and,
- Finding the right data or combinations of data that will answer a business or scientific question from a very large volume of data in real-time.

Coupled with the above more generic big data challenges is the reluctance of financial services participants to move from their trusted existing IT infrastructure. Thus, an organisation such as a bank may have multiple parallel systems that have different internal data structures. Something different is required and this is what our research work is aiming towards. More specifically our aim is to allow for these existing platforms to be kept in place and it allow a common definition of the data within the systems to exist enabling a convenient and efficient way to access this data using new tools and methods. We refer to our approach as the Big GRC (Governance, Risk and Compliance) Data FIORD platform

2 Motivations and Background for the Research

Our research work will look into understanding the causes that could generate unwanted market anomalies as the events that trigger financial market drops and then to prevent such drops reoccurring. As all financial markets are interconnected within massive collaborative networks, any anomaly can be propagated in seconds within multiple markets This was the case of the United States stock market crash on May 6th, 2010 (known as the 'Flash Crash') in which the Dow Jones Industrial Average (DJI) plunged by 1,000 points (about 9%) only to recover those losses in 20 minutes; and the twitter inspired precipitous drop in the Dow of 140 points on April 23, 2013 that recovered in a single minute. Our research work would contribute to the forensic review increasing the reliability, transparency and security of markets leading to heightened investor confidence and greater market liquidity. It should be able to anticipate and perhaps even protect the global economy from unexplained precipitous swings in market prices as was the case with the previously mentioned "Crash Flash"

of 2010. [3]. In this later respect we envision providing semantically augmented technology solutions to market participants' infrastructure and regulators' monitoring capabilities which would gain and apply knowledge of the industry and the use of data there-in. This knowledge will be captured in open source versions of ontologies and semantic algorithms, which in turn will be put forth as draft standards for the financial services industry to upgrade and update the technology of this globally important but volatile industry. The core of our research (see Fig. 1) relates to the use of semantic technologies, ontologies and taxonomies. Industry-standard tagging of data are emerging as one of the technologies to enable the financial industry and its regulators to observe financial transactions by computer means through their ability to precisely define data content, facilitate classification, represent data rules and add dimensions of intelligence onto data structures more efficiently and at a lower cost than conventional technologies.

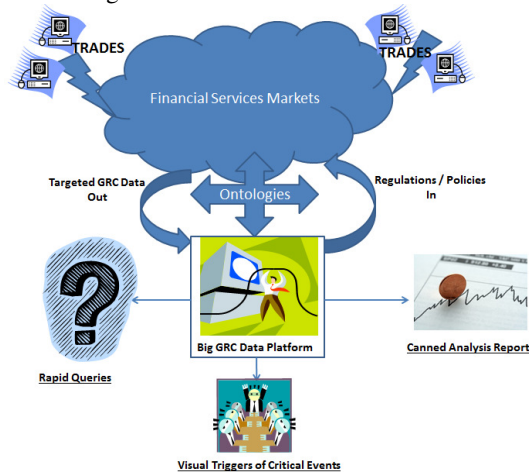


Fig. 1. FIORD High level View

Therefore, a regulator like the recently proposed EU Single Supervisory Mechanism or a large market participant would be able to:

- Perform complex and rapid queries on the large volume of trades that can be generated in just seconds;
- Use these queries to analyse market activity;and,
- Provide collated data to a suite of reporting tools that would allow rapid analysis and understanding of the changing position in the market.

3 State-of-the-Art

3.1 State-of-the-Art on Big Data and Collaborative Networks: Big Data is an emerging field where innovative technology offers alternatives to resolve the inherent problems that appear when working with huge amounts of data while

providing new ways to reuse and extract value from information. All this is done mainly within collaborative networks environments as data is reused, improved and created only by cooperation and not in isolation. Big Data Public Private Forum (BIG) is working towards the definition and implementation of a clear strategy that tackles the necessary efforts in terms of research and innovation while providing a major boost for technology adoption and supporting actions for the successful implementation of the Big Data economy. **Volume, Velocity and Variety** were originally posited by *Gartner's Doug Laney* in a 2001 [4] research report that provides terms to describe some of the issues that affect big data. State of the art in these areas includes:

- **Volume:** Big data companies such as EMC can now offer storage solutions such as the 4 petabyte VMAX array. Meanwhile in the financial services domain, the world's financial markets and financial services companies generate and store petabytes of data on a daily basis. However financial services companies struggle to structure, query, analyse and act on this data mainly as it is stored in disparate data sets.

- **Variety (Big Data and Collaborative Networks):** Big data is any type of data - structured and unstructured data such as text, sensor data, audio, video, click streams, log files and more. New insights are found when analysing these data types together, acquired from multiple networks.

- **Velocity:** For time-sensitive processes such as catching fraud, big data must be used as it streams into the enterprise in order to maximize its value. High Frequency Trading (HFT) and real-time risk management are some of the most time sensitive data applications in financial services. In these strategies, computer scientists rely on speed to gain minuscule advantages in arbitraging price discrepancies in some particular security that is trading simultaneously on disparate markets.

3.2 State-of-the-Art on Regulation: Regulations in the financial services industry are becoming increasingly complex (Fig. 2) and can involve more than one regulator for a firm as well as different products, regulations and jurisdictions. However they play an important role in ensuring the stability of Europe's markets and economies and protect its citizens. Risk and controls for compliance are becoming more complex and more are needed in real time. Proposals have been made to use solutions that are based on semantic web technologies.

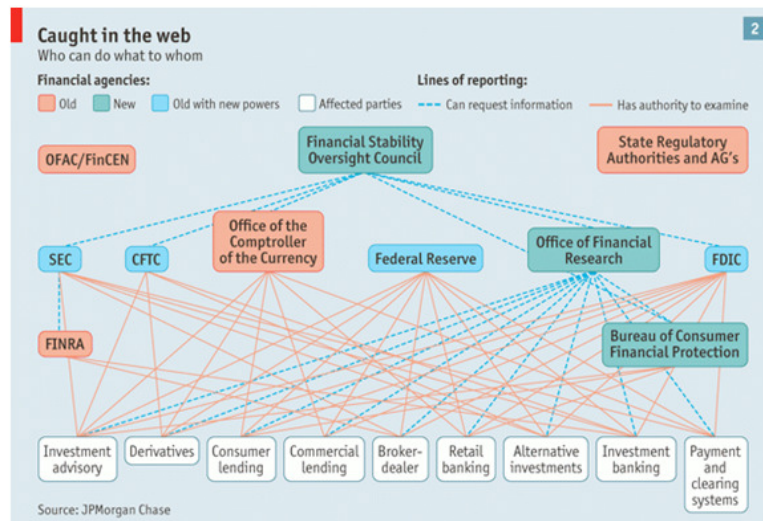


Fig. 2. The complexity that is Financial Regulation – An Example from the United States

An initiative related to trading is the Financial Industry Business Ontology (FIBO) [6]. This is a collaborative effort engaging participants from the financial industry, the regulatory community, academic institutions and semantic technology vendors to model and build a set of open standard ontologies that will deliver transformational benefits to the financial system. The Enterprise Data Management Council (EDMC)[7] and the Object Management Group (OMG)[8] are partnering to lead this effort. Following the 2008 global financial crisis it became clear that the absence of certain data standards led to a lack of awareness of the magnitude of risk that the industry was exposed to. The goal of FIBO is to introduce semantic capabilities that will better enable: common terminology for business entities, financial contracts and instruments; improve data transparency; improve data integration and linkage; and improve financial reporting and analytics, so that greater "data" quality and health can be introduced into the financial system.

Other activities use collaborative groupings such as the Fix Protocol Ltd organization and the XBRL International Federation [9] to leverage eXtensible Mark-up Languages (XML) [10] and its offspring such as XBRL that tag data through standard protocols to identify data elements that computers can query and retrieve. Big GRC Data would take this collaborative approach and apply it to the space between regulators, market participants and data.

3.3 State-of-the-Art in Trading and Risk Management

Beginning in the early 1970's entrepreneurs and established exchanges began to introduce electronic trading systems, the CAT System of the Toronto Stock Exchange, private initiatives Instinet and Autex, and the Nasdaq market are from that

era. In the 1980's trading centers became increasingly electronic with electronic futures markets such as Intex and the forerunner to CME's Globex arising, along with equity markets in the UK, such as SEAQ and SETS; and in Europe such as Xtra in Germany and the Paris Match system (see Past, Present and Future: The Evolution and Development of Electronic Financial Markets [11]). In the United States, National Market System I and then II and in Europe the Market in Financial Instruments Directives I and II furthered the development of electronic trading markets and precipitated high-speed trading. First, it was enough to simply get a trade done without human hands or voices being involved. Later, it became a matter of getting to the execution facility's order book before others. Direct market access to these executing facilities with limited credit limit checking became a way of eliminating delays in round-trip time. Latency busting co-location facilities, fiber networks and stream processing, all came together to process multiple data feeds in parallel in real-time streams, again to cut down on round-trip time. Further speed advantage was made possible by a new family of multi-core processors - symmetric multiprocessors making multiple central processing units (CPUs) available to complete individual processes simultaneously. Clusters of multiprocessors made massively parallel processing possible within a single machine or across multiple machines. Also becoming increasingly prevalent is asymmetrical processing, which uses separate specialized processors for specific tasks, like the Graphical Processing Unit (GPU) that moved from the video gaming to the financial district. Real-time trade risk management is becoming more of a possibility and a demand by regulators with the furtherance of a pervasive global network of almost unlimited bandwidth and with massively parallel, almost unlimited computing capabilities. This takes the form of shared facilities available on demand in the form of computational utilities provided as a service, referred to as cloud computing.

4 The proposed FIORD Architecture Platform

FIORD initiative (driven by an EU consortium made of the authors of this paper) will introduce an innovative approach, based on existing platform elements from the big data, semantic, and cloud areas as described in next Fig. 3.:

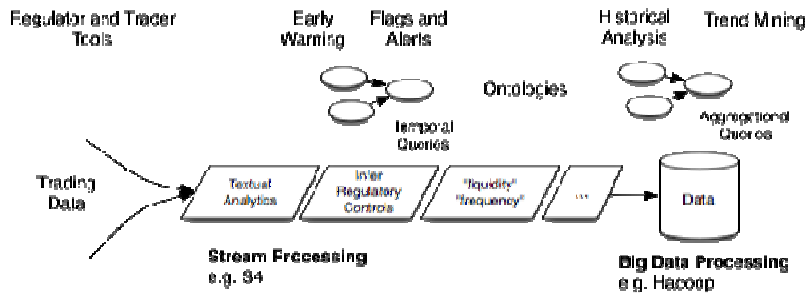


Fig. 3: FIORD Platform Architecture

The Big data elements will be required to store and analyze huge amounts of data. This will be done in three parts.

4.1 Stream Computing: This is a high-performance computer system that analyzes and collates multiple data streams from many sources. Industry users will then be able to use their existing data subscriptions in the final implementation of the platform. Where regulators can provide access to live real-time data we would avail ourselves of such; as well as internal feeds of data from the individual industry users. The Stream Computing elements of the platform would process the data using an inference engine [17] to find patterns in the data and stream it back out as a single flow. Stream computing uses software algorithms that analyze the data in real time as it streams in to increase speed and accuracy when dealing with data handling and analysis. Using this element of the platform FIORD would be able to uncover patterns of trading that will give regulators insight into triggers of systemic risk exposures (a suggested focus for the FIORD project is HFT systemic risk exposures), establish those triggers, or at least propose them, then install them in our designs.

4.2 Historical Analysis: In addition to processing data in real time, the streams will be stored in vast storage arrays to allow their review. For example, FIORD could take the triggers identified for systemic risk and see how they stand up to the historical data record as well as use them to proactively give signals of preventive measures. The addition platform element for this may include Data Storage. A potential solution may be EMC hardware, for example the Symmetrix VMAX, Teradata processors, SAPs in memory data bases, Kove's Dram accelerators, etc. It may be practical in today's technology environment that federated networks of servers across a vast globally interconnected network as envisioned by the LEI initiative of the G20's Financial Stability Board [12] which may provide possibilities to use hundreds of servers as a clustered virtual computing environment.

4.3 Semantic Platform The above approach will require a web semantic platform to allow the platform to link data from diverse sources together, reason over it, federate it, and query it. The key platform capabilities may include such techniques as :

- Querying: SPARQL[13] is an RDF [14] query language, that is, a query language for databases, able to retrieve and manipulate data stored in Resource Description Framework format.
- Ontologies: OWL[15] or the Web Ontology Language, a language for describing and sharing ontologies on the World Wide Web. This will allow the ontologies developed as part of FIORD to be authored and distributed.
- Rules: RIF[16] is the W3C Rule Interchange Format. It's an XML language for expressing rules which computers can execute.
- Taxonomies: RDF is a set of classes with certain properties using the RDF extensible knowledge representation language, providing basic elements for the description of ontologies, otherwise called RDF vocabularies. These resources can be saved in a triplestore to reach them with the query language SPARQL.
- Reasoner: This is a piece of software able to infer logical consequences from a set of asserted facts or axioms. There are various options available for this element of the platform and the most appropriate will be investigated.

5 Conclusions

FIORD aims to provide tools and methods, and a pilot of both, for detecting systemic risk. It will leverage existing programs that are already moving forward in both the public and private sectors. For example systemic risk analysis will require an external globally unique identification system for all financial market participants and the products they trade, own and/or process. It will also require standardized data tags to surround the analyzed data in order to enable efficient computer query and analysis.

Benefits are expected to be significant as real-time monitoring by regulatory supervisors will enable stress events such as a High Frequency Trader algorithms going rouge to be easier and much earlier detected. FIORD would also contribute to systemic risk analysis across multiple financial institutions. Here a cascading trigger event could potentially be detected due to concentration and liquidity risk exposures observed across the interconnected financial system. Increased safety and security of markets would result leading to heightened investor confidence in those markets and in financial institutions in general..

References

1. IBM, 2012, What is big data ? <http://www-01.ibm.com/software/data/bigdata/>
2. MarketData Peaks ,2013 - www.marketdatapeaks.com
3. Reuters, 2010 –‘Single trade helped spark May’s flash crash’ <http://www.reuters.com/article/2010/10/02/us-flash-idUSTRE69040W20101002>
4. Doug. L, 2012 <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
5. Akamai, 2013 <http://www.akamai.com/html/technology/dataviz1.html>
6. FIBO, 2013 – Financial Industry Business Ontology <http://www.omg.org/hot-topics/fibo.htm>
7. EDM Council, 2013 www.edmcouncil.org
8. OMG working group www.OMG.org
9. XBRL, 2013 <http://www.xbrl.org/>
10. XML, 2013 – W3C <http://www.w3.org/standards/xml/>
11. Grody, L. et al, 1993 Grody, Levecq; NYU Working paper Series, Nov. 1993 at <http://ssrn.com/abstract=1284845>
12. Financial Stability Board, 2013 www.financialstabilityboard.org
13. SPARQL, 2008 – Query Language for RDF - <http://www.w3.org/TR/rdf-sparql-query/>
14. RDF, 2013 – Resource Description Framework <http://www.w3.org/RDF/>
15. OWL, 2007 <http://www.w3.org/2004/OWL/>
16. RIF, 2013 - <http://www.w3.org/TR/rif-overview/>
17. Singh, S et al 2010 – A comparative Study of Inference Engines http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5501457&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5501457