Knowledge Representation and Information Management for Financial Risk Management

Report of a Workshop

Arlington, VA

July 21-22, 2010

Mark Flood, University of Maryland

Albert S. (Pete) Kyle, University of Maryland

Louiqa Raschid, University of Maryland

Knowledge Representation and Information Management for Financial Risk Management

Report of a workshop

Arlington, VA
July 21-22, 2010

CONTENTS

Contents

The financial and CS communities seek to bring about a new era of quantified external and internal oversight of financial activity, activity evolving under continuous competitive acceleration. We can only succeed if the foremost financial and computer science theorists and practitioners can meet, as we did here, to understand one another's frontier perspectives, problems, and vocabularies.

> − *Workshop participant Leonard Nakamura, Assistant Vice President, Federal Reserve Bank of Philadelphia*

EXECUTIVE SUMMARY

Overview

The National Science Foundation and The Pew Charitable Trusts co-sponsored a *Workshop on Knowledge Representation and Information Management for Financial Risk Management* on July 21 and 22, 2010 in Arlington, Virginia (see Flood, Kyle and Raschid, 2010). The goal of the workshop was to initiate a research discussion about the knowledge representation challenges for effective financial information management. Over fifty invited academic researchers, financial regulators and industry practitioners participated in the event. The participants brought diverse perspectives and expertise in economics, computer science, finance, and information science, resulting in an interdisciplinary exchange of ideas. The specific topics covered a broad range, including financial risk management, ontologies for knowledge representation, formal logics, schema mapping, systemic risk, constraint languages, networks, simulation techniques, data integrity, operational risk and data security, to name a few. This report describes the discussions flowing from the workshop and its immediate aftermath in greater detail. We hope, however, that this is only the beginning of a much longer conversation.

The case for the workshop: information requirements in finance

Information occupies a central place in financial theory and practice. Core concepts include: the efficient markets hypothesis, which implies that securities prices alone can provide a sufficient statistic to guide investment decisions; market completeness, which measures whether enough distinct prices exist to span all the significant information in the state space; and the role of asymmetric information in organizational efficiency and contract design. Despite the central place for information in financial modeling, many core concepts from knowledge representation, information management and computational methods seem to play only a limited role in financial research and practice. It is our judgment that computer and information scientists have much to add to the quality and reliability of data processes and information available to financial practitioners, regulators and researchers.

To illustrate the potential, imagine the following hypothetical scenario occurring a few years hence: a large, developed economy experiences economic stress arising from an unhappy confluence of factors – public-sector liabilities mount for health and retirement benefits of an aging population; meanwhile tax revenues lag during a cyclical recession, and a significant crop failure arises from an unusually hot and dry summer. Although the elements of this "perfect storm" have been an active topic of political debate for months, the newly elected government's move actually to reschedule a portion of sovereign debt payments takes the markets by surprise: no major industrial economy in living memory has resorted to this expedient. The resulting credit event reveals that two large international banks, both based in the U.S., are badly overexposed as sellers of sovereign credit default protection for the rescheduled debt. As ugly rumors about the two banks begin to swirl on Wall Street, federal banking regulators (still painfully aware of the 2008 Lehman crisis) move to confront the burgeoning panic. They want to know the implications of this shock first and foremost for the troubled institutions themselves. Emergency stress tests are ordered: Which contractual positions are affected? What is the extent of the exposures? Will the banks have sufficient liquidity during the coming week to meet their immediate obligations? Should the Fed step in as lender of last resort – i.e. do the banks' credit losses dwarf their short-term liquidity difficulties? Will the banks' "living will" provisions have to be invoked, and will these resolution mechanisms work as intended? As they contemplate worst-case dynamics, regulators' thoughts soon turn to the implications of a systemic crisis: what will be the price impact as the troubled institutions dump assets to raise cash? How might their potential failures propagate through the counterparty network? To whom do they have net short exposures? Are their netting agreements in place? How are these contracts collateralized? In principle, much of the information to answer all of these questions – even if only approximately – exists in the vast network of legal and business relationships that defines the modern financial system. The actionable question is whether the system will have the information toolset in place to extract the necessary answers in a timely way.

Systemic risk and the Office of Financial Research

The workshop was conceived and organized in the aftermath of the financial crisis of 2008-09.[1] The policy imperative to address "systemic risk" was clear, even if there is not yet a general consensus on a precise definition of the term. One working definition is the "risk that an event will trigger a loss of confidence in a substantial portion of the financial system that is serious enough to have adverse consequences for the real economy;" see Taylor (2009). The crisis and the ensuing recession have shone a light into the esoteric world of financial data processing, laying bare the inadequacies of the information infrastructure supporting the financial system. The episode demonstrated painfully that theory and practice of good financial data management are critical issues for risk assessment, with significant consequences and systemic implications. The lack of rigorous research and best practices for standards and ontologies, data sharing procedures, quality metrics and mathematical modeling and reasoning have been an important contributor to the instability of the system.

The crisis had at least one very tangible and significant outcome of direct relevance to the issues under discussion at the workshop. President Obama signed the *Dodd-Frank Wall Street Reform and Consumer Protection Act* into law on July 21, coincidentally the opening day of the workshop (see

---

[1] The bibliography to this report includes references to a number of recommended papers, articles and speeches that describe and analyze the crisis, its causes and possible remedies.

U.S. Congress, 2010). Among its many regulatory changes, the Dodd-Frank Act created an Office of Financial Research (OFR) with the mandate to establish a sound data management infrastructure for systemic risk monitoring. The OFR should be a focal point for defining and implementing many of the responses to the challenges sketched above.

In §154(c), the Dodd-Frank Act mandates that the new agency contain a Research and Analysis Center (OFR/RAC). The OFR/RAC must maintain computational and analytical resources to: measure and report on changes in systemic risk, evaluate institutional stress tests, advise fellow regulatory agencies, investigate disruptions and failures, analyze proposed policies and promote best practices in risk management.

To provide the information required for these systemic risk analyses, §154 of the Dodd-Frank Act also mandates that the OFR contain a Data Center (OFR/DC) to manage data for the new agency. Under the terms of the Act, the OFR/DC must:

- Publish financial instrument reference data,

- Publish legal entity reference data,

- Publish data reporting standards and formats, and

- Collect contractual positions and transactions data.

These requirements are specified in §154(b) of the Dodd-Frank Act itself. The Act further requires that the OFR/DC maintain data security and confidentiality. In addition to these data resources mandated explicitly by the Dodd-Frank Act, the measurement and analysis performed by the OFR/RAC will, as a practical matter, require access to market prices and related data for the valuation of positions, development of models and scenarios and the measurement of micro-prudential and macro-prudential exposures.

Institutional risk and practical challenges

The management of financial risk information is not only a concern at the systemic level, of course. Data and information are collected, managed and analyzed at all levels of the institutional hierarchy, and much of the implementation activity therefore occurs at the narrowest institutional level. To emphasize this point, the first day of the workshop included two parallel breakout sessions, one on "Risk Management in the Large" and another on "Risk Management in the Small." There are significant shortcomings with current data implementations in finance; there are large volumes of mission-critical data involved, so that implementing well can imply very large reductions in operating costs and operational risk.

At the institutional level, technical challenges around large-scale data systems interact with significant economic forces involving innovation, transparency, confidentiality, complexity and organizational change to create a very difficult problem. Today, corporate managers are uncertain of their internal risk and accounting numbers; academic researchers lack information needed to analyze market operations and behavior; and regulators and analysts are denied the depth of understanding of market activities needed to forge good financial policy.

While there were myriad interacting causes, a central theme in the crisis was the proliferation of complex financial products that created a gap between the quality and quantity of data needed for an appropriately diligent analysis of the risks involved, versus what was available. In many cases,

the information available about products and counterparties was inadequate. The data quality gap in finance is an evolutionary outcome of years of mergers and internal realignments, exacerbated by business silos and inflexible IT architectures. Difficulties in unraveling and reconnecting systems, processes and organizations – while maintaining continuity of business – have made the problem intractable. Instead, data are typically managed on an ad-hoc, manual and reactive basis. Workflow is ill defined, and data reside in unconnected databases and spreadsheets with multiple formats and inconsistent definitions. Integration remains point-to-point and occurs tactically in response to emergencies. Many firms still lack an executive owner of data content and have no governance structure to address organizational alignment or battles over priorities.

Financial risk and information managers are gradually recognizing the concepts of metadata management (managing precise data definitions based on semantic models), ontologies (formal representations of concepts within a domain, along with their meanings and interrelationships) and knowledge representation (representations structured specifically to support automated inference) as essential strategic objectives. These same concerns apply with equal urgency to the financial regulators tasked with understanding individual firms and the overall system. A sound data infrastructure and open standards are necessary, both for effective regulation and for coordinating industry efforts. The history of patchwork standards and partial implementations demonstrates the tremendous obstacles to consensus over shared standards in the absence of a disinterested central authority.

The workshop and subsequent developments

The workshop program and participant list appear in the appendix to this report. The primary discussion of the workshop took place in eight breakout groups. Each group considered the financial and risk management challenges from the viewpoint of a finance or economics researcher and the computational and information management challenges.   The eight breakout topics were:

1. Risk management in the small (i.e. institution-level risks and information).

2. Risk management in the large (i.e. systemic risks and information).

3. Knowledge representation frameworks (ontologies, schemas, models, formal logics) to describe complex financial instruments.

4. Managing risk models: schema mapping, data exchange and model comparison and reliability.

5. Languages (operators and rules) for specifying constraints, mappings and policies governing financial instruments.

6. Financial networks, agent-based simulation and architectures for large scale computation.

7. Data integrity, data quality and operational risk.

8. Privacy, confidentiality, security and trust in managing financial data.

This was an interdisciplinary workshop, and a key goal was to show the promise of collaboration to both finance researchers and computer/information science researchers.

In the aftermath of the workshop, the report-writing committee chose to crystallize the themes and discussions in six financial use cases. These are intended as concrete examples of how data management research and expertise might help address particular financial challenges and

application areas:

1. Visualization and analysis of a financial counterparty network.

2. Knowledge representation of a financial contract.

3. Implementation of a "living will" for a large financial firm.

4. Fostering an ecosystem of credit analysis.

5. Reasoning over financial contracts for completeness and integrity.

6. Privacy and trust: multiparty sharing of confidential financial data.

These use cases are elaborated in greater detail below. For each, we present the context and objectives, the potential users and stakeholders and a detailed usage scenario.

In addition to the research challenges addressed in the workshop, there was a strong interest among the participants to engage in the specification of open standards, the development of open source software solutions and curriculum development activities.

- A discussion group and corresponding wiki site have been set up for the discussion of open standards and open source software solutions for financial information management and information exchange.

- The Dodd Frank Act has triggered significant activity around financial information management. The Act mandates the preparation of specific feasibility reports by various regulatory agencies, in many cases with short production deadlines (see Alston & Bird, 2010). The NSF workshop has acted as a catalyst for this task, and a subset of the workshop participants have begun to examine some of the specific reporting topics required by the Act.

Recommendations

There is enormous economic value to be gained by improving the efficiency and robustness of information systems for financial institutions and their regulators. The hypothetical crisis described above suggests some of the important ways that good information, delivered in a timely way, can be vital. Beyond that, good risk information management will improve outcomes in less spectacular ways. For example, improved diligence in the loan origination process, fewer broken and failed trades due to inadequate documentation and more efficient hedging of established risk exposures will all reduce the costs and increase the value of the financial sector.

Our use cases, as well as the reports of the breakout groups, address the challenges and best practices for financial information management, as well as the appropriate representations, models, algorithms and systems for managing a crisis. The research questions in this report represent an essential first step toward defining a research agenda and enhancing practices and methodologies for the management of financial risk information.

If we were to summarize the main recommendations that flowed from the workshop deliberations, they would be:

1. Financial risk and information managers across the industry and regulatory community

should: establish semantic models that reflect best practice in knowledge representation; establish and adopt precise data definitions based on sound ontologies for all basic financial data; and promote sound standards for all metadata management. A sound data infrastructure and open standards are necessary, both for effective regulation and industry coordination. The OFR should fully implement its legally mandated Data Center (OFR/DC), including the regular and timely reporting of: financial instrument reference data, legal entity reference data, positions/transactions data and prices and related data.

2. The data reporting standards and recording architectures for the data collection required for regulatory purposes should be defined with care.

- o There is likely to be pressure to make minimum changes to existing systems, and this is likely to lead to a collection of "silos." Such solutions will not facilitate data integration, analysis and the derivation of knowledge to determine systemic risk.

- o The involvement of data management and data integration experts early in the process, as well as a system-wide acceptance of open standards, will facilitate an appropriate level of inter-operability. The ability to combine information from multiple sources is important to support adequate regulatory decision making.

- o Prior successes stand as examples in information sharing in the biomedical community, and there are ongoing efforts for the sharing of patient health information. Lessons learned from these prior endeavors can and should be successfully applied to systemic risk and financial information management.

3. To engage the academic research community, the OFR should seek out grand challenge problems while making the associated datasets available. Model building can advance only if test data are made available to the community of model builders.

- o There is a strong history in financial research – particularly in the area of securities valuation – to verify models with data. For example, databases such as the Center for Research in Securities Prices (CRSP) at the University of Chicago spawned a generation of productive scholarly research into investment portfolio management for exchange traded equities. Such data sources make use of the well organized public trade tickers provided by securities exchanges. Similar data resources should be made widely available for contracts that trade off the exchanges, including foreign currencies, mortgage-backed securities (MBS), interest-rate swaps and over-the-counter derivatives.

- o To support scholarly investigations into microstructure issues such as high-frequency trading, trade-by-trade and tick-by-tick data should be recorded. To protect the confidentiality of participants and their trading strategies, such data should be suitably anonymized and delayed as appropriate.

- o For more complex securities, such as MBS that aggregate information over hundreds of underlying mortgages, or structured notes with significant embedded optionality, there is a need for more sophisticated data representation beyond simple instrument identifiers.

- o To support systemic risk analysis, such trade data repositories should identify trade participants in a consistent and valid format, so that network relationships can be

established and tracked over time. Again, such data might be suitably anonymized and delayed as appropriate.

4. The OFR should promote and fund necessary research in multiple disciplines, including knowledge representation, information integration and extraction, data and text mining, financial modeling and simulation and result presentation and visualization. While the evaluation of system risk ultimately requires the construction of a financial model, doing so effectively requires advances in multiple disciplines. Challenges in multiple disciplines must be overcome and it is important not to be too narrow in defining the problem to be solved or in circumscribing the relevant areas of research.

Executive Summary

Financial research has been seriously limited due to inadequate access to data and a limited range of analytical tools. I believe that the next "big leap" in financial research can only happen with the capabilities that computer science in general and information management in particular bring to the effort. I hope that this workshop will play a pivotal role in jump starting this most important interdisciplinary collaboration.

− Workshop participant Allan Mendelowitz, former Chairman of the Federal Housing Finance Board

FINANCIAL CONTEXT

Information management in finance

Information occupies a central place in financial theory and practice. An important rationale for the structure of liberal capital markets is that they provide mechanisms and incentives to process information effectively, so that investment decisions can allocate capital to its most productive uses at the margin. This is the essence of the *efficient markets* hypothesis, which posits that current financial market prices incorporate "all available" information, so that the prices alone provide a sufficient statistic to guide investment decisions. In the practical realm, this translates into stock exchanges, trading networks, price tickers and other mechanisms for communicating prices, quantities, delivery times, allowable counterparties, etc. for both pre-trade bid and offer quotes and post-trade transaction details. A related concept is *market completeness*, which measures whether enough distinct prices are available to carry all the information needed to explain every important aspect of the state space – i.e. the domain of core random variables that governs the ex-post profitability of ex-ante investment choices. An important practical manifestation is certain forms of financial innovation, whereby new contracts are devised to isolate particular outcomes in the state space for a customized payoff structure to the counterparties.

In many typical situations, markets are highly liquid and competitive, approaching the efficient ideal. However, practical reality can also fail to live up to the theoretical ideal of complete and efficient markets. In short, there are capital market *imperfections*, including *information asymmetries*.[2] A simple example is insider trading, whereby managers trade securities in their own

---

[2] In the case of *external capital markets*, the communication of information occurs largely between firms, as they create contracts and trade securities. For example, in many financial markets there are information asymmetries between borrowers who are relatively well informed about their investment prospects, and lenders who are less well informed. The latter typically assume that applicants for loans are painting a rosier picture than is justified by their information sets. This is the problem of the so-called "market for lemons," and lenders typically charge a lemons premium to compensate themselves for the possibility of such misrepresentations. See Hubbard (1998) for a survey of some of the information issues in external capital markets. In the case of *internal capital markets*, information sharing and processing occurs within firms – among managers, divisions, and employees. In principle, managers should have better information about their firms' investment prospects than do outside investors,

firms based on private knowledge of emerging events. Similarly, in a *principal-agent* context (see Salanié, 2005), managers may choose investments to maximize job security or minimize managerial effort, rather than maximize investment return. The crucial point for data management is that capital market imperfections can create – inherently in the structure of the system – incentives for participants to "get the information wrong." Aside from such structural asymmetries and intentional misrepresentations, there are a variety of historical and institutional reasons that the financial system does a poor job of generating, managing and disseminating information.[3]

Financial data are used to value securities, manage operations and measure risk.[4] Financial institutions and their data suppliers acquire information from hundreds of sources including prospectuses, term sheets, corporate filings, tender offers, proxy statements, research reports and corporate actions. They load the data into master files and databases providing access to prices, rates, descriptive data, identifiers, classifications, credit information, etc. Participants in financial transactions use this information to derive yields, valuations, variances, trends and correlations; they feed the raw material and other derived works into pricing models, calculation engines and analytical processes as the material and works travel from process to process among multiple global participants.

For many years, financial firms have been afflicted by "data anarchy." The deployment of personal computers has dispersed access, ownership and control of data, creating a mare's nest of data silos. Globalization and industry consolidation imposed new requirements, while the time frame for producing results was shrinking.   The data quality gap is an evolutionary result of years of mergers and internal realignments, exacerbated by business silos and inflexible technology architectures. The core requirement of continuity of business frequently prevents the proper unraveling and reconnecting of systems, processes and organizational environments. Instead, data management typically occurs on an ad-hoc and reactive basis. Workflow is not well defined and end-user computing abounds; data integration is largely point-to-point, performed tactically in response to immediate crises. Many firms still have no executive owner of data content and no governance structure to help with organizational alignment, funding challenges or battles over priorities. The concepts of metadata management, precise data definitions and semantic models are only now beginning to emerge as essential strategic objectives.[5]

The shortfalls in information management have traditionally been addressed as a kind of

---

because managers see the same markets as investors in addition to the detailed inner workings of their firms. Unfortunately, along with this additional information comes another information asymmetry, which managers may exploit to benefit themselves at the expense of investors.   See Stein (2003) for a discussion of the issues for internal capital markets.

[3] See Hu (1993) for a discussion of legal issues around information management for derivatives contracts.

[4] This section draws in part from the Committee to Establish the National Institute of Finance (CE-NIF) (2009a). In promoting the OFR, the CE-NIF produced a series of white papers: CE-NIF (2009a, 2009b, 2009c and 2009d).

[5] These issues are taken up below in a use case (#2) on knowledge representation of financial contracts and a related use case (#5) on automated reasoning. The Enterprise Data Management Council's work on a semantics repository is an example of an implementation in this area; see Bennett (2010).

operational risk.[6] It is no coincidence that operational risk audits often focus on information and accounting systems. Many of the problems engendered by data anarchy produce operational losses. The complexity of data processes inevitably gives rise to human errors, software bugs and breakdowns in systems and controls. Moreover, many principal-agent problems are caused or enabled by operational hazards (see Jarrow, 2008). For example, the failure in 1995 of Barings Bank to alleviate informational asymmetries by monitoring internal accounts led to large losses by their (rogue) trader in Singapore, Nick Leeson. It is well understood and standard practice in banking to verify traders' contractual exposures in the trading book via a properly segregated back office function; Barings' failure to do so was an invitation to internal fraud.   Another example is the recent controversy over missing documentation for many loans currently entering the foreclosure process. Much of the processing – at both origination and foreclosure – was delegated to specialty subcontractors, whose compensation depends significantly on the volume of deals booked. In the rush to handle deals, it appears that corners were cut; in some cases fraud may have occurred.[7] In general, many operational risk management practices evolved over the years as a patchwork of responses to specific incidents and organizational malfunctions and do not necessarily scale well. As the volume of data and information processed through financial institutions continues to grow, the system can benefit enormously from a rigorous and comprehensive analysis of its data operations.

Systemic risk and the OFR

The management of data and information is not just a technology problem at the level of individual institutions. As the current crisis reveals, it is a systemic issue as well. There is not yet a complete consensus on what is meant by, "systemic risk" (see Lo, 2009 for a discussion). One widely used definition is the "risk that an event will trigger a loss of confidence in a substantial portion of the financial system that is serious enough to have adverse consequences for the real economy;" see Taylor (2009). Nonetheless, there is broad agreement on some of the larger components of the term. In particular, it is generally agreed that systemic risk monitoring should at least capture the following:

- Forward looking risk sensitivities to stressful events, e.g. what would a 1% rise in yields mean for my portfolio?

- Margins, leverage, and capital for individual participants, relative to market liquidity, e.g. how large a liquidity shock could an institution absorb before defaulting?

- The contractual interconnectedness of investors and firms, e.g. if Lehman Bros. fails, how will that propagate to me?

- Concentration of exposures to common or correlated market factors, e.g. how many banks are simultaneously deeply exposed to California real estate? Alternatively, how large of a contingent exposure should a single institution be allowed to take on?

---

[6] Operational risk is typically defined as "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events" (see BIS, 2001).

[7] Problems arising from delegation of duties show up again in our use case (#4) on credit analysis.

In all of these dimensions, but especially regarding interconnectedness and deleveraging, feedback loops represent an important dynamic.

Apropos of concentration and connectedness, the case of AIG Financial Products (AIGFP) is instructive. AIGFP was able to sell default protection in the market for credit default swaps (CDSs). It was primarily insuring collateralized debt obligations (CDOs), effectively taking on the credit risk for many of the subprime mortgages originated at the height of the housing boom. Unfortunately, AIGFP concentrated risks by selling protection well beyond the capacity of its own capital should the default experience increase significantly, which it did in 2008.[8] The ultimate failure of AIGFP after the collapse of Lehman Brothers in the fall of 2008 ramified to other firms. Lehman's failure revealed how intensely interconnected a single large institution can be with rest of the financial network.[9]

Most of the data necessary to support these systemic risk metrics is encompassed by a sufficiently broad set of terms and conditions (T&C data) for individual financial contracts, together with historical data on key macroeconomic and market indicators, such as interest rates and security prices. T&C and market data are the inputs to the analytics software that can calculate risk exposures, stress tests and capital adequacy. Notably, these same crucial risk management activities are also of immediate relevance to financial firms in isolation (i.e. these are valuable not just for systemic risk calculations.) Captured at the transaction level, T&C data can inform questions on market liquidity and concentration. If the T&C data include counterparty references, then the interconnectedness of the contractual network can be reconstructed and examined. As an important benefit, T&C data have precise meanings, hammered out in the numerous security master agreements and thousands of instances of successful contractual relationships in the system.

To comply with all their legal, contractual, and reporting obligations, financial industry participants must understand all the ownership hierarchies and counterparty and supply-chain relationships. They must keep up with changes and corporate events occurring continually among the thousands of entities around the world. Because this is realistically achievable only via automation, the reference data must be generally available in strictly standardized formats. These data are linked to accounting, trade execution, clearing, settlement, valuation, portfolio management analysts, regulators and market authorities. The same information on legal entities and their corporate events is also relevant for monitoring systemic risk for many of the same reasons.

The *Workshop on Knowledge Representation and Information Management for Financial Risk Management* took place in the wake of the crisis and the passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act). Indeed, the presidential signing ceremony for the legislation occurred on July 21st, just as the workshop was beginning, to the applause of many in attendance. Sections 151-156 of the Dodd-Frank Act create an Office of Financial Research (OFR) with a mandate to monitor systemic risk. In particular, §154(b) of the Dodd-Frank Act

---

[8] See Ashcraft and Schuermann (2009) for a description of the mechanics of mortgage securitization. We argue in the first use case (#1) that a relatively simple data visualization exercise might have revealed the growing concentration of risk at AIGFP much earlier.

[9] See Bernanke (2009) for a description of AIG's interconnectedness with the system. The use case on living wills (#3) outlines a possible technique for safely excising a failing firm from the network.

requires that the OFR will contain a Data Center (OFR/DC) to manage data for the new agency. Under the terms of §154(b)(2)(A) of the Act, the OFR/DC must:

- Publish financial instrument reference data: information on the legal and contractual structure of financial instruments, such as prospectuses or master agreements, including data about the issuing entity and its adjustments based on corporate actions;

- Publish entity reference data: identifying and descriptive information, such as legal names and charter types for financial entities that participate in financial transactions or that are otherwise referenced in financial instruments;

- Publish formats and standards especially for reporting financial transaction and position data; and

Under §154(b)(2)(B)(iii) of the Act, the OFR/DC has the authority to:

- Collect positions and transactions data: i.e. terms and conditions for both new contracts (transactions) and the accumulated financial exposure on an entity's books (positions).

The Act further requires that the OFR/DC maintain data security and confidentiality.[10]

The OFR is not merely a data collection agency. In addition to the OFR/DC, the §154(c) of the Dodd-Frank Act also mandates that the new agency contain a Research and Analysis Center (OFR/RAC). The OFR/RAC must maintain computational and analytical resources to:

- Measure systemic risk;

- Monitor and report on changes in systemic risk;

- Evaluate institutional stress tests;

- Advise and support fellow regulatory agencies;

- Investigate system disruptions and institutional failures;

- Analyze the systemic risk implications of proposed policies; and

- Promote best practices in risk management.

To achieve these goals, the OFR will also need access to:

- Selected prices and related data: transaction prices and related data used in the valuation of positions, development of models and scenarios and the measurement of micro-prudential and macro-prudential exposures.

The combination of these data can resolve the fundamental questions of who (i.e. which specific legal entity) is obligated to pay how much to whom, on which future dates and under what contingencies.

---

[10] The final use case (#6) deals with issues of privacy and confidentiality.

Financial Context

FINANCIAL USE CASES

In the wake of the workshop, the reporters developed a group of six use cases to illustrate in concrete terms some of the potential for tools and techniques from computer and information science to play a valuable role in the financial field. A salient factor in selecting use cases was a desire to identify challenging application areas that would benefit from a research focus:

1. Visualization and analysis of a financial counterparty network

2. Knowledge representation of a financial contract

3. Implementation of a living will for a large financial firm

4. Fostering an ecosystem of credit analysis

5. Reasoning over financial contracts for completeness and integrity

6. Privacy and trust: multiparty sharing of confidential financial data

Each use case adheres to the following outline:

- CONTEXT AND OBJECTIVES: Summary of the financial need and forces impinging on possible solutions

- USERS AND STAKEHOLDERS: Key participants in the process

- USAGE SCENARIO: Step-by-step example of a possible solution implementation

1. Visualization and analysis of a financial counterparty network

Context and objectives

The counterparty network represents the entities (firms, individuals, organizations, contracts, etc.) and their interactions, as well as events and workflow within the network. The objective is to represent those features of a complex financial network that are crucial for understanding systemic risk(s). Visualization of the network provides a better understanding of the systemic (or macro-prudential) risk; this risk comprises those hazards arising from the linkages among financial firms. We note that macro-prudential risk is built upon, and must represent, the micro-prudential risks that are contained within the individual firms. A parsimonious representation of the network level macro-prudential risk would have the advantage of being simpler to visualize and comprehend.

The most important linkages and workflow in the counterparty network are the cash flow obligations defined in specific financial contracts that occur amongst the participants in the system. While the typical contract represents a binary relationship between two entities, multiparty contracts are also possible. Roles in a contract may be restricted to particular subclasses of entities, based on considerations such as exchange membership, legal jurisdiction, professional licensing, institutional charter type, etc. A non-exhaustive list of some of the key attributes that should be modeled is as follows:

- The number and size of individual legal entities (nodes) in the system.

- The number and size of contractual relationships between nodes. Special attention will be required to manage multiparty contractual relationships, since these are not isolated binary edges in the graph.

- Events in the underlying state space that can trigger payoff clauses in specific contracts. In financial terms, events are typically modeled as exogenous random forcing functions defined on a linear state space. Contractual payoffs are then explicit functions – either directly of the state variables or indirectly as functions of other contracts.

- Capitalization and liquidity buffers for each firm or legal entity. These buffers represent key constraints on agent behavior in the system. A firm is economically insolvent when its capital is exhausted, i.e. when the net present value of contractual liabilities exceeds that of assets. A firm is legally insolvent when its liquidity is exhausted, i.e. when current cash resources are inadequate for immediate contractual payment obligations.

There is a wide range of additional information beyond the previous base features, e.g. Securities and Exchange Commission (SEC) filings, geospatial data, jurisdictional details, etc. that may be identified to be relevant to specific scenarios.

A visualization technology should highlight the key risk attributes in the system, where the desired emphasis may be context dependent. For example, an analyst may wish to focus only on the largest firms or may wish to reveal the presence of broad-based, diffuse imbalances. Attention may focus only a certain subset of triggering contingencies or on a diverse range of stress scenarios. In short, an analyst using the system will likely require a variety of specialized views that can only be specified at run time, implying a need for a versatile and interactive user interface.

More generally, given the heterogeneity of both the questions and possible solutions, an

evolutionary approach to creating visualizations would potentially be more successful than a centrally designed solution. Specifically, this could be realized as an ecosystem of visualization providers, with many players adding value to the contributions of others in a mix of competitive and cooperative models. Such a development ecosystem could be facilitated via the Internet, with techniques for sharing both data – perhaps anonymized or sandboxed for confidentiality, if necessary – and computational/storage resources.

An important caveat is that there are important limitations to the effectiveness of visualization tools for data exploration and decision support in this context. Visualization analytics as described here are not a panacea for monitoring systemic risk, in part because not all systemic risks are readily measurable on the network graph of contractual connections among firms. For example, the run on the money market funds industry following the Reserve Fund's losses on Lehman debt in the fall of 2008 is better modeled as a discontinuity in investor behavior that hit the entire money-fund industry, rather than as a liquidity shock propagating through the edges of the contractual network.

Moreover, as with any data intensive solution, the quality of the outputs (e.g. policy decisions) is governed by the quality of the data inputs. For example, risky positions may not be properly disclosed to the monitoring system: both the then-CEO and then-CRO of AIG testified that they themselves were unaware that AIG's derivatives with Goldman Sachs were subject to margin calls, until Goldman Sachs made a $1.5 billion margin call. Visualization notwithstanding, an important side benefit of constructing the counterparty graph could be the ability to enforce better data integrity constraints, via double entry bookkeeping. If both AIG and Goldman Sachs had reported their derivatives contracts in a consistent format amenable to comparison, it might have been clear to AIG, Goldman and their regulators that these contracts were marked to market. If, in addition, both parties to a contract report mark-to-market or mark-to-model prices, these too can be checked for consistency. Where these "marks" (i.e. market or model settlement prices) come from is important. Counterparties might use tools to document how they mark contracts to market or mark contracts to model, although defining a standard representation for settlement prices is another research challenge.

Finally, we offer a caution in using visualizations for decision support in systemic risk monitoring. Images can be powerful, and there is therefore the danger that decision makers relying on visualization tools will labor under an illusion of omniscience. In the absence of other monitoring technologies, they may easily be blindsided by risks that are obscured by the limited scope or inadequate data foundations of their toolsets.

Users and stakeholders

The immediate user(s) of the application would consist of systemic risk supervisors who are examining the system to detect accumulating or incipient hazards revealed by the network of contractual obligations. In addition to hazards, they may also be interested in understanding the level of activity in the system, the degree of complexity of the network, etc. Additional stakeholders include the following:

- Regulators receiving risk reports from the analysts using the visualization tool.

- Regulated entities (and their employees, shareholders and investors) depicted as nodes in the network.

- Financial researchers and policymakers whose recommendations and decisions would be affected by the reports of the systemic risk supervisor using the tool.

Usage scenario

A typical use case of a network visualization tool would involve a systemic risk analyst who needs to produce a report on one or more questions of financial network fragility. She would either select an existing counterparty network (or create a new network) and specify criteria including the time period of interest. As suggested above, the following diverse aspects of analysis might inform her exploration and report:

- She may wish to focus only on the largest firms, or may wish to reveal the presence of broad-based, diffuse imbalances. The tool offers a way to filter the displayed graph based on node size, or other basic attributes.

- Attention may focus only a certain subset of triggering contingencies, or on a diverse range of stress scenarios. Again, the tool offers a way to indicate which stress scenarios are relevant to the analysis, filtering the rest of the edges from view.

- She may define clusters of institutions (e.g. all state-chartered insurance companies, all qualified thrift lenders or all futures commission merchants) to examine the attributes of the cluster as a whole, rather than the individual component firms. The tool offers a way to define criteria for various clusters. Given cluster definitions, the tool creates "meta-nodes" in the graph to represent aggregates of individual firms (and their contracts) satisfying the criteria.

- The tool enables navigation through the graph, drilling down to view the attributes of individual nodes and contracts. It is also possible to change viewpoints, scale, and resolution, and to re-center the view of the graph to highlight a specific node.

- An analyst may wish to see a snapshot of the network at one point in time or an evolution of the network over a period of days or weeks. For an evolution of the network, the tool provides a simple animation of sequential snapshots over the interval.

- The tool provides facilities for printing views, or saving them to disk in a variety of standard formats for later sharing or reuse.

2. Knowledge representation of a financial contract

Context and objectives

The valuation and risk analysis of individual securities, or of a portfolio comprising multiple securities, is a fundamental task. The strength and validity of the analysis relies upon the application of multiple analytical packages. Each such package of risk analysis models is typically complex, intricate and highly non-linear, and the fact there are often many such models reflects the reality that most valuation/analytic methodologies are not settled science. Some models will perform relatively well on certain asset classes, or for certain areas of the state space, but may do less well on others. The principle of comparative advantage suggests that each model will have some asset classes or state space where it will dominate all other competitors. In addition to the complexity and diversity of the models, there may be both incomplete specification of the underlying problem, as well as other sources of uncertainty. Incomplete information and uncertainty more reflect missing data, human error (e.g. invalid input or data entry errors), or software limitations (e.g. a parameter value or range is incorrect.)

More importantly, there is a fundamental tension between modeling for risk management and modeling for trading. Simplistically, risk management models might be based only on the "risk neutral" probability measure (i.e. probabilities implicit in information reflected in traded prices), while trading models could be seen as seeking "alpha" (i.e. profiting from discrepancies between risk neutral probabilities and the "true" probabilities). The simplistic view presents a false dichotomy, since proper risk management also needs to consider the possibility that the market's risk neutral measure is wrong, while alpha models cannot ignore risks: trading and risk management are intertwined. Unfortunately, this tension can be exploited: a fundamental problem in the financial crisis was that firms had sold insurance in excess of their ability to pay out should the insured outcomes occur. Examples include Citigroup's off-balance sheet holdings of highly rated subprime securities, AIG's derivatives, and the leveraged portfolio holdings of Bear Stearns and Lehman. Depending on the stakes and probabilities involved, this strategy can make sense for a trader who collects a bonus if the insured event(s) do not occur and is looking for another job if they do. The risk manager, however, should be concerned with making sure the firm survives, even if unanticipated events do occur. If risk managers ignore bad outcomes because they believe those events simply will not occur, it suggests they have been co-opted by traders. To prevent being trapped in a set of assumptions built up by the traders, risk managers should use a robust diversity of models to think outside the box that traders want to put them in.

Thus, for both technical and strategic reasons, the standard practice is to try several competing and complementary models and to reconcile any discrepancies in their output manually, using professional judgment. The objective of this use case is to mark up a financial contract with respect to a shared ontology so that multiple analytical packages (tools, software suites) may be applied. The markup will allow the analyst to compare the results (output) of various tools more easily. The markup may also assist the analyst in resolving discrepancies. Further, the resolution of discrepancies may in itself represent knowledge that can be encoded in the shared ontology.

For a very simple example, consider syntactic variation. One model may specify 5½% interest rate as 5.500, while another expects .05500. A more complex variation may involve semantics, e.g. one model may represent the tenor of a loan using start date and end date, while another may use start

date and days to maturity, etc. In a more sophisticated example, one model might accept an interest rate cap as a single security, while another might require that it be decomposed into a collection of individual interest rate options (so-called "caplets"). Much effort is involved in mapping security or portfolio input data to each analytical package, independently and manually, and validating the input data against the specification of each package (input schema). However, a set of input data that validates against the input schema of one package does not provide any guarantee (or even increase the likelihood) that it can be validated for any other package. Further, validation of the input data and input schema for each package is independent of, and typically may not contribute towards, the comparison of the output data (output schema) of the multiple packages or the resolution of inconsistencies. To summarize, current solutions are manual, laborious, error prone and not scalable.

A desirable solution would be to create a shared semantic representation of the securities in the portfolio and then generate valid model-specific exports (input data compliant with each input schema) that are tailored to each analytical package. This economizes on the effort that must be expended to apply multiple analytical packages. More important, the shared representation may simplify the task of sharing results (output data and output schema) across multiple packages and may aid in resolving discrepancies. An appropriate solution is an ontology (and accompanying knowledge representation and markup) for financial contracts. One benefit of such an approach is that the knowledge representation might be extended beyond simply a way of capturing the contractual terms and conditions, encompassing also the various cash flow patterns that are generated by the particular security types; see Brammertz, et al, (2009).

The process might even be reversed: instead of starting with the legal specification (terms and conditions for a particular financial instrument) and deriving the cash flow patterns, one could start with a financial engineering specification (desired cash flow patterns) and work back to find a workable legal specification that generates those cash flows. In this way, the knowledge base could become a financial engineering design tool.

Users and stakeholders

The primary user in this scenario is a desk quant or risk manager who needs to analyze a security or portfolio in several analytics packages, to gain assurance that his valuation or risk analysis is reliable.

Usage scenario

A trading desk quantitative analyst has a portfolio of short-term loans with embedded optionality that he must analyze in each of three separate software models (A, B and C) to control for "model risk." He has a knowledge base software tool to assist in this task.

- He begins by launching the knowledge base application and entering or importing the details (terms and conditions) of each contract.

- He invokes the data integrity checks built into the knowledge base to ensure that the descriptions of the securities are acceptable.

Financial Use Cases

- He exports the descriptions of the securities to the input schema for Model A. He starts Model A and imports the portfolio description into the software. He uses Model A to run a pre-programmed battery of standard tests and analytics, saving the results.

- He repeats the previous step for Model B and Model C.

- Given the three sets of results, he imports them into a spreadsheet to perform simple manipulations and comparisons. He writes up the results of this reconciliation.

3. Implementation of a "living will" for a large financial firm

Context and objectives

The Dodd-Frank Act, in §165(d)(1), requires large financial firms – both bank holding companies and non-bank financial firms – to develop "living wills." These are plans for the rapid and orderly resolution of the company in the event of financial distress. Among other things, living wills will include the following: descriptions of the firm's ownership structure and contractual obligations; information on risk mitigation provisions protecting the firm from its own nonbank subsidiaries; and identification of the firm's major counterparties and collateral pledges. These plans will be updated periodically and filed with regulators for review and approval.

The connection amongst living wills, knowledge representation and systemic risk is both an important public policy issue and an important computational challenge. Ideally, a financial firm would "collapse" gracefully – i.e. undergo a regulatory resolution process – over a single weekend to prevent the reality or uncertainty of real or possible losses from propagating through the counterparty network when business resumes on Monday morning. While the FDIC insurance fund would interpose its creditworthiness where necessary, the afflicted firm's shareholders would be wiped out, and the losses imposed on unsecured creditors.

Defining the elements of a living will as well as the mechanisms and protocol for implementing one is an important intellectual challenge, requiring insights from economics, finance, statistics and computer science. One key ingredient will be availability to regulators of data that are much more detailed (contract level terms and conditions, including counterparty identifiers), up to date (daily) and accurate (cross-checked to ensure validity) than ever before. It will also require algorithms for dividing the assets of the firm and assigning them to the various creditors – i.e. for cutting the graph. Part of this will involve simply identifying the counterparties. Then there will be the difficult question of assigning present values to the failing firm's various incoming and outgoing net financial claims. Finally, there will be the problem of assigning actual payoffs to the creditors. In a failure, the aggregate net outgoing claims will exceed the net incoming claims, so the former might be assigned a "score" or "priority in resolution," giving the claims a position in a pecking order.

Users and stakeholders

The immediate user(s) of the application would consist of:

- A resolution team working for a financial regulator who is charged with efficiently resolving a failing institution over a weekend.

- Risk managers and accounting specialists working for the failing firm.

Usage scenario

A typical usage of a resolution toolset would involve the following steps:

- The resolution team consults the living will plans previously submitted by the failing firm. This documentation provides context and guidance for the rest of the process.

- The resolution team extracts a listing of the firm's contracts from a data repository mandated by the living will process. This data extract unambiguously defines the counterparties for each contract and sufficient information to calculate a mark-to-market or mark-to-model net present value (NPV) for each contract. It also identifies any specific collateral and/or blanket liens against the assets of the failing firm.

- The accuracy of the data extract is confirmed by the internal experts at the failing firm.

- Based on seniority and collateral provisions, the toolset generates a pecking order of the creditors for all of the assets of the failing firm. Claims of shareholders and unsecured creditors fall to the bottom of the list.

- The toolset assesses net present values for each contract, a process requiring a full portfolio revaluation using preapproved pricing models. To the extent it is legally admissible to net the contractual obligations against a given counterparty, netting is applied.

- The toolset calculates the total resources available from positive-NPV exposures and the total obligations from negative-NPV exposures. Positive-NPV contracts will be assumed by a receivership corporation, which pays in cash to fund payouts on the negative-NPV contracts.

- The toolset calculates the distribution of cash resources and specific collateral through the pecking order to the firm's counterparties.

- If there is a shortfall of resources due to secured creditors, the resolution authority steps in to fund the discrepancy. Unsecured creditors and shareholders are not made whole. On the other hand, unsecured creditors and shareholders will be paid to the extent that excess resources are available after paying off secured creditors.

4. Fostering an ecosystem of credit analysis

Context and objectives

In the aftermath of the recent financial crisis, a number of commentators opined (in some cases self-servingly) that "no one saw this coming." In reality, some analysts did foresee the collapse of the mortgage market. The challenge is to know ex ante which analysts are correct. More realistically, what is needed is not perfect foresight, but a vigorous debate over assumptions, methods of analysis and conclusions about creditworthiness.

In practice, much of the formal credit analysis of securities is provided by credit rating agencies. In the U.S., regulators have designated 10 firms as "nationally recognized statistical rating organizations" (NRSROs), but in reality three firms control the vast majority of this business. There are conflicts of interest in this game, since most debt instruments are typically "issuer-pay," meaning the NRSROs are compensated by the issuers of the securities, who naturally prefer a higher rating. These conflicts are exacerbated by the fact that NRSRO ratings are conferred special regulatory status in some cases (e.g. only NRSROs can confer the prized "investment grade" status). The Dodd-Frank Act includes requirements for government studies on ways to improve agency ratings and the role of the NRSROs. An easy first step toward expanding the discussion of particular issuers would be to remove the special status conferred on NRSROs, but this alone would not provide any incentive for issuers to pay for unswervingly honest ratings. It is quite possible that credit analysis is an inherent component of the investors' due diligence process, and that this task can never be productively delegated to third-party rating agencies due to conflicts of interest (or what economists call agency costs). Note that the problem here is not that the agencies' analytics were opaque. Indeed, these models were typically widely shared with industry participants (especially issuers) to facilitate the construction of collateral pools that could satisfy an issuer's rating target. Rather, it is the securities themselves that are difficult to understand. The problem is a system that encourages investors to delegate their pre-purchase diligence to third party agencies that may not be properly incented to provide thorough analysis.

Regardless of the conflicts associated with the issuer-pay model, it seems likely that one key reason for delegating credit analysis to third-party providers is that the cost and complexity of ingesting the requisite data are prohibitive for anyone except specialists. These are technological constraints that could be relaxed. For example, for complex securities such as tranched mortgage-backed securities (MBSs), a diligent analyst must consider a large amount of information on the loans in the underlying pool. Loan level attributes such as property size, geographic location, borrower credit rating, existence of secondary liens, etc. are all relevant to rating an MBS. With current technologies, ingesting this detailed information is per se a technical challenge that significantly raises the cost of creating a credit rating and therefore discourages analysis. A standardized markup for expressing the vital attributes of a transaction in a machine-digestible format would greatly facilitate the process. Issuers are currently required to publish these details in a deal prospectus and other offering documents, but these are lengthy tomes of legalese that do not invite automated risk analysis. If the cost of doing analysis were lower, more analysts might step forward, without having to be paid by the issuers. Similarly, analysts and rating agencies might adhere to a standardized taxonomy to describe their assumptions and methodologies, making it easier to compare recommendations across analysts and over time.

Users and stakeholders

The immediate user(s) of the application would consist of:

- investors interested in the creditworthiness of current or potential investments

- regulators interested in the quality of portfolios relative to capital and other prudential limits

- issuers who want their securities analyzed

While the set of users interacting directly with the system is limited to these immediate producers and consumers of the deal descriptions, the indirect beneficiaries extend to a wide range of market participants. In the case of mortgage securities, for example, these would include:

- homeowners, whose house prices are affected by the flow of investment into the housing sector

- housing market participants (mortgage brokers, real-estate agents, appraisers, homebuilders, etc.), whose business is affected by the flow of investment into the housing sector

- security issuers in other economic sectors who must compete with MBSs for the aggregate flow of funds

Usage scenario

A typical usage of a deal-reporting system would involve both an issuer who creates a deal and an analyst who appraises it.

Issuer perspective:

- The issuer creates a traditional prose prospectus. He extracts from the prospectus (and other sources) the vital details required by a reporting standard and creates a markup and tags those details according to the data markup and exchange standard.

- He posts the marked up document, along with the traditional prose prospectus, either in a central repository or on the issuer's web site.

- Access to these documents might be limited to authenticated users (such as eligible investors).

Analyst's perspective:

- She downloads the marked up document and maps the tagged information into a variety of analytical software packages that can provide a comprehensive "work-up" on the quality of the security; she then runs those analytics.

- The analysis will typically consider the performance of the security under a variety of exogenous stresses and scenarios, to understand where its strengths, weaknesses and trade-offs lie. Attention may focus on a small set of specific contingencies, a diverse range of stress scenarios, or a random (but plausible) distribution of exogenous factors evolving over time.

- She performs an automated textual analysis of the traditional prospectus, looking for special features that might not be captured by the structured schema and inconsistencies between the structured data and the textual presentation.

- She maps the tagged security information, along with the other intermediate results described above, into an analysis package that produces one or more scalar credit ratings, indicating the estimated probability of default on the security at various time horizons.

Having collected the necessary information, the analyst produces a written research report, describing the entire process of analysis, along with her conclusions and recommendation.

5. Reasoning over financial contracts for completeness and integrity

Context and objectives

Even a "simple" 30-year fixed rate mortgage can be a remarkably complicated financial instrument. At settlement, the buyer and seller will sign a number of important documents (for example, see Fannie Mae, 2010), most of which they will not have reviewed in advance, and which will be described only very briefly by the closing attorney. This may be reasonable diligence in the case of a plain vanilla mortgage conforming to Fannie/Freddie underwriting standards, where the loan contract is a standard document that has been vetted in tens of thousands of previous transactions over time. For more complicated loans – involving, for example, custom prepayment penalties, hybrid adjustable rates, negative amortization, interest rate caps and floors, teaser rates, pick-a-payment options, etc. – this reliance on "community vetting" or "crowd-sourced diligence" becomes more questionable.

Now consider the case of a collateralized debt obligation squared (CDO2), a product innovation in the mortgage securitization market. A CDO is a vehicle that pools a set of MBSs and issues tranches of debt securities against them, where the tranches are ranked by creditor priority, so that junior tranches absorb the blow of credit losses on the underlying MBSs first, before exposing any of the more senior tranches. A CDO2 is a second securitization layer that pools a set of underlying CDOs, again creating prioritized tranches of debt to finance the pool. The underlying MBSs at the first securitization layer will each typically contain hundreds of mortgages. As a back-of-the-envelope approximation for the documentation page count, if each securitization layer (CDO2 > CDO > MBS > mortgage) pools roughly 100 contracts of its underlying type, and there are 100 pages of documentation for an average mortgage, then a typical CDO2 would ultimately involve approximately 100 million pages of legalese. This is clearly beyond the capacity of ordinary human analysis, so diligence must either be forgone or automated. Indeed, it seems likely that many investors fail to read the fine print on any of the offering documents, preferring instead to take a bet based on issuer reputation and views about prospective market movements. In fact, the prospectus for even a single layer in this securitization chain can run to many hundreds of pages.

One important task when analyzing a contract is to verify that the cash flow streams it actually obligates for the counterparties are precisely those cash flows that the counterparties think they are agreeing to. Ideally, a contract would be represented in such a way that its cash flows can be simulated under many different model assumptions. It should be unambiguously feasible to calculate risk measures from the contract representation. (In some cases, this requires the contract to specify fairly intricate mathematics to determine cash flows.) For many contracts, especially so-called "structured" securities, cash flow streams are contingent on the future behavior of one or more state variables outside of the control of the immediate parties to the deal. For example, the payoffs from issuer to investor on a mezzanine tranche of a CDO on adjustable rate mortgages depend both on the index interest rate as well as on the default behavior of individual mortgage borrowers at the bottom of the securitization hierarchy. With so many pages of legalese, it is a challenge even to ensure that cash flows have been specified (exactly once) under all significant contingencies. However, it might be possible to structure and tag the contract to identify all relevant state variables and their possible combinations, as well as those clauses in the contract that specify cash flows. If so, it should be possible to build a reasoner to walk through the tagged

representation of the contract, verifying that it specified (exactly once) a valid cash flow rule for each possible future state at each future date. Given such extensive structuring, other simple validity checks and reasoning rules are readily imagined: Are all required disclosures included? Are definitions provided for all terms of art? Is a party to the contract listed in an inappropriate role? An analyst might also wish to highlight clauses that affect or involve specific contingencies to ensure that the contract handles these as expected.

On the issuer's side, if the formal structure is sufficiently expressive, it might even be possible for him to generate programmatically large chunks of the legal language, by including, excluding or overriding various standard contractual clauses as appropriate. Note that this is similar to the role served by standard master agreements, which are typically imported into a contract by reference, with various clauses then overridden or amended as required. All of this would require a fairly general contractual language for specifying state variables and their contingencies, as well as legal clauses that specify cash flows relative to those state variables. Another practical outcome of a sufficiently expressive language might be the ability, perhaps via reasoning or pattern matching, to benchmark a complex contract relative to a similar but well defined contract. This would leave some residual or "basis" risk that is not carefully specified, and dealing with this basis risk would then be a task left for the risk managers. An advantage of a benchmarking technology is that firms could at least be required to declare a similar benchmark for a complex contract when a complete cash-flow definition is not provided.

Users and stakeholders

The immediate users of the system are the following:

- the issuer, who must tag the contract and other relevant offering documents with all of the necessary structural information, and

- the prospective (or actual) investor, who wishes to verify the integrity of the agreement.

Usage scenario

In this use case, we imagine that the issuer has already tagged a contract, perhaps via a software tool that allows him to select state variables and typical contingent cash flows. Such an issuer tool then programmatically generates a skeleton of a contract populated with boilerplate and standard clauses describing the contingent cash flows. The issuer then amends and modifies this skeleton until it forms a full-blown contract and associated offering documents.

The attention here is on the prospective investor, who wishes to verify the agreement. From her perspective, the process involves:

- She receives the contract, perhaps as a download from the issuer's Internet site.

- She loads the contract into a verification tool on her local machine. The contract embeds a reference to the specification version that was used to tag its structure. The tool uses this information to load the correct specification version as well.

- The investor first checks the contract for simple completeness. Are cash flow obligations defined for all future values of all the state variables? Are all required clauses present in

the contract? Are any forbidden clauses present? Note that each investor may have idiosyncratic rules about which clauses are required or forbidden.

- The investor is also interested in whether her requested provision was included, that her payments should never exceed 10%, regardless of the value of the floating rate interest index. She asks the tool to highlight (in a garish color) all clauses involving the floating rate payout function.

Having thus verified that the basic parameters of the contract are satisfactory, she sits down to read through the document once from start to finish, in hopes of moving soon to settlement.

6. Privacy and trust: multiparty sharing of confidential financial data

Context and objectives

When the SEC or the Commodity Futures Trading Commission (CFTC) investigates adverse events in the financial market, they may demand confidential information from firms about themselves and their customers. This information includes account numbers, names on the accounts and data on individual transactions (execution time stamp, account number, price, quantity, broker, clearing firm, broker time stamp and trade ID for matching buyer and seller). Such information is typically considered "business confidential," and by law should be protected from public disclosure.

On the other hand, in many cases the public generally would benefit from disclosure. We consider a potential recommendation regarding the public reporting of trades in regulated securities markets as follows:

- Pre-trade transparency, in the sense of making all bids and offers publicly available.

- A central limit order book, in the sense that customer A and customer B can post bids and offers and trade with each other if their prices are compatible with a trade.

- Post-trade transparency in that the price and quantity (and perhaps additional details) of each trade are revealed publicly.

- Single price auctions for the determination of settlement prices. All traders, both customers and bankers, should be able to participate in such single-price auctions. Current practice in many markets, especially fixed income, is to use a poll of dealers and this practice is subject to abuse by the dealers.

In the case of the SEC, for example, the law requires publicly traded firms (which are within the SEC's ambit) to reveal publicly all material information about themselves. Paradoxically, the SEC itself would typically be disallowed from revealing that same information. The situation is complicated, naturally, because disclosure would typically also generate wealth transfers (i.e. specific winners and losers), even if it is a positive-sum transaction overall. For example, if a publicly traded bank is on the FDIC's problem list, disclosure of this fact per the "material information" rule might make the markets more efficient, but it would also imply an immediate drop in the share value, effectively transferring wealth away from the bank's stockholders.

Public reporting as described above would clearly allow the construction of a log or history of individual traders. There are clearly multiple arguments, both for and against disclosure. Legally, some argue that there is a fundamental property right in trading data, and its owners may therefore maintain confidentiality if they choose. Disclosure of trades or positions might allow reverse engineering of trading strategies or other trade secrets. This could entice bucket shops or front-runners into the market, discouraging investment in otherwise profitable trading technologies. Or it might simply encourage complexity, as traders work to hide their activities in a mare's nest of accounts, each holding only a fraction of the overall strategy. On the other hand, in some cases, trade secrets may simply be protecting the status of a middle man who brokers a trade but adds no value to the net transaction. Disintermediating (i.e. circumventing the middle man) might in this case be economically preferable but legally questionable.

In some cases, the regulators attempt a balancing act. For example, the CFTC attempts something

like K-anonymity by revealing information only after aggregating traders into groups of accounts large enough to prevent analysts from reverse engineering the behavior of any one trader. Alternatively, delayed disclosure (e.g. until after a trade settles) might in some cases empower analysts without disadvantaging the trader(s) involved.

All of these issues will have special importance for the newly created OFR, which will have access to more and better data than current regulatory agencies. Given that certain data must legally be kept confidential, there is a strong case for such data being collected and protected by an agency that has the expertise in data security. Regulatory employees without special training in data security will likely not be reliable stewards for the large volumes of confidential data the OFR will handle. Note that the OFR will manage not only transaction and position data but also instrument reference data describing the contracts themselves. In some cases, these contracts will directly represent confidential trade secrets – e.g. how to set up and price a specialized deal – and this too will need protection. The fact that the OFR will need this information in one place (a potential single point of failure) further raises the security stakes.

The benefits of an OFR to centralize and standardize the data are potentially enormous. For example, the CFTC and the SEC are currently completing a joint study of the "flash crash" of May 6, 2010. Had the OFR existed, the post-event analysis would have taken one day or less. The types of issues being studied require reconstructing from audit trails which traders were buying and selling stocks, futures contracts on stocks and perhaps even options on stocks. To do this effectively requires matching the ticker tape with the identities of traders on both sides of all trades and then merging this with data concerning the identities and types of the trading accounts. The CFTC obtains reasonably good audit trail data from the CME, as is evident from the initial report of the joint study, but the SEC does not have good data on the identities of traders. Currently, the SEC is mostly limited to tick data without identities of individual accounts. Obtaining such data from dozens of venues in different time zones with local reporting conventions will be an enormous task.

To learn whether operational factors were important in the flash crash, it is not only necessary to have data on trades but also data on other actions, particularly the placement and cancelation of orders. For example, some have argued that traders may have jammed the order handling system with meaningless orders, similar to a denial-of-service attack, or that various systems at stock exchanges were operating more slowly than they should have, as if there was a malfunction or an overload. To explore these various theories, one also needs data on margin calls, the amount of equity in various trading accounts, and particular counterparty exposures. In the stock market crash of 1987, important issues included the disintermediation of the futures exchanges by traders transferring positions among themselves, logjams in the payments systems, the difference between five day settlement in some markets and next day settlement in other markets, the role of short-sale constraints in preventing index arbitrage from functioning effectively, the role of bankruptcy of a clearing firm for floor traders reducing liquidity in the options markets, the inability of the ticker to keep up with trading and the failure of NASDAQ dealers to answer their telephones.

In January 2008, a rogue trader scandal forced the French bank Societe Generale to liquidate tens of billions of Euros in equity futures. The problem surfaced on a very tumultuous day in the world's equity markets, during which the Fed made a large emergency cut in interest rates. Studies of this incident were inadequate. Had the OFR existed to force better data standards and good reporting, the rogue trading probably would not have happened in the first place. The rogue allegedly concealed his positions from the firm by entering offsetting trades which were classified as errors

and later reversed. Detailed reporting of clean data would have revealed this activity to regulators much earlier.

Users and stakeholders

It is apparent from the foregoing context discussion that a complex interplay exists between the legal, economic and security forces at play here. To keep matters simple, this use case will focus on a relatively narrow aspect of the problem – accessing data from a number of diverse trading venues (e.g. stock exchanges) to the OFR. The message bundles consist of the all daily trading activity on the exchanges, including trade times, amounts, prices, counterparty details, etc.

The immediate user(s) of the application would consist of the following:

- exchange IT staff from multiple trading venues, charged with reporting the data

- OFR IT staff, responsible for receiving it

- a malicious user (e.g. a terrorist), trying to disrupt the process

Usage scenario

The usage scenario is straightforward: the exchanges report their trade data, and the OFR confirms receipt, while the malicious user attempts (unsuccessfully) to prevent this. The key issue is whether it is preferable from a security point of view to centralize the reported data in a single repository at the OFR, which would focus security resources efficiently but create a single point of failure in the event of a security breach. Alternatively, the data could be physically federated at the exchanges and accessed remotely by the OFR.

SMALL GROUP DISCUSSIONS – KNOWLEDGE REPRESENTATION AND INFORMATION MANAGEMENT

The goal of the breakout groups was to bring computer and information scientists together with economics and finance researchers and practitioners to brainstorm promising research avenues at the intersection of the disciplines. The groups ran on two parallel tracks, with roughly half of the workshop attendees participating in each breakout group to facilitate input from everyone. Each breakout group was chaired by a group secretary.

The six breakout sessions on the second day focused on a variety of techniques and approaches from computer and information science as possible implementation solutions:

1. Knowledge representation frameworks (ontologies, schemas, models, formal logics) to describe complex financial instruments – *Andrea Cali*

2. Managing risk models: schema mapping, data exchange, and model comparison and reliability – *Dennis Shasha*

3. Languages (operators and rules) for specifying constraints, mappings, and policies governing financial instruments – *Leora Morgenstern*

4. Financial networks, agent-based simulation, and architectures for large-scale computation – *Michael Wellman*

5. Data integrity, data quality and operational risk – *H. V. Jagadish*

6. Privacy, confidentiality, security and trust in managing financial data – *Benjamin Grosof*

The chairs facilitated the discussions and took the notes that formed the basis for the reports that follow. Each report follows the same broad outline:

- OVERVIEW: Summary of the goals and topics for the session.

- FINANCIAL AND RISK-MANAGEMENT CHALLENGES: What were the interesting problems from the viewpoint of a finance researcher or economist that were identified?

- COMPUTER AND INFORMATION SCIENCE CHALLENGES: What interesting CS challenges need to be addressed to solve the previous problems?

Note that these discussions are not intended to be read sequentially or in any particular order. Each breakout group discussed an independent nexus of topics and comprised a unique combination of participants. The sections below may be enjoyed "à la carte."

1. Knowledge representation frameworks (ontologies, schemas, models, formal logics) to describe complex financial instruments

Overview

This session focused on envisioning new lines of research in modeling financial instruments in a formal way, which would be suitable for automated analysis. This is crucial to forecast systemic risk. At the moment, a good part of the information regarding financial contracts is specified in natural language documents. This leads to potential and sometimes intentional obfuscation of key information. Automated deduction about financial contracts in real world financial scenarios is still far from being a reality. The goal will be to specify technical details so as to build a solid mathematical and formal framework for applications and investigations. A decision was made to tackle the problem of understanding and reasoning about financial markets by thinking about how to model and reason about financial contracts. Such a bottom-up approach for predicting systemic risk and systemic phenomena appeared to be well suited for several macroeconomic scenarios. As in many computer simulations, predictions can be made by simulating the building blocks of the large-scale phenomena; the building block here is certainly the financial contract.

Financial and risk-management challenges

One of the first issues to consider is where data come from, and their level of detail. It is evident that, apart from reasoning and prediction techniques, we need to have a consistent and reliable data set on which to reason. On top of the data, meta-data, often in terms of rules, are needed; these constitute the ontology. A problem well known in the research area of ontologies is the potential mismatch between data objects and real-world objects; the same problem exists at meta-data level. In more practical terms, this is a data quality issue that needs to be addressed case by case.[11] This problem links to that of mislabeling, that is, using (intentionally or not) a wrong term for a certain concept or notion; addressing this issue could be useful to tackle the not-so-infrequent deliberate obfuscation in financial contracts.

One key tool for evaluating the results of future research is a good set of case studies. An interesting case study, on which everyone agreed in the breakout session, is that of mortgage contracts (or better, mortgage-loan contracts). Mortgages have played a central role in the past crisis, as regulators and governments were unaware of certain facts in the financial market (for instance, second mortgages). More crucial case studies are to be considered during the future research on the topic.

One central issue is whether current off-the-shelf knowledge representation formalisms and languages suffice for modeling complex scenarios such as the financial markets. It is important to notice here that most such formalisms assume a representation of the world in terms of sets of objects (also called classes, entities and concepts) and relations among sets of objects (also called associations, relationships and roles). This is an assumption and not an established truth. Most of

---

[11] For an introduction to ontologies, see Gómez-Pérez, Fernández-López, and Corcho (2004) or Staab and Studer (2004).

the cases will be treated under this assumption, which is reasonable in most cases that were foreseen in this session. On the other hand, it is evident that current models cannot be used right away and will need to be extended (more on this later in this report).

Computer and information science challenges

Representation

Generally speaking, it was agreed that the representation of financial contracts as the building block of financial markets should be done by using formal theories. Formal logics should be the core mathematical tool, but this depends on the scenario.[12] Certainly, there need to be several extensions and variants to model, even in the simplest contracts. As in every attempt of modeling real-world scenarios, the choice of terms is crucial. In general, there is no agreement on terms in different contracts, markets, or countries. For instance, the term "borrower" has different meanings in different contexts. A shared vocabulary, with links between related terms, is necessary; two terms can be related if they are synonyms, if one is more general than the other, if they denote sets of objects that are disjoint and so on. Several off-the-shelf ontologies, equipped with reasoning services, are nowadays available for the basic task of representing correspondences among terms. The issue of the meaning of terms is of utmost importance when using data coming from heterogeneous and unreliable sources, such as web sources. While web data have been successfully extracted and analyzed for financial forecasting purposes, the web context inevitably has errors and imprecision, and these have to be addressed in a sound way. The key notion here is precision and the level of detail to be considered.

Ontology matching and merging

The heterogeneity that is expected in most financial scenarios requires that different ontologies, independently developed, are merged into larger ones. This is a well-known problem (hardly solved in general), which can be very complex, as we will probably need formalisms with high expressive power in order to model financial instruments. Different organizations and markets will have different vocabularies and relationships among terms, but terms in different ontologies are likely to overlap significantly, thus creating the need for a unified representation. In the integration/merging scenario, we will also face different needs and perspectives in the merging itself. The perspective of regulators, for instance, is different from that of a third party. This is likely to be represented with metadata which will be taken into account during the merging. It is important to note that the overall integration needs to be based on atomic operations on atomic concepts. At the same time, several concepts that need to be distinct in one ontology might be lumped together in another one. This would require a sort of reverse engineering on the lumped information, for which we will need a formal framework, as declarative as possible. Finally, we observe that we will need to model and manage the evolution of ontologies, according to the changes in the real world setting (regulations, facts, behaviors and so on).

---

[12] Description logics encompass a fairly broad class of formal logics with applicability to databases. See, for example, Baader, Horrocks, and Sattler (2004), Borgida, Lenzerini and Rosati (2002), and Baader, et al (2002), and esp. Nardi and Brachman (2002).

Reasoning, rules and complexity

It is immediately seen that nothing is static in financial markets; at the same time, most knowledge representations are based on static models of knowledge, which are able to represent a snapshot of information but not its evolution. Certainly, modeling financial contracts needs to take time into account and the state of contracts at any given time; for instance, a mortgage-loan contract evolves over time as agreed by the terms. Rules have been employed in knowledge representation to infer new (static) information; rules can also be used to model the evolution of financial scenarios and the behavior of different parties of the contracts over time.

There is a large corpus of research on rules and reasoning, applied to several contexts. In general, even in rule-based formalisms that are not too expressive, basic reasoning services are often undecidable (i.e. reasoning might not terminate). Also, decidable formalisms are often intractable, that is, their computational worst-case complexity is too high in terms of space and time. In order to achieve decidability and in some cases tractability, several variants and restrictions of common rule languages have been proposed and studied. As for our context of financial markets, it is likely that the size of the input data (and possibly of the ontology itself in terms of rules) is large; therefore, tractability will be crucial if we want to devise effective and efficient prediction and analysis tools.

Reasoning can be applied not only to predict the evolution of the state of affairs in a contract, but also to compare contracts. If we had a way to compare contracts in a formal and automated way, a registry of financial products would be possible, where the characteristics of contracts would be clear. In the presence of standards Governments could, for example, enforce additional charges depending on the risk associated to different products.

Workflows are another well-developed formalism to represent systems that evolve over time. Workflow operators, broadly and thoroughly studied in the literature, are definitely useful as rules to model financial contracts. We should not forget that here we will have to face a trade-off between transparency (essential to avoid, for instance, obfuscation) and privacy of the parties involved in the contract.

Finally, it is obvious that rules have to be sent to automated reasoners in the form of software tools; at the same time, sets of rules representing ontologies have to be somehow human-readable, directly or indirectly and a domain expert, not necessarily knowledgeable in the technical aspects of reasoning, has to be able to design behaviors and ontologies.

Approximation and probabilistic reasoning

Every financial contract is in fact a promise. This is indeed the only fact: all the rest is inherently uncertain. Innumerable examples of this have been seen in the recent history of financial institutions. For instance, whether the government is going to bail a bond issuer out is an uncertain fact, the probability of which is difficult to evaluate. A formal representation of uncertainty and imprecision is surely necessary in our context of financial markets. Several formalisms have been developed, which are able to capture probabilistic information. At the same time, quantitative information (such as prices) will have to be modeled. Representing uncertain numerical values in financial contracts represents a fundamental challenge in our research.

Risk evaluation

Generally speaking, evaluating the risk associated with a certain asset will amount to querying a

suitable knowledge base in the presence of data regarding the asset itself and other related data. Querying ontologies in the presence of uncertainty will be one of the main challenges of this research group. This will also be applicable to systemic risk. This task is highly complex in most cases. For instance, a regulator might want to know how liquid certain instruments are; this will need an understanding and reasoning on business processes, and the information will be inherently uncertain. Timeliness and quality of data will be essential to ensure an effective and valuable evaluation of the risk. Moreover, the risk has to be properly represented; the uncertainty can reside at the mere data level and also at the rule level, and this needs to be formally represented. A crucial aspect in risk evaluation seems to be completeness, that is, the fact that no relevant information is missing from the model. While this seems obvious, it is also hard to evaluate the completeness of a certain representation in general. For now, we might conclude that this is a task to be carried out by finance experts. Therefore, methodologies need to be developed and studied in order to guarantee the most complete representation; if relevant aspects are missing, the whole risk evaluation might be useless. Finally, the evaluation of risk, needless to say, should be available to finance professionals and experts, and not necessarily to IT experts. This opens the way to research on suitable human readable representations of data, rules and results of risk evaluation.

2. Managing risk models: schema mapping, data exchange, and model comparison and reliability

Overview

Currently, there is a paucity of global disaggregated data. Different financial institutions represent data differently. There is no global view. Data aggregators like Bloomberg do provide data, but it is expensive, allows only restrictive access, and is often incomplete (e.g. it doesn't fully describe bonds). Issues include determining how to change incentives to make data available, which data to access, and how to manage historical data.

Financial and risk-management challenges

Research questions:

- Incentives (economics research question): data has an economic value. Which economic incentives can be given to data owners/aggregators to make the data more widely available? Old economic data has a lower economic value than newer data, so perhaps incentives can be provided to make older data more widely available. Can we imagine a situation where different data types may have different release dates? How can agency fiat authority be combined with incentives to make data more widely available?

- Useful data (economics research question): Determine which data should be gathered and at which level of granularity. Regulators will have fiat authority and therefore can ask for certain data types to be given fixed tags. The research question is which data should be tagged. Answering this question will entail showing the value of that data.

- Identifiers (economics/computer science): Currently, a single identifier (such as CUSIPS for short term commercial loans) can be used for several instruments and sometimes multiple identifiers refer to the same instrument. The research need is to determine which objects should have identifiers, to track identifiers across mergers and separations, to cross-reference among different identifier standards and to include checksums in identifiers to catch data entry errors. The computer science aspect is to define a protocol for the evolution of identifiers. (See Kent, 1991, for an early discussion of the issues.)

- Simulation (economics/computer science): How to simulate loans/hedges/pricing models etc. under normal business cycles and rare event business cycles. (See Jobst, Mitra and Zenlos, 2006, for example.)

Computer and information science challenges

Research questions:

- Global synthesis (economics/computer science): Connect data sets across institutions given that the semantics of fields differ. Ontologies and schema matching techniques should help here, as would sampling and estimation to accommodate fields having overlapping but distinct meanings. The state of the art includes tools to help a database designer create

precise matchings and generating a semantic matching with or without validation is important here.[13]

- New data type management (computer science/software engineering): Because finance is a dynamic field, new instruments are continually being invented. Similarly, as financial analysis advances, new data types may be needed. For this reason, regulators may demand that new data types be provided. The research issue is to coordinate the provision of the data types: the development of a common vocabulary and perhaps a language to describe the data types, the specification of the timeliness requirements (e.g. how much time between a trade and the production of data), the tracking of development progress among different providers and the integration of the data from different data providers. This involves both standards and software.

- Time (computer science): Data changes, instruments change, and analysis programs (e.g. risk management) change. It is sometimes necessary to recreate the data and software that was present at a previous time in order to study historical events. A research issue is to make this possible in an intuitive and efficient way. A second research issue is to create time series even as instruments change. Snodgrass and Jensen's work on bi-temporal databases should help here (see Jensen and Snodgrass, 1999, and Snodgrass, 1999).

---

[13] The issues of schema matching are covered in a series of papers by Bernstein (2003), Bernstein and Rahm (2000), Madhavan, Bernstein, Domingos and Halevy (2002), and Pottinger and Bernstein (2008). Flood (2009) discusses some of the issues in the context of finance.

3. Languages (operators and rules) for specifying constraints, mappings and policies governing financial instruments

Overview

Financial instruments are often complex, arcane and unwieldy. For example, a mortgage may have 1000 features; a credit default swap contract may run 200 pages. Moreover, there may be multiple representations for the same class of financial instruments, which are not necessarily mutually consistent.

Representing and reasoning with these financial instruments – especially with regard to specifying constraints, mappings, and policies, and particularly when they use different representational schemes – is therefore a significant research challenge. In our breakout session, we discussed several research problems engendered by this overarching challenge, and when possible, suggested an available technology or approach that has the potential to solve these problems. These research problems are presented below.

Financial and risk-management challenges

Although we know from anecdotal evidence that financial instruments and policies are large and complex, a repository of completely and unambiguously specified exemplars that have been developed for the research community does not exist. As a first step, we suggest creating a body of such exemplars.

This would entail the following steps:

- First, creating the repository (e.g. on a shared wiki space), with fully specified and unambiguous exemplars of financial instruments, such as ordinary shares, or written call options and of financial policies, such as those governing living wills and bankruptcy;

- Second, developing tutorials on the wiki that could be used by novices and experts to learn about these financial instruments and policies;

- Third, developing a scheme for annotating such examples;

- Fourth, constructing an adequately expressive representation for these examples, and formalizing the examples in that representation;

It is important to include in this repository both typical examples – e.g. of a mortgage default or of a small bank failure – and boundary cases – e.g. of Lehmann Brothers (a large bank failure that was allowed to happen) and Citigroup (a potentially large bank failure that was not allowed to materialize but instead resulted in a bailout by the U.S. government). It is also important to represent associated legal processes, such as the foreclosure of a house or a federal bailout.

A related research problem is the representation of linkages among financial entities, financial instruments, counterparties and obligors. Our breakout session recommended first focusing on such representation for the repository exemplars.

Computer and information science challenges

Mapping between ontologies and models (related to data integration)

Because there are multiple ontologies in the financial community, it is important to consider how we can map between different ontologies or models. For example, there are at least 7 different ontologies/models for financial instruments, including the ISO 20022 data dictionary, the ISDA (International Swaps and Derivatives Association) model, the EDMC Semantic Repository, the FISD FIX (Financial Information Exchange) protocol, FpML (the Financial Products Markup Language) and EFAMA (the European Fund and Asset Management Association). It is important to be able to characterize how concepts in these ontologies and models interrelate.

This problem arises in its simplest form when different terms are used in different ontologies to refer to the same entity. For example ISO 20022 and the EDM council may have different ways to represent a mortgage loan. If the same underlying structure is used to represent a mortgage, then the mapping is a simple renaming. If the underlying structure is not the same, the mapping must take into account how the mortgage components of one standard map into the mortgage components of another. The mapping becomes more difficult when the concepts are related but not identical, or when there are fundamentally different concepts involved. For example, one ontology may represent the processes and events related to a mortgage, such as defaulting on a mortgage or foreclosing a home.

It may be especially worthwhile to determine if it is possible to define concepts from different ontologies in terms of a common set of primitives and a common set of ontological commitments. For example, it may be possible to define mortgages in terms of various loan-related primitives and a set of events and processes. Often, this process facilitates the development of a bridging vocabulary between several ontologies.

It is also worthwhile to determine if ontologies and vocabularies developed for more general areas of study can be re-used to formalize financial information. For example, MATLAB may be useful to describe financial phenomena. Using MATLAB in such a way, however, would entail solving this mapping problem (as well as dealing with issues of scale, since MATLAB is intended for use on smaller problems).

Collaborative architecture

Because so many organizations are involved in the development of financial instruments and protocols, and with technologies to support them, it is important to develop a collaborative architecture that facilitates synergy between different groups.

Such architectures will need to support partly open data, open source data and open models. It will also need to support multiple layers, on the meta, conceptual, semantic and physical levels. Tools will be necessary for prediction, validation and parameter fitting; it will be important to have support for rapid prototyping and multiple, possibly competing, modeling teams.

An important component of such architecture will be accepted standards for the interchange of rules and ontologies. We expect that the new Rule Interchange Format (RIF) standard as well as the OWL2 RL standard will be incorporated into any such architecture (see Morgenstern, Welty, and Boley, 2010, and Motik, Patel-Schneider and Parsia, 2009).

Scaling

The sheer mass and complexity of financial data poses another challenge for this endeavor: that of scaling up. Some approaches that may work well on a small scale may not scale up to larger problems. For example, as mentioned above, we discussed the suitability of MATLAB for representing financial instruments. However, it is unclear whether MATLAB would be efficient enough to use on a large scale. Being able to translate from MATLAB into a more efficient framework would be essential. For example, it might be possible to integrate techniques used for reasoning with "Very Large Knowledge Bases."

More on Integration

There will undoubtedly be new opportunities for data integration, particularly as new types of data are introduced. Indeed, we expect that this research will stimulate and support the development of further standards for knowledge base development and languages.

4. Financial networks, agent-based simulation and architectures for large-scale computation

Overview

This breakout group discussed in general terms the enterprise of modeling "the financial system." Although the financial system is a broad and imprecisely defined scope for a modeling effort, capturing systemic issues calls for an inclusive approach: stretching to capture elements not traditionally covered by financial models. The following describes many of the points raised and issues discussed in the 1.5 hour conversation we devoted to the topic. Fleshing out these points in a more systematic and scholarly manner would be a valuable exercise in preparation for a potential research initiative in this area.

Goals

The goals of modeling the financial system are perhaps obvious in the wake of the recent financial crisis. This experience not only revealed the inadequacy of existing regulatory and macroeconomic control mechanisms, it also exposed significant gaps in our understanding of how all of the elements in our complex financial system interact, how vulnerabilities can build up over time and how the incentives of individuals and firms may misalign with the objective of a well functioning overall system.

The pure intellectual merits of deepening our economic understanding of the financial system are themselves substantial. The ability to characterize aggregate behavior of systems characterized by intricate interdependencies, dynamic uncertainty and asymmetric information structures would be a major achievement in its own right. Progress on this goal would also undoubtedly produce broadly useful analytic concepts, methodological techniques and economic insights bearing on many other sectors.

More practically, large scale financial system models would be directly useful to policy makers and regulators for the purpose of designing mechanisms, rules and institutions that govern operation of the financial system.

Financial and risk-management challenges

Model Characteristics

In what ways does the enterprise of modeling the financial system stretch our existing approaches to financial modeling? The challenges become clear when we consider the general characteristics that such models would strive to exhibit.

By definition the scope of a system-level model is unusually broad. The objective would be to maximize comprehensiveness, in terms of including actors and activities that bear on the financial system. Another way to express this is that the model strives to endogenize as much as possible.

Inclusiveness entails accounting for the heterogeneity in kind and shape of relevant financial actors. In effect, we require fine-grained models, in order to capture the differential behaviors and interactions among these heterogeneous actors.

Capturing the linkages mediating interactions means that the models are inherently networked. Relating network structure to aggregate behavior is a key objective.

Given the importance of information structure—who knows what and when, and what do the parties know about each other's information holdings—explicit modeling of information is essential. This includes describing inherent information asymmetries, as well as dynamic flows of information throughout the system.

Information flow is just one example of dynamic phenomena that are important to capture in financial system models. Economic actors adapt over time, and in reaction to each other this induces a co-evolutionary system. Some would characterize the system-level properties arising from fine-grained dynamic interactions as "emergent," and the ability to recognize such phenomena is a central goal of any system-level modeling effort.

Relating properties of behavior at multiple levels is yet another distinguishing characteristic of this modeling enterprise. Examples include geographic scope (local / regional / national / global), actor aggregation (individuals / firms / households / industries / economies), activities (payments / deals / investments / market strategies), or time scales.

What might we want to ask of models with these characteristics? Example reasoning patterns include:

- Prediction. What are the likely trajectories from a given initial state?

- Counterfactual reasoning. What would happen if X occurs, where X is some exogenous event or policy option? Counterfactual reasoning is central to inference in support of mechanism design and policy making more broadly.

- Explanation and diagnosis. Given some (perhaps surprising) observation, what was the underlying cause? In other words, what just happened and why?

- Identification of macroscopic properties. For example, assess the stability of the system in a given state.

Computer and information science challenges

Approaches

To some degree models with these characteristics can be developed using established techniques of financial modeling (or incremental extensions), including both theoretical and empirical methodologies. To the extent this is feasible, we should of course promote such developments.

Some of the desired characteristics, especially taken together, may call for methods qualitatively different from those more traditionally employed in the finance domain.

The "agent-based modeling" approach analyzes a complex social system through simulation of fine-grained interactions among the constituent decision makers, or agents, described and implemented as (usually simple) computer programs. This approach has developed an expanding community of researchers in the social sciences, who typically appeal to one or both of the following reasons (among others):

- Agent-based models enable study of dynamic environments that are intractable for analytic methods. Deductive theoretical analysis often requires extreme simplification or stylization, which may sacrifice critical detail. Simulation modeling accommodates rich environmental description and supports exploratory investigation where the modeler is initially unsure about which details are essential.

- Modeling the decision makers as computational agents can avoid the strong rationality assumptions often invoked, particularly in economics. The computational methodology also facilitates employment of alternatives to classical equilibrium-based reasoning, for example evolutionary approaches.

The "network science" approach focuses on the structure of relationships in a large-scale dynamic system. It employs techniques from graph theory and statistics (among other fields) to develop models of how networks evolve and to characterize behavior with respect to position and interactions in a network.

Methodological Issues

This modeling enterprise is nothing if not challenging, and particularly when pursued with not-(yet)-well-established techniques, it is important to anticipate the difficulties of achieving desired results, and devoting research efforts specifically to overcome them. In a nutshell, the challenge is how to produce believable results from system-level financial models, particularly those employing agent-based simulation. Believability may be in question for several reasons, including:

- Fine granularity and fidelity lead to a multiplicity of agents, possible agent behaviors, and possible environment configurations. This means that there are enormous degrees of freedom and so are perhaps too easy to support any particular conclusion.

- The evidentiary value of simulation results is not traditionally well established in economics and finance disciplines.

There are also several promising approaches for enhancing believability of agent-based modeling. One broad strategy is to ground the simulation descriptions in well-established theoretical or econometric models. This enables model alignment to some degree. Another way to connect agent-based simulation to accepted economic practice is to appeal to the general principles (e.g. rationality assumptions) that carry force in the discipline. Although as noted above agent-based modeling is often motivated by deviation from rationality, this is not a necessary step. Indeed, it may be possible to justify choices of agent behavior through standard game theoretic analysis concepts, albeit using simulation results rather than analytic models to characterize the game.

Another broad strategy is to promote calibration of simulation models with real-world data, and improve reliability through replication of key studies by multiple research groups.

Further Discussion

The breakout group touched on several other topics relevant to establishing a research agenda for financial system modeling. This includes the challenges of model management and communication, data requirements and potential applications. Despite the ambitious nature of the central problem, the enthusiasm underlying the discussion suggests that there is great interest and potential for attempting to tackle these challenges. Given the magnitude of the importance of this domain, even partial progress on the fundamental problem can have enormous public benefit.

5. Data integrity, data quality and operational risk

Overview

If financial systems are to be modeled, and systemic risk evaluated based on the data, then the results we obtain can only be as good as the data we start with. In this section, we consider some of the ways in which the data obtained may not meet the desired quality standards and some ways in which we can work to address these issues. We also consider how data are recorded and represented, since this has a significant impact on its suitability for downstream use.

Financial and risk-management challenges

Quality Measurement

There is much debate within the financial industry on how we define and measure data quality. One can find different opinions, typically driven by differing application needs, on the definition of quality, the method of measurement, and even the function served by the data. Nevertheless, one can at least record a taxonomy of seven different dimensions in which to evaluate quality of data:

- completeness,
- conformity,
- precision,
- accuracy,
- timeliness,
- association, and
- consistency.

Most of the terms above are self-explanatory. We add a few words of explanation, where confusion may be possible. Conformity refers to standards. If data are not recorded in conformity with standards, they are liable to misinterpretation, and likely to lead to incorrect conclusions when rolled up with other data conformed differently. Precision differs from accuracy in that the former has to do with the granularity and the rounding, whereas the latter has to do with incorrect values being recorded, for instance on account of a data entry error. When data from multiple sources, or data about multiple real-world entities, are brought together, there is always a question of correctly associating the entities referenced. Finally, consistency refers to changes in the way data are recorded over time, across recording agents, etc.

For any financial model or system risk computation to be evaluated against this data, it is important to state clearly what quality of data are required, along these seven dimensions, for the results derived to be valid. One can then see to what extent the data available actually meets these assumptions. If it does not, then the derived conclusions should not be trusted.

Risk Representation

In simple financial models, risk is only an output variable: it can simply be stated, as a beta, a VAR,

or a risk premium, and we are done. For systemic modeling, it is usually necessary to take risks that have been computed at a granular level for individual instruments, for example, and to use these as the basis to compute risks at the next level. The problem we have here is that most risk measures, when stated as a single number, e.g. VAR cannot be rolled up or manipulated. In contrast, a full state-space representation, showing the complete risk distribution, does indeed carry enough information to be able to derive the next level of risk and price, after aggregation, and other operations; however, such a representation is prohibitively expensive in space to store, even if all the information required were available. This suggests the need for research into suitable risk representations at suitable intermediate points between the two extremes: what risk information must we minimally capture so that risk can correctly be aggregated, without requiring a full distribution of possibilities to be maintained.

A second challenge with risk is how to show risk factors to decision-makers in a form that helps them to act. It is very easy to show someone a single boiled-down number. However, that in itself, even aside from the composability and aggregation problems discussed above, may be insufficient information even as a basis for action. We require research on how to show the different components of risk, on how to support the decision-maker to drill down from aggregate data based on risk factors, on how to support "what-if" analyses based on specific risk scenarios.

A third challenge with risk is to include factors that are not normally included in risk analysis. For example, there can be risk exposure simply on account of the accounting system used. A research direction is to characterize such risk factors and evaluate their impact.

Computer and information science challenges

Data Recording

Technology trends towards electronic execution of contracts, along with new regulations likely with the new OFR, will both suggest that much data must be recorded and made public very quickly. In many circumstances, there are likely to be errors in the initial recording of data. (We have heard estimates as high as 40% for the fraction of trades today where some correction is required before final settlement. To the extent that much of this error is human-driven, one can expect that the error rates will be much lower at electronic exchanges. Empirically, this does seem to be the case, though error rates are still significantly greater than zero.) If nothing has been done with the recorded data and if no one has "seen" it yet, then applying a correction is straightforward. However, if there is a public record, and/or if downstream computations depend on the recorded data, then a suitable protocol must be developed to install corrections. This protocol must not only correct the error in the recorded data itself but also compensate for downstream computations that depend on it.

In addition to recording data, we must represent and record associated metadata: the who, what, when, etc. associated with the data record itself. Banks already record audit trails and so have systems that keep track of at least some metadata. This provides a natural place from which to start. Where there are fast-moving phenomena, there is a question of time stamp granularity and precision. For example, if an analysis of some sequence of computerized trading is required, we may need to know precisely at what time something occurred, down to a small fraction of a second. Delays in propagation of the data record can cause the recorded history to become misleading and possibly even inaccurate.

While most analysis of recorded data may take place at human time scales, there is likely to be a need for fast alarms, at least in some scenarios. These will need to be designed to be particularly robust to prevent false alarms being issued without human approval. Research is required to develop such alarms based on data that is "sniffed" on the wire, as it were.

Information Representation

There are many crucial decisions to be made with regard to the representation of recorded information. For example, see discussion of risk representation above. The choices made can have significant impact on the suitability of recorded information for downstream analyses. In many cases, the challenge is not simply a choice between well-known alternatives but rather the development of new representation alternatives.

Mathematical formulae frequently occur in financial systems. When there is derived information, knowing the formulae used to derive it is often valuable. It is straightforward to record a formula as a text string, and this is a minimum level of formula recording that we recommend. However, more is useful. If the system could understand and interpret recorded formulae, then it becomes possible to change a formula and to have values derived from it change in consonance. Doing this correctly, and with sufficient generality, is an area of research.

The accounting system used should be recorded not just as uninterpreted metadata, but as queriable data. This will enable analyses that rely on the specifics of the accounting rule used. It will also be the first step towards enabling "what-if" analyses associated with changes to accounting rules.

Integrity Maintenance

As data are recorded, errors can be detected and corrected, through the development of integrity maintenance software that checks data for consistency and "reasonableness." Checks at the low level are quite easy to develop and may often be put in place automatically by system developers. For instance, the price of a security must be a number and cannot have any non-numeric characters in it. However, research is required to develop application level integrity constraints. We believe, when correctly specified, that these can be particularly effective in detecting and correcting errors. However, specifying these rules is hard. If there is enough data, we may be able to develop research efforts that will learn such rules.

Discrepancies across data sources can be helpful in catching errors. To the extent that new regulations will bring together information from multiple sources, this in itself can become a means for error reduction. Patterns of data across multiple sources may be particularly helpful in detecting fraud. For example, trades that do not have counterparties, or trades where the value is differently recorded by the parties involved.

Unlike most other data management applications, fraud is a significant concern in financial systems. When an error is detected, it could well be fraud rather than an innocent mistake. In light of this, logic is required to determine quickly whether a detected error possibly falls into a fraud pattern. If it does, we may prefer to delay correcting it, so as not to alert the perpetrator, and set in motion an evidence collection mechanism instead. Since in other circumstances we would want to correct errors as quickly as possible, there is a need for an automated classification technique that can identify fraud and begin suitable forensic action.

6. Privacy, confidentiality, security and trust in managing financial data

Overview

Trust and confidentiality are important for many social and organizational activities, but this is particularly true in finance. Confidentiality and privacy are issues among organizations, including between regulators and financial market participants. Trustworthiness and security – especially in the senses of honesty and accuracy – are crucial prerequisites for markets to function well. Trust is essential for financial leverage; without trust, the extension and expansion of credit could not occur. On the other hand, misplaced trust can result in excessive defaults and overleveraged institutions. Enhancing the characteristics of honesty and accuracy in financial data should be a prime goal of financial risk management and regulation overall.

Financial and risk-management challenges

Privacy and Confidentiality

There is a basic need for policies regarding information sharing. This should include definitions of rights for regulators, as well as for market participants. It should also include definitions of rights to information for researchers, including those at the various regulatory agencies. An important practical question will be to define what public-use subsets of regulatory data should be published, and what sorts of obfuscations can be employed to facilitate this process. Computer and information science techniques can help here. Ontologies and semantic rules (including deontics) are a good way to represent many kinds of policies. The issues here are similar to the representation of contracts, instruments, regulations, laws, and policy-based processes discussed elsewhere in this report.

Policy management is a more general issue. Both regulators and financial institutions must manage a variety of policy-related tasks. First, they must specify, validate, and update the policies themselves. Second, given a set of policies, organizations must operationalize, enforce and monitor them in practice, including the associated transparency and governance activities. Finally, they must integrate the policies with each other, both within organizations and across them.

The "privacy inference" problem represents a specific challenge. This arises when multiple pieces of information – when put together – reveal something that was intended to be private, or is undesirable to disclose. For example, outsiders assembling disparate bits of information may gain the opportunity to reverse-engineer proprietary trading or investment strategies. There has been some computer/information science work on the privacy inference problem. A sandbox specific to the financial domain would be useful for supporting research. One basic approach is to query an overall integrated knowledge base and see what can be inferred by a complete (as well as correct) reasoner. When and how might techniques of masking and aggregation work? Higher-level dependency analyses can be conducted (automatically) to derive guarantees in some cases.

Another specific challenge involves establishing property rights to data, e.g. rights held by exchanges or dealers (see Mulherin, Netter and Overdahl, 1992). There is a need to analyze the economic value of these kinds of rights. A related issue is the need for government to clarify the relevant law and regulations; this was not directly in the workshop's mission but affects it. Those

rules are in some places murky, e.g. what are the powers OFR will have to share data with other regulators? Case law sheds doubt in some situation on rights to information even as basic as URIs. Regulators might potentially cut out the previous business of private data providers.

Confidentiality represents an emerging challenge in the area of ontologies. Pro forma definitions (e.g. for a new variety of derivative security) and valuations often constitute a sort of ontology that is or should be confidential. The possible confidentiality of ontologies has not been widely addressed in previous research. Policies and techniques to handle confidential ontologies need to be developed. Incentives for appropriate disclosure are important as well.

Security and Trust

There is a need – and opportunity – to construct a provenance map or "geography" for data sources in finance. That is, data sources should include a record of who provides the information, where they get it, how they get it and who owns it. Also important are techniques for managing provenance information in the context of aggregated data. Techniques in computer and information science for network analysis and data provenance can provide approaches for this. The need for, and issues around, data feed mapping are immediate with respect to data repositories and formats for several of the financial regulatory agencies.

There is a need for controls and audits of how information is used (i.e. that it only be for its intended or permitted purposes.) This is more about uses of data than authority to collect – it is about accountability. For example, a bookie will often disclose his illegal income to the IRS, knowing that the IRS is institutionally constrained from sharing the facts with the FBI.

There is also a need for accuracy in both data and metadata. Inaccurate data are of prime importance to regulators. Bad data can indicate processes that are broken or failing or simply suspicious (e.g. when a series diverges significantly from some average.) Examples of such signals and techniques include: "liar loans" or contracts without proper documentation; provenance on valuations that point to a dealer's own valuation model; a drill down in bank stress tests to reconcile the bank's numbers with regulators' calculations; and correlations of data accuracy to time of submission (e.g. do fraudulent reports typically arrive early versus late). Timestamps are often significantly inaccurate. For example, timestamps from back-end clearance providers are frequently aggregated or suffer from a lack of synchronization of clocks among order receipt and execution steps; they can thus misreport transaction time by as much as 30-40 seconds.

Computer science techniques for data mining and machine learning techniques can help here. There has been substantial experience in industry applying these techniques to certain kinds of financial fraud and money laundering, as well. A goal should be to detect inaccuracies and inconsistencies in real time as events are recorded.

It is important to model and reason about how specifically how things can go wrong. Previous work on malfunction models in life science and electronic/mechanical device troubleshooting can provide approaches here. Techniques and incentives from the capital markets for insuring accuracy can and should be extended to the rest of the financial system. For example, and very importantly, a decline in underwriting standards underlies much of the recent credit crisis, starting with residence appraisals and the ratings process for mortgage-backed securities. What can one do to detect such situations, and how can one get beyond impressionistic and anecdotal explanations (e.g. taking cues from delays and broken processes)?

Overall, we should try to learn from previous relevant efforts in financial reporting and compliance (e.g. XBRL) but also including large-scale data collection efforts from other domains, such as the census, or health care (HIPAA). In addition, new kinds of data sources are emerging. For example, there may be a way to "crowd-source" mortgage data. Online services such as Zillow (www.zillow.com) are suggestive of the possibilities here.

Computer and information science challenges

Several of the financial challenges discussed in the previous subsection correspond directly to challenges and/or technology areas within computer/information science. Much applied research work is needed, however, to combine and apply general techniques and understanding to the specifics of financial risk management and information integration. Some of the promising directions mentioned in the last subsection include:

- Semantic rules and ontologies for representing and reasoning about policies

- The privacy inference problem

- Data mining and machine learning

- Network analysis

- Data/knowledge provenance

- Computational mechanism design

- Mal-behavior modeling

Information integration is a basic underlying aspect for all of the above, of course. We next elaborate on the first three of these directions.

Knowledge representation for policy management of confidentiality, trust, property rights, audit and control

Semantic technology, especially based on rules and ontologies, is already used for many kinds of policy management, including all of these tasks (confidentiality, trust, etc.), and many areas within the financial services industry. It appears that semantic technology could fruitfully be applied to the financial systemic risk management arena, as well. Key technical knowledge representation challenges include representing and reasoning about:

- Negation and prioritized handling of conflicts between policies (i.e. defeasibility of rules. There are recently developed general techniques here, but they have yet to be applied extensively on a large scale; performance optimization and policy debugging tools will need research work. Semantic technologies have a number of advantages compared to legacy non-semantic business rule techniques including: easier/greater reuse and integration, more flexibility and expressiveness, more transparency and more sophisticated reasoning techniques (e.g. scaling up computationally).

- Quantitative aspects in tight and efficient combination with logical aspects. For example, equations, inequalities, constraint systems, differentials and integrals are all types of quantitative calculations that sometimes need to be part of the overall reasoning. As

another example, spreadsheet information needs to be integrated with rules and ontologies and RDF/relational databases.

- Probabilistic and approximate/fuzzy/vague aspects in tight combination with logical aspects. This is a basic and fundamental research challenge in knowledge representation today.

KR for the Privacy Inference Problem

Automated knowledge base querying, using semantic technology and database techniques and theory, can be used to address the privacy inference problem. A set of queries can be developed and tested against integrated information bases.

Data mining and machine learning

Data mining and machine learning techniques extensively developed already in computer and information science can be used to predict inaccuracies, find correlations, detect suspicious patterns and occurrences of them. Associated aspects include handling event flows and data warehousing.

SMALL GROUP DISCUSSIONS – RISK MANAGEMENT

The goal of the breakout groups was to bring computer and information scientists together with economics and finance researchers and practitioners to brainstorm promising research avenues at the intersection of the disciplines. The groups ran on two parallel tracks, with roughly half of the workshop attendees participating in each breakout group, to facilitate input from everyone. Each breakout group was chaired by a group secretary.

The first two breakout sessions introduced the financial and risk management issues at the level of the individual firm, and at the level of the system as a whole:

1. Information for risk management in the small – *Dan Rosen*

2. Information for risk management in the large – *Charles Taylor*

The chairs facilitated the discussions and took the notes that formed the basis for the reports that follow. Each report follows the same broad outline:

- OVERVIEW: Summary of the goals and topics for the session.

- FINANCIAL AND RISK-MANAGEMENT CHALLENGES: What were the interesting problems from the viewpoint of a finance researcher or economist that were identified?

- COMPUTER AND INFORMATION SCIENCE CHALLENGES: What interesting CS challenges need to be addressed to solve the previous problems?

1. Risk management in the small

Overview

The scope of discussion for the session on risk management in the small was enterprise risk management (ERM) data at the level of the individual firm. In other words, the issues discussed should generally have implications only within a single firm and not extend to multiple firms or systemic issues. The objective of the session was a report on topics appropriate for a research agenda related to data requirements for enterprise-wide (firm-level) risk management applications and the connection between firm-level data management and operational risk. Typical ERM applications include processes and software for measuring and managing: market risk, credit risk, operational risk and liquidity risk. They also include tools for asset-liability management.

Financial and risk management challenges

Background

The discussions worked from the presumption that the OFR would collect four principal kinds of data in a standardized and systematic manner from financial institutions:

- Contractual terms and conditions data, with sufficient detail to allow the derivation cash flows and eventual value, income, sensitivity and risk. This will require clear definitions of contract types.

- Counterparty attributes for each contract (as far as privacy allows), including unique entity identifiers and hierarchical relations among entities. These are vital for measuring individual credit exposures for the firm.

- Actual position and portfolio information (i.e. which contracts are held by the firm).

- Market data, such as interest rates and other market prices, describing key attributes that are not specific to individual contracts. In most cases, these data can be sourced from existing providers.

This regulatory mandate would impose a discipline on financial firms. For example, each firm would map its financial contracts (within the boundaries of its own systems) into one of the standard contract types and link its counterparties properly to identifiers defined in the OFR's legal entity reference database. While this mapping will enable systemic analytics at the level of the OFR, it will also facilitate institution-level analytics, such as portfolio stress tests.

Objectives of research

Research efforts into data management for financial firms must acknowledge at the outset that many firms have been collecting data and building enterprise risk systems for the last 20 years. Many lessons have been learned, and many production systems are already operating effectively. Rather than re-inventing the wheel, research should start with the foundation of current knowledge and reach to difficult problems. There is no shortage of difficult problems; a productive approach will be to look for the "pain points" in institutions' current processes.

Several high-level goals were identified that can guide research in this area:

- Well functioning data systems should bring transparency to all stakeholders of the firm, including senior management, risk managers, shareholders, debt holders, regulators and ratings agencies. To a significant extent, many of these constituencies currently rely on traditional accounting statements, which do not have a forward-looking, risk-centric perspective. This is a shortcoming of the status quo.

- Different information should be available at different levels within the organization. Business line managers require data at a level of detail that would overwhelm senior management or shareholders. This implies that data should be recorded at the most granular level – ordinarily the level of an individual contract or position – with provisions for roll-ups and aggregations for high-level views. It is typically trivial to sum or average granular data to create an aggregate; disaggregating a composite, on the other hand, tends to be difficult, error-prone and assumption-dependent.

- Well functioning information systems should reduce operational risk. This should be a significant motivator for firms to invest in data technologies. At a basic level, there is an organizational "drag" created by ineffective data processes and simple clerical errors. For example, failed trades due to the inability of back offices to confirm contractual details represent a significant opportunity cost to many dealer institutions (see, for example, Garbade, et al, 2010). More significantly, information systems can affect strategic decision making at the highest levels of the firm. Moreover, some of the most spectacular instances of internal fraud – the large-scale rogue trader cases – have been typically facilitated by ineffective data systems.

The quality of data definition, entry, and mapping at the lowest level will constrain the quality of data systems at all levels. Therefore, a significant focus of regulation at the institutional level should be on data entry to ensure quality. This has ramifications for a number of data requirements and operational risks that should be studied at an early stage:

- Living wills – The Dodd-Frank Act requires systemically important financial institutions (SIFIs) to establish so-called "living wills," or predefined resolution protocols that can be activated in a timely manner in the event of financial trouble, so that institutions can be unwound quickly and gracefully, with minimal impact on the system. This mandate implies significant information management requirements in terms of the accuracy, timeliness and granularity of information to be maintained. In addition to data infrastructure needed to support living wills, regulatory scrutiny to assure the adequacy of the process will impose additional requirements.

- Contractual descriptions – The Dodd-Frank Act also moves the industry toward standardization of information on instruments and positions to facilitate reporting and validation. Such data must be shared internally, of course, but also with regulators and counterparties in many cases. The OFR is mandated to maintain standard representations and identifiers for both legal entities and financial instruments in public reference databases. Firms must be able to determine that their contractual descriptions are valid, and that they contain sufficient information to describe cash flows in all relevant circumstances. A typical application is the ability to simulate portfolio performance under various risk events (e.g. probability of default and loss given default under risk scenarios).

- Extra-contractual descriptions – There are a number of important contextual features, in addition to contractual terms and conditions that must be tracked to support the process of risk management and securities valuation. For example, external macroeconomic and market factors affect the performance of financial instruments. Thus, systems, data, and representations are needed for these events, conditions, and reactions. Moreover, there is a need for languages or representations for important financial business processes, including references to entities, actions, and consequences.

- Decision processes – Similarly, structured descriptions of business decision-making are valuable for supporting controls, audits and the Sarbanes-Oxley reporting process. Required representations include the business processes triggered by decisions (actions, time to completion, objectives, obstacles, recourse, etc.). In addition, there are questions of who decides, what they can decide and under what circumstances, and who is affected and their recourse. Decision support and recording should include protocols and technologies for access and process control.

Computer and information science challenges

Representation and standardization of firm-wide data

The foregoing suggests some infrastructure requirements to support portfolio and enterprise risk management. While many of the basic techniques and technologies are well understood, current implementations vary widely across the industry. Traceable representations of firm-wide wide data are required for: instruments, positions, market data and model requirements and configuration parameters. These specifications must also be shared in many cases with outside entities, such as regulators and counterparties. A rapidly changing business environment imposes special challenges.

Even at the level of the individual firm, inter-firm data integration and reporting across multiple sources remain an issue. For example, alternative investment institutions, such as hedge funds, must interface with third-party providers, such as prime brokers. The resulting coordination and standardization has ramifications for integration with firm-specific risk management systems. This is especially true for firms that engage in extensive product innovation, as this generates dynamic and unstable metadata.

Many of the basic techniques such as data warehousing are well known in other contexts (e.g. auto or department store inventory and supply chain management) but must be tailored to a financial context. Other techniques will be more specific to finance and require more focused research. Conceptual modeling is needed to describe the state space of possible triggering events and conditions in contracts, as well as to describe the behavior and reactions of market participants. This would facilitate systematic simulation of market events. Ideally, contracts would be mapped to computable models or a domain-specific language or ontology. Such a representation could support automated reasoning, for example to determine the internal consistency and completeness of a contract. Alternatively, it might support automated analysis, aggregation, visualization, and reporting of contracts for a variety of stakeholders, both internal and external.

Financial risk modeling and management

Financial risk management presents a number of specific research and implementation challenges.

Management of counterparty credit risk requires accurate identification of legal entities, a task that is frequently poorly managed in practice. Such a service needs to record ownership (and other) relationships among firms and to manage changes in these over time. The OFR's mandated legal entity reference database is anticipated to provide many of these services as a public good.

An important aspect of cash flow management is the measurement and monitoring of liquidity risk. This requires an understanding of counterparty creditworthiness in both ordinary market conditions and extrapolated into stressed circumstances. One way a legal entity repository can support liquidity monitoring is via "look-though" analyses of indirect counterparty exposures. The creditworthiness of counterparties can be tracked in a variety of ways, including traditional accounting statements and more detailed network information that maps out the status of collateral positions and flows among the firm's counterparties.

Data management and operational risk

Data management is a key ingredient in operational risk. Inadequacies in centrally managed processes and methodology have led to a heavy reliance on end-user computing at many firms. There is a need for centralized management of metadata, perhaps through ontologies or other tools, to give business users (rather than corporate IT departments) flexible control over semantic content as an alternative to spreadsheets and local databases. A "data management maturity" process for certification and staff training – similar to the SEI's capable maturity model for software development – would be a welcome innovation. Part of this should be economic and empirical research on the benefits and costs of data management techniques, including reduction of operational risk.

In addition to data management as a component of operational risk, there is a need to capture data *about* operational risk. Databases are currently available with details of operational events from a variety of firms, but there are concerns – perhaps insurmountable – about completeness and selection bias. In addition to information about ex-post operational events and outcomes, there is a need for tools and standardized language for operational process and contingencies ex-ante. A structured representation of core financial processes could help with analysis and reporting of operational vulnerabilities, including public reporting of industry best practices.

2. Risk management in the large

Overview

Risk management in the large is about managing the risks to the financial system as a whole. Of course, the behavior of individual agents also contributes to risk in the large. Healthy institutions are less likely to fail. Super-spreader institutions and those with systemically important functions, like clearing and settlement, custodian services or prime brokerage, should be especially well managed for the good of the overall system.[14]

The opposite can occur too: good institution-level practices can produce bad outcomes in the aggregate (for example when common thresholds or triggers induce herding behavior, or when positive feedback loops accelerate instability.) Systemic risks are relatively rare and beyond the influence of individual institutions. Stockholders – or even debt holders – may not benefit directly from the expenditure of resources on systemic risk management at the institutional level. To a significant degree, systemic stability is a public good and regulators have to be its stewards. The session on risk management in the large discussed not only what the OFR should do but also the broader research community.

The purpose of research

Systemic risk management involves three activities that research should illuminate:

- Strengthening the system – research into the vulnerabilities of the system and ways to encourage private behaviors and policy changes that can make it more resilient;

- Identifying and responding to emerging threats – research to deepen our understanding of systemic instability and ways to identify and respond to early warning signs; and

- Managing crises when they occur – research into crisis management strategies, options, information needs, planning and preparation for both individual firms and regulators.

Building a vibrant research community

The research community must be multidisciplinary, embracing not only data management, finance and economics but also network theory, complex adaptive system theory and other branches of mathematics. So far as possible, data should be openly available to this community. Regulators might offer sandboxes for academic use, and fellowships and secondments to the OFR are mandated by the Dodd-Frank Act. Aggregated and anonymized data should be released early, with progressively more granular data made available with progressively greater lags. In Singapore, there is a wiki-style experiment underway to establish Asian credit ratings in an open community and perhaps lessons for the U.S. can be learnt from that experience. Constraints will be freedom of use of commercial data, proprietary value of analyses and data generated and confidentiality and

---

[14] A great deal has been written about systemic risk since the crisis. See, for example, Acharya, Pederson, Philippon, and Richardson (2010), Brunnermeier (2009), Buiter (2009), Committee on the Global Financial System (2010), Engle and Weidman (2010), French, Baily, Campbell, et al (2010), Geanakoplos (2010), Haldane (2009 and 2010), Hammond (2009), Hendricks (2009), Johnson (2009), King (2010), Large (2010), Rajan (2010), and Taylor (2009).

security. Also, the possibility of adverse feedback affecting system dynamics may limit open access.

Making research industrial strength

How can we make analysis industrial strength? Issues are data standardization and accessibility, third party validation, teaming analysts effectively (and incentives to work in teams on such things as data management). Other disciplines manage these things well but not economics and finance. If this effort succeeded, it might just possibly pioneer a wider change in the economics profession.

The audience for research

To the extent that research is aimed at influencing policymakers, the ultimate audience at any point in time is quite small. However, institutions and agents at different levels throughout the financial system may benefit greatly from this research too where it sheds light on risk management in the small, for example. Given how young this field of study is, there will also be an important academic audience who should question findings and challenge approaches at every turn.

Financial and risk-management challenges

In this section, we consider several possible research topics for systemic financial risk.

Analytical approach and focus

There is no dominant model of system risk and consequently, no dominant analytical approach. Broadly speaking, there are three sorts of analysis in the literature today:

- *Macro-economic analyses*: These focus on a few macro-economic causes and effects. Analysis relies on a few key phenomena such as leverage, capital adequacy, maturity transformation, asset price bubbles and credit expansion to explain the emergence of systemic instability and the subsequent collapse. Data volumes are modest but data quality – completeness, relevance, timeliness, accuracy and reliability - is an abiding issue. By and large, these are analyses that can be conducted in the short term.

- *Institutional analyses*: These concentrate on specific features of markets and institutions, processes and practices and how details of the structure of the financial system can exacerbate stresses and strains when they emerge. Data requirements tend to be greater for this type of analysis than for macro-economic analysis – dramatically so for real-time monitoring – for example, to understand the risks of another flash crash. These sorts of analysis can also be started in the short term – indeed many ad hoc studies of this kind have been conducted in the recent past. For the moment at least, data requirements may be so idiosyncratic that quality is likely to be difficult to systematize and is likely to stay variable.

- *Granular analyses*: These tend to look for the seeds of instability in the characteristics and behavior of individual agents and in the patterns of individual transactions.

All three approaches address all three purposes for research and analysis – developing insights on how to strengthen the system; developing methods to monitor and report on emergent systemic risks; and developing approaches to crisis management.

Short-term research topics

- *Exposure maps*: The Dodd-Frank Act mandates collection of credit exposure data from systemically important financial institutions (SIFIs). Exposures between institutions take forms other than credit exposures – liquidity exposures, ownership exposures, risks of reputational contagion, and inter-dependent portfolio exposures. Research is needed to establish how these are related, and their potential systemic importance, not just for SIFIs but for other financial institutions as well.

- *Capital and liquidity adequacy*: Basel III establishes guidelines for minimum national capital and liquidity adequacy standards. These focus on individual institutional soundness and ,– especially when the institutions in question are SIFIs – both the concepts and the standards may need to be re-thought in the systemic context. Here the issue is not so much whether the individual institution may fail as the extent to which it may act as a catalyst for a contagious systemic collapse. Standards may have to be higher, so that as institutions enter a systemic stress episode, and asset prices are depressed and become more volatile and correlated, there is a buffer that can be used up and capital and liquidity nevertheless remain adequate. The research questions are how to characterize these different states of the financial system and how large are the required additional buffers.

- *Concentration*: Banking in the United States has grown progressively more concentrated. Per King (2010), "[The assets of] the top ten banks amount to over 60% of GDP, six times larger than the top ten fifty years ago. Bank of America today accounts for the same proportion of the U.S. banking system as all of the top 10 banks put together in 1960." But, at the same time, the financial system has been disintermediated. The research issue is to improve understanding of trends, causes and consequences of concentration, not just at the aggregate level in banking but also within specific markets and activities across all of the major segments of the financial system. How contestable are these markets and activities, and how might they be made more so?

- *Designation*: What are the consequences of institutions being designated SIFIs? There is speculation that moral hazard will increase, reducing funding costs and increasing risk-taking. The research question is to understand the size and significance of these effects.

- *Failure rates*: Financial institutions, markets and processes fail occasionally in normal times and at an accelerated rate during periods of material financial distress. Something is known about the causes and consequences of bank and insurance company failures. Failures of money market mutual funds, repo markets, and ABS markets during the recent crisis have been the subject of intensive study. The research question is to understand better the causes and consequences of failures and the way they are handled for the overall stability of the system – bearing in mind that Schumpeterian creative destruction is, in normal circumstances, a healthy thing for industries.

- *Key markets and activities*: Our financial system – including some of the remedies of the Dodd-Frank Act – assumes that some activities must be extremely safe and be sheltered from normal competition. One of these is wholesale payments and settlement activities. Other candidate activities are custodian services and prime brokerage. It is arguable that liquidity markets are more crucial than capital markets. The research issue is to sharpen our

ideas of what makes an activity systemically "key," to understand how to lessen systemic vulnerability toward them.

- *Leverage and credit cycles*: A distinction can be drawn between moderate business cycles, which are often associated with changes in credit conditions, and more violent cycles which seem to be strongly associated with changes in leverage. Some ideas are well established about how to measure credit and leverage cycles – and therefore how to monitor them – in the banking system. The research issues is, how should they be measured (and aggregated) across the rest of the financial system, and also across the household, business and public sectors to provide a reasonable (and forward-looking) measure of systemic vulnerability?

- *Internalizing systemic risks*: Systemic risk may sometimes be a public "bad:" while it is in the interests of everyone to avoid systemic peril, individuals lack incentives to shun actions that may, in concert with the actions of others, help precipitate disaster. CoVaR – the measure of systemic externalities generated by large financial institutions – is a possible basis for capital charges to help institutions internalize systemic risks (see Adrian and Brunnermeier, 2009). The research question is whether a "tax" could be placed on strategies (other than gigantism) that contribute to systemic risk? For example, might trading strategy imitation be taxed to align private and public costs? How high might such taxes have to be and how might they be implemented?

- *Distributed analysis*: How much of systemic risk management can, in some sense, be distributed or delegated across the system to individual private players? Can public and private risks be sufficiently aligned? Can society at large trust private market participants more, or less, than government entities, to be assiduous in working to avoid, mitigate or manage very low probability / very high cost events? Is there any way to divide the issue naturally into the parts that can be tackled most effectively by individual private actors, private collective action (which accounts for the existence of most utilities and standards in the financial system) and public regulation and provision of services (such as wholesale payments services)?

- *Forensics*: One of the functions suggested for the OFR is a forensics capability -- to analyze ex post the details of any potentially significant systemic event. In the long term, such a capability should be built upon the collection of routine detailed data about the financial system. In the near term, an analysis of the recent crisis focused on aspects complementing the work of other experts (e.g. the FCIC) might be extremely helpful. In particular, it would be useful to examine issues relevant to the design of a systemic risk monitoring and reporting regime or a future "crisis control room." For example, what was the range of observable indicators that things were going awry in the build-up to the crisis? Was there something about the confluence of signals that might have warranted a particularly vigorous call for action? Was there a window of opportunity in which the indicators of heightened exposure to systemic risk were there to be clearly seen and yet there was still time to make a course correction to avoid the crisis? What form might that course correction have taken if the FSOC (and the international coordination mechanisms of the G20 and the FSB) had been in place at the time?

Long-term research topics

Longer term, the systemic risk research agenda expands considerably. As more data and computing

capacity become available, some methods of analysis such as agent-based simulation and other methods for thinking about complex adaptive systems are likely to become more practical and promising. Many short term investigations will inevitably generate additional ideas over time. Some questions, such as credit exposure mapping that can be tackled in the near term in an approximate way by using institution-level data reported, say every quarter, should be amenable to much more thorough, continuous, complete and accurate analysis as more consistent and granular data becomes available from across the financial system.

For this reason – many sorts of short term analyses are likely to be improved over time with more refined methods and better data -- the line drawn here between short and long term analyses is fairly arbitrary. Not all workshop participants would agree with how topics have been sorted here.

- *Valuation*: One task that the OFR may choose to tackle is to value positions and portfolios independently across the financial system. The argument is that, otherwise, there can be no assurance that valuations on which systemic risks are assessed are consistent, and events such as AIGFP might not again arise undetected. It could also provide an independent check of private valuation models which, to the extent that good risk management at the institutional level reinforces good systemic risk management, would be helpful in the systemic risk context. Creating a public/private partnership to develop and apply a comprehensive set of public models might make this problem more tractable than it would otherwise be.

- *Heterogeneity*: From the point of view of stability, one important aspect of the evolution of any population is how much heterogeneity waxes and wanes. When population members are dissimilar, the chance of a random exogenous shock that kills one member killing many others may be reduced. Fashion and herd reactions may become more muted. A loss of confidence in one member is less likely to infect the reputation of others if the population is materially diverse. We know very little about the heterogeneity of institutions, customers, processes, markets, portfolios, business strategies and services across the financial system. Some of what we do know is worrisome – the relentless increase in concentration of activity in SIFIs and their growing organizational and strategic similarity. Measures and data both need to be developed and explored and the relationship between heterogeneity and system stability needs to be better understood.

- *Complexity*: There is no doubt that complexity has been increasing in the financial system for several decades as technology, globalization, new analytical possibilities and customer demand have driven and facilitated change. It has become clear in recent studies that complexity in instrument design, securitization and organization structure all contributed to a widespread failure of institutions and regulators to understand, measure and manage risk well in the decade leading up to the recent crisis. When the crisis struck, complexity (as well as a lack of information) made an accurate assessment of the situation and evaluation of options almost impossible. Since then, it has made managing the after-effects far more costly and difficult than it would otherwise have been – think of the Lehman bankruptcy which involves court actions in 76 jurisdictions and over 8,000 subsidiaries and affiliates of the holding company.

  Unmanageable complexity increases the appeal of management fashions in good times and the risk of herd behavior in bad. Much is to be done to characterize and track changes in

complexity and to understand their relationship to institutional, market and system stability.

- *Network structures*: A good deal is known about the relationship between network structure and network stability in other fields of science, ranging from the engineering of the internet and the power grid and the dangers of breakdowns and blackouts through to social networks and the dangers of emergent radicalism and terrorism and to networks of personal contact and the danger of contagion spreading epidemics. The prominence of specific nodes, the size of the network, the tightness of different links, the extent of redundancies, variation and levels in capacity utilization, and the dynamics of network change all contribute to systemic stability. The banking system and the payments and settlement systems have been studied as networks in recent years. Research is needed to extend this sort of analysis to nonbanks and to explore the implications of viewing the system as a network of processes, contracts, and markets as well as institutions.

- *Beliefs and behaviors*: Sociologists and neurologists know quite a bit about panic while economists, with their tradition of using a model of behavior that is ultra-rational, know relatively little. Panic is a feature of late stages in a crisis and is just one example of a type of behavior of importance to the stability of a system that is, fundamentally, about holding and trading beliefs, hopes and promises. In the first instance, the research issue is to assess the potential for drawing productively on the methods and insights of these other sciences to understand how beliefs and behaviors can be observed and how they can contribute to systemic fragility.

- *Positive feedback*: Positive feedback effects occur in many parts of the financial system in different situations, contributing to individual market failures, normal economic contractions and the emergence of systemic instability. They are typically the result of interactions between beliefs/states of mind (euphoria, depression) behaviors (buying/selling), asset prices (booms/busts), balance sheet values (prosperity/insolvency) and liquidity (funding/market-making). Sometimes they reflect specific institutional arrangements like margin requirements in particular markets. Sometimes they are the product of good risk management practices at the institutional level producing perverse effects at the systemic level. In some circumstances, they may arise because of particular regulatory arrangements – the pro-cyclicality of capital requirements for banks is a much discussed example of current policy interest. And in some instances, such as the credit cycle in normal times, positive feedback loops explain parts of the cycle, but their force wanes and is replaced "automatically" by negative feedback loops in other parts.

  By some lights, positive feedback loops are among the most important phenomena explaining emergent systemic instability – booms followed by busts. Research is needed to understand the relationships between positive and negative feedback loops at different levels of the financial system – to see how low-level loops can feed high-level ones; to explore the scope for creating automatic negative feedback loops; to avoid unintended perverse regulatory effects; and to devise strategies for monitoring the system for emergent positive feedback loops.

- *Reaction functions*: It is one thing to understand how systemic instability might arise in a static regulatory environment and another to figure out the possible effect of changes in

policies on the financial system. If monitoring revealed a systemic threat, what should the regulatory agencies and the administration of the day try to do and how vigorously? Calibrating policy responses depends on understanding the reaction functions of the private sector to changes in public policy changes. If capital standards were eased in the face of weakening economy, would institutions really take more risk with less capital? Would they be more likely to do so if reserve requirements were lowered and capital requirements were kept steady? Reaction functions to policy are an important topic for systemic research.

- *Cumulative risks*: While in portfolio theory and the fields of credit and market risk management there is an enormous body of literature based on the insight that risks can offset each other, there are other risk types that tend to reinforce one another. Operational risk, control theory, engineering reliability and system stability are all areas where a number of small errors, exceptions or breakdowns can presage massive failure. The same would appear to be the case in many sorts of financial systemic breakdown: witness the unresolved controversies about which factors contributed how much to the recent crisis. This accumulation of risks may be quite as important as the observed breakdowns in established correlations between positions during periods of significant system distress. Research is needed to explore whether it may be possible to apply the idea of small risk accumulation at nodes in the system to develop a useful approach to monitoring and managing system fragility.

- *Self-Criticality*: One of the current puzzles of systemic risk discussion is the seeming increase in the frequency of systemic crises. In particular, banking crises since World War II have been vastly more common than they were before. Increases in interconnectedness, complexity and concentration have all been cited as possible explanatory factors. But is there a more general way to look at this phenomenon that would allow the OFR to understand and monitor the interplay between factors in a way that might help in developing strategies to reduce the increase the fight this trend?

  One possible approach would be to apply the general idea that some systems can tend by themselves toward states in which they are especially fragile or critical. This idea of self criticality has been widely applied in other sciences. Static economic systems tend to be self critical – no excess profits under perfect competition must make every firm extremely vulnerable to small outside shocks – and, while the financial system is certainly not static, it exhibits structural stability in different markets and organization for protracted periods of time. Its interconnectedness is a mechanism for the domino effects characteristic of collapse in many other self-critical systems. The research issues is, then, whether looking at the interval between crises as a period when the system is trending back toward self-criticality will help in identifying ways to monitor the threat of systemic fragility and to design strategies to lengthen periods of stability.

- *Evolution*: Evolutionary theory may provide a useful general framework for systemic analysis. The mathematical underpinnings of evolutionary theory are ripe. It is a theory that is capacious enough to accommodate neo-classical analysis; complex adaptive system analysis, agent-based simulation, network analysis and other ways of thinking of the financial system can be viewed as ways of looking at different aspects of evolution or understanding different circumstances as special cases. Heterogeneity, complexity and

instability are the bread and butter of evolutionary theory. But the question remains, what tangible advantages would accrue from looking at systemic instability and macro-prudential policy through the lens of evolution? Does theoretical consistency and completeness make much practical difference? This is another subject that needs further research.

- *Accounting issues*: Although accounting practices – particularly mark-to-market and fair market accounting – have been implicated as a contributory factor in the build-up to the crisis, they have tended to be viewed as a side-bar by serious analysts of systemic risk, rather than a topic of central importance. This may be doing them an injustice. Representation and determination of value are interlinked. While accounting principles should be invariant, the effects they have on the representation of changes in values and therefore on how people react to those changes is central to course of any financial crisis; and different accounting conventions are likely to influence that course in different ways, depending on how much information they convey and how well key actors understand and interpret them. More research is needed to understand these interdependencies better.

- *Transparency*: Financial economists usually treat transparency like an ethical standard -- something that every civilized person should believe in unquestioningly. Accurate, complete, low-cost information about prices, values (and policies) will usually improve decision-making, reduce moral hazard and information asymmetries, and help restore equilibrium in markets.

  But transparency may not always be a public good, if, for example, markets are near a destabilizing threshold or in the grips of some perverse positive feedback loop. It is not clear, for example, that more information on possible future values of institutions and assets will assuage all anxieties that are created, at least in part, by complexity and fundamental uncertainty. Transparency about the policymaking process can sometimes inhibit thoughtful policy deliberation. Transparency and privacy may sometimes be at odds. Sometimes central bankers have favored "constructive ambiguity" about policy as a way of blunting moral hazard and avoiding perverse private market reactions. And it is not completely inconceivable that a high level of transparency about the workings of the financial system might not be used by an unscrupulous investor or a foreign power to devise a strategy to profit from destabilizing the U.S. financial system. These are issues that need further investigation.

- *Incentives*: To the extent that SIFI stability and financial system stability are mutually interdependent, aligning incentives of executives and other risk-takers within the institution with the shareholders and holders of debt who are invested in the institution is likely to lead to better trade-offs of risk and reward. At least that will be the case in normal times, provided that policies combat effectively the moral hazard that surrounds SIFI status by internalizing systemic risks within SIFIs and through a credible prospect of resolution without bail-out in the event of failure. Incentives are also very important in determining how institutions treat their customers. There is current interest in compensation arrangements (partnerships, for example) and structures. More needs to be learnt about the determination of compensation regimes, to understand how legal rules and social issues affect risk-taking behavior and how this behavior can change during periods of heightened financial system distress.

Computer and information science challenges

Systemic financial risk is a new area of inquiry in many ways. Certainly nothing on the scale and scope of the OFR has been attempted before. Therefore, the research agenda described above implies an array of data requirements that themselves need further research. This subsection organizes these data requirements for systemic risk measurement according to strategic themes.

Short-term data imperatives

In the short term, systemic risk monitoring must necessarily rely on existing data sources, especially those available to regulators. Much of this information is in the form of traditional financial reports. Although corporate accounting statements are not tailored for systemic risk management, they can still be mined for useful details. For example, one might calculate various measures of industry or subsector concentration based on total assets, deposits, geography, etc, on the theory that excessive concentration creates "single points of failure" within the system.

Although there is a danger in an exclusive focus on the housing markets as the most recent and familiar trouble spot, it is nonetheless important to maintain an awareness of this as a particularly fragile market sector. Therefore, measures of stability in housing may be useful gauges of systemic stability. For example, one might track housing indices, such as new construction starts, repeat and/or new sales indices, or price-to-rent ratios for hints of a bubble. One might also track data at the level of individual households, perhaps via stratified sampling.

In some cases, new regulatory data collections might be relatively easy to implement as possible bases for measures of systemic risk. For example, SIFIs might be required to report aggregated (or granular) exposures by counterparty to other institutions in the system, enable the construction of a rudimentary network map of credit risk in the system. Other key relationships might also be identified and measured, including formal relationships such as ownership hierarchies, or simple parallels such as portfolio similarities that might be suggestive of herd behavior or shared exposures. Industry surveys are another technique for exploratory research into the landscape of risk. (See, for example, the Foreign Exchange Committee of the Federal Reserve Bank of New York, 2010.) Such surveys could cover a variety of key areas, including margins, haircuts, triggers, or the behaviors, beliefs and incentives of dealers.

Building standard reference databases should begin in the short term, even if many of the benefits will be experienced over the long term. In addition to creating reference databases for legal entities and financial instruments, as mandated by the Dodd-Frank Act, design, agreement and implementation can begin for standardized reporting and messaging formats/syntax, identifier symbologies and ontologies and other semantic resources.

Systemic risk monitoring is a young field, without well defined standard metrics for reporting and decision-making. Poorly designed high-level metrics can be misleading and may be worse than supplying no aggregate measure at all. A truly robust set of metrics may take months or even years to refine. Work should begin early on to develop data visualization tools, dashboard reports and non-dashboard system monitoring protocols for recording system status. Certain critical nodes in the system (e.g. clearinghouses) may ultimately require special attention or instrumentation.

Long-term data imperatives

Data organization will be an ongoing challenge. Systems in which the basic abstractions are not

designed well (or not designed at all) tend not to scale well. The OFR in particular runs the danger of getting lost in a sea of data without a robust and structured conceptual foundation for its activities. An information model that is optimized too intensively for the status quo may prove to be fragile when confronted with new requirements. It is therefore important to design the strategy with flexibility and open-endedness in mind. Issues to consider include:

- disciplined procedures for testing and releasing new versions of data symbologies and ontologies

- rules and processes (such as propagation of detailed schemas) for ensuring the accuracy and reliability of data as it is reported to the OFR; and

- the trade-offs in terms of security, convenience, reliability and record retention involved in collecting data into a centralized repository.

In all of this, the OFR will need to balance private value against private cost. Data governance should be formalized at the highest level of the agency. Rules and procedures for sharing the costs of the system and control over reported data must be established. Significant subsets of the OFR's database, especially including the positions and transactions data, will be business confidential, imposing important security requirements on the system. The infrastructure must be able to ingest data securely while maintaining data integrity and confidentiality and also support research and analysis activities. Legal boundaries will need to be established for what information may be published, when and by whom. Research is needed into techniques to anonymize and aggregate information in ways that cannot be reverse-engineered or inferred (e.g. the "mosaic" effect).

The long-run data strategy should be driven ultimately by the requirements of the analytical models in use. Daily or monthly reporting may be adequate for some series, but not others. For example, the recent "flash crash" highlights the growing importance of high-frequency trading, suggesting that real-time feeds from the exchanges could ultimately be needed to ensure that regulators have timely information in these markets. As the research develops techniques and metrics for systemic heterogeneity, complexity, and feedback loops, specialized derived data series will need to be defined and managed. Economic and risk forecasting research will define time series of interest that should be collected and stored. Historical analysis of past crises and market scenarios will establish requirements for record retention and archiving of data and metadata.

Discussions – Risk Management

REFERENCES

Acharya, Viral V., Pedersen, Lasse, Philippon, Thomas, and Richardson, Matthew, 2010, "Measuring Systemic Risk", Working Paper, New York U Stern School of Business, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1573171 .

Adrian, Tobias and Brunnermeier, Markus, 2009, "CoVaR", Staff Report, 348, Federal Reserve Board of New York, http://ssrn.com/abstract=1269446.

Alston and Bird, LLP, 2010, "Studies and Reports Required by the Dodd-Frank Wall Street Reform and Consumer Protection Act," technical report, http://www.alston.com/files/docs/Dodd-Frank_studies_and_reports.pdf.

Ashcraft, Adam B. and Schuermann, Til, 2009, "Understanding the Securitization of Subprime Mortgage Market Credit", Staff Report, 318, Federal Reserve Bank of New York, http://www.newyorkfed.org/research/staff_reports/sr318.pdf.

Baader, F., D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider (eds.), 2002, *Description Logic Handbook*, Cambridge University Press.

Baader, F., I. Horrocks, and U. Sattler, 2004, "Description Logics," in: *Handbook on Ontologies*, S. Staab and R. Studer, eds., Springer Verlag, Berlin, pp. 3-28.

Bank for International Settlements (BIS), 2001, Operational Risk, consultative document, Basel Committee on Banking Supervision, January, http://www.bis.org/publ/bcbsca07.pdf.

Bennett, M., 2010. "Enterprise Data Management Council Semantics Repository," Internet resource (downloaded 26 Sep 2010). http://www.hypercube.co.uk/edmcouncil/.

Bernanke, Ben, 2009, Testimony on American International Group before the Committee on Financial Services, U.S. House of Representatives, Washington, D.C., March 24, 2009, http://www.federalreserve.gov/newsevents/testimony/bernanke20090324a.htm.

Bernstein, P., 2003, "Applying Model Management to Classical Meta Data Problems," *Proceedings of the First Biennial Conference on Innovative Data Systems Research (CIDR)*, Asilomar, California, January 5-8, 2003.

Bernstein, P., and E. Rahm, 2000, "Data Warehousing Scenarios for Model Management," *Proceedings of the 19th International Conference on Entity-Relationship Modelling, (Lecture Notes in Computer Science)*, Springer Verlag, Berlin.

Borgida, A., M. Lenzerini, and R. Rosati, 2002, "Description Logics for Data Bases," in: *Description Logic Handbook*, F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, eds., Cambridge University Press, pp. 472-94.

Brammertz, Willi, Akkizidis, Ioannis, Breymann, Wolfgang, Entin, Rami, and Rustmann, Marco, 2009, *Unified Financial Analysis : the missing links of finance*, Wiley , http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470697156.html.

Brunnermeier, Markus K., 2009, "Financial Crisis: Mechanisms, Prevention, Management", working paper, NABE, http://www.nabe.com/pc09/documents/Brunnermeier02b_NABE_2009_Brunnermeier.pdf.

Brunnermeier, Markus and Oehmke, Martin, 2009, "Complexity in Financial Markets", working paper, Princeton University, http://www.princeton.edu/~markus/research/papers/Complexity.pdf.

Buiter, Willem, 2009, "Lessons from the global financial crisis for regulators and supervisors", working paper, Financial Markets Group, http://ideas.repec.org/p/fmg/fmgdps/dp635.html.

Committee on the Global Financial System, 2010, "Macroprudential instruments and frameworks: a stocktaking of issues and experiences", CGFS Papers, 38, Bank for International Settlements, http://www.bis.org/publ/cgfs38.pdf?noframes=1.

Committee to Establish the National Institute of Finance (CE-NIF), 2009a, "Argument for Data Management in support of the NIF mission for improved oversight and safeguarding of the modern financial system," technical report, http://www.ce-nif.org/images/docs/ce-nif-generated/nif_data_management_executive_summary_final.pdf.

_____, 2009b, "Argument for Research and Analytics in support of the NIF mission for improved oversight and safeguarding of the modern financial system," technical report, http://www.ce-nif.org/images/docs/nif_research_and_analytics_executive_summary_final.pdf.

_____, 2009c, "Data Requirements and Feasibility for Systemic Risk Oversight," technical report, http://www.ce-nif.org/images/docs/ce-nif-generated/nif_datarequirementsandfeasibility_final.pdf.

_____, 2009d, "Data Security considerations in support of the NIF mission for improved oversight and safeguarding of the modern financial system," technical report, http://www.ce-nif.org/images/docs/ce-nif-generated/nif_data_security_white_paper_final.pdf.

Engle, Robert F. and Weidman, Scott, 2010, *Technical Capabilities Necessary for Regulation of Systemic Financial Risk: Summary of a Workshop*, National Research Council of the National Academies, National Academies Press , Washington, DC, http://www.nap.edu/catalog.php?record_id=12841.

Fannie Mae, 2010, Single-Family Forms: Selling & Servicing Guide Forms, Internet site, https://www.efanniemae.com/sf/formsdocs/forms/index.jsp?from=hp.

Flood, M., 2009, "Embracing Change: Financial Informatics and Risk Analytics", Quantitative Finance, 9(3), April, 243-256.

Flood, M., A. Kyle, and L. Raschid, 2010, "Workshop on Knowledge Representation and Information Management for Financial Risk Management," Internet site, downloaded 21 July 2010, http://www.nsf-fiw.umiacs.umd.edu/index.html.

French, Kenneth R., Baily, Martin, and Campbell, John, et al., 2010, *The Squam Lake Report: Fixing the Financial System*, Princeton U. Press, http://press.princeton.edu/titles/9261.html.

Garbade, K., F. Keane, L. Logan, A. Stokes, and J. Wolgemuth, 2010, "The Introduction of the TPMG Fails Charge for U.S. Treasury Securities," *Federal Reserve Bank of New York Economic Policy Review*, October, 45-71.

Geanakoplos, John, 2010, "Solving the Present Crisis and Managing the Leverage Cycle", Cowles Foundation Discussion Paper, 1751, Yale U., http://cowles.econ.yale.edu/~gean/crisis/d1751.pdf.

Gómez-Pérez, A., M. Fernández-López, and O. Corcho, 2004, *Ontological Engineering*, Springer Verlag, London.

Haldane, Andrew G., 2009, "Rethinking the Financial Network", Speech at Financial Student Association, Amsterdam, Bank of England, http://www.bankofengland.co.uk/publications/speeches/2009/speech386.pdf.

_____, 2010, "The $100 Billion Question", Speech at Institute of Regulation and Risk, Hong Kong, BIS, http://www.bis.org/review/r100406d.pdf.

Hammond, Ross A., 2009, "Systemic Risk in the Financial System: Insights from Network Science", Briefing Paper, 12, *Pew Financial Reform Project*.

Hernandez, M. A., H. Ho, G.Koutrika, R. Krishnamurthy, L. Popa, I. R. Stanoi, S. Vaithyanathan, S. Das, 2010, "Unleashing the Power of Public Data for Financial Risk Measurement, Regulation, and Governance," IBM Technical Report # RJ10475, November.

Hendricks, Darryll, 2009, "Defining Systemic Risks", Briefing Paper, 1, *Pew Financial Reform Project*.

Hu, Henry, 1993, "Review: Misunderstood Derivatives: The Causes of Informational Failure and the Promise of Regulatory Incrementalism," *The Yale Law Journal*, 1993, 102, 1457-1513.

Hubbard, R. G., 1998, "Capital-market imperfections and investment," *Journal of Economic Literature*, 36(1), 193–225.

Jarrow, R. 2008, "Operational Risk," Journal of Banking and Finance, 32, 870-879.

Jensen, Christian S. and Richard T. Snodgrass, 1999, "Temporal Data Management," *IEEE Transactions on Knowledge and Data Engineering*, 11(1), January/February, 36-44.

Jobst, Norbert J., Gautam Mitra, and Stavros A. Zenlos, 2006, Integrating market and credit risk: a simulation and optimization perspective, Journal of Banking and Finance, 30(2), February, 717-742.

Johnson, Simon, 2009, "Systemic Risk: Are Some Institutions Too Big To Fail and If So, What Should We Do About It?", Testimony submitted to the House Committee on Financial Services Hearing, 21 July 2009", working paper, U.S. House of Representatives, http://www.house.gov/apps/list/hearing/financialsvcs_dem/simon_johnson.pdf.

Kambhu, John, Weidman, Scott, and Krishnan, Neel, 2007, *New Directions for Understanding Systemic Risk: A Report on a Conference Cosponsored by the Federal Reserve Bank of New York and the National Research Council of the National Academies*, National Academies Press , Washington, DC, http://books.nap.edu/catalog.php?record_id=11914.

Kent, William, 1991, The Breakdown of the Information Model in Multi-Database Systems, SIGMOD Record 20(4) December, 10-15.

King, Mervyn, 2010, "Banking: From Bagehot to Basel, and Back Again," Speech by the Governor of the Bank of England: The Second Bagehot Lecture at the Buttonwood Gathering, New York City on Monday 25 October 2010.

Large, Sir Andrew, 2010, *Systemic policy and financial stability: A framework for delivery*, Centre for the Study of Financial Innovation , http://bookstore.csfi.org.uk/action_150_106.aspx.

Lo, Andrew, 2009, "Regulatory reform in the wake of the financial crisis of 2007-2008," *Journal of Financial Economic Policy*, 1(1), 4-43.

Madhavan, J., Bernstein, P., Domingos, P., and Halevy, A., 2002, "Representing and Reasoning about Mappings between Domain Models," *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, AAAI Press, Edmonton Canada, pp.80-86.

Morgenstern, L., C. Welty, and H. Boley, 2010, *RIF Primer*, W3C Note, http://www.w3.org/2005/rules/wiki/Primer.

Motik, Boris, Peter F. Patel-Schneider, and Bijan Parsia, eds., 2009, *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax*, W3C Recommendation, 27 October 2009, http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/. (Latest version available at http://www.w3.org/TR/owl2-syntax/.)

Mulherin, J. H., J. Netter, and J. Overdahl, 1992, "Prices Are Property: The Organization of Financial Exchanges from a Transaction Cost Perspective," *Journal of Law & Economics*, 34 (2), October, 591-644

Nardi, D., and R. Brachman, 2002, "An Introduction to Description Logics," in: *Description Logic Handbook*, F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, eds., Cambridge University Press, pp. 5-44.

Pottinger, Rachel, and Philip A. Bernstein, 2008, Schema merging and mapping creation for relational sources, 11th International Conference on Extending Database Technology (EDBT), 73-84.

Rajan, Raghuram G., 2010, *Fault Lines: How Hidden Fractures Still Threaten the World Economy*, Princeton U. Press , Princeton, http://press.princeton.edu/titles/9111.html.

Salanié, Bernard, 2005, *The Economics of Contracts: A Primer*, MIT Press.

Snodgrass, Richard T., 1999, *Developing Time-Oriented Database Applications in SQL*, Morgan Kaufmann Publishers, Inc., San Francisco , ISBN 1-55860-436-7.

Staab, S., and S. Studer (eds.), 2004, *Handbook on Ontologies*, Springer Verlag, Berlin, pp. 93-113.

Taylor, Charles, 2009, "Managing Systemic Risks", Briefing Paper, 11, *Pew Financial Reform Project*, http://www.pewfr.org/project_reports_detail?id=0025.

U.S. Congress, 2010, *Dodd-Frank Wall Street Reform and Consumer Protection Act*, Pub. L. no. 111-203, 124 Stat 1376 (2010). http://www.gpo.gov/fdsys/pkg/PLAW-111publ203/content-detail.html.

APPENDIX: CONFERENCE ORGANIZATION

Conference Program

Full details of the conference are published on the Internet, at:
 http://www.nsf-fiw.umiacs.umd.edu/index.html.

Wednesday, July 21, 2010

12:00 p.m. - 1:15 p.m. Buffet lunch and presentation

- Acknowledgments: Introductions and thank-yous – *Louiqa Raschid (U. of Maryland) and Frank Olken (National Science Foundation)*

- The Context of the Workshop: A light introduction to why information management and knowledge representation matter for transparency, operational risk management, and systemic risk – *Mark Flood (U. of Maryland) and Pete Kyle (U. of Maryland)*

1:15 p.m. - 1:45 p.m. Plenary session

- What We Know Now: An inventory of some of the things we need, and the things that are currently in place for information management and knowledge representation, including financial standards and data modeling techniques. – *Mike Atkin (Enterprise Data Management Council)*

1:45 p.m. - 2:15 p.m. Plenary session

- Information Modeling: Semantics, representation formats, and visualization tools in finance – *Mike Bennett (Enterprise Data Management Council and Hypercube) and Dave Raggett (W3C)*

2:15 p.m. - 2:30 p.m. Plenary session

- Textual modeling: Integrating and reasoning over textual financial documents – *Shiv Vaithyanathan (IBM) [video supplied]*

2:45 p.m. - 3:15 p.m. Plenary session

- Risk Management in the Small: Data requirements for firm-level risk management applications (including market, credit, liquidity, etc.); the connection between data management and operational risk – *Sumeet Malhotra (UNISYS) and Bill Nichols*

3:15 p.m. - 3:45 p.m. Plenary session

- Risk Management in the Large: Data requirements for systemic risk monitoring; the outcome of the June CE-NIF workshop on systemic risk modeling – *John Liechty (Pennsylvania State U.) and Alan King (IBM)*

4:00 p.m. - 4:30 p.m. Plenary session

- Breakout Ground Rules: What we expect and hope for from the breakout sessions.

4:30 p.m. - 5:30 p.m. Breakout Sessions

- Breakout Group 1: Information for risk management in the small – *Sec'y: Dan Rosen (U. of Toronto, Fields Institute)*

- Breakout Group 2: Information for risk management in the large – *Sec'y: Charles Taylor (Pew Trusts)*

Thursday, July 22, 2010

9:30 a.m. - 11:00 a.m. Breakout Sessions

- Breakout Group 3: Knowledge representation frameworks (ontologies, schemas, models, formal logics) to describe complex financial instruments – *Sec'y: Andrea Cali (U. of Oxford)*

- Breakout Group 4: Managing risk models: schema mapping, data exchange, and model comparison and reliability – *Sec'y: Dennis Shasha (New York U.)*

11:15 a.m. - 1:00 p.m. Breakout Sessions over Working Lunch

- Breakout Group 5: Languages (operators and rules) for specifying constraints, mappings, and policies governing financial instruments – *Sec'y: Leora Morgenstern (New York U.)*

- Breakout Group 6: Financial networks, agent-based simulation, and architectures for large-scale computation – *Sec'y: Michael Wellman (U. of Michigan)*

1:15 p.m. - 2:45 p.m. Breakout Sessions

- Breakout Group 7: Data integrity, data quality and operational risk – *Sec'y: H. V. Jagadish (U. of Michigan)*

- Breakout Group 8: Privacy, confidentiality, security and trust in managing financial data – *Sec'y: Benjamin Grosof (Vulcan Inc.)*

3:00 p.m. - 5:00 p.m. Plenary session

- Summary of Breakout Sessions

Rapporteurs

The following participants contributed to the writing of this report:

- Andrea Cali, University of Oxford
- Mark D. Flood, University of Maryland
- Benjamin Grosof, Vulcan Inc.
- H.V. Jagadish, University of Michigan
- Albert "Pete" Kyle, University of Maryland
- Leora Morgenstern, New York University
- Louiqa Raschid, University of Maryland
- Dan Rosen, University of Toronto
- Dennis Shasha, New York University
- Charles Taylor, Pew Charitable Trusts
- Michael Wellman, University of Michigan

Participants

In addition to the rapporteurs, the following attended and participated in the workshop:

- Lewis Alexander, U.S. Treasury
- Richard Anderson, Federal Reserve Bank of St. Louis
- Mike Atkin, Enterprise Data Management Council
- Mike Bennett, Hypercube
- David Blaszkowsky, Securities and Exchange Commission
- Willi Brammertz, Brammertz Consulting
- Sandra Cannon, Board of Governors of the Federal Reserve
- Theresa DiVenti, Department of Housing and Urban Development
- Michael Donnelly, U.S. Treasury
- Darryl Getter, Congressional Research Service
- Le Gruenwald, National Science Foundation
- Walter Hamscher, Securities and Exchange Commission
- Alan King, IBM Research
- Andrei Kirilenko, Commodity Futures Trading Commission
- Joe Langsam, Morgan Stanley
- Adam Lavier, U.S. Treasury

- John Liechty, Pennsylvania State University
- Jorge Lobo, IBM Research
- Hank Lucas, University of Maryland
- Sumeet Malhotra, UNISYS
- Allan Mendelowitz, Committee to Establish the National Institute of Finance
- Leonard Nakamura, Federal Reserve Bank of Philadelphia
- Bill Nichols, RCube Information Management
- Frank Olken, National Science Foundation
- Francis Parr, IBM Research
- David Raggett, World Wide Web Consortium
- William Rand, University of Maryland
- Rick Ross, Committee to Establish the National Institute of Finance
- Lemma Senbet, University of Maryland
- Nitish Sinha, University of Maryland
- Arthur Small, VentiRisk, Committee to Establish the National Institute of Finance
- Chester Spatt, Carnegie Mellon University
- Vivek Srivastava, University of Maryland
- Anthony Tomasic, Carnegie Mellon University
- Kenichi Ueda, International Monetary Fund
- Shiv Vaithyanathan, IBM Almaden
- Vish Viswanathan, Duke University
- Susan Wachter, University of Pennsylvania
- Nancy Wallace, University of California at Berkeley
- Katherine Wyatt, FDIC