

Saints and Sinners: An Introduction to the Workshop on Records and Information Management for Financial Analysis and Risk Management

Financial records, information and data can be viewed, at one and the same time, as both “saints” and “sinners”. On the sinning side, the financial crisis of 2007-2008 has drawn attention to how weaknesses in the quality and management of records, information and data have led to operational risks in financial institutions, flawed bankruptcy and foreclosure proceedings following the Crisis, and inadequacies in financial supervisors’ access to records and information for the purposes of a prudential response. Lack of access to information on a timely basis, data quality issues, missing evidence, poor record keeping leading to operational risks, and the use of new channels of information such as social media have all been implicated in the build up of financial risk and the spread of contagion.

On the saintly side of things, it is clear that good quality and well managed records, information and data are cornerstones of effectively monitoring risks, making good risk decisions, and holding individuals and institutions to account when rules on risk tolerance levels are breached. This duality of records, information and data – the potential for both good and evil, if you will – raises many questions concerning what it is about records, information and data that produces a good result versus a bad one. Research into the characteristics and conditions that contribute to good versus bad results in the creation, management and use of financial records, information and data will support the development of standards, best practices and tools aimed at securing a future in which results will be more good than bad.

Conducting such research is no small task. Fields that carry out research on records, information and data – records management, archival science, information science and computer science, for example – each have their own perspectives, research questions and communities and do not always communicate with one another nor collaborate. It is rarer still for these researchers to join forces with finance and economics researchers, as is necessary for effective research on the use of records, information and data in the context of financial analysis and risk management. Even rarer is the coming together of academic researchers with financial regulators, industry experts and financial technology developers to translate theory into practical and socially beneficial results. This is the unique place of this workshop and what I hope will be the research partnership that forms from it, as it brings together all of these groups with a focus on a single challenge – the challenge of effectively dealing with the management of records, information and data for financial analysis and risk management.

This workshop was inspired by the Workshop on Knowledge Representation and Information Management for Financial Risk Management organized by Mark Flood, Pete Kyle and Louiqa Rashid July 21-22, 2010. The report on this workshop is available here: <http://www.nsf-fiw.umiacs.umd.edu/docs/FIWreport-FINAL.pdf>. As stated in the final report on the workshop:

The goal . . . 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. . . We hope, however, that this is only the beginning of a much longer conversation. (p.6).

Our smaller-scale Workshop on Records and Information Management for Financial Analysis and Risk Management seeks to continue this conversation on three focussed themes: governance, analytics and long-term digital preservation. In doing so, it introduces two new themes to the conversation: visual analytics and long-term digital preservation and seeks to expand on the themes of ontologies for knowledge representation, formal logics, schema mapping, data integrity, and operational risk based on developments since the Workshop on Knowledge Representation and Information Management for Financial Risk Management as well as on novel research. This workshop also broadens the interdisciplinary exchange of ideas on the subject of financial data by introducing the perspectives of records managers, archivists, cognitive psychologists, behavioural economists and sociologists to the ongoing dialogue.

A few words about terminology are in order before proceeding further, especially since each discipline and field of research defines terms differently. Coleman et al (2011) note that:

Crucial though records are to the efficient functioning of organizations worldwide, there is surprisingly little consensus on exactly what constitutes a 'record' in a corporate context, or on how the keeping of records relates to the management of documents, data or information. Some people, especially in the ill-defined world of 'information management', see records, data and information as very much the same thing; all of them are 'assets' of a financial institution, and all need to be managed effectively if the institution is to flourish. Others draw a distinction, emphasizing that data and information may be changeable over time, while records are stable and unchanging (or need to be made so). In the records management community, records are commonly associated with the particular activities and transactions that an institution carries out; each activity or transaction takes place at a moment of time, but the record ensures that a representation of the activity or transaction persists and can be used to provide evidence of, or knowledge about, what has occurred in the past. Records are created or received in the course of the institution's activities, and if they are retained by the institution and managed in a way that supports their continuing integrity and accessibility, the institution will be able to call on its records whenever it has a need for the evidence they supply. In a world where financial institutions are increasingly required to demonstrate accountability for their past actions, the ready availability of such evidence is essential.

The use of the term record also varies considerably in different legal contexts. Some laws and regulations pertaining to the financial sector refer to 'records', some use the term 'books and records', and some refer only to 'documents'. Many legal professionals prefer to use the term 'documents' as a broad term covering a wide range of materials, but in other disciplines the word 'document' carries narrower connotations. Computer scientists, for example, may use the term document to mean a narrative text (in hard copy or digital format) and may exclude data in databases from their interpretation of 'documents' (Coleman, et al, 2011).

The term 'data' can also give rise to confusion. In technological circles this word can be used specifically to refer to the content of databases; it can also be used more widely to refer to any kind of digital content. Within financial institutions there is often a perception that certain kinds of data fall outside the ambit of record-keeping systems. Many who use e-mail and database applications will confidently say that they produce data, but not records. Yet evidence of organizational activity is often to be found

within database applications (e.g. trading and accounting systems) and e-mail systems (Coleman et al, 2011).

Nor is the term information any less varied in use and conceptualization of meaning. Shannon (1993) observes that the word information has been given many different meanings by various writers in the general field of information theory and further notes that "It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field" (Shannon [1993], p. 180). Luciano Floridi (2011) refers to information as a "conceptual labyrinth." It is easy to see that any discussion on the subject of records, information, and data therefore is bound to be a tricky one, especially one that is interdisciplinary. Nevertheless, in this workshop, we will go where "angels fear to tread."

Returning to the themes of the workshop, the governance theme focuses on issues of accountability and decision making in relation to the valuation, creation, storage, use, archiving, and deletion of records, information and data. It also includes the processes, roles, standards, and metrics that are needed to ensure that financial institutions have the records, information and data they need for effective risk management and that regulators, in turn, have what they need. The papers by Willi Brammertz, Sherry Xie, and Alexandros-Andreas Kyrtis tell of the challenges of information governance that arise from what Kyrtis aptly describes as the `messiness and fogginess` of record keeping and records management in financial institutions. Kyrtis goes on to argue that financial bubbles and collapses must be understood, therefore, not solely in terms of problems of market risk (i.e. of inflated values which may collapse because of unfavourable market events) but also as organizational problems and problems of the way financial technologies are developed and applied in operations, in financial information management and in subsequent decision-making processes. Anya Savikhin`s discussion of her work on the application of Visual Analytics (VA) to asset pricing bubbles reinforces Kyrtis` hypothesis.

The governance theme raises broad questions of policy such as what is the right balance between creating and keeping records, information and data for risk analysis and management and the need to ensure that the operations of financial institutions are flexible and nimble enough to respond to changing market imperatives. Over-regulate and operations run the risk of bogging down the delivery of financial services in overly bureaucratic regimes. Under-regulate and the senior management, boards, shareholders, clients and regulators of financial institutions lack the transparency they require to properly assess an institution`s levels of risk. These are the questions that many financial regulators around the world are wrestling with today as they seek to put in place new regulatory frameworks, including frameworks for the governance of records, information and data, following the financial crisis of 2007-2008.

In an article entitled "What is Information Governance and Why is it So Hard?", Gartner analyst Debra Logan writes: "There is something buried in the definition which I think is at the heart of what I consider to be the problem that most of us face when we start talking about information governance. The word that matters most is *accountability* [author`s emphasis]. The root of all of our problems with information, and we do have lots of problems with it, is the fact that there is no accountability for information as such." She goes on to note that IT claims responsibility only for infrastructure and software, not the data. Rightfully so, she asserts, as the responsibility for data properly lies with business owners, but they are often too busy to address its management. More than this, however, they may not be properly incentivized to do so, as the workshop paper by Lemieux, Woo, Limonad and Monu suggests in analyzing the impact of the rising demand for private label residential mortgage-backed securities on the documentation of residential mortgages. Their paper prompts questions about

whether it is better to separate governance from the day-to-day management of records, information and data because of the problem of agency, a problem that Alexandros-Andreas Kyrtis also acknowledges in his paper. Given the problem of agency, it seems reasonable to ask whether information governance, as distinct from daily management of information, should be placed in the hands of more disinterested parties (i.e. independent committees within financial institutions or external regulators) than the business owners of the data themselves.

The papers presented by Mark Flood and Allen Mendelowitz and Willi Brammertz touch upon these and other issues of information governance, asking, and suggesting answers to, many important questions about how the way in which information is governed can either support or undermine effective risk management. In his paper, Brammertz tackles the Operational Risk that the Office of Financial Research, itself, will face if inundated with large amounts of data from financial institutions, calling for new approaches that combine what he refers to as the “mechanical” and the “subjective” components of finance. Drawing on many years of practical experience with designing risk analysis systems for financial institutions, Brammertz observes that systems in many of these institutions lack a key construct needed to accurately calculate risk; that is, the notion of “contract type.” Moreover, Brammertz proposes that contract type must be linked to the algorithms used to calculate their cash flows, with variations for market conditions and behaviour, so that calculations can become mechanistic and simplified. This, he suggests, will provide a model that the OFR can use to “look through the windshield”, as Senator Dodd proposed, to identify and address risks that could threaten financial stability.

The second theme with which this workshop engages is analytics. This, in itself, is an enormous area, but the workshop hones in on two aspects of analytics: 1) the ontologies, schemas and models that can be used to represent financial domains and financial risks and 2) the visual analysis of financial data. Since the Workshop on Knowledge Representation and Information Management for Financial Risk Management huge strides have been made in the area of representing financial domain knowledge. The Enterprise Data Management Council, a non-profit industry association seeking to enable financial institutions to share information on the business strategies and practical implementation realities associated with achieving enterprise-wide control over data content, and to raise the profile of data management at the senior levels and represented at the workshop by the Council’s Managing Director and our facilitator, Michael Atkins, has developed its Semantics Repository after many months of consultation and work with financial industry experts. The goal of the Semantics Repository, which has now been labelled the “Financial Industry Business Ontology (FIBO)”, is to standardize the terms and definitions of all reference data attributes stored in the master files of financial institutions and passed among supply chain partners. Precise nomenclature translates into a common language between systems and sources, reduces the cost of doing business, and promotes confidence in data among business users, according to the EDMC. As the Council points out on its website, the Semantic Repository seeks to provide a foundation for effective data management and an essential building block for business process automation (EDMC, 2011). The EDMC, in collaboration with OMG, the computer industry standards organization, has since built a Proof of Concept of the FIBO for the Commodities and Futures Trading Commission which represents OTC Derivatives. Specifically, it has been working on semantic representations for interest rate swaps and has demonstrated a preliminary capability of using FpML swap confirmations to categorize swaps. The group is now extending this representation to include the ISDA Master Agreement and its various schedules and annexes. In a statement on the collaboration between the two groups to build a common modelling language and symbology for the financial domain, they noted the value of doing so as the need to address the lack of high quality data, interoperability and portability across platforms and requirements to support the provision of accurate

information flows in near-real time to a wide spectrum of stakeholders and systems that is a necessary for efficient capital markets. A common but extensible language and symbology of terms, semantics to define complex inter-relationships, hosted in an open standards based semantic repository furthers the cause of finding solutions to data management challenges. Eric Chacon, Global Head of Data Standards at Citi and Dr. Harsh W. Sharma of Citi Chief Data Office, and OMG Finance Task Force Co-Chair have said that this work will help Citigroup and other financial services organizations better address the business and operational challenges emerging from the financial reform regulations such as the Dodd-Frank Act. Assessment and analysis of systemic risk not only requires Legal Entity and Financial Instrument Identifiers but also a deeper understanding of their complex hierarchies and traceability across front-middle-back office systems. Such capability can be better achieved by modeling the business events and concepts, their contextual nuances and inter-relationships supported by standard interchange formats across heterogeneous IT systems.”

Regulators also recognize that better managed data within financial institutions supports better management of financial systems and of systemic risk. The CFTC has set up a sub-committee to develop standards for describing, communicating and storing data on complex financial products. Explaining the plan in May, CFTC commissioner Scott O’Malia, said: “The data and reporting mandates of the Dodd-Frank Act place the CFTC in the centre of the complex intersection of data, finance and the law. There is a need and desire to go beyond legal entity identifiers and lay the foundation for universal data terms to describe transactions in an electronic format that are identifiable as financial instruments and recognizable as binding legal documents.” The CFTC has now appointed members from the EDMC (Mike Atkins), Citi, Barclays Capital, ICE, Google, ISDA, Finra, the DTCC, Goldman Sachs, Bloomberg, Thomson Reuters, NFA, ISO, Swift, CME Group, MarkitServ, Blackrock, docGenix, and the FIA. Federal agency observers from the SEC, Treasury/Office of Financial Research and Ferc have also been appointed. Chaired by CFTC chief economist Andrei Kirilenko it will hold three sessions over the course of 2011, with four working groups: product and entity identification, semantic representation of financial instruments, machine-readable legal documents, and storage and retrieval of financial data. The sub-committee will then present a report with its recommendations to the Technology Advisory Committee, which will then consider it and make recommendations to the CFTC.

Though huge strides have been made, there remains much work to be done, both in terms of articulating the theoretical framework of the FIBO as well as in translating the theory into more practical Proofs of Concept for different types of derivatives and other financial products, their relationships, and their flow through capital markets. Willi Brammertz’s paper suggests that the construct of Contract Type which, at the same time, incorporates the exact algorithm needed to calculate a contract’s cash flow, with adjustments for risk factors and behaviour, must be a foundational construct in any ontology for representation in the financial domain. The paper by Lemieux, Woo, Limonad and Monu also engages with this theme. Drawing on research work concerned with tracing the relationship between records and information and the build up of risk in financial institutions and systems during the financial crisis of 2007-2008, the work of these researchers points to the value to be gained by drawing on a diverse range of conceptual modelling approaches to the further develop and expand on the FIBO (e.g., Monu’s Organizational Actor Method which seeks to represent the cognitive state of domain actors may help to represent financial actors’ behaviour, for example) and to gain a clearer understanding of causality in relation to the dynamics of financial crises – as a ‘technology of thinking’ (Stenning, 2002) about our representations of these dynamics.

Another workshop theme addressed is visual analytics (VA). VA is defined as “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas and Cook, 2005). Many observers of the

financial crisis have found that traditional computational approaches to risk analysis let us down. In a recent paper, one of the most vocal critics of traditional risk models, Nassim Nicolas Taleb, the author of *The Black Swan*, writes:

In real life we do not observe probability distributions, we just observe events. So we do not know the statistical properties — until, of course, after the fact . . . Given a set of observations, plenty of statistical distributions can correspond to the exact same realizations—each would extrapolate differently outside the set of events on which it was derived. The inverse problem is more acute when more theories, more distributions can fit a set of data—particularly in the presence of nonlinearities or nonparsimonious distributions . . . The problem comes from the structure of the decline in probabilities for larger deviations and the ease with which the tools at our disposal can be tripped into producing erroneous results from observations of data in a finite sample and jumping to wrong decisions. (Taleb, 2011, p 8).

VA interfaces harness the power of human visual perception and cognition to overcome some of the limitations of traditional computational approaches, such as that described by Taleb. Typically, VA is used when the analytic problem is not sufficiently well-defined for a computer to handle it algorithmically. The purpose of VA is not to replace quantitative analysis, but instead to allow quantitative analysis to be more focused or to extend it (Grinstein and Ward, 1997).

VA is premised upon the human remaining a key part of the analytic process, blended with computational analysis. In conducting VA, people use dynamic interaction with visual representation of datasets to “ . . . generate new hypotheses when exploring a completely unfamiliar dataset, to confirm existing hypotheses in a partially understood dataset, or to present information about a known dataset to another audience” (Munzner, 2009). Further, VA allows analysts to see patterns and recognize relationships in large-scale data sets, a challenge that the financial institutions and financial regulators surely face with the volume of transactions through interconnected global capital markets. In a forthcoming book chapter on the visual analysis of Lemieux, Fisher and Dang (forthcoming) discuss many examples of how the visual analytics approach has been used in finance to scale up the information representation capabilities of traditional financial charts and graphs to deal with massive quantities of high-dimensional financial data sets.

The workshop papers by Anya Savikhin and Thomas Dang address the application of visual analytics to large sets of financial data for financial analysis and risk management. In her paper, Savikhin discusses VA’s application to interactively discovering information from large information sets, to improve financial decision-making and finds that VA reduces the cost of obtaining information, improves decisions, and increases confidence of consumers in decision tasks involving risk, such as choosing assets for a portfolio and identifying asset price bubbles. Thomas Dang’s paper addresses the goal of creating a functional evaluation framework of VA techniques with regard to investment analysis problems, as well as a table of existing products that is capable of supporting common investment analysis problems. Dang observes that with a functional evaluation framework and a table of off-the-shelf solutions, more effective and theoretically grounded feasibility and cost-benefit analysis could be performed to justify and plan applications of the visual analytics approach in financial organizations. A secondary goal of his paper is to explore and initiate an in-depth discussion on the problems of file formats and information management, which pose considerable challenges to the implementation of VA solutions as in the deployment of other automatic and semi-automatic data analysis techniques. Both Savikhin’s and Dang’s papers suggest that visual analytics presents many opportunities to see data in new ways that can support more effective risk analytics and risk management. Both papers identify many promising avenues of future research in the theory and application of VA in financial decision-making.

The final theme of the workshop is long-term digital preservation. Often a forgotten aspect of the management of records, information or data, long-term digital preservation is critical to creating the capacity for longitudinal studies of market dynamics and of risk in financial institutions and financial systems. Here, it is important to make a point that long-term does not, as one might expect, just mean the preservation of records, and information and data so that they will be accessible hundreds of years from now, although this is an important objective. Many institutions are experiencing trouble right now with retrieving and accessing data in as little as three to five years from the point of its creation. This is due to technological obsolescence and change, as well as to a failure on the part of institutions to take measures to ensure that digital records, information and data are created in forms that will persist over time and space and that such records are properly archived. Many institutions have relied upon backup tapes to archive data, but this has proved to be a universally ineffective strategy as backup tapes are susceptible to loss and deterioration and their format makes it particularly difficult to retrieve specific and demonstrably reliable records. In recent years, largely in response to financial regulation (e.g., SEC rule 17a-4) and high-profile litigation (e.g. Zubulake v UBS Warburg), many financial institutions have introduced new technologies to ensure that archived records, information and data can be reliably maintained and rapidly retrieved when needed. Yet, very little work has been done to address the risk factors that can lead to long-term deterioration of reliable and authentic digital records, such as work to address file format obsolescence, ensuring digital records can still be read even after changes in data structure, improving system documentation, ensuring records can still be retrieved from decommissioned systems, improving audit trails in data migration, and managing uncontrolled accumulation of records. Consequently, as a worst case scenario, even in the face of rivers of records and data, financial institutions and regulators may still be unable to rely on having access to critical information beyond a three year window. Further, as we move to establish new standards for domain representation, we need to consider how data and systems created using earlier versions of these standards can be accessed or interpreted when newer versions have been released. Arguably, at every stage of managing the life cycle of records, information and data we should be considering how our present choices affect future accessibility and reliability of records and information.

Much archival research on long-term preservation of digital records has focussed on unstructured data, while many of the approaches to the long-term preservation of structured data have arisen from research communities little concerned with the preservation of digital objects that must provide impartial and reliable evidence of business transactions on which decision-makers, risk analysts, and historians of the future can rely. A merging of these different approaches is needed in order to develop practical and workable standards and strategies for financial institutions and financial regulators to use in preserving financial records, information and data in digital form for the long haul. The paper by Sherry Xie discusses some of the research on long-term digital preservation of authentic and reliable records that has taken place as part of a ten-year research initiative involving an international research consortium led by Luciana Duranti of the University of British Columbia's School of Library, Archival and Information Studies. The word InterPARES, which is used as an acronym for International Research on Permanent Authentic Records in Electronic Systems (the name chosen by the consortium for the project), means "among colleagues" and serves as a model of a successful and sustained research partnership that, it is my hope, we can emulate to address the issues associated with managing records and information for financial analysis and risk management. As one of the key findings of the InterPARES project attests, a deep understanding the processes and practices of record keeping – what Bruno Latour (Latour, 1986) calls the "everyday techne" of organizational records making and keeping practices - will be necessary for the development of effective standards and strategies to achieve long-term digital preservation of financial records. Alexandros-Andreas Kyrtis points out that the meaning

or ``signals`` in records derive as much from the communities of practice within organizations and ``financial techno-organizational gestalts`` as they do from their content. Thus, meaning is fluid and socially-negotiated. In order to have permanent and authentic records, digital preservation must, unpick socio-technical ``black boxes`` to implement measures that halt and fix the inexorable state change of financial data as it flows through systems and markets.

Records, information and data are not, in themselves, saints or sinners, as the title of this introduction to the workshop might suggest. No – it is what we do or fail to do with them that will produce either a good result or a bad one. The very fact that this workshop is taking place acknowledges the insufficiency of what we, to date, have or have not been doing to create and manage records, information and data to meet the challenge of effectively analyzing and managing financial risks. More must be done, but what that is still remains open to discussion. The papers presented at this workshop move that discussion forward, suggesting many fruitful avenues for further basic and translational research to provide the financial services industry and financial regulators with the concepts, the knowledge, the methods and the tools to tackle complex records, information and data management challenges

Victoria L. Lemieux

August 20, 2011

References

Coleman, L., Lemieux, V. Stone, R. and Yeo, G., Eds. (2011). *Managing Records in Global Financial Markets: Ensuring Compliance and Mitigating Risk*. London: Facet Publishing.

Enterprise Data Management Council. (2011). ``Website.`` Available at:
<http://www.edmcouncil.org/default.aspx>

Floridi, L. (2011). "Semantic Conceptions of Information." In Zalta E.R. Ed., *Stanford Encyclopedia of Philosophy* Available at: <http://plato.stanford.edu/archives/spr2011/entries/information-semantic/>>.

Flood, M., Kyle, P. and Rashid, L. (2011). "Knowledge Representation and Information Management for Financial Risk Management." Report of the Workshop, University of Maryland, 21-22 July, 2010. Available at: <http://www.nsf-fiw.umiaccs.umd.edu/docs/FIWreport-FINAL.pdf>.

Grant, Jeremy. (June 28, 2011). ``Global Derivatives Lexicon Marches On.`` Financial Times. Available at: www.ft.com/intl/cms/s/0b170fa4-a00c-11e0-a115-00144feabdc0,Authorised=false.html?_i_location=http%3A%2F%2Fwww.ft.com%2Fcms%2Fs%2F0%2F0b170fa4-a00c-11e0-a115-00144feabdc0.html&_i_referer=http%3A%2F%2Fwww.ciferresearch.org%2Fresearch%2Fnews_and_press%26cid%3D261#axzz1VXI3nEa4

Latour, Bruno (1986). "Visualization and Cognition: Thinking with Eyes and Hands." In Henrika Kuklick and Elizabeth Long, eds. *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, JAI Press, 6.

Logan, D. (2010). `What is Information Governance and Why is it so Hard?` Available at:
http://blogs.gartner.com/debra_logan/2010/01/11/what-is-information-governance-and-why-is-it-so-hard/.

OMG. (March 16, 2011). ``OMG and Enterprise Data Management Council to Collaboratively Develop a Sustainable Framework and Data Standards for Financial Reform.`` IT Briefcase. Available at:
<http://www.itbriefcase.net/omg-and-enterprise-data-management-council-to-collaboratively-develop-a-sustainable-framework-and-data-standards-for-financial-reform>

Shannon, C. E., (1993) *Collected Papers*, edited by N. J. A. Sloane and A. D. Wyner, New York: IEEE Press.