SWIFT INSTITUTE

SWIFT INSTITUTE WORKING PAPER NO. 2014-009

THE ROLE OF BIG DATA IN GOVERNANCE: A REGULATORY AND LEGAL PERSPECTIVE OF ANALYTICS IN GLOBAL FINANCIAL SERVICES

DR. DANIEL GOZMAN
PROFESSOR WENDY CURRIE
DR. JONATHAN SEDDON

PUBLICATION DATE: 01 DECEMBER 2015

.

The Role of Big Data in Governance: A Regulatory and Legal Perspective of Analytics in Global Financial Services

Dr. Daniel Gozman
Henley Business School
University of Reading

Prof. Wendy Currie
Audencia Nantes School of Management

Dr. Jonathan Seddon
Audencia Nantes School of Management

Abstract

This paper discusses how global financial institutions are using big data analytics within their compliance operations. A lot of previous research has focused on the strategic implications of big data, but not much research has considered how such tools are entwined with regulatory breaches and investigations in financial services. Our work covers two in-depth qualitative case studies, each addressing a distinct type of analytics. The first case focuses on analytics which manage everyday compliance breaches and so are expected by managers. The second case focuses on analytics which facilitate investigation and litigation where serious unexpected breaches may have occurred. In doing so, the study focuses on the micro/data to understand how these tools are influencing operational risks and practices. The paper draws from two bodies of literature, the social studies of information systems and finance to guide our analysis and practitioner recommendations. The cases illustrate how technologies are implicated in multijurisdictional challenges and regulatory conflicts at each end of the operational risk spectrum. We find that compliance analytics are both shaping and reporting regulatory matters yet often firms may have difficulties in recruiting individuals with relevant but diverse skill sets. The cases also underscore the increasing need for financial organizations to adopt robust information governance policies and processes to ease future remediation efforts.

Table of Contents

1. Introduction

Business analysts expect the big data market to grow to $32.1 billion by 2015 and to $53.4 billion by 2017, where 2.5 quintillion bytes of data is produced daily, with 90% of world data created since 2012 [92.]. Correspondingly, within global markets we have seen the extensive adoption of technology, the globalization and consolidation of industries as well as increasingly unpredictable and dynamic business environments [21, 36, 112]. Types of organizations affected include exchanges, banks, brokers, insurers, data vendors and technology and services suppliers. One feature of this environment is an increasing focus on rules and regulations designed to protect a firm's employees, customers and shareholders as well as the economic wellbeing of the state in which the organization resides. Another is the growth of analytics and data pertinent to the enforcement of such rules and laws [4, 64, 117]. For example, a United States regulator, the Securities Exchange Commission, has used big data analytics to estimate risk metrics for funds, leading to six enforcement actions by 2012 [119]. Specifically, our study focuses on how global financial institutions are using big data compliance analytics[i] within their governance operations to manage legal and regulatory obligations [120]. Thus, the study examines the broader issues of how new business is being created in the advent of big data, where more legal services are required to understand and interpret financial regulation (law firms), how government and other interested parties hire third party firms to search for patterns in data (eDiscovery), and how software vendors are increasingly developing compliance systems and wrap around consultancy services to help their financial clients meet regulatory mandates (IT vendors).

Our aim is to critically evaluate the implications of the pervasive use and commercialization of big data analytical technologies in capital markets from a legal and regulatory perspective,[ii] that is to meet regulatory obligations and handle related litigation, and thereby, seek to identify both contradictions and complementary factors arising between the post-crisis regulatory and legal landscape and the current socio-technical environment to guide both policy makers and practitioners. Consequently, our data collection focuses on two key areas: the use of big data analytics to facilitate day to day compliant trading, and the use of big data analytical tools when serious legal and regulatory breaches occur. We are guided by the following high-level research question and more granular sub-research questions:

- How do big data technologies intervene in the management of regulatory breaches?
    - What are the implications of post-crisis global regulation on collection, usage and maintenance of data in global financial organizations for daily trading activities?
    - How can big data tools intervene when serious but rare legal/compliance breaches occur to analyse structured and unstructured data across the enterprise?

[i] Compliance analytics or just analytics hereafter refers to calculative functions for meeting regulatory obligations which utilise algorithms and draw upon data sets with volume, variety velocity and veracity [4, 118]. Visualization software (e.g. dashboards) may then be required to present the outputs in a way where it is easily understandable to humans [90].

[ii] Although it is not our intention to provide insight into the application of specific laws or regulatory mandates.

This paper is structured as follows. First, we present our review of the literature on big data in financial services, with contributions from information systems and financial sociology. Research on big data is attracting interest from information systems [1, 3,] accounting and finance [8, 9] and general management [44] fields. We combine this work with recent contributions on the topic of big data. Second, we draw on risk management perspectives to frame the research context and highlight the relationship between our case studies. Third, we discuss our methods, data collection and analysis. We then present two illustrative case studies where we gathered in-depth interview data from fifty three interviews and secondary data sources including white papers, press releases and speeches, regulatory mandates, marketing materials and commentary from legal and accounting firms. Next, we consider the implications of our findings from policy and practitioner perspectives. Then we consider future avenues of research. Lastly, we draw the study to a close with some concluding comments.

2. Literature Review - Big Data in Financial Services

KPMG suggest vast increases in regulation and compliance, with big data seen as an important part of the narrative [67]. Furthermore, the financial services sector has always been a heavily regulated yet a data and technology driven industry [22]. For example, it is estimated that upwards of seven billion shares are exchanged daily. Approximately two-thirds of this figure is traded by computer algorithms based on mathematical models which analyse vast quantitates of data in order to maximize gains and reduce losses [76]. Big data in financial services covers many areas, including regulation and compliance, customer data, transactions between institutions and networks and risk management. Sub-sectors of the financial industry are asset management, banking (commercial and retailing), capital markets and insurance. Currently, vast amounts of digital data are being continuously created through the rise in use of digital devices and interfaces [77]. Select examples include: indexing and benchmarking, referencing data for different securities (e.g. ratings and classifications), pricing (including real-time, snapshot and end of day), issuer details and operational data (e.g. logs on order activity, system and network performance). Key factors contributing to this phenomenon include the widespread availability of the internet with increased download speeds, the use of mobile smartphones and the cloud, coupled with falling storage costs [76]. Related fundamental properties of digital technologies are re-programmability and homogeneity [112]. These socio-technical changes have resulted in services and related analytics becoming pervasive and globally accessible [113]. The potential business value of such data is increasingly being recognised, captured and harnessed by financial services businesses and regulatory agencies. As a consequence, many financial services businesses are seeking to understand how to best collect, aggregate and exploit their own and commercially available data to the maximum without falling foul of legislation outlining intellectual property or data privacy rights [34, 64]. Data privacy in particular has been highlighted as providing challenges for big data analytics as concerns around surveillance and integrity grow [62, 74].

Defining Characteristics of Big Data

As volumes of data have increased correspondingly academic and practitioner interest in big data has grown in recent years where multiple definitions have emerged. One definition states, 'big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time' [101] Another common definition, the four Vs, has focused on what differentiates big data from common analytics suggesting the volume of data sets, the speed of data creation and availability (velocity), the variety of data types (e.g. social media, emails, videos, GPS signals) and the trustworthiness and integrity of the data (veracity) collectively define this phenomenon [24, 34, 77, 113]. In addition, the topic of big data spans numerous sectors (variety), e.g. finance, healthcare, manufacturing, social media. Our purpose here is to focus on the financial services industry since the 2008 financial crisis.

Related Studies and Research Motivation

The 'pace, volume and origin' of the change affecting the financial services industry is unprecedented [116], and this is linked to a range of factors, including: high frequency trading (HFT), money transfers; payments technology; peer-to-peer finance and portfolio analysis, all of which rely on the speed and accuracy of data flows in an increasingly networked and international financial industry. Big data offers rich opportunities and challenges for information systems' researchers [1, 3] and more generally for practitioners across the range of industry and not-for-profit sectors. In a recent editorial in a leading management journal, the editors noted that big data is now widely used in the business community but, 'there is very little published management scholarship that tackles the challenges of using such tools – or, better yet, that explores the promise and opportunities for new theories and practices that big data might bring about' [44].

What is new about big data and, do we need new theories and concepts to help us understand it? A good starting point is the topic of structured and unstructured data in relation to the financial industry. While both types of data in the banking industry are growing, concerns about how to understand vast amounts of unstructured data are emerging as an increasing concern for regulators [36, 43]. Furthermore, how to regulate the financial industry in the light of algorithmic and high frequency trading [117] where traders operate in geographically diverse and fragmented financial markets is also of interest to policy-makers and commercial firms. Analytics are becoming increasingly important when rare but serious compliance breaches occur. More and more specialist tools, such as eDiscovery tools are being utilised to traverse large volumes of structured and unstructured data held within organizations but across borders to help evaluate compliance breaches and assist with litigation. Business analysts [125] suggest, 'Big data has been a reality for eDiscovery for longer than it has in most other application areas. The volume of information collected in response to legal and regulatory challenges has grown from thousands, to hundreds of thousands, to millions of documents over the last few years.'

While information systems' researchers focus on the IT artefact, often at the level of the organization, financial sociology has much to say about the policies and practices of traders operating within increasingly regulated financial organizations [66]. While the practice of managing large data has been a perennial topic for information systems for decades, few studies are situated within financial services which link important topics of regulation, compliance, technology and the professional practices of individuals, such as lawyers, compliance managers, fund managers and traders. Prior work on managing technology in financial services has widely addressed data and information issues around trading [18, 114, 115] and more recently, on analytics and inter-organizational standards in the mortgage industry [75]. The move from manual based to electronic trading following the 'Big Bang' in 1986 has generated interesting studies about the use of technology and data by fund managers and traders [88, 89]. A study on regulation and IT following the financial crisis observed the scope of the credit crisis and resultant great recession (marked by the collapse of Lehman Bros and actions required to save Northern Rock) extended well beyond the corporate failures of the dot.com era [51]. However, there are relatively few studies from the information systems' community that focus on the wider policy issues relating to financial regulation, technology and data.

So how can literature from information systems and financial sociology inform our research enquiry on big data in the financial industry? Our review suggests that a good way to gain a deeper understanding of big data in the financial industry is to combine work from other disciplines. In this research, we draw from information systems and the sociology of financial markets [15]. These disciplines operate within silos in management research. While both these fields offer interesting and insightful studies on various aspects of information systems or the financial services industry, the topic of 'data' and especially, 'big data' is under-represented and under-theorized. Alongside issues about volume, velocity and variety of big data, our interest in starting our research enquiry was not simply to absorb the hype about 'big data' as a worthy topic for empirical investigation but to identify the issues and concerns (if any) facing people in the financial industry. Another motivation for our study was to position the topic against the background of the post financial crisis, where banks and other financial firms now face stringent requirements to meet regulatory mandates.

Key Concepts: Technological Affordances, Neutrality and Performativity

Regulations and laws are not objective but require social interpretation [28]. Compliance systems and related analytics underpin internal controls and risk management efforts as interpretations of rules, norms and logics become encapsulated within IT artefacts [51, 82]. The analytics they provide create their own world view which alters the perceptions of those decision makers the system was designed to inform [54]. In this way, information systems and underlying data play a key role in underpinning governance practices by both affording and constraining actions [46, 124]. The concept of technology affordance relates to the potential actions which an individual or organization (with a particular purpose) can perform with a technology or information system[73]. Big data analytics provide affordances for automating operational

and strategic decision making essential to knowledge roles (including compliance managers, traders and lawyers) [74]. Thus, compliance related technologies which draw from big data analytics, 'might authorize, allow, afford, encourage, permit, suggest, influence, block, render possible, forbid...' actions on a daily basis and thus have the potential to both help and hinder desired outcomes [68]. Correspondingly, technological constraints and affordances are viewed as composite of intertwined human agency, 'the ability to form and realise goals', and material agency, 'the capacity for non-human systems to act on their own apart from human intervention' [69]. Such studies highlight how analytical technologies, utilised by financial organizations are not neutral in the information and affordances they provide and the responses they elicit [123]. An associated literature stream within the social studies of markets, termed, 'the technological constitution of financial markets', addresses how technological arrangements may define boundaries and delineate domains of activity, thereby legitimizing and institutionalising them [88, 89].

This work focuses on how various mechanisms, devices, and technologies not only represent markets and finance through a form of passive recording but how they actively shape and constitute the markets [118]. Such properties are referred to through the term 'performativity'[14]. Performativity is a term used in social science and refers to the quality of a theory or calculation (in this case analytics) to not only describe a phenomenon but to also shape it. For example, Callon[14] suggests, 'Economics does not describe an existing external 'economy' but brings that economy into being: economics performs the economy, creating the phenomena it describes. For example, Mackenzie and Millo[72] conducted research at the Chicago Board Options Exchange and found that option pricing theories (e.g. Black Scholes) shaped derivatives markets. We build on this concept to suggest that big data analytics are not only used to describe compliance but are also shaping regulatory responses and new mandates.

Related studies have focused on the performativity of formulae and models enacted through technology to construct economic activity [71, 72]. Correspondingly, IS scholars have highlighted how big data does not merely describe social media for example, but also, through automated analytics, shapes the reality of social media. Thus, big data analytics are viewed as also having performative characteristics [19, 122] Complex analytical technologies are integral to financial markets and in turn structure and influence transactions and also the rules and laws which govern them. Thus, technologies may be seen as cultural tools which enact the markets [7]. Data analytics are seen to represent, 'the material and discursive assemblages that intervene in the construction of markets' [80]. From this perspective, markets may be viewed as technological arrangements composed of artefacts and formula which project their own paths of action to create 'calculative agencies' [14]. Consequently, such analytical tools may be viewed as having their own agency and ability to exert both constraining and constitutive effects, they co-exist with human actors and so are co-participants in socio-technical networks. To summarize, the pervasive adoption and use of big data analytical tools may create performatvities, in the form of calculative agencies, which may produce unintended as well as intended outcomes.

The Post-Crisis Landscape

The financial crisis of 2007-2009 and the resultant Great Recession has highlighted how the failure of financial organizations may have dire economic and social consequences at a national and global level. As a result, the G20 and regulatory bodies worldwide enacted regulatory change focused on plugging the gaps in regulatory systems that have become apparent as a result of the crisis and also post-crisis governance failures, such as the Libor or Foreign Exchange (FX) rate rigging scandals. Correspondingly, trust in financial and regulatory organizations has seriously diminished in recent times [95, 111]. Despite the extensive use of mathematical models within capital markets which give an aura of impartiality and reliability, finance is not physics and to a large degree operates on trust and faith ultimately underpinned by the availability and accuracy of underlying data [112]. The FCA's Risk Outlook for 2014, which outlines the major risks the industry is facing from the regulator's perspective, highlights asymmetric information as an ongoing risk: 'Information asymmetries – when one party in a transaction has more or better information than the other party – are common in most retail and wholesale financial markets' transactions. They potentially affect outcomes along the distribution chain, causing mis-selling and reduced trust and can affect market integrity if used to benefit the firm at the expense of one or more conflicted clients' [36]. Thus, at a time where volumes of digital data are increasing exponentially, the ability of big data analytical tools to provide transparency into financial organizations' daily operations and decision making is becoming increasingly important and thus deserves scrutiny and research.

3. Framing the Study's Context through Operational Risk

The Basel Committee on Banking Supervision [5] defines Operational Risk as, 'the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events.' While a related category of risk, termed 'Compliance Risk', addresses, 'the risk of legal or regulatory sanctions, material financial loss, or loss to reputation a bank may suffer as a result of its failure to comply with laws, regulations, and rules.' Often firms organise their compliance function within their operational risk function as there is a close relationship between compliance and operational risk. A third relevant risk category is termed 'Regulatory Risk', which refers to the risk that a change in regulatory rules and laws may impact a business [5, 6, 97]. These definitions provide us with a useful point of departure from which to consider the use of big data technologies for managing compliance and investigating breaches. In a paper for the International Monetary Fund [61] Jobst suggests, 'the typical loss profile from operational risk contains occasional extreme losses among frequent events of low loss severity (see Appendix A). Hence, banks categorize operational risk losses into expected losses (EL), which are absorbed by net profit and unexpected losses (UL), which are covered by risk reserves through core capital and/or hedging.' The LIBOR and FX rate rigging scandals and rogue trader malpractice are examples of rare operational risk events leading to considerable fines and reputational damage [12, 47, 48, 78].

We build on Jobst's representation of operational risk in order to frame our study and illustrate the relationship between our two otherwise very distinct case studies, see Figure 1. The two case studies

collectively illustrate how analytics are used in investigating and managing expected (case study 1) and unexpected (case study 2) regulatory breaches at both ends of the operational risk spectrum.

The first case study addresses regulatory breaches which occur on a mundane basis and are predictable to the extent that technologies have been developed to specifically manage these breaches which occur in organizations engaged in similar business practices around the world on a daily basis. We focus on an Investment Trading Platform (ITP) which manages day-to-day compliance of trading practices. Such systems deal with vast amounts of data in the form of market pricing, benchmarks, compliance rules and risk calculations, all of which are constantly shifting and changing. Such systems must also maintain an audit trail of all transactions occurring within this data swirl.

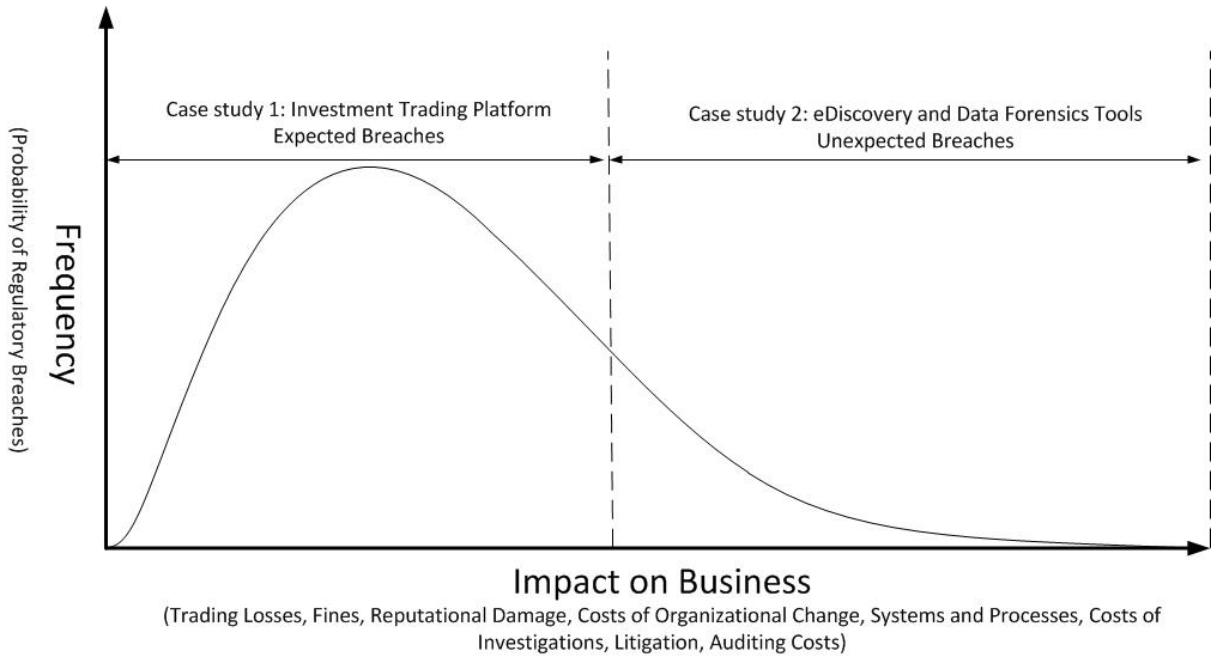


Figure 1. Frequency and Impact of Regulatory Breaches

The second case study addresses low probability breaches which occur much more rarely and are often distinguished by huge fines and substantial changes and refinements to regulatory frameworks. A report authored by the UK Government Office for Science [50] titled 'High Impact Low Probability Risks' addresses ways to mitigate risks which are unlikely but yet may have catastrophic consequences. The review provides useful insight into the nature of such risks suggesting, 'The identification of low probability risks and the subsequent development of mitigation plans is complicated by their rare or conjectural nature, and their potential for causing impacts beyond everyday experience.' [50]. Within this paper we apply the concept of high impact low probability risks in financial services to regulatory matters. These events may be characterised by regulatory authorities instigating complex investigations, perhaps operating across multiple jurisdictions and countries, often across multiple-organizations, each with global operations. Consequently, financial firms' subjected to regulatory investigations and litigation are increasingly required to perform their own internal investigations into the vast amounts of structured and unstructured data held within their organization.

Drawing from the four Vs definition, both cases' analytical capabilities are reliant on sets of data which are voluminous, rapidly changing and varied in type. While the analytics they generate are, in the first case study, trusted by financial firms, regulators and investors to maintain ongoing compliance or, in the second case study, by regulators and legal counsel to uncover crucial evidence which may prevent or lead to multi-billion dollar fines.

4. Methodology

While theory development is usually a key priority for information systems (and management) researchers [2], our reading of the current big data literature convinced us to adopt an inductive approach (theory building) rather than a deductive (theory testing) approach[100]. This is because it seems inappropriate to construct a theoretical straight jacket around a topic which is currently ill-defined and heavily promoted as the latest 'buzzword' in management and IT consulting.

To fulfil our research goal, we selected one case study on a leading financial software vendor, headquartered in the US, with a London-based office serving clients ranging from very large international banks to small hedge funds. For our second case we selected an eDiscovery and data forensics consultancy, also based in London and also serving a variety of financial organizations worldwide. The study used semi-structured interviewing techniques with 53 interviews conducted across both cases, with managing directors, senior business managers, relationship managers, software developers, sales personnel, lawyers, data forensic experts, project managers and eDiscovery consultants. Further interviews were carried out with clients of both firms to explore issues on regulatory compliance and handling financial data. These informants included compliance officers, traders, fund managers and IT engineers. Interviewees provided insightful responses to questions about the post-crisis regulatory environment and about the use of information technology for data governance and compliance.

Our inductive (theory-building) approach allowed us to build our case studies initially from a series of pilot interviews with informants from the software vendor and consultancy. From the outset of this study it was important to develop a working definition of the concept of 'big data' relevant to the financial industry and the technology under investigation. The results of these interviews with business and IT managers showed that big data was characterized in three ways. First, informants discussed big data in terms of increasing volumes where lawyers, compliance managers, fund managers and traders now work with granular data (reported on an item-by-item basis). Second, the velocity of data has grown where data is frequently updated and analysed in real time. Third, the variability of data has increased where data can be structured or unstructured (i.e. text, video).

To control the scope of our study, our interview schedule situated 'big data' around how the software vendor and consultancy was changing products/services and client requirements for meeting regulatory compliance mandates and conducting regulatory investigations. Our aim to impose discipline on our research design by carrying out open-end interviews on a more narrow range of areas and topics in

an attempt to avoid some of the methodological pitfalls facing qualitative researchers. A common problem being that qualitative interviews generate numerous amounts of data which is 'messy' and difficult to organize [99]. The result is often an over-scoping of the study, where the phenomenon becomes lost in translation as the situations and contexts to which informants refer are not well defined.

As big data is one of the latest buzzwords in management research, our interest was to interview a range of informants using a semi-structured interviewing approach, as this would enable them to describe and reflect on their thoughts and perceptions about 'big data' and also to consider the extent to which today's financial data offers new potential for research enquiry than was previously the case. Data analysis was conducted through long established interpretive techniques for analysing data through the recursive identification of patterns, first through categorization and then abstraction [45, 52, 79, 95, 99, 102, 103]. During the process of data analysis, primary and secondary data were closely reviewed to determine points of importance and interest [79]. Common themes were identified and categories assigned for each case independently [100]. Thus, long interviews were simplified through the adoption of simple categories [91]. The analysis adopted a two cycle approach to coding. The first cycle adopted a 'Descriptive Coding' approach for summarizing segments of data. This method is appropriate for inductive studies utilizing semi-structured protocols [95]. This approach requires the application of a content phrase to a segment of data representing a topic of inquiry, and so related to the risks and challenges being faced for example 'Regulatory Investigations', 'Unstructured Data' or 'Changes in Data Volume'

The second cycle adopted a 'Pattern Coding' approach to identify major themes by searching for causes and explanations from the data. Such an approach builds on the first cycle of analysis and are, 'explanatory or inferential codes, that identify an emergent theme, configuration or explanation. They pull together a lot of material into more meaningful and parsimonious unit of analysis' [79]. Examples include 'Performativity', 'Affordances' and 'Data Heterogeneity'. Scope, depth and consistency were achieved by discussing key concepts, constructs and terminology with each of the informants and triangulating the findings across primary and secondary data sources [41, 96]. Secondary data included white papers, press releases and speeches, regulatory mandates, marketing materials and commentary from legal and accounting firms. For example, interviewee references to particular areas of regulation were triangulated with the original regulations and industry commentary to ensure key points were fully understood and consistent across sources.

Summing up our approach, we agree, 'the real strength of qualitative research is that it can use naturalistic data to locate the sequences (how) in which participants' meanings (what) are deployed. Having established the character of some phenomenon, it can then (but only then) move on the answer 'why' questions by examining how that phenomenon is organizationally embedded'[100]. Our case studies provide a practice oriented narrative of how big data technologies are deployed to meet regulatory mandates. Our intention is not to treat technology as a 'black box' but to provide a detailed description of how technology is developed and applied in conjunction with regulatory change.

5. Case Study 1: High Probability Regulatory Breaches

Charles River Development: Compliance Processes and Automated Systems

Charles River Development (CRD) (www.crd.com) is one of the leading providers of front and middle-office investment management systems (CRIMS). Its roots can be traced back to the 1990's when, working with Putnam Investments in Boston, it developed a system which was able to capture order details and provide pre-trade compliance checks in addition to those made at the end of each trading day. Whilst other vendors offered compliance capability, what differentiated CRD was how it integrated its software with a company's existing trading system. Capturing and checking a fund manager's orders against compliance regulations before sending them to the trading desk provided a valuable service which could identify a potential breach before it had occurred.

Any investment management company will have to check up to three different sets of regulatory requirements. At the national level, directives such as MiFID must be followed and whilst not necessarily the most restrictive, heavy fines will be incurred if they are not followed. The second collection of constraints will come from the investment management company which will have guidelines supporting its own blend of policies, such as the only sectors where it will invest. The third set of regulations come from large investors themselves (for example the trustees of a pension) who will stipulate how their money is to be managed (for example, no tobacco or alcohol). In order to be able to validate these rules, Charles River has developed different types of compliance checks which can be run individually or combined with each other:

- Exclusion rules will prevent a security from being held, based upon a condition such as its type or the country where it was issued
- Counting the number of entities that are held, for example no more than 10 securities from Japan
- Concentration tests such as no more that 7% of holdings will be held as cash. These rules can become quite complex, for example the UCITS 5/10/40 rule only allows more than 10% of assets in a single issuer if the total value held in issuers that invest in more than 5% of assets does not exceed 40%
- Logic tests around 'if-then-else' conditions
- Use of customer defined variables in the calculations of numerator and denominator values

Whilst the regulator will not recommend a specific investment system to automate workflows, they will impose a fine on a company if it persists in using manual systems (such as Excel) if they feel that the regulatory risk of such a practice is unacceptable. One compliance officer commented:

> *'We were fined because of the high risks that the regulator had associated with us running $1bn fund on Excel. The fund managers refused to use other tools because of the unique way that they modelled. Six months after the fine, they were still using Excel, and the threat of another much bigger fine forced them to use an automated product.'*

The head fund manager at another company was fired and a heavy fine was imposed after his use of Excel to manage his holdings had failed to price positions at the correct value. The portfolio had lost over 60% of its market value but this was not reflected in how the holdings were managed. When first purchased any compliance system will replace the existing workflows that were used to check current holdings, providing a more accurate and complete picture. Inevitably, this will un-earth breaches that were hidden because of limitations in earlier tools and the adjustment may also lead to penalties as investor holdings are corrected for breaches in mandates. Fines will vary but will generally be only a few thousand pounds, in addition to the cost of making the holdings compliant. For example, the costs incurred in selling out of a position that was not permitted, together with the costs of investing in new securities and any loss that had not been realised must be borne by the investment company. A passive breach occurs when market prices move and push holdings out of tolerance and these are generally picked up during the overnight checks. These failures are picked up at the start of the next business day and will form part of the fund manager and compliance officers' workflow.

Initially, the OMS (Order Management System) was just used for daily activities when at the start of each day past positions were cleared out and new data was fed from the accounting system (a process called flush and forget). Over the past decade, data volumes have grown phenomenally. Now, every activity that touches upon an order needs to be recorded and this is then used for reporting actions. New modules that require the storage of historical data are used for historical 'what-if' compliance functionality. The delivery of performance measurement and risk tools by definition now need vast sets of historical data for the analysis and calculations that are required. One of the other more recent data intensive processes is a module which allows the system to operate for many days without direct updates from an accounting system (the Investment Book of Records, or IBOR).

The volume of data (the first of the four Vs mentioned earlier) needed by CRIMS has increased following the development of functionality that is now provided, not only from the new modules mentioned above but also from the additional rules that now need to be checked. Every new type of test is often accompanied by new data that is required for its evaluation. The speed of data creation, its velocity, has also seen a marked change with the increased use of algorithms to process orders. It is far more common for the trader to process automatically the more liquid security whilst working on the stock that is harder to trade. The third V, variety, can be seen to have increased with new types of derivatives as securitization has increased the complexity of tests in the post-crisis regulatory environment. These rules now look to integrate forward price curves and risk measures as part of their calculation and new reports which provide further management information on breaches are now expected. The fourth V, veracity, or the trustworthiness and integrity of the data have also increased as a necessity. Now, numerous data providers are needed because of the wide spectrum of instruments that are traded and the dual pricing of securities that is often used. As the complexity of the compliance activities increases, the confidence in what is being supplied and mapped has to move in step.

During the early part of the 2000's, CRD embarked on three simultaneous projects. The first was to re-write the entire system using Microsoft .NET, removing functional and visual limitations imposed by the original language that was used (PowerBuilder). Not only did this offer access to the next generation of tools but it also enabled the full development of web services. The second was to continue adding new functionality required by its existing user base and industry. The product had developed from a dedicated compliance system to an integrated order generation and trading platform for most types of instrument. To have stopped any development would have caused future sales problems as by now other competitors were in this space offering rival products. CRD had to remain competitive by offering better or improved functionally to that delivered by its rivals, some of whom had developed their order management systems (OMS) on more recent programming languages. The third area, and for some customers the most controversial, was the development of additional systems that were tightly linked with the trading platform and offered an enterprise solution. Given all of the tasks that CRD was trying to juggle, some customers were worried that focus would be lost, development would grind to a halt and the product would simply become less competitive over time. By the end of this decade, CRD had successfully re-written their product (with more than 7,000,000 lines of code) and was offering additional functionality to run what-if scenarios on historical data, complex derivative analysis and check compliance rules at any stage of the trading process. To achieve all of this additional functionality the data that was now needed for the system had increased significantly. Not only had the OMS database increased in size to accommodate the audit of every operation that had occurred during an orders life cycle but so too was the volume of data required to support models, benchmarks, indexes and derivative calculations.

In addition, the system also included new modules that could be used for investment related activities, namely, IBOR, performance measurement, risk and settlement. Each of these systems ran from the OMS database that was used during its day to day operations. The benefit of having just one source for all of the information that was used by each system is obvious. However, the additional increase in the database size and its ability to continue operating efficiently meant that the knock-on effects of any part of the process failing could be catastrophic for any company.

Today, CRD employ over 500 people and has offices located in the Americas, Asia-Pacific, Europe, the Middle East and South Africa. Its product is used by over 350 clients in 44 countries. CRD has divided its business between the following sectors: institutional asset and fund management; private wealth management; banking; hedge funds (alternative investments); insurance and pension funds. No one sector dominates its sales' focus and this ensures that product development does not become skewed or that CRD face excessive risk should any sector of that market change (for example the crisis that began with hedge funds and investment banks with unregulated derivatives in 2007-9).

Evolving Analytics and Regulatory Complexity

It used to be that if an accounting system was used to run all of the compliance checks, then there was a limit to the complexity of mandates that could be run and the ease at which they could be customized. One of the main challenges with earlier accounting systems was with the technology used. Whilst they were capable of calculating values for concentrations or checking for securities that should not be held, changing their structure to add new security data types, let alone calculating the necessary analytics for a conditional rule, was expensive and took a long time to deliver. In some systems, it was simply not possible. Regulations themselves also became more complex in the conditions that they were looking to evaluate. Whilst an accounting system was used to generate general ledger reports, it was the software tools that were written with more modern languages that provided the solution for intra-day compliance calculations. Within CRIMS, not only was it now possible for customers to write far more of their required rules but new rule types or enhancements to existing rules could be included much more quickly in an upgrade. The following table shows the main changes to the CRD compliance system since early 2000.

Early 2000's	Mid-Late 2000's	Early 2010's
Now possible to have an order checked by another fund manager before it is sent to the trading desk	Improved compliance action time by adding emails functionality following any checks failure	Ability to capture all of the compliance data to allow for a historical re-run using what-if scenarios
Improved compliance functionality for a fund manager and a trader	The ability to evaluate a check in any currency, and not just that used by the portfolio added additional compliance functionality	Logic allows for a compliance check at any stage in an orders life cycle, allowing transactions to occur at the boundaries of regulatory rules
More user defined fields that could be included in compliance tests	Improved audit of the compliance process by additional failure comments	In addition to the systems rule structure, now user defend logic can also get embedded
Additional templates for different country rules added to system's library	Much more comprehensive derivative and debt security information	With new and better data feeds mapped into the OMS, improved capability for advanced analytical calculations

Table 1. Changes to Compliance Functionality

Every major system release would deliver additional compliance logic, data fields and internal calculations that could then be used in tests. It would also add to the rule library that provided templates for the rules that were required in different countries. In the early 2000's, the compliance functionality was focused on checking the orders that were being proposed by the fund manager (pre-trade) and how the trader was filling them (post-trade). Even if there were tens of thousands of accounts and rules that needed testing, running all of the data into the system before running the compliance engine was easily done before the start of the trading day. Within the decade, this system was capable of processing hundreds of thousands of accounts before trading began.

The changes that were seen during the mid-2000's coincided with the rapid increase in regulations both in the US (Sarbanes-Oxley Act 2002) and Europe (Undertaking for Collective Investments in Transferable Securities - UCITS, Markets in Financial Instruments Directive - MiFID and Capital Requirements Directive – CRD Alternative Investment Fund Managers Directive - AIFMD).

Following the Sarbanes-Oxley Act, it became obligatory for any non-US company that was trading US securities to follow all of its regulations, as well as those that were required within Europe. Not only did the data required for the regulators increase the volumes of what was required for a test, but so too did the complexity of the data that was needed. The regulator has also added to the types of data that are required to satisfy new checks. For example, in the Solvency regulation tests, the OMS is expected to calculate exposure values from forward risk curves. Whether it is the capability of the technology that forces the changes to the regulations, or whether they are simply reflecting the changes in how investments are made, compliance products continuously have to evolve to incorporate the new tests that are being created.

In order to assist with managing the growing complexity and volume of rules CRD offers a compliance advisory service where their experts will spend time to review and offer recommendations to rule changes that need to be made if the checks are to be as efficient and effective as possible. Given the sheer number of compliance checks that need to be run and the overwhelming number of regulations and mandates that have to be interpreted, optimizing this process is a crucial part of the systems' successful operation. One of the problems that occur at all customer sites is that client's own rule library simply grows as new rules and accounts get added. Underlying data may change or be updated, rules can often become duplicated or the conditions that will fire a warning may no longer be appropriate. Some clients are using this service and letting their own customers know that they have done so:

> *'We use this service because customers are more than ever before wanting to know that we are properly monitoring all of our positions, and this service is effectively a rubber-stamp from the software vendor that we have done this properly.'*

As the regulatory landscape becomes more complex, IT vendors need to develop their products taking into account multi-jurisdictional financial regulations. One of the strengths of the CRD compliance tool is the ability for customers to write their own tests with an English-like syntax developed by the company. This, combined with the library of 1,700+ templates for 35 regulatory bodies (including rules from regulations such as UCITS, MiFID, Dodd-Frank and Sarbanes-Oxley) meant that each customer could quickly trade new types of security and comply with differing requirements. These compliance checks can now be run at any stage of the trading cycle and can even be used to take a holding in a particular security to the limit of what was allowed. There are always rules that are subject to a company's own interpretation (such as the fair distribution of a partial executed order) however a full audit trail of every change to the order during its life is also recorded in the system.

> *'How you interpret fair-allocation is one of UCITS requirements that no-one has ever defined. We have written our own report that is fed from the Charles River database because all of the data that we need is saved in the order history table.'*

At the time of writing, IT vendors need to take into account the burgeoning financial regulations, particularly those surrounding new EU financial services regulation on data protection. Measures in this area include banking capital requirements and rules on the derivatives markets, among others. Since the

2007-9 financial crisis, over 40 new laws have been proposed with many coming from the G20 discussions. One Compliance Manager stressed:

> *'The proliferation of new rules and regulations mean that we are constantly playing catch-up. We are currently focusing on the data privacy and security legislation as we know this is going to become even more important in the light of recent events.'*

Compliance as a Service

Having evolved the product from offering only compliance, in the 1990's, to an enterprise solution, CRD now also offers Compliance-as-a-Service (CaaS) which is a standalone web-based solution. Such a service is fed by client trading inputs (e.g. positions, account details) and provides reports on the status of each test. As a managed service using CRD's data, this offering represents a commoditization service which is targeted towards small sized customers which do not have the in-house skill sets to manage their regulatory checks.

Originally the system was sold to be run at each customer's site. All of the hardware and infrastructure was maintained and supported by each company's own IT team. For smaller sized customers (e.g. fewer than 10 people) this was often the only system that they had. Whilst it did not represent the best tool for every type of instrument that could be traded, it was a leader in the areas of compliance and equity. Not only was the cost of owning multiple systems seen as problematic but integrating them together was seen as cost ineffective. The limitations of this OMS were viewed very differently by those companies who already had in-house expertise to manage multiple vendor tools. Larger companies with mixed asset portfolios offered their fund managers and traders the best tools that they needed. An example of how an OMS could be integrated in a large company is given below. Typically, there would have been separate systems for the different trading classes (e.g. equity, derivative, debt) as well as additional systems that were used for performance measurement, risk and order matching.

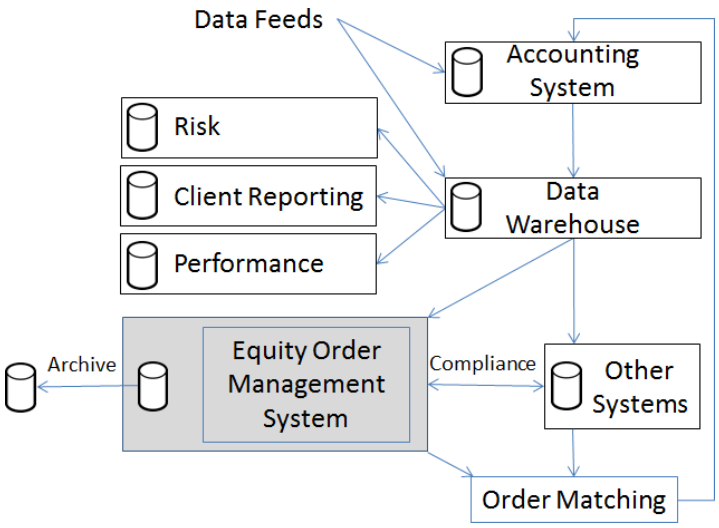


Figure 2. The Integration of several Order Management Systems

For medium and large customers the cost of implementing and supporting all of these systems was high. These costs were from running and maintaining the databases used, (different products might not even use the same database provider), support for issues that occurred during the trading cycle, customization of workflows and upgrading to later versions. Even if the product worked as required and no new functionality was needed, customers would still need to upgrade because all the other software that was necessary to run the system would eventually stop being supported. For example, Microsoft stopped supporting its server NT 4 software in 2004, having replaced it with Windows 2000. Oracle would replace and retire its database products every five years. Having incurred the costs of installing and running the system, additional cost had to be planned. The risks associated with running an un-supported product, in an environment where compliance is one of the key factors in winning customer confidence meant that every three or so years the OMS had to be upgraded. The complete suite of different trading systems ultimately became a patch-work quilt of integrated products. The same data was stored in numerous locations and the problems associated with ensuring that the correct value was being used for a calculation or a report only increased over time. The original compliance issues that were faced when CRIMS was first introduced (timeliness, accuracy, quality of compliance rules, issues with 24x5 trading) changed by the end of the decade (data quality, increased reliance upon analytics, new instrument types, global regulatory requirements).

Faced with the increased complexity of managing all parts of their OMS, CRD now sold its product as a software as a service (SaaS) package. CRD does not yet offer a SaaS in the same way that Google or Salesforce do. Those firms have products which are neither as configurable nor as customizable as CRIMS. Although their user bases are vastly bigger, the complexity of their product is much lower than that of CRD. CRD managed all aspects of the systems' day to day operations whilst the customer would access the product across the internet or a private network connection. Using such architecture, this vendor was able to manage its software and data feeds whilst the client would focus on what it did best – managing its clients' portfolios. An example of this structure is shown below, see Figure 3.

The OMS was now able to offer functionality that could compete with that offered by other systems (such as modelling a derivative product), but all of the data that was now required was stored on a single database. Now the same data that was created by the risk module was also used in compliance reports or a fund manager's model. This database has to allow every user fast access whilst at the same time scaling to the increased data volumes that need to be stored. These volumes would only ever increase. Each of CRD's hosted customers will have their own version of CRIMS running in a private cloud. They will have their own 'space' and can feel confident that the firewalls will protect any unauthorized access to their data. The size of the database required for the OMS had increased fivefold between 2000 and 2012. Adding to this, the massive data sets that were required for all of the historical compliance analyses, performance measurement, risk and the IBOR reinforced the benefit of outsourcing

all of this data to the cloud. Not only would this become the responsibility of CRD to ensure that it was adequately managed but so too would tasks such as archiving and recovery.

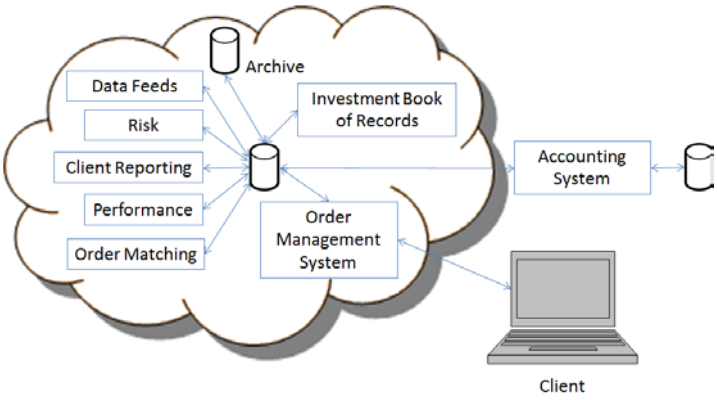


Figure 3. Example of a cloud hosted OMS

The idea that the vending company would be better at managing the tools they have developed is not new, but this has only become a realistic proposition as the speed and capability of the internet has increased, combined with the improved capability of software tools and internet security. CRD provides two variations on SaaS. The first is where the clients system is maintained in the cloud and accessed across a secure private network. The second is where the client hosts their own software whilst engineers at CRD manage its day to day operation. Both of these offerings are combined with a data service that guarantees to feed the correct numbers into the system's reference database.

Data Management

One of the greatest issues with any financial system is underlying data quality. End-users were concerned if they received incorrect analytical calculations, for example if a bond yield was wrong or its coupon payment was not included in the opening day's cash balance. The fact that either the data was wrong (incorrectly mapped or translated) or the security had been incorrectly set-up for calculations (e.g. its duration), was something that was often ignored. Under the time pressure to trade, looking at the 'detail' of what might be wrong was simply not of interest. If the system could not be used to create an order because it did not show the right details, then the view from the traders or fund managers was that the product was simply not good enough. It was up to the operations and back-office team to feed the system correctly and make sure that it was all working but with multiple data feeds and continuous security additions capturing all of the data become a complex struggle. This problem had always been recognized by CRD and in the mid 2000's they provided 'scrubbed' data that could be fed into their system. What changed was a new direction by CRD to use cloud technology to remotely manage or host the software thus ensuring that the system would be maintained as it should be.

Prior to CRD launching its data service, companies would receive their feeds from many different sources and the files that were imported often went to multiple systems which required the same data. Differences between each provider's data files meant that multiple import translations would need

to be run to ensure that the values that were saved in the target databases matched. For example, one provider might show a security price in cents and another might use dollars. In small companies all of the data would have been fed into CRIMS as this represented their secure master copy of the data.

There are many definitions about what big data represents. Common to every one of them is that the volumes of data that are available for analysis and strategy are vastly bigger today than they were a few years ago. With CRD, not only has the volume of data increased for the OMS activities but the data that is required to support the new applications (such as performance measurement) has also added to this. The major issues that surround data relate to when it is delivered (any compliance checks need to be made on the latest values), what translations need to be made (similar securities from different countries could have different precision for trading values), and how each value is mapped into the CRIMS master copy of the data. This issue is compounded by the continual changes that are made with new securities and data feeds. Moreover, the high volumes of data that are required for decision and risk systems, not forgetting how the data is stored and quickly accessed, adds additional complexity to this problem. Providing a single end to end solution is only realistic when all of the required underlying data is provided.

Independent research commissioned by CRD, (with a sample size of 4,941 users), showed that not only did 62% of the respondents believe that the number of data sources they were going to use would increase, but also that 50% of their current data had to be scrubbed. Thus, not only did companies face high costs in collecting and using this data, but looking forward, this was only expected to increase. In another survey (where 42% of the respondents had a portfolio management system), 75% of their responses indicated that improved speed (as well as time to market) and access to data was their number one priority. Many financial companies have used Excel or Access simply because it is quicker to write a desk-top application that is able to use new data than to wait for this to be fed into a master database (or data warehouse) and accessed via a trading system. Knowing what additional data is going to be required (e.g. OTC derivative values, new compliance rule needs or advanced analytics) has only increased the pressure on the technology department to support the business.

The CRIMS has an open SQL database, (i.e. it is not a black-box into which feeds are stored but users are unable to run their own queries against them), with interfaces to numerous data sources, (e.g. Markit, Bloomberg and Thomson Reuters), and it has in the past few years developed its own data service. In partnership with a data provider, CRD offers a ‘white labelled’ data service. This is sold as a single point solution where all of the sourcing (validation), aggregation (normalization), enrichment (augmentation) and mapping have been done for the client.

Real-time pricing from over 135 global exchanges offers additional capability, such as capturing prices across an orders’ lifecycle, depth of market and trends. There are over a 1,000 fields available for the vast array of different security types that are traded (e.g. derivatives, debt, equity as well as indexes

and benchmarks). Looking at just the active bonds, there are well over 1,000 issues from just 27 countries. Not only are the data volumes large but they are also set to grow in in conjunction with the growth of securitized products.

Due to the breadth of functionality that CRIMS offers, improved data veracity (the fourth V) had benefits with users from different disciplines. Integrating this service with SaaS is an attempt to provide both a turn-key solution as well as maintaining an existing client base of unique implementations. The impact of poor quality data has a massive impact on the banking and financial services' industry. High quality data will reduce errors throughout the investment lifecycle, adding confidence to the compliance events that are run. It is hard for a firm to justify any IT investment project which does not have a direct impact on regulatory work, risk or customer satisfaction. Given how data within CRIMS influences all of the stages between decision support, compliance, trading and post-trade actions, the significance of proper data handling cannot be under-estimated.

Although improvements to other asset classes had begun well before the middle of the 2000's, it was only towards the end of this decade that the system started to gain traction as a multi-asset product. Problems that initially accompanied this increased capability could be seen as a consequence of the low quality data that it was fed and how instruments were set up (for example if a debt bond is to calculate any accrued income that is due, then the instrument needs to have all of its factors correctly set up). Attempts had always been made to ensure that, if the data was correctly mapped to a security, then the analytics would be correct. However, the number of ways that a security could be set up meant that if the fund managers or traders did not see the value that they expected, (they would often use other systems such as Bloomberg to check that the figures which CRIMS had calculated were the same – irrespective of whether that number was correct), they would assume that the OMS simply did not work. Any ability to gain traction and build up a user group of satisfied debt users was always going to be a more difficult task when compared to the equity and compliance user base. Key to the success of the system was a drive to encourage end users to use the data that was supplied by CRD. These data services began by maintaining specific interfaces to the main data provider feeds. This ultimately became part of the data service where CRD simply provided all of the data.

6. Case Study 2: Low Probability Regulatory Breaches

Millnet: eDiscovery and Legal Document Consultancy

In our second case, we study a full service eDiscovery firm. Millnet is one of the UK's largest legal data services and document solutions providers, with clients in over 60 countries. The firm was incorporated in 1996 and has evolved from providing traditional legal print services to providing electronic document consulting, processing and review. Millnet's clients' include Legal 500 firms and FTSE 100 companies. Unlike the previous case, Millnet is not a software vendor but instead utilises best in class eDiscovery software to provide consultancy, infrastructure and expertise to support the investigation and review of structured and unstructured electronic data (including, emails, voice recordings, video streaming, chat rooms, spreadsheets and text based documents) held within financial organizations, which may relate to serious internal investigations, litigation or regulatory breaches. Specifically, the firm works with a number of software vendors. Appendix C summarises Millet's service offering within financial services and the technology vendors it partners with to support those activities. Millnet recently moved premises and invested £1M in a new facility. This investment allowed them to double their square footage to facilitate growth in personnel, allowed for the integration of purpose built forensic and server rooms, and upgrades to their data network security and biometric entry control systems for quarantined areas.

Gartner suggests that the eDiscovery market is in a stage of high growth where maturity, innovation and consolidation are all increasing. They estimate that the size of the enterprise eDiscovery software market was $1.8 billion worldwide in 2014, with a five-year compound annual growth rate of 12%. A key driver of this growth is the growth of digital data expanding the scope and size of eDiscovery cases (Zhang & Landers, 2015). Data preservation for compliance purposes holds particular challenges in relation to data growth. As data sources grow in variety and volume, organizations will have to deal with increasing complexity during data collection and processing. This presents an increasing cost centre for regulated financial firms not least as the number of regulatory investigations has grown considerably since the financial crisis. The increasing commercialization of such tools and their growing importance is reflected in Millnet's double digit growth and increase in headcount for its eDiscovery services.

Millet's senior management view their expertise and access to multiple technologies, and correspondingly their understanding of the strengths and weaknesses of each tool set, as an essential foundation for their eDiscovery and document services consulting business. Rather than focus on a single vendor partnership, Millnet focuses on multiple vendor's technologies. However, they view kCura's tool, Relativity, is the current market leader. The firm's Managing Director, views the firms competitive advantage as stemming mainly from the firm's employees' experience and capabilities with different technologies. He commented,

'Our focus over the last couple of years has been to really understand the technology, all of the technologies whether it's a Relativity or a Nuix. They've got all these clever pieces associated with it, use predictive coding, use this, use that. Well, what does that actually mean? What can you actually do with it? And so by working with the technologies and understanding how we can bring different capabilities into our workflows, we can improve our results and save money for our clients whilst improving the quality.'

While the different technologies supplied by a variety of vendors have different strengths and weaknesses, their broad purpose is the same. As the descriptions in Appendix C show, the boundaries between data forensic tools and eDiscovery tools are becoming blurred. However, data forensic software focuses on preserving evidence in its most original form, while eDiscovery tools search through vast amounts of data and through different data types held by organizations to identify relevant documents, to be disclosed in the course of a regulatory investigation, legal case or internal investigation. Appendix D outlines further case studies showing how eDiscovery tools have created value in legal matters.

Prior to the use of eDiscovery tools, organizations, in partnership with their legal teams, were required to review paper documents and print outs of a relatively small number of electronic documents and to disclose relevant documents to the courts or regulators. However, as data has grown exponentially this approach has become increasingly problematic. A Millnet eDiscovery Consultant outlined the genesis of such systems,

'It's now called eDiscovery and had grown from what was called Litigation Support. The original was premise was simply IT support for lawyers and the legal team, either like say regulatory investigation dispute, any legal dispute, which involved collating loads of documents together to assist with the legal teams' review process. Over 15 years eDiscovery has evolved from scanning paper documents into systems to delve through someone's laptop and going into data such as in pulling out every single email for the last five years, so it's a complete change from paper to electronic.'

Regulatory Investigations

The Financial Services Act 2012 requires the UK regulator, the FCA, to conduct an investigation into a potential regulatory failure and subsequent report where both parts of a two part test are met.

Part 1	Part 2
Where events have occurred in relation to a regulated person or others which indicated a significant failure to secure appropriate consumer protection, or had or could have had a significant adverse effect on our integrity or competition objectives.	The events might not have occurred or the adverse effect might have been reduced but for a serious failure in the system established by the Financial Services and Markets Act 2000 (as amended) (FSMA) or the operation of that system.

Source: (FCA, 2013)

Table 3 Regulatory Investigations: The two tests

Regulatory investigations may often incorporate 'dawn raids.' Such raids are defined as searches of individuals and businesses offices, often carried out in the early hours, by the FCA under warrant and in the presence of a police officer. The FCA undertakes these raids in order to prevent the removal of laptops, desktops, PDAs and mobile devices and the destruction of electronic documents and paper files. A key motivation is to obtain a complete list of customer records so that they can be contacted. From

2012 to 2013, the number of dawn raids conducted by the FCA almost doubled from 11 to 20 raids (RPC, 2014).

Dawn raids represent a source of revenue for Millnet. Its website states, 'We have been instructed in numerous cases to assist firms in responding to dawn raids instigated by regulatory bodies. This type of investigation would commence by trying to establish what information the regulator has taken. For instance the same computers or servers that the regulator had an interest in would be forensically captured. We would then analyse the data, for example search the data set using key words, date ranges and/or key individuals. The documents would then be reviewed by the law firm and corporate to try to ascertain what the regulator may be looking for and to establish what exposure the corporate may face.' Millnet's Director of eDiscovery Advisory & Business Development commented further,

> *'There's definitely a lot more activity for sure, a lot more dawn raids than there were before, there's a lot more regulatory activity. One, because of the regulation but I think also probably because of public pressure as well, after the financial crash. The public want, to see them be a lot more active and so I think they're under pressure themselves to be more active in the marketplace.'*

However, regulatory investigations do not always take the form of dawn raids. Regulators have the power to require financial organizations to conduct internal investigations and report back. Where regulators suspect that inappropriate actions may have occurred or want to clarify that they have not, they may instruct financial organizations to conduct an investigation and submit relevant analysis and comments in prescribed format. For example, when the UK regulators wish to enlarge the scope of its investigation perhaps on the basis that analysis of other firms' records (e.g. chat room data) suggests that further organizations have been involved in rate rigging (e.g. LIBOR or FX), they may instruct organizations to conduct investigations into individuals and specific sets of data over a defined period of time. Where such malpractice is thought to be widespread, the regulator may require firms to prove they have not been involved. Such investigations may be costly as the regulator may come back to the firm and ask them to widen the scope by including more individuals, more data types or lengthen the time periods reviewed. Often the timeframes for reporting back are tight. In such cases, financial organizations often look to their general council who in turn, may look to external legal firms and eDiscovery consultancies for additional resource and expertise.

A key challenge faced by such firms is meeting tight deadlines for disclosing information to regulatory authorities, which our respondents highlighted as often being tight, particularly if things do not run to plan initially, yet are non-negotiable. Consequently, law firms with eDiscovery capabilities as well as consultancies are much in demand. Simmons and Simmons, a major global law firm with over 1,500 people and 22 offices worldwide, have sought to develop their own eDiscovery capabilities in-house. However, they often require support from consultancies such as Millnet, not least as currently firms may struggle to find individuals with the correct mix of legal and technical knowledge.

Regulatory Challenges

Where regulatory matters involve a multi-jurisdictional element, the use of eDiscovery analytics raises challenges in information gathering and collation whilst observing each jurisdiction data protection and privacy. Due to the types of information financial organizations hold data privacy laws are highly relevant, particularly when financial information may be used in correlation with big data analytics, An eDiscovery consultant commented,

> *'I think with financial institutions it's often more important than other industries because a lot of people are quite sensitive about how much money they earn, what they spend it on. With the right analytics your spending habits can reveal your habits, your health, your gambling or whatever it might be. Through big data all sorts of things can be revealed in that data. But there is a difference between the volumes of data and the usefulness of the data. Correlation is not the same as causation. Systems are only as good as the algorithms that are used to work out what it's all about, because unless you've got some effective way of interpreting data, picking up those patterns, deciding what it is, you're not going to get any sense out of it. So there might be intellectual property or licensing issues around some of that as well as confidentially and data privacy.'*

EU directives prevent personal data leaving the EU however member states have applied more stringent data protections laws. For example, Germany goes further by prohibiting personal data leaving the German Federation. French law prevents (subject to international treaties or conventions and applicable laws and regulations) disclosure of documents or information to be used as evidence in foreign administrative or judicial proceedings. Conflicts may exist between two sets of regulatory regimes, where financial regulators require information to be disclosed by a global financial organization from its foreign operations, which is prohibited under that countries data protection regime. Nicola Fulford, a Head of Privacy at law firm Kemp Little LLP who specialises in Data Protection laws commented,

> *'Well it kind of depends where the companies head quartered often, and so for example if you have a US head quartered company in US litigation and they want discovery of information in Germany or France that relates to French or German employees, the data protection authorities have said no you can't have it, and I have heard of people using that as a shield, kind of deliberately storing information in those countries so that it wasn't discoverable and what have you, but it tends to be because the courts in the US can issue such huge fines and exemplary and different types of damages to punish people as well as to deal with the losses, so the US head quartered company will generally rather have the* [data protection] *fine than a US court fine. It's very much a situation of between a rock and a hard place and there's no easy way to fix it. You're going to be in breach of one or the other, which fine is worse? Which regulator is tougher?'*

Fines levied by data protection regimes have been increasing in recent years yet are still dwarfed by those levied by financial regulators. For example, in the UK fines levied by the Information Commissioner's Office (ICO) totalled £160,000 in 2010 but by 2014 reached £1,520,000 (IT Governance Institute, 2015). However, proposed future changes to EU data protection laws will increase fines to up to 5% of annual worldwide turnover, or €100m, with the possibility for individuals and associations to also bring claims for non-compliance (Long, 2013). In addition, firms will be required to enhance their governance of the personal data they hold and document their processes. Other related potential conflicts

exist between the requirement for financial organizations to know their customers' details for anti-money laundering purposes and sharing that data with authorities overseas in breach of local data protections laws. However, all the legally trained practitioners we interviewed all concurred that anti-money laundering obligations would normally 'trump' data privacy laws.

Data Management

The variety, velocity, and volume of the data integral to regulatory investigations pose specific challenges. As Millet's website states, 'Banking matters tend to involve vast amounts of information and can often include unusual file types such as Bloomberg messaging and audio files.' A key challenge for those conducting regulatory investigations is reviewing a vast 'universe' of structured and unstructured data and then narrowing down the amount of files which are actually passed on to be reviewed by expensive legally trained individuals, whose time should be maximised. Caroline Hunter-Yeats, Partner at law firm Simmons and Simmons and client of Millnet commented, on how the volume and variety of files has grown in recent years at a rapid pace along with the technology employed,

> *'About six, seven years ago now, electronic data was becoming more of a challenge previously when it was all hard copy lots of paper files came in and we had to deal with it manually. We could just print the emails. Over the last two, three years the volumes of data have gone through the roof. You're no longer dealing with data sets that tend to bulk out to about 20 to 30,000, you're talking about millions. So from a lawyer's perspective, they are going from, 'I got a box of files or maybe on a bad day I got ten boxes of files, to, I've suddenly now got a warehouse full. A conceptual warehouse full and you're obviously not going to print them all out. So big data for us, or what counts as big, is things in the millions. And actually to be honest, things in the hundreds of thousands, anything where you're not going to be able to have a whole bunch of humans looking at it. The last two years have seen developments in the infrastructure, both the software and the hardware that enable people to do a lot more a lot quicker. We're talking days for hundreds of gigabytes, days rather than weeks.'*

The need for eDiscovery systems to deal with a variety of file types has become increasingly important. The need to investigate chat data has become common in regulatory investigations, particularly those involving multiple organizations, for example the recent investigations into rigging the LIBOR and FX benchmarks. Yet, several of the study's participants highlighted chat data as providing particular challenges. A Millnet eDiscovery consultant commented,

> *'We are seeing more of chat room data because people are not just using emails, they're using chats, they're using their internal chat programmes, and they're using the Reuters and Bloomberg chatrooms. Chat data are big, long streams of text, maybe 800 pages. It comes out in long transcript and is not pretty on the eye and is not easy to review. More often than not it's got hundreds of hits and somebody just has to sit there and go through it. Also, you see a lot of noise so everybody coming in and out you see, everyone's email, everyone's company disclaimers, and you've got to wade through all of this and within that there may be something dreadful going on. But how, as a human being, you're going to find it? The other challenge for chat rooms, it's the phraseology people use. So it's not text searchable easily because people don't say, 'I'm going to go and manipulate x.'*

Our participants highlighted how technologies allow for the reduction of 'noise' essential to allowing human reviews. An experiment conducted by Simmons and Simmons using two individuals to review

the same set of five documents revealed that chat data could be reviewed 40% quicker using an eDiscovery platform which removed the 'noise'.

In addition to unstructured data, structured data (data held within relational databases) also presents challenges. Financial organizations often have large numbers of bespoke, vendor and legacy systems containing vast amounts of structured data. Examples include customer relationship management tools, accounting tools and trading and risk platforms. Data schemas inherent in such systems allow the data held to be accessed quickly and easily to facilitate business as usual processes. The first case is an example of such a system. The foundation of eDiscovery tools is the ability to turn unstructured data into structured data. That is, to identify, analyse, search and present vast quantities of unstructured data. In order to do so; the system creates a database of structured data populated by unstructured data. Thus, eDiscovery tools ensure that the data held within the database is searchable and can be presented in a format which is easy for humans to understand. Consequently, it may be assumed that taking data which is already structured and importing it into an eDiscovery tool might be easier. An eDiscovery Project Manager commented,

> *'Structured data is a strange one because it feels like it should be the Holy Grail. All of eDiscovery is about taking unstructured data and turning it into structured data, that's what the damn process is all about. And the data is already structured, it should be easy. You should be able to run your queries and find all your relevant events or client log activities or whatever it is. And my experience is that you almost never can.'*

There are several reasons why analysing structured data, held within information systems, through eDiscovery tools presents additional challenges. Often, the information systems implemented by financial organizations contain structured data not designed for eDiscovery purposes but are instead designed for people to conduct their day-to-day work, for example, systems which maintain customer data. This often creates problems when conducting eDiscovery searches, where the data schema of the database is not designed to facilitate related queries. Another reason cited, was that it is often not easy to mine the data from the system. Software vendors may not include functionality to allow the extraction of the data as it is not usually necessary and the inclusion of such functionality may provide opportunities for data theft. These challenges are eased where organizations use well known systems from vendors such as Microsoft or Oracle. However, further challenges occur, where the eDiscovery team may not have access to the vendors' license and their data schema or related design documentation, or where the system in question was bespoke, and the design is not obvious or is a legacy system no longer supported by the vendor. An eDiscovery Consultant commented,

> *'So extraction doesn't exist to a huge degree, which is really bizarre and it means that, on the occasions that we do end up doing structured data in a huge way, it ends up being treated much more like forensics because you are having to piece together a system, quite often from its back end without its interface, which you normally don't have a licence for, or perhaps an installer for, or just perhaps an environment in which you can install them. So you're picking to bits a database which, it's much, much worse than unstructured data because the unstructured data is basically a load of formats that we deal with every day. Yeah, the data schemas are difficult to recreate. But decoding these structures, if it's noisy or not obvious how to recreate something that's useful can be difficult.'*

A common challenge across both structured and unstructured data types includes the need to understand what constitutes duplication and so to remove irrelevances in the data. For example, email trails are often duplicated where individuals forward or reply to existing email trails. Duplication complexity is increased where emails are held in different formats across numerous devices, including the exchange server folder, local inboxes on desktops and laptops and mails stored on mobile devices. Furthermore, while each email may look similar to a human, each mail's meta-data relating to author, recipient, date and time will also differ. An eDiscovery consultant provided an example of the problems meta-data can cause,

> *'I can give you a real world example, which is if you created some documents in 2012 and today you copy and paste them onto a USB stick, actually what you'll do in doing that is you will reset the creation date of the copy documents to today's date. Now you'll get some people that will do a collection where they say, right, we want all documents, I don't know, related to miss-selling between 2009 and 2011. If the IT department has gone at some stage and copied the documents from one system onto another they have basically reset the creation dates, so there'll be great chunks of documents there that actually aren't within the search'*

Other complexities occur in defining and applying keyword searches which run the risk of being, 'both over- and under-inclusive in light of the inherent malleability and ambiguity of spoken and written English' (Sedona, 2007). Simple keyword searches when used in isolation may only reveal 20% of relevant evidence in a large, complex dataset, such as an email collection. Instead, search terms should be thoroughly tested for efficacy part of which would include sampling to ensure that categories are neither over nor under inclusive and that there exists an iterative feedback loop to ensure that terms are refined appropriately (Gonsowski, 2010).

More recently, eDiscovery vendors have sought to incorporate more automation in order to assist with the increasing data complexities. Where key word searches are unable to deal with the variety and volumes of data being considered, predictive coding is increasingly used when there is a need to investigate large volumes of varied structured and unstructured data in a cost effective manner. Predictive coding involves using sophisticated algorithms to determine the relevance of documents based on feedback from a human. Instead of junior staff reviewing large volumes of data, the senior partners will review and code a 'seed' set of documents. As this process continues, the system learns more about the coding approach and begins to predict the reviewers' coding. At the point where the reviewers and systems coding are sufficiently similar, the system is deemed to have learned enough to make confident predictions regarding the remaining documents. In 2012, a landmark judgment in Ireland allowed predictive coding as part of the discovery process. The judgement explains that, 'The plaintiff's initial scoping exercise involving a key word search yielded 1.7 million potentially relevant documents. By September, following removal of duplicates (deduplication) and documents in other languages, that number had reduced to 680,809 documents suitable for predictive coding. [It was expected that] less than 10% of the 680,809 documents would need to be reviewed if predictive coding is employed [and] estimated that a traditional linear review, using a team of 10 experienced reviewers, would take 9 months

at a cost of €2 m leaving supervision and technology costs aside, whereas, the use of predictive coding would enable the plaintiffs to make discovery within a much shorter timeframe and at substantially lower cost.' However, a senior forensics consultant commented that predictive coding also had limitations particularly where different languages have been employed,

> *'And so the analytics piece* [in predictive coding] *doesn't just look for the words but it looks for the content around the words, so it looks for the placing of the words, how it sits within, with other words and so I can imagine it would be challenging where you've got somebody flipping between languages so, which we see, so the start of the email chain will be in English and then suddenly it moves into another language and then it moves into another language. To be very honest they're the documents you want to look at first...'*

The final challenge the participants highlighted was the management and inclusion of paper documents into the process. Where investigations need to include documents which are held in paper form from several years ago, additional challenges exist as these documents must be scanned and turned into electronic documents before being analysed. Crucially, during the scanning process optical character recognition (OCR) must be used to make the newly created electronic searchable. However, the use of this technology also creates problems. A Solicitor and eDiscovery Consultant commented,

> *'We had a team of people out in the, somewhere near the Ukraine, and they'd spent three months scanning a warehouse full of papers. If you've got a whole load of material that has been scanned at some point, what you've done is you've created photographs of that, the documents, and you then run an OCR process to extract the text out of the documents. Depending how that's done, the quality of the OCR will vary, so if the thing is badly scanned, it won't work very well. If there are manuscript comments on the side of the documents and someone isn't manually reviewing what's been scanned, those manuscript scribbles won't get picked up at all. So actually if you gave five different organisations a million documents to scan, I would actually guarantee to you that they will not remotely have the same results by the end of the process. And I'd say that paper is definitely the original problem. Often the OCR process is not brilliant, even if it's only the white space changing. Paper is also hard to de-duplicate, and scanned docs won't automatically fall into the right concept categories that your electronic documents will, and paper just has this whole raft of issues. Lawyers have been coping with them for so long they often forget they exist. And some find the technology a bit daunting and run home to paper, which is a huge problem.'*

7. Implications for Policy, Practice and Research

Analytics and Performativity

The Economist [106] recently noted that, 'The new masters of the financial universe are neither bank bosses nor hedge-fund titans. They are the regulators whose job it is to make finance safer. [They] may not have the salaries, egos or profiles of Wall Street superstars, but the decisions they and people like them make are shaping the industry.' Our study points to a closer coupling between regulation and compliance analytics, with such analytics increasingly facilitating regulatory agendas [4, 119]. A distinguishing factor between big data analytics and regular analytics is the performative nature of Big Data and how it goes beyond merely representing the world but actively shapes it [9, 121].

The cases show how financial firms are facing demands to meet regulatory mandates using the latest and most effective forms of analytics and how regulators are increasingly requiring organizations to conduct vast searches of their organizational data (structured and unstructured) to prove a negative and thus avoid sanctions or instead disclose levels of malpractice. Evolving systems and analytics may reflect changes in financial products, which in turn create regulatory refinements (e.g. securitization). Conversely, as analytical capabilities have developed regulators have been able to demand that firms monitor and report closer to real-time. New analytical capabilities may also enable financial innovations and the introduction of new products and services. Analytical outputs may provide a basis for strategic decision making by regulators, who may refine and adapt regulatory obligations accordingly and then require firms to use related forms of analytics to test for compliance [119]. Furthermore, analytics may allow financial organizations to understand implications in markets and their levels of exposure to sanctions and litigation more quickly. These examples illustrate how compliance analytics are not simply reporting on practices but also shaping them through accelerated decision making changing strategic planning from a long term top down exercise to a bottom up reflexive exercise [19, 62, 120]. Consequently, compliance analytics and the algorithms and data which underpin them collectively constitute their own calculative agency and performativities [14, 80].

Our cases shows that data and how it is collected, stored and used is becoming more important, particularly as firms enter into contracts with third party providers (e.g. software vendors, data providers, legal firms and technology consultancies) to maintain compliance, defend against sanctions and litigation and develop their business. The complexity and heterogeneity of financial data is increasing where big data volumes, velocity and variability impact trading which is now a 24/7 activity. Our informants were concerned that increased regulatory rules and laws were forcing firms to invest more time and resources into fire-fighting activities just to keep pace with the changes. The growing complexity of data (structured and unstructured), the introduction of new technologies (e.g. Cloud) and the need to source data from third parties (e.g. Bloomberg) from the 'big data universe' adds further complexity. While complexity facilitated by financial technology is viewed as a contributory factor in financial crises [65], increased inter-connected system architecture and applications will only add to this complexity and heterogeneity.

Our findings underline how formal regulatory mechanisms are underpinned through analytics and information, where the regulator can evaluate whether pre-defined goals have been met (outcome control) or whether compliance with prescribed methods and standards has been achieved (behaviour controls) [110]. Both cases show how regulations and laws, however, are not apolitical but require social interpretation and embedding within operational practices [28]. Yet the increasingly pervasive reliance on analytics has led some to caution against an 'automation bias' which wrongly assumes technological neutrality [58]. Compliance analytics provide visualizations of complex structured and unstructured data, but in doing so related algorithms may privilege certain facets of information over others and so prevent

visibility of indicators of important market shifts or malpractice [90]. Thus, compliance analytics are not neutral in the data and information they provide and the responses they elicit [123] nor are the benchmarks and indices upon which they may draw [23]. Recent cases (LIBOR, FX) have highlighted how such benchmarks are open to manipulation. Predictive coding functionality within eDiscovery tools are also not neutral as they deliberately build human bias into the analytical process as the system is trained to understand and follow human interpretations.

An important implication of the non-neutrality of analytics is that the decisions and actions they afford and prohibit will also be prejudiced. Yet analytics underpins compliance and control practices which afford and constrain actions [46, 68, 69, 73]. Ultimately, both systems seek to establish a binary outcome, compliant or non-compliant, with varying degrees of consequence for being non-compliant. In addition, OMS will either afford or prohibit actions depending on the system's configuration. Thus, reducing regulation in this way also embodies particular affordances and discounts others [4]. Individuals may not stop to question whether a personally lucrative trading strategy is ethical or even in the best interests of their client but merely if it is possible within the imposed (regulatory) binary rules of the game against which they will be measured and monitored, now through the OMS or, in the future if an investigation takes place. Compliance rules are embedded within technologies and so an individual need look no further than whether the system affords the trade, thus giving assurance that they are operating ethically [68, 69]. Consequently, analytics may legitimise and locally institutionalise inappropriate practices outside the interests of other stakeholders including their clients. Conversely, technologies which implement surveillance and monitoring capabilities may also create self-disciplined behaviours through a pervasive suspicion that individuals are being currently observed or may have to account for their actions in the future [42].

Our analysis shows how the complexity and heterogeneity of underlying data and related analytics provides a further layer of technical complexity to banking matters and so adds further opacity to understanding controls, behaviours and misdeeds. For example, one must understand the nature of eDiscovery search capabilities and related data issues to run effective searches. Predictive coding affords the automation of operational practices for discovery and so shapes this process iteratively as the system initially learns from human input and eventually takes over [74]. In this way human and material agencies become further entangled and performative [14, 6]. However, the pervasive and performative nature of analytics may create additional risks as the heterogeneity of data increases. As Yoo [121] notes, 'Increased heterogeneity, in turn, makes the behaviour of the system less predictable. Furthermore, these individual components evolve over time, sometimes intentionally and sometimes by error, driving change in the system. As such, changes in one part of the system may cause a cascading sequence of events throughout the system, propagating a complex set of changes leading to emergent system behaviours that are hard to anticipate.' Design decisions are embedded within technologies shaped by underlying analytics and further underpinned by data. Both cases highlight how there are few software vendors in each market. If

firms use just a few vendors, then systemic risk occurs because all of the buyers and sellers are using the same calculative engines where logic flaws will become a standard part of the process. This issue may become exasperated where systems are constantly being developed and updated to incorporate new regulations or new complex data types. Data accuracy may also act to unduly influence outcomes. Consequently, this further underscores the need to understand big data analytics at the level of micro practice and from the bottom up. In summary, the increasingly widespread adoption and commercialization of OMS and eDiscovery tools means their performative effects for better or worse are likely to become amplified as their use becomes more common and institutionalised.

Information Control and Privacy

The use of analytics is now part of a wider compliance regime in financial institutions where the risk of sanctions and reputational damage are ever present if malpractice is uncovered [17]. Big data analytics allows firms to use the technology to build up a precise profile of an individual's behaviour and practices [87]. As technology increasingly scrutinizes trader activity and potential malpractice, which can lead to disciplinary action, intrusions into individual privacy must be proportionate and comply with regional data protection laws [11]. Our research of the two case studies highlights how big data technologies and related analytics differ considerably, even within the same industry. However, the two cases illustrate how the growth in the variety, volume and velocity of data, along with uncertainties created by an increasingly regulated environment, has influenced analytics employed to manage compliance risk. This uncertainty creates challenges in structuring systems, through data schemas, architectures and algorithms which may have to adapt to increasing levels of volume, velocity and variety, whilst retaining a high level of veracity. As numerous every day processes and communication tools brush up with one another on an increasingly regular basis, potential heterogeneity within the systems may be increased, particularly as new socio-technical objects are introduced [122]. In the first case, OMS illustrates this point as the system has had to adapt to increased trading volumes, complexity and diversity in financial products, growth in regulatory rules and also technological paradigm shifts (i.e. hosting the OMS internally to Compliance as a Service), As the data requirements for the system have evolved, the vendor has responded to increasing data heterogeneity by becoming a 'single point' data provider. By doing so, they are seeking to control the variability and structure of underlying data. As Constantiou and Kallinikos [119] note, 'it makes a great deal of difference whether data is gathered through a carefully laid out cognitive (semantic) architecture or, by contrast, is captured and stored without such a plan and on the assumption that it may be variously used *a posteriori*.' Our second case also well illustrates these concepts. The purpose of eDiscovery tools is to manage heterogeneous data created in haphazard fashion and to apply and impose a clear structure upon it so that it can be searched and analysed. Where new data types, such as chat room data, become relevant to regulatory investigations and litigation such systems must be flexible enough to incorporate such variety. An important function of such systems is to create structured data out of unstructured data. eDiscovery systems classify and assemble data which has been generated as part of everyday working practices and communications and

stored at the point of creation with little view as to how such data may be structured to support future regulatory investigations and litigation.

Building on this perspective we suggest that organizations may seek to apply order across haphazard data and thereby reduce related complexities by implementing proactive data and information governance practices. The information governance element of the EDRM model advocates integrating clear policies and transparent processes for information governance in order to improve security and privacy, IT efficiency, legal and compliance risk and thereby get your 'electronic house in order'. However, edrm.net also notes that, 'There is a genuine need for a general-purpose, broadly applicable reference framework for the industry at large (end users, vendors, influencers, and other market players). No such model currently exists,' and so a study of the effectiveness of such practices may be a fruitful topic of further research [30].

Respondents across both cases felt that future compliance pressures and risks could be somewhat mitigated through proactive categorization and management of data by financial organizations, yet often information and data governance within financial organizations was felt to be not well implemented and not a current priority. Consultants often found that when they interviewed compliance managers they had little understanding of where relevant data was held, on what servers or in which countries. This is perhaps unsurprising in the post financial crisis environment where operations' budgets are often consumed with meeting new compliance practices and where there exists little residual appetite or resource for implementing proactive measures aimed at improving or gold platting existing compliance measures [51]. However, we suggest that firms which proactively organize and manage their data will find the pain of compliance and managing breaches easier in the years to come. Where firms are faced with increasingly complex data to be managed through systems, such as CRIMS, the introduction of complex new products and new regulatory rules may be eased through a better understanding and governance of the organization's data and information. As regulatory investigations and related litigations becomes increasingly common, financial organizations which are likely to have to undertake future eDiscovery projects may use information governance techniques to reduce the need to rely on costly eternal resources.

Where information can be found quickly and easily organisations can react more quickly. Our respondents suggest that one of the key challenges in responding to regulatory investigations was the tight timeframes set by the regulatory bodies. Tight deadlines for responses may create further challenges where financial organizations see eDiscovery searches as simplistic and so do not appreciate the intricacies involved at the micro/data level, including reducing 'noise', accessing and managing structured data, preserving metadata and approaches for scanning, analysing and indexing paper documents. Consequently, they may leave interacting with eDiscovery experts too close to the deadline. The eDiscovery consultants interviewed felt that was often because, initially, the scope and complexity of the investigation was misunderstood or that the ability of technology to automate work and reveal in the

early stages the impact of the investigation or litigation was underestimated. Consequently, we would advise financial services practitioners conducting eDiscovery projects to engage with technical experts early on who understand the issues at the micro/data level. Firms which understand the impact of litigation and regulatory investigations may formulate appropriate strategies. For example, in litigation cases firms may wish to settle early where the case against them is strong. In regulatory investigations early determination of whether the firm is likely to be subject to fines and further litigation allows organizations to segregate funds appropriately and put strategies in place to mitigate reputational damage. Furthermore, regulators have previously reduced fines for organizations which have been the first to come forward and highlight a problem.

As data privacy regulations impose increasing levels of administration and sanctions, we expect policy makers at the global level to be placed under increased pressure to mitigate regulatory conflicts and multijurisdictional tensions between data privacy and financial services' regulations. Currently, the existing environment creates a type of regulatory arbitrage where financial organizations may refuse to disclose information citing local data protection regimes. Overall, there exists a dichotomy between technology and law. Technologies such as social media or cloud computing facilitate data sharing across borders, yet legislative frameworks are moving in the opposite direction towards greater controls designed to prevent movement of data under the banner of protecting privacy. This creates a tension which could be somewhat mediated through policy makers' deeper understanding of data and analytics at a more micro level and thereby appreciate how technical architectures and analytics are entangled with laws and regulations. Technology has the potential to both amplify regulatory and jurisdictional conflicts, for example where individuals deliberately locate servers in countries where data protection laws prevent easy transfer of data outside their borders. Conversely, technology may mediate such conflicts. For example, where Millnet cannot receive data from overseas to be processed in its London HQ due to foreign data protection laws, their consultants often travel with their eDiscovery hardware and software to the client's foreign office where the data is processed onsite in the presence of a data regulator, meeting local data protection requirements.

However, our respondents suggested that persuading organizations who have not yet been implicated in regulatory investigations to invest resources in proactive information governance programs might be challenging where there is no current perceived issue or need. Yet the growth of organizations such as Millnet and the rise in the number of regulatory investigations suggest that financial organizations are increasingly required to account for their actions. This is also true for OMS which provides an auditable trail should investors or regulators question how mandates have been applied. Furthermore, the imminent introduction of data protection laws will further require organizations to account for how they manage information, requiring much more responsibility from data controllers. Firms are likely to be required to understand the privacy impact of new projects and correspondingly assess and document perceived levels of intrusiveness. They will also be required to, ensure that data is

held for appropriate purposes, that data collection is not excessive and deletion occurs within prescribed time frames.

Implementing an Information Governance Strategy

For firms that are willing to engage in exercises to improve data governance we offer several actionable recommendations. Firstly, firms should work towards developing an enterprise wide information governance strategy with related policies. A key element of the strategy should address what data quality means for the organization. Data quality may be analysed across three interrelated dimensions. The first dimension is content and relates to the relevance of data and, the context in which the data is used will influence its relevance to specific users. Completeness is another aspect of this dimension. The content of data should also be concise (not over detailed) and within scope. The second dimension relates to time. Data should be available when required but also cover appropriate periods and should be supplied at regular intervals. The third dimension of data quality addresses the form of the data. The way data is presented must be clear, sufficiently detailed and available in appropriate formats and media. Firms may wish to consider how each of these dimension can be translated into appropriate metrics/KPIs for monitoring data quality. Firms should also seek to learn lessons from other organizations' successes and failures in governing information both inside and outside of the financial services industry.

The management of meta-data and its preservation, so that it can be evidenced to regulators and courts, should be considered when formulating strategies and tactics. Firms may wish to conduct impact analysis and consider the different use cases by which compliance analytics may support the business and prioritise related risks accordingly. It should be recognised that the relevance of the same information will differ across business units. Managers should also evaluate what data quality means to them within their specific sphere of operations. Firms should allow discrete business units to develop their own tactics to implement related policies and strategies. Thus, policies should be high-level enough to be relevant across the firm while allowing each unit to interpret them according to their own circumstances. Yet, firms may wish to centrally govern the data most critical to compliance and which is commonly referenced across the firm. Information shared across applications may be governed within regions or business units, while information held within single applications may be governed by local teams. We suggest that such initiatives are seen as a business project rather than an IT project and are sponsored by senior business managers. Yet, the IT function should play an important role, not least in evaluating supporting technologies (needed to apply controls derived from policies) and to prevent business units applying costly siloed technologies within discrete business units. Firms should develop a business lexicon to support analysis and documentation and to provide a common nomenclature across business and IT employees.

Individuals may be required to alter or change processes and workflows as a result of the introduction of new policies. Correspondingly, the application of a successful information governance

strategy will require some level of cultural change to ensure individuals understand the potential value and risks inherent within the firms' structured and unstructured data. Education of staff is also recommended to ensure that all individuals responsible for managing information understand their own responsibilities and are aware of the reasons the business is adopting a holistic strategy for information governance.

8. Further Research and Education

One identifiable problem, and indeed opportunity with the 'big data' literature is the vast potential of research topics it offers, particularly those linked to past and current debates within information systems, such as big data's relationship to socio-technical theory, institutional theory, the sociology of financial markets (e.g. looking at concepts of performativity, calculative agency) and many more. While we do not aim to extend theoretical debates in this study, we do offer two comprehensive cases where our respondents have given thoughtful and reflective comments on 'big data' issues. As over 40 new laws have been proposed since the 2007-9 financial crisis, this extensive body of regulation has increased the supervision of the actions and behaviour of financial market participants, including traders, investors and IT providers.

Symbolized by the four V's (volume, velocity, variety and veracity) there is no 'one-size-fits-all' template for all organizations and institutions. As we see from our two cases, both companies operate in very different segments of the financial services' industry. However, the common theme is the need for each company to keep pace with the ongoing legal and regulatory onslaught, where new directives, laws and rules are coercively applied by multi-jurisdictional regulatory bodies. By providing empirical examples of how companies operate within their own big data landscape, it is apparent that many of the examples we discuss range from highly strategic, where each firm has to interpret, develop and implement a big data strategy, to the very mundane, by considering how each rule or guideline applies to their own operations. While much of the current academic literature looks at the strategic impact of big data, we caution that in many regards, the 'devil is in the detail', and that a minor infringement of data caused by a company (such as the loss of personal data) can have significant repercussions, resulting in reputational damage and large fines.

Faced with new EU directives and laws, companies operating within the EU region must be aware of their specific rights, obligations, procedures and oversight mechanisms for controlling and processing big data. They also need to understand the 'rules of the game' for undertaking these activities outside the EU. Many of the thorny issues surrounding big data are at the micro-practice level [44] which is less often researched than macro-levels (industry-wide) or meso-levels (across and within companies). We believe that future research which considers big data in the context of financial services and other areas, such as healthcare, may consider multi-level studies which link policy and strategic issues with more granular practices.

As we have seen from our case studies, however, the proliferation and reach of big data means that even looking at a single case study, such as a site within a company, poses significant research challenges. This is because the global reach of data now extends well beyond a single site and involves the interventions, decisions, and applications of multiple participants, including regulators, industry professionals, vendor partners, and customers. These challenges are exacerbated by the characteristics of structured and unstructured data, where the latter is vastly more prolific than the former. Ring-fencing or defining boundaries of big data is difficult. The vastness of the data pool exemplified by our CRD case, where the system now has 7,000,000 lines of codes, coupled with the granularity of data, which may focus on only a segment of the code, suggests two very different types of research approach. In the past, much research in the information systems field has considered the organizational aspects of how the 'IT artefact' is introduced. Today, however, the notion of the IT artefact as something which can be unpacked by identifying a set of key variables is less viable. As we have shown with our case studies, companies working with multiple clients across different jurisdictions, where data and algorithms have expanded exponentially, now face additional challenges of interpreting complex financial regulations in an environment of inter-connected systems. Yet our understanding of the performative nature of algorithms [121] is unsophisticated. This is evident in a recent case in the financial market, where the media has considered how 'out of control algorithms' in high speed trading in Chicago has come under the scrutiny of the regulator [83]. While the issue of algorithms is pertinent to big data, the research challenge is whether it is possible to shine a light on the workings of complex algorithms to better understand their effects and impacts on society and organizations and activities, including regulatory compliance. While studies on the granularity of big data may overlook the socio-technical complexities facing contemporary financial services, an evolutionary ontology [121] which extends the research focus to include a wider array of factors (such as the methods used in evolutionary systems biology) is only likely through a cross-disciplinary and cross-country study. While we have considered only two companies in our sample, our results raise important issues for further research. Scholarship needs to take a view, not simply on the strategic or operational implications of big data within single firms, but also how societal and institutional factors help to define and shape big data effects and processes across multiple jurisdictions.

There is an opportunity for information systems' researchers to explore how financial trading is undergoing further transformation which may result in the need for even greater regulation. This topic is discussed in the finance discipline, yet the technology artefact remains under-theorized. However, regulation is enacted following technical change and this creates an almost permanent 'catch-up' scenario. Furthermore, we suggest that future research may identify companies which are 'high-performers' in information governance and so review related successful strategies and practices.

As financial innovation (i.e. new products and services offered) is introduced into the market, regulators need to understand not only how these offerings are facilitated by analytics but also whether existing regulations are still relevant. In the case of data infringements, regulations are often changed

after a rare but high profile event, so a challenge for researchers and policy-makers is to anticipate potential problems which may occur as a result of big data technology. We suggest that opportunities exist for researchers to carry out longitudinal studies on regulatory change which may impact on big data technology within firms. In the area of financial trading, concerns over market manipulation are forcing regulators to impose new rules and mandates on how firms manage their data with a view to preventing market abuse [116].

Finally, the participants in both cases highlighted how recruiting individuals with the appropriate skillsets was a significant problem. The work of both compliance managers and eDiscovery consultants is cross–disciplinary in nature touching on project management skills, computer science (databases and niche technologies) as well as legal and regulatory knowledge, financial products and the way capital markets function. Many of the respondents have built the required variety of skills as their careers have, often by chance, developed in this way. However, both cases underpin how technology is becoming pervasive and influencing traditional disciplines, such as law. Consequently, we suggest that educational institutions consider more multi-disciplinary courses which integrate technological knowledge with capital markets and regulation and law.

9. Concluding Remarks

Regulatory failures are increasingly the subject of national news and related commentary from politicians and other policy makers. Yet little research has been conducted which makes transparent the role of technology and specifically analytics in meeting regulatory obligations and conducting investigations where malpractice is suspected. Our study shows how the commercialization of big data analytics is pervasive within the financial services' industry and is increasingly underpinning compliance practices. We feel that one of the strengths of this study has been to unpack two illustrative cases and to understand the specific issues at the micro/data level.

The study has illustrated the performative nature of compliance analytics [14, 71, 80] and that such performativities may have unforeseen consequences. The philosophy of reacting to organizational and regulatory failures by introducing ever more controls and rules means that regulated activities will become increasingly reliant on compliance analytics. Yet such automation comes at a price by limiting the scope of regulatory structures and analytical processes and does not address deep rooted unethical behavioural practices beyond providing accountability and surveillance of existing rules. Our study has shown how analytics and compliance are becoming increasingly cohesive and that whilst data volumes are growing, heterogeneity across structured and unstructured data is also increasing. Differing levels of regulatory supervision and new mandates are partly informed and shaped by analytics which also apprise regulators and managers of daily compliance positions and the nature and severity of breaches when they occur. When seen as impartial and objective, such technologies have the potential to provide assurance that appropriate outcomes and behaviours occur. This may go some way to help restore the faith and trust

in the financial system eroded by the financial crisis and other failures and scandals. However, we should also recognise that such analytics are not inherently objective and include inevitable biases and limitations. Thus, we should proactively seek to understand the implications of underpinning design decisions and the technical architectures upon which they depend, rather than wait for comprehension to emerge as the result of a failure.

Our findings show that multijurisdictional challenges and regulatory conflicts exist between cross-industry regulations in the form of data privacy laws and industry specific financial services' regulations. Different data protection requirements across the EU and worldwide may create challenges where regulatory investigations transcend borders. Compliance analytics has the potential to both exasperate and mitigate related challenges. Our study illustrates how the use of analytics is becoming increasingly common place and important for managing related operational risks at both ends of the risk curve. Our study suggests that across the industry there is a paucity of individuals with an appropriately diversified skill set to develop and support compliance analytics. Arrays of skills are needed to apprehend the technologies and regulations at a suitably micro/technical level, whilst also understanding the complex nature of markets and the collective impact on the business. Individuals need to understand legal rules and technical innovations along with the resultant implications for a product line or a specific area of the supply chain. This mirrors a wider challenge the industry is currently facing, to recruit knowledgeable and experienced compliance professionals to meet an increasingly burdensome regulatory environment. Industry here can benefit from implementing training programs to fill this gap for individuals whose backgrounds most match the needs as well as to work with higher learning institutions to create programs to meet this need.

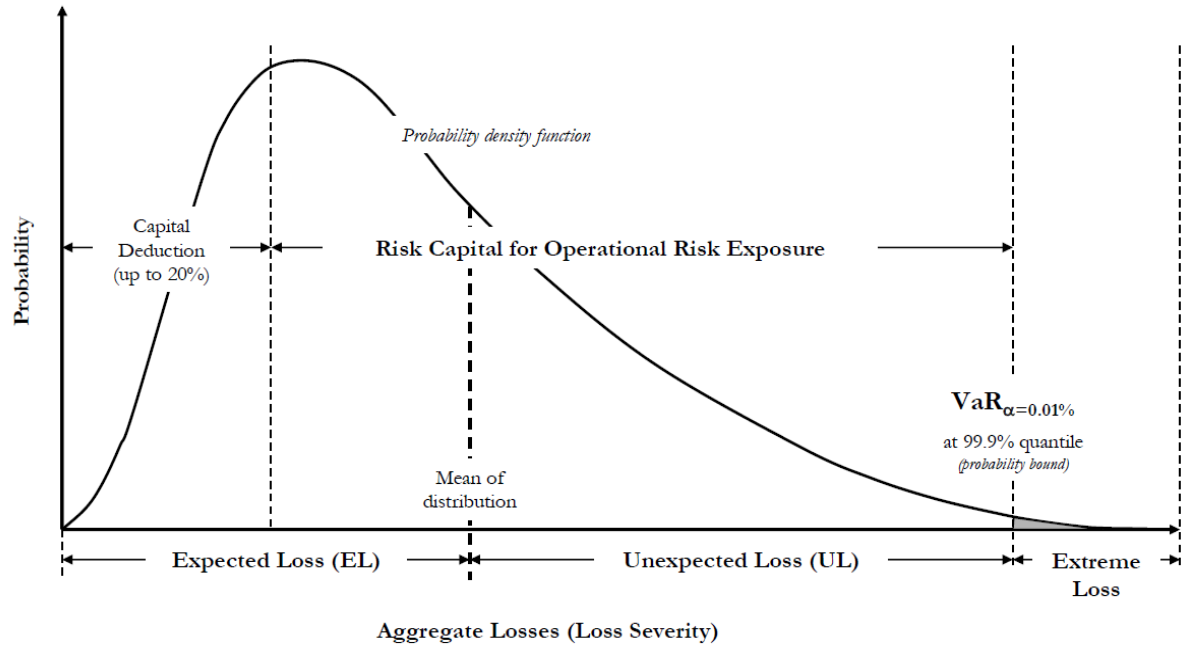
As the use of big data analytics within financial firms becomes further embedded and institutionalised, the ability of firms to facilitate analytics and reduce related costs and overheads through information governance will become increasingly important. Yet, our study shows that proactively structuring and managing data is of a low priority for many managers as the volume and variety of regulatory rules increases along with related costs and overheads. A further contribution is made in reviewing the complexities of dealing with different data types and how paper documents may still present challenges to those conducting regulatory investigations. Many discussants of big data overlook the fact that large volumes of important documents (e.g. financial records) are often still held in paper form and that transferring these to searchable electronic documents may not be as straight forward as assumed. The regulatory implications of historical data, held in paper form, needs to be thoroughly analysed and budgeted for as well. As regulatory obligations (e.g. MiFID II, data privacy regulations) increase, organizations should engage with information governance practices to control data categorization and storage. By doing so, they may not only better meet existing obligations but also reduce the operational burden of changing systems when new regulations come into force or new

products are introduced. Thus, firms will be able to respond more strategically to litigation and regulatory investigations.

Harnessing the power of analytics to better understand organizational operations may have many additional benefits beyond compliance. Through better understanding and control of the data their organization holds, firms will be much better placed to reap the benefits of big data analytics. For example, analytics may help firms identify areas where duplication of effort and systems are occurring and so improve processes and consolidate software licences. Improved understanding of operational risks may also allow firms to reduce their requirements to hold higher levels of regulatory capital. Furthermore, analytics may help organizations better understand how individuals in the firm interact with one another and thereby act to improve lines of communication. Analytics may also assist organizations in vital strategic decision making and related efforts to recruit and retain necessary staff. As a consequence, firms which embrace information governance techniques are better placed to exploit big data analytics and related future innovations. To conclude, firms which are able to become masters of their own data and conquer challenges related to volume, velocity, veracity and variety will be able to draw a competitive advantage through enhanced strategic decision making and increased operational efficiency.

Appendixes

A. Loss Distribution Approach for Operational Risk



Source: (Jobst, 2007)

B. Selected Recent Financial Scandals

	LIBOR Benchmark Rigging	Foreign Exchange Benchmark Rigging	Money Laundering Scandals	The Financial Crisis and Mis-selling mortgage backed securities
Summary	In 2012, an investigation into the London Interbank Offered Rate, or Libor, which underpins over $300 trillion worth of loans worldwide, revealed collusion across multiple banks to manipulate interest rates for their own profit from 2003.	Similar to the LIBOR scandal in 2013 an investigation by UK, USA, and Swiss regulators, assisted by authorities in Hong Kong, revealed they were scrutinizing 15 banks for manipulating a benchmark for setting the price of major currencies from 2006. This market is the world's largest where turnover is over $5 trillion a day.	In 2012, two financial firms agreed to settle with US regulators and signed statements acknowledging their role in facilitating illegal financial transfers on behalf of Iran, Sudan, Myanmar and Libya. In 2014, another firm also pleaded guilty to concealing billions of dollars' worth of transactions for clients in Sudan, Iran and Cuba. A fourth firm was also rebuked for allowing money from Mexican and Columbian drug gangs to enter the US. In 2015, Swiss prosecutors raided the firm's offices to investigate claims of money laundering. Authorities have been loath to reveal the details of breaches in case loopholes in other firms' processes may also be open to exploitation.	In 2013, a financial institution was sanctioned for miss-selling mortgage backed securities in the run up to the financial crisis (2007-2009), together with the two financial organizations it acquired. When the institution purchased these firms during the crisis, it also took on their legal liabilities. Due diligence failures identified workers vetting mortgages were encouraged to process as many as possible. An email revealed that one case worker had to review 1594 loans in 5 days.
Fines	Fines have been levied across multiple regulatory bodies in the UK, USA and the EU, currently more than $9 billion for rigging Libor. From 2015 investigations are continuing with other institutions expected to be implicated and related fines and civil lawsuits likely to ensue.	Multiple banks have paid a total of $5.6 billion. The FBI has described the scandal as involving criminality on a massive scale. Further regulatory investigations and law suits are expected as are criminal charges.	In 2012, a financial firm agreed to pay $327million on top of an earlier $340m fine. In 2014, the same firm agreed to pay $300m to the New York state department of financial services (DFS). In 2012, another firm paid $1.9 billion, a record fine at the time. In 2014, a third firm agreed to pay $8.9 billion to the US Justice department. In 2015, another financial firm was fined $30.9m by Swiss authorities for 'organisational deficiencies' that had enabled money laundering.	The institution agreed a record settlement with the US Department of Justice and state authorities to pay $13 billion. At the time the firm stated that it had prepared a $23billion war chest to meet this obligation and the tidal wave of related litigation it expected.
Regulatory Changes	Proposed new 'benchmarking' regulations focus on increasing sanctions for criminal and market abuse and on strengthening the governance, reliability and robustness of benchmarks used for pricing of financial instruments.		The 4th EU anti-money laundering directive came into force in June 2015. This directive includes new obligations to report transaction and record payments, as well as strengthening operational controls.	In the US the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010 included securitization process reforms focused on risk retention and increased disclosure to investors.

Source: (Council on Foreign Relations, 2015; European Commission, 2015; Financial Times, 2013a, 2013b, 2015a, 2015b; Out-Law.com, 2015; The Economist, 2013, 2014, 2015; The Guardian, 2015; The Washington Post, 2013, 2014)

C. Summary of Millnet's Services and Technology Partners

	Data Services					Print Services
Service	*Digital Forensics*		*eDiscovery*		*Virtual Data Room*	*Financial Print*
Service Overview	Digital forensics is the process of uncovering and interpreting electronic data. The goal of the process is to preserve any evidence in its most original form while performing a structured investigation by collecting, identifying and validating the digital information for the purpose of reconstructing past events. eDiscovery software facilitates the identification, collection, preservation, processing, review, analysis and production of electronically stored information (ESI), while meeting the mandates imposed by common-law requirements for discovery. These demands may be due to civil or criminal litigation, regulatory oversight or administrative proceedings.				Virtual data rooms (VDR) allow access to strictly confidential data and documents with restrictions and controls on viewing, copying or printing. A VDR allows documents to be accessed by regulators, lawyers and investors to view documents without need for physical copies or even a physical meeting room.	Millnet's Financial Document Services provide services to investment banks, corporate brokers, law firms, accountants and public companies in both the UK and international markets. They also support corporate finance transactions including flotation, hostile takeovers, recommended offers, rights issues, shareholder circulars, report and accounts.
Tech Vendor	*Relativity*	*Nuix*	*Index Engines*	*EnCase*	*Ethos Data*	N/A
Solution Overview	The most supported eDiscovery solution in the market, used by over 7,500 organisations worldwide for litigation cases, internal investigations, and responding to regulatory and government requests.	The Nuix Engine makes it possible to index and search large volumes of unstructured data in eDiscovery, digital investigation information governance and cybersecurity cases.	Index Engines' platform supports the disposition of data from migration, defensible deletion, and archiving to policy audits and automation of governance rules.	Encase allows efficient, forensically sound data collection and investigations using a repeatable and defensible process.	Ethos data allows organizations to exchange confidential information securely and efficiently through virtual data rooms. Set times are scheduled for log-on and viewing the documents and data.	Typeset and printing of financial documents in a format acceptable to regulator. Includes Secure distribution of electronic documents via email, ISDN, and virtual data rooms.

Source: (Ethos Data, 2015; Gartner, 2015; Guidance Software, 2015; Index Engines, 2015; Janssen, 2015a, 2015b; Kcura, 2015; Nuix, 2015)

D. Select eDiscovery Case Studies

Investigation	Subject Access Request	Potential contentious matter	Financial Services Market Abuse Investigation	Using analytics to solve a problem
Background	A former employee in a global organisation made a subject data access request under the UK data protection act.	A potential contentious matter for a global organisation.	Confidential market abuse investigation with a very broad Information request.	Disclosure had already taken place and Simmons and Simmons were trying to establish 'who knew what, When, around documents that were ultimately executed at various board meetings.
Facts	Searching by key words returned c.110,000 potentially relevant documents. Following standard review of the documents most likely to return "personal data", Simmons and Simmons identified 79 responsive documents from over 12,000 which were reviewed. They determined that the Analytics engine, in particular Relativity's Technology Assisted Review, would be the most efficient way to analyse the remaining 98,000 keyword responsive documents.	Email accounts belonging to 23 potentially relevant people in three jurisdictions covering an 18 month period This returned over 3.6 million documents. By locating the documents relied upon by the senior individual and documents highlighted by witnesses and applying the "email threading" functionality, we quickly identified 1,198 highly relevant documents. A number of complicated, targeted, custodian and key word searches (in English, German and French), refined by specific deduplication searches to overcome the challenge of email address fields not always being identical when processed, reduced the dataset to just over 7,400 documents which required human review.	Original collection of 20,000 documents (T1) – review completed. Second collection of 300,000 documents (T2) with less than 4 weeks to review. Using the relevant documents from T1 + 'good example/key documents' and subject matter experts to train Relativity Assisted Review on what makes a document relevant vs not. Relativity Assisted Review provides a complete audit trail on every decision the computer makes based on what is deemed to be a seed document as reviewed by the subject matter expert.	Difficulty in working out the document trail that led up to the execution of documents at board Meetings. 31,900 disclosed documents. 40 document trails to investigate.
Result	Reduced number of responsive documents requiring human review to 2,722. 145 man days saved-approximately £100,000 (87%) cost savings.	It cost £145,000 to review this dataset. Over 350 man days saved - approximately £263,000 (60%) cost savings compared to a traditional keyword driven document review.	85,000 documents were reviewed at first pass and 9,000 reviewed at the second pass. Total cost of review to production exercise was £175,000 compared with traditional document review at a document by document level would have cost £465k and taken 4662 review hours to complete first pass review alone.	Simmons and Simmons developed a workflow using analytics to find conceptually similar documents and then were able to track back to determine the people involved in document creation and how the documents developed over time. Textual analysis was the only way of deciphering the trail without manually reviewing all 31,900. This resulted in a cost saving of 82%.

Source: (Cases courtesy of Simmons and Simmons LLP)

E. The eDiscovery Process

Within the UK and the USA, the legal profession has been transformed through a combination of technological advancement and related alterations in the legislative landscape. In 2006, the USA's Federal Rules of Civil Procedure (FRCP) and in 2013, the Jackson Reforms were brought into effect in the UK. Both sets of legislation address how technology may be used to support civil cases. A crucial development is that electronically stored information (ESI) has been accepted as being of equal evidentiary weight and value as conventional paper documents. Deloitte [25] suggests that, 'It is often the case that an entire business dispute, regulatory investigation, or multimillion pound litigation may hinge on identifying when a single piece of data was communicated, generated, altered or deleted, by and to whom and under what circumstances.'

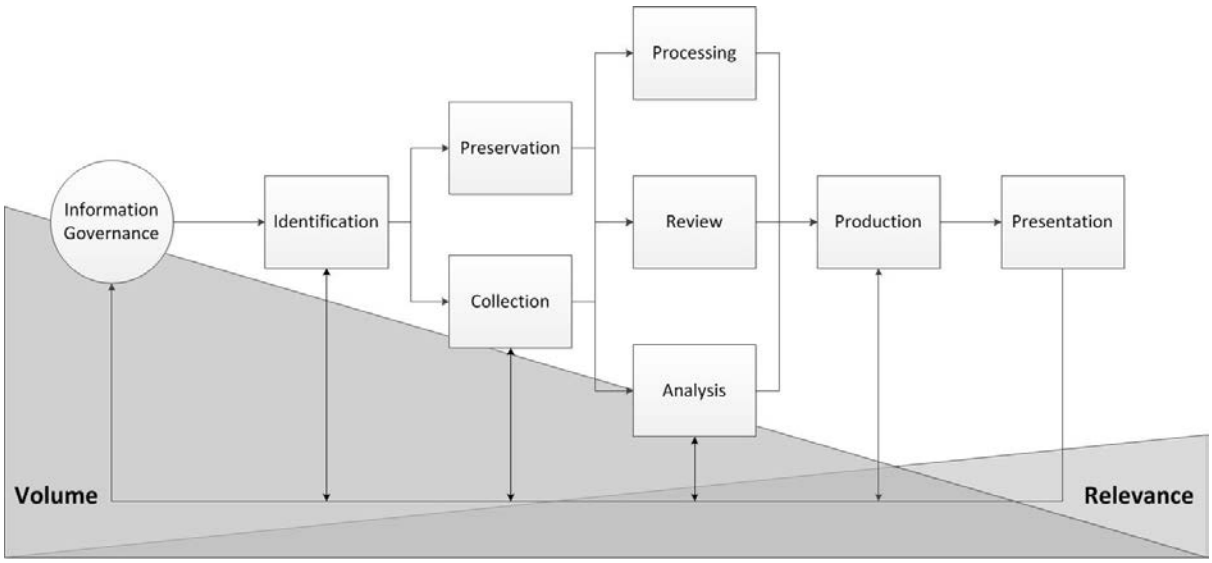


Figure 4. Electronic Discovery Reference Model v3.0 (edrm.net, 2014).

The Electronic Discovery Model (EDRM), Figure 4, represents a conceptual presentation of the eDiscovery process (edrm.net, 2014). The model should not be interpreted as a as a literal, linear or waterfall model. Systems and firms may facilitate discrete elements or the whole model, particularly as software vendors begin to consolidate functionality across the EDRM. The process depicted should be viewed as iterative. The same activities may be repeated many times to create an increasingly accurate set of results. It may also be necessary to cycle through earlier steps to define the approach being adopted as investigators obtain a better understanding of the data or the context regarding the investigation

The *Identification* component of the model refers to the need to ascertain sources of information relevant to the regulatory or legal investigation. A key aim of this activity is to establish who the 'custodians' are, individuals who are responsible for data types or repositories or are in possession of pertinent data. Examples of the different sources and types of data that eDiscovery tools can analyse are summarised in Table 4.

Data Types	Data Sources
Email	Laptops / Desktops
Calendar	Mobile Devices & Tablets
Text (Word, PowerPoint, Lotus Notes, PDFs etc.)	GPS & Digital Camera
Webmail (Gmail, Yahoo etc.)	CD / DVD/ USB
Access Databases	Mainframes
Spreadsheets (e.g. Excel, SPSS)	Servers Including Virtual Servers
Voice Recordings (Mobile And Internal)	Cloud Platforms
Chat (Bloomberg And Reuters)	Application Data
Windows / Mac / Linux / Android Files	Relational Databases
Intranets (e.g. SharePoint)	Paper Files
Paper Documents	Tape Archives

Table 4. Examples of eDiscovery data types and sources.

This process is facilitated through interviews with key individuals, not least to develop awareness of jargon and acronyms which may have been used in communications and documents. Another key aspect of this process is to determine the time frames relevant to the investigation, to further narrow the search. In the case of regulatory investigations, the regulatory may identify the timeframes and individuals to be investigated. An important outcome of this activity is 'key words' which can be used in later stages. The Preservation element of the model refers to the need to isolate and protect data in ways that are legally defensible, proportionate, auditable and cost effective. Key to this is the utilization of appropriate data forensic techniques. The third component, Collection, refers to the acquisition of the electronically stored information (ESI) defined in the Identification stage and may provide feedback to this stage to refine its effectiveness to hone in and better target relevant data. Again it is important that the data is collected in a way which is legally defensible, proportionate, auditable and cost effective. The Processing element of the model refers to the need to ascertain exactly what data exists within the scope, or 'data universe,' previously identified and to select and prepare and thereby reduce the number of items to be reviewed. This analysis is conducted at the level of individual items and so includes recording meta-data items for each item before the meta-data is altered by processing. At this stage, data items may need to be processed to allow for further work. For example, pst or zip files may need to be extracted and legacy files and mail formats may need to be converted. Archives and backup tapes may need to be accessed. All items need to be catalogued and their meta-data captured. Approaches for identifying duplicates and exceptions are then applied to remove redundancies and irrelevancies. Items are hashed and indexed and search terms applied to reduce the data considered. Testing samples of results of search terms may be conducted to refine the accuracy and value of the terms being applied. Samples of the outputted data for review may also be tested. The *Review* component of the model aims to identify which documents need to be disclosed and which privileged documents to withhold. This is normally conducted by the firm's legal team to understand the factual issues related to the investigation and formulate responses to courts and regulators based on the identified facts. Consequently, this process is essential in formulating legal strategies and is heavily reliant on technology to deal with the large volumes of data requiring review. Edrm.net suggests, 'Electronic discovery, with its enormous volume of data, can seem daunting. The good news is that significant improvements in data storage, database and search

technology, and review application functionality are providing increasingly efficient options for handling the volume of data and streamlining the review process. In addition, emerging search technologies that use methods like concept-based searching, linguistic pattern recognition and other areas that move beyond traditional keyword searching are now being used for initial culling of data as well as to provide supplemental search capabilities for different stages of the document review. A general knowledge of tools and trends has become an important part of the job responsibilities for those charged with preparing for a document review.'

The related *Analysis* component refers to analysing ESI for content & context, including key patterns, topics, people and discussion. This component is defined, in Figure 4, as succeeding the Review element yet this activity may occur at different stages within the model and may take different forms, including fact finding, and refining and enhancing the search and review process. Consequently, this process focuses not only on finding the relevant documents and facts pertinent to the investigation but also on improving the overall effectiveness and efficiency of supporting activities. Table 5highlights the different analysis activities and related sub-tasks.

Type of Analysis		
Fact Finding	***Search Enhancement***	***Review Enhancement***
Information Management	Inputs	
Litigation Readiness	Roles	
	Metrics	
Data Assessment	Tools and Technology	
Collection	Outputs and Desired Outcomes	

Developed from: (erdm.net, 2015)

Table 5 eDiscovery Forms of Analysis

The *Production* and *Presentation* stages of the model refer to the need to produce files and deliver them to third parties through appropriate mechanisms and in agreed forms. Regulators typically stipulate the format and presentation of data items. Where the investigation involves archived paper files, organizations may be required to scan them in order for them to be identified, reviewed, analysed and produced in an electronic format. The final component of the EDRM model is termed *Information Governance* and represents proactively organizing information within the firm to ease future eDiscovery projects and so mitigate future risks and reduce costs. This is especially important for large banking corporates who are increasingly becoming the targets of regulatory investigations and related litigation.

F. Sample Interview Guides

Technical Questions

1. What is the size of your team, and the size of CRD?
2. What is the direction of the company (movement towards Hosting/Managed services)?
3. How does CRD plan to tackle the challenges presented by the cloud?
4. Does the increase in data increase the products costs?
5. Has the equipment needed become more expensive as requirements have changed?
6. How is the company looking to support its future development strategies?
7. What are the functions within the new offices?
8. How has the managed client base grown over the past 5 years?
9. How has the hosted client base grown over the past 5 years?
10. How has CRD managed the SaaS proposition?
11. How has the increase in client services impacted the traditional operations?
12. What is the success of data services in EMEA/US?
13. What is the success of FIX in EMEA/US?
14. Has there been any change in the structure of the UK support team?
15. How have you improved the quality of data supplied by CRD across its product types?
16. Has data storage impacted performance?
17. Has there been increased pressure from SimCorp/Markit/Bloomberg…BlackRock?
18. NoSQL and BigData – is this being reviewed?
19. What are the potential impacts from NoSQL and how will this change SQL logic?
20. Is the CaaS product offered as a monthly application?

Compliance Questions

1. What is the size and structure of the US development team?
2. What is the process that you use to interpret the rules?
3. Do you have a legal team to help with your interpretations?
4. How do you focus on the EU and not just the US regulations?
5. How often do you visit other countries?
6. Do you discuss anything with the regulator?
7. How do you make sure that's the library built on what the clients will be looking for?
8. How do you code rules such as Volcker when the market can't agree?
9. How have reporting formats changed with non-SQL data such as email and a need to report on this. Are there any plans to change what is reports are made available?
10. Is Compliance As A Service (CaaS) finding a market in US/EMEA?
11. What is the size of the library, countries covered, which regulations are reviewed?
12. Performance difference between cloud and on-site runs?
13. Reports for historical activities – what are the big data requirements?
14. Is compliance delivery determined by the next software delivery? Can you export new libraries to older versions?

References

[1] Agarwal, R., & Dhar, V. (2014). Editorial—big data, data science, and analytics: The opportunity and challenge for is research. *Information systems research*, *25*(3), 443-448.

[2] Avison, D., & Malaurent, J. (2014). Is theory king?: questioning the theory fetish in information systems. *Journal of Information Technology*, *29*(4), 327-336.

[3] Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2014). Transformational issues of big data and analytics in networked business. *MIS quarterly*, *38*(2), 629-631.

[4] Bamberger, K. A. (2010). Technologies of Compliance: Risk and Regulation in a Digital Age. *Texas Law Review*, *88*(4), 669-739.

[5] Bank for International Settlements. (2001). Basel Committee on Banking Supervision: Operational Risk. Retrieved 6th June, 2015, from https://www.bis.org/publ/bcbsca07.pd

[6] Bank for International Settlements. (2005). Compliance and the compliance function in banks. Retrieved 2015, 8th July, from www.bis.org/publ/bcbs113.pdf

[7] Barry, A., & Slater, D. (2002). Introduction: the technological economy. *Economy and Society*, *31*(2), 175-193.

[8] Bennett, M. (2013). The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, *14*(3), 255-268.

[9] Bhimani, A., & Willcocks, L. (2014). Digitisation,'Big Data'and the transformation of accounting information. Accounting and Business Research, 44(4), 469-490.

[10] Bhimani, A. (2015). Exploring Big Data's Strategic Consequences. *Journal of Information Technology*, *30*(1), 66-69.

[11] Bholat, D. (2014). Big Data and Central Banks. Bank of England. Retrieved 5th June, 2015, from http://www.bankofengland.co.uk/research/Documents/ccbs/bigdatawriteup.pdf

[12] Bloomberg. (2015). The London Whale,. Retrieved 9th June, 2015, from http://www.bloombergview.com/quicktake/the-london-whale

[13] Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: work, progress, and prosperity in a time of brilliant technologies*: WW Norton & Company.

[14] Callon, M., & Muniesa, F. (2005). Peripheral vision economic markets as calculative collective devices. *Organization studies*, *26*(8), 1229-1250.

[15] Carruthers, B. G., & Kim, J.-C. (2011). The sociology of finance. *Annual Review of Sociology*, *37*, 239-259.

[16] Cetina, K. K., & Bruegger, U. (2002). Global Microstructures: The Virtual Societies of Financial Markets1. *American Journal of Sociology*, *107*(4), 905-950.

[17] Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045-2084.

[18] Clemons, E. K., & Weber, B. W. (1990). London's big bang: a case study of information technology, competitive impact, and organizational change. *Journal of Management Information Systems*, 41-60.

[19] Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57.

[20] Council on Foreign Relations. (2015). Understanding the Libor Scandal,. Retrieved 6th July, 2015, from http://www.cfr.org/united-kingdom/understanding-libor-scandal/p28729

[21] Crook, S., Pakulski, J., & Waters, M. (1992). Postmodernization: Change in Advanced Society. London: Sage. London: Sage.

[22] Davenport, T., Davenport, T. H., & Horváth, P. (2014). big data@ work. *Harvard Business Review Press, Boston*.

[23] De Goede, M. (2005). *Virtue, Fortune, And Faith: A Geneaology Of Finance* (Vol. 24): U of Minnesota Press.

[24] De Mauro, A., Greco, M., & Grimaldi, M. (2015). *What is big data? A consensual definition and a review of key research topics.* Paper presented at the AIP Conference Proceedings.

[25] Deloitte. (2015). Analytic and Forensic Technology. Retrieved 18th June, 2015, from http://www2.deloitte.com/jp/en/pages/risk/solutions/frs/analytic-and-forensic-technology.html

[26] Denzin, N. K., & Lincoln, Y. S. (2000). *The Sage handbook of qualitative research*: Sage.

[27] Economist, T. (2015). One regulator to rule them all. Retrieved 8th August, 2015, from http://www.economist.com/news/leaders/21660534-officials-have-been-given-enormous-discretion-corral-finance-has-costs-one-regulator
[28] Edelman, L. B., & Suchman, M. C. (1997). The legal environments of organizations. *Annual Review of Sociology*, 479-515.
[29] edrm.net. (2014). EDRM Stages. Retrieved 4th June, 2015, from http://www.edrm.net/resources/edrm-stages-explained
[30] edrm.net. (2015). Information Governance Reference Model (IGRM). Retrieved 8th June, 2015, from http://www.edrm.net/projects/igrm
[31] erdm.net. (2015). Analysis Guide. Retrieved 3rd july, 2015, from http://www.edrm.net/resources/guides/edrm-framework-guides/analysis
[32] Ethos Data. (2015). Virtual Data Room. Retrieved 7th June, 2015, from http://www.ethosdata.com/dataroom/
[33] European Commission. (2015). Benchmarks. Retrieved 4th June, 2015, from http://ec.europa.eu/finance/securities/benchmarks/index_en.htm
[34] Executive Office of the President. (2014). BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES. Retrieved 2015, 8th July, from https://www.whitehouse.gov/issues/technology/big-data-review
[35] FCA. (2013). How the Financial Conduct Authority will investigate and report on regulatory failure. Retrieved 8th July, 2015, from https://www.fca.org.uk/static/fca/documents/how-fca-will-investigate-and-report-regulatory-failure.pdf
[36] FCA. (2014). FCA Risk Outlook 2014. Retrieved 4th May, 2014, from http://www.fca.org.uk/static/documents/corporate/risk-outlook-2014.pdf
[37] Financial Times. (2013a). Foreign exchange: The big fix. Retrieved 2015, 6th May, from http://www.ft.com/cms/s/2/7a9b85b4-4af8-11e3-8c4c-00144feabdc0.html#axzz2sG2XPJec
[38] Financial Times. (2013b). JPMorgan agrees $13bn settlement over mortgage securities. Retrieved 7th July, 2015, from http://www.ft.com/cms/s/0/0a76c1ae-512e-11e3-9651-00144feabdc0.html#axzz3gjnbK56V
[39] Financial Times. (2015a). Six banks fined $5.6bn over rigging of foreign exchange markets. Retrieved 6th July, 2015, from http://www.ft.com/cms/s/0/23fa681c-fe73-11e4-be9f-00144feabdc0.html#slide0
[40] Financial Times. (2015b). Swiss prosecutor raids HSBC's Geneva premises,. Retrieved 7th July, 2015, from http://www.ft.com/cms/s/0/6ee57092-b74c-11e4-8807-00144feab7de.html#axzz3gjnbK56V
[41] Flick, U. (1998). *An introduction to qualitative research*. London: Sage.
[42] Foucault, M. (2008). Discipline and punish: The birth of the prison.
[43] Gartner. (2015). E-Discovery Software. Retrieved 2015, 3rd June, from http://www.gartner.com/it-glossary/e-discovery-software
[44] George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321-326.
[45] Gibbs, G. (2007). *Analysing Qualititave Data*. London: Sage.
[46] Gibson, J. J. (1986). *The ecological approach to visual perception*: Psychology Press.
[47] Gillet, R., Hübner, G., & Plunus, S. (2010). Operational risk and reputation in the financial industry. *Journal of Banking & Finance, 34*(1), 224-235.
[48] Girling, P. (2013). *Operational Risk Management: A Complete Guide to a Successful Operational Risk Framework*. Hoboken: NJ: Wiley Finance.
[49] Gonsowski, D. (2010). The good, the bad and the ugly of e-discovery keyword search. Retrieved 17th June, 2015, from http://www.geonlegal.com/news-details.php?id=65
[50] Government Office for Science. (2012). High impact low probability risks: Blackett review. Retrieved 4th May, 2015, from https://www.gov.uk/government/publications/high-impact-low-probability-risks-blackett-review
[51] Gozman, D., & Currie, W. (2014). The role of Investment Management Systems in regulatory compliance: a Post-Financial Crisis study of displacement mechanisms. *Journal of Information Technology, 29*(1), 44-58.
[52] Guest, G., K, M., & E, N. (2012). *Applied Thematic Anlaysis*. Thousand Oaks: Sage.

[53] Guidance Software. (2015). EnCase eDiscovery 5. Retrieved 14th June, 2015, from https://www.guidancesoftware.com/products/Pages/encase-ediscovery/overview.aspx?cmpid=nav
[54] Heidegger, M. (1954). The question concerning technology. *Technology and values: Essential readings*, 99-113.
[55] HM Government. (2014). Emerging Technologies: Big Data. Retrieved 2015, 13th July, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/389095/Horizon_Scanning_-_Emerging_Technologies_Big_Data_report_1.pdf
[56] Index Engines. (2015). Enterprise Information Management & Data Governance for an Unstructured World Retrieved 6th June, 2015, from http://www.indexengines.com/
[57] IT Governance Institute. (2015). Data Protection Penalties. Retrieved 2015, 1st August, from http://www.itgovernance.co.uk/dpa-penalties.aspx
[58] Itami, H., & Numagami, T. (1992). Dynamic interaction between strategy and technology. *Strategic Management Journal, 13*(S2), 119-135.
[59] Janssen, C. (2015a). *Digital Forensics*. Retrieved from http://www.techopedia.com/definition/27805/digital-forensics
[60] Janssen, C. (2015b). Virtual Data Room (VDR). Retrieved 2015, 4th June, from http://www.techopedia.com/definition/24342/virtual-data-room-vdr
[61] Jobst, A. A. (2007). Operational Risk—The Sting is Still in the Tail but the Poison Depends on the Dose. *IMF Working Paper, 07*(239).
[62] Kallinikos, J., & Constantiou, I. D. (2015). Big data revisited: a rejoinder. *Journal of Information Technology, 30*(1), 70-74.
[63] Kcura. (2015). Relativity. Retrieved 18th May, 2015, from https://www.kcura.com/relativity/
[64] Kemp Little LLP. (2013). Big Data - Legal Rights and Obligations. Retrieved 17th January, 2013, from http://www.kemplittle.com/cms/document/Big_Data_Legal_Rights_and_Obligations.pdf
[65] Kirilenko, A. A., & Lo, A. W. (2013). Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents. *The Journal of Economic Perspectives, 27*(2), 51-72.
[66] Knorr Cetina, K., & Preda, A. (2004). *The sociology of financial markets*: Oxford University Press.
[67] KPMG. (2014). Taking the Lead of Big Data. Retrieved 8th July, 2015, from https://www.kpmg.com/UK/en/IssuesAndInsights/ArticlesPublications/Documents/PDF/Advisory/taking-the-lead-on-big-data.pdf.
[68] Latour, B. (2005). Reassembling the social-an introduction to actor-network-theory. *Reassembling the Social-An Introduction to Actor-Network-Theory, by Bruno Latour, pp. 316. Foreword by Bruno Latour. Oxford University Press, Sep 2005. ISBN-10: 0199256047. ISBN-13: 9780199256044, 1*.
[69] Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS quarterly, 35*(1), 147-167.
[70] Long, W. (2013). EU Data Protection Regulation: fines up to €100m proposed,. Retrieved 8th June, 2015, from http://www.computerweekly.com/opinion/EU-Data-Protection-Regulation-fines-up-to-100m-proposed
[71] MacKenzie, D. (2006). *An engine, not a camera: how financial models shape markets*: Mit Press.
[72] MacKenzie, D., & Millo, Y. (2003). Constructing a market, performing theory: The historical sociology of a financial derivatives exchange1. *American Journal of Sociology, 109*(1), 107-145.
[73] Majchrzak, A., & Markus, M. (2013). Technology Affordances and Constraints Theory (of MIS): Thousand Oaks, CA: Sage Publications.
[74] Markus, M. L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology, 30*(1), 58-59.
[75] Markus, M. L., Steinfield, C. W., & Wigand, R. T. (2006). Industry-wide information systems standardization as collective action: the case of the US residential mortgage industry. *MIS quarterly*, 439-465.
[76] Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt.
[77] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data. *The management revolution. Harvard Bus Rev, 90*(10), 61-67.
[78] McConnell, P. J. (2013). Systemic operational risk: the LIBOR manipulation scandal. *Journal of Operational Risk, 8*(3), 59-99.

[79] Miles, M. B., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). Thousand Oaks CA: Sage.
[80] Muniesa, F., Millo, Y., & Callon, M. (2007). An introduction to market devices. *The Sociological Review, 55*(s2), 1-12.
[81] Nuix. (2015). The Nuix Engine. Retrieved 3rd June, 2015, from http://www.nuix.com/nuix-engine
[82] Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the" it" in it research—a call to theorizing the it artifact. *Information systems research, 12*(2), 121-134.
[83] Orol, R. (2012). Chicago Fed laments 'out-of-control' algorithms. Retrieved 6th June, 2015, from http://www.marketwatch.com/story/chicago-fed-laments-out-of-control-algorithms-2012-09-18
[84] Out-Law.com. (2015). New EU anti-money laundering directive to come into force from 26 June,. Retrieved 3rd July, 2015, from http://www.out-law.com/en/articles/2015/june/new-eu-anti-money-laundering-rules-to-take-effect-from-26-june/
[85] Patton, M. (1990). *Qualitative Evaluation and Research Methods*. Beverley Hills, CA: Sage.
[86] Pentland, A. (2014). *Social Physics*. New York: Penguin.
[87] Preda, A. (2006). Socio-Technical Agency in Financial Markets The Case of the Stock Ticker. *Social Studies of Science, 36*(5), 753-782.
[88] Preda, A. (2007a). The sociological approach to financial markets. *Journal of Economic Surveys, 21*(3), 506-533.
[89] Preda, A. (2007b). Technology and boundary-marking in financial markets. *economic sociology_the european electronic newsletter*, 33.
[90] Pryke, M. (2010). Money's eyes: the visual preparation of financial markets. *Economy and Society, 39*(4), 427-459.
[91] Punch, K. F. (2005). *Introduction to Social Research: Qualitative and Quantitative Approaches* (2nd ed.). London: Sage.
[92] PWC. (2012). Big Data: A new way to think about data - and a new way of doing business. Retrieved 5th July, 2015, from http://www.pwc.com/us/en/analytics/big-data.jhtm
[93] Ragin, C. (1982). Comparative sociology and the comparative method. *International Journal of comparative sociology, 22*, 102-120.
[94] RPC. (2014). Number of FCA dawn raids almost doubles in just one year. Retrieved 18th June, 2015, from http://www.rpc.co.uk/index.php?option=com_flexicontent&view=items&cid=51:media-centre&id=20358:number-of-fca-dawn-raids-almost-doubles-in-just-one-year&Itemid=48
[95] Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks: Sage.
[95] Sants, H. (2010). UK Financial Regulation: After the Crisis. Retrieved 17th March, 2010
[96] Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry, 5*(4), 465.
[97] Securities Institute. (2004). *Operational Risk Official 6th Edition IAQ Workbook* London: Centurion House.
[98] Sedona. (2007). Retrieved 18th June, 2015, from https://thesedonaconference.org/system/files/sites/sedona.civicactions.net/files/private/drupal/filesys/publications/Best_Practices_Retrieval_Methods___revised_cover_and_preface.pdf
[99] Silverman, D. (2001). *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction* (2nd ed.). London: Sage Publications.
[100] Silverman, D. (2014). Taking theory too far? A commentary on Avison and Malaurent. *Journal of Information Technology, 29*(4), 353-355.
[101] Snijders, C., Matzat, U., & Reips, U.-D. (2012). Big data: Big gaps of knowledge in the field of internet science. *International Journal of Internet Science, 7*(1), 1-5.
[102] Spiggle, S. (1994). Analysis and interpretation of qualitative data in consumer research. *Journal of consumer research*, 491-503.
[103] Symon, G., & Cassell, C. (2012). *Qualitative organizational research: core methods and current challenges*: Sage.
[104] The Economist. (2013). Pay up, move on. Retrieved 4th May, 2015, from http://www.economist.com/blogs/schumpeter/2013/11/jpmorgan-chase-s-legal-troubles
[105] The Economist. (2014). Bank, fix thyself. Retrieved 5th June, 2015, from http://www.economist.com/news/finance-and-economics/21598678-bank-england-faces-questions-over-its-role-rigged-forex-deals-bank-fix

[106] The Economist. (2015). Too big to jail. Retrieved 5th July, 2015, from http://www.economist.com/news/finance-and-economics/21568403-two-big-british-banks-reach-controversial-settlements-too-big-jail
[107] The Guardian. (2015). HSBC money-laundering procedures 'have flaws too bad to be revealed. Retrieved 7th July, 2015, from http://www.theguardian.com/business/2015/jun/05/hsbc-money-laundering-procedures-flaws-too-bad-to-be-revealed
[108] The Washington Post. (2013). Everything you need to know about JPMorgan's $13 billion settlement. Retrieved 7th June, 2015, from http://www.washingtonpost.com/blogs/wonkblog/wp/2013/10/21/everything-you-need-to-know-about-jpmorgans-13-billion-settlement/
[109] The Washington Post. (2014). France's BNP Paribas to pay $8.9 billion to U.S. for sanctions violations. Retrieved 7th June, 2015, from http://www.washingtonpost.com/business/economy/frances-bnp-paribas-to-pay-89-billion-to-us-for-money-laundering/2014/06/30/6d99d174-fc76-11e3-b1f4-8e77c632c07b_story.html
[110] Tiwana, A. (2010). Systems development ambidexterity: Explaining the complementary and substitutive roles of formal and informal controls. *Journal of Management Information Systems, 27*(2), 87-126.
[111] Turner, A. (2009). The Turner Review A regulatory response to the global banking crisis. from http://www.fsa.gov.uk/pubs/other/turner_review.pdf
[112] Turner, A. (2012). Banking at the cross-roads: Where do we go from here? Retrieved 2012, 22nd December, from http://www.fsa.gov.uk/library/communication/speeches/2012/0724-at.shtml
[113] University Alliance. (2015). What is Big Data? Retrieved 8th June, 2015, from http://www.villanovau.com/resources/bi/what-is-big-data/#.VbOGArO6eUk
[114] Weber, B. W. (1999). Next-generation trading in futures markets: A comparison of open outcry and order matching systems. *Journal of Management Information Systems*, 29-45.
[115] Weber, B. W. (2006). Adoption of electronic trading at the International Securities Exchange. *Decision Support Systems, 41*(4), 728-746.
[116] Wheatley, M. (2014). The Technology Challenge. Retrieved 7th July, 2015, from http://www.fca.org/news/the-technology-challenge
[117] Williams, J. W. (2012). *Policing the markets: Inside the black box of securities enforcement*: Routledge.
[118] Williams, J. W. (2013). Regulatory technologies, risky subjects, and financial boundaries: Governing 'fraud'in the financial markets. *Accounting, Organizations and Society, 38*(6), 544-558.
[119] Wixom, B., & Ross, J. W. (2012). The US Securities and Exchange Commission: Working Smarter to Protect Investors and Ensure Efficient Markets. Retrieved 23rd July, 2015, from http://cisr.mit.edu/blog/documents/2012/11/30/mit_cisrwp388_sec_wixomross-pdf/
[120] Woerner, S., & Wixom, B. H. (2015). Big data: extending the business strategy toolbox. *Journal of Information Technology, 30*(1), 60-62.
[121] Yoo, Y. (2015). It is not about size: a further thought on big data. *Journal of Information Technology, 30*(1), 63-65.
[122] Yoo, Y., Richard J. Boland, J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for Innovation in the Digitized World. *Organization Science, 23*(5), 1398-1408. doi: doi:10.1287/orsc.1120.0771
[123] Zaloom, C. (2003). Ambiguous numbers: Trading technologies and interpretation in financial markets. *American Ethnologist, 30*(2), 258-272.
[124] Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization Science, 18*(5), 749-762.
[125] Zhang, J., & Landers, G. (2015). The State of E-Discovery in 2015 and Beyond. Retrieved 7th June, 2015, from https://www.gartner.com/doc/2984917/state-ediscovery-