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.

- 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.