The Report Committee for Peter Austin Moore
Certifies that this is the approved version of the following report:

OATS, CAT, and CARDS:

Financial Regulation in the Era of Big Data

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Kenneth Flamm

Paul Von Hippel

OATS, CAT, and CARDS:

Financial Regulation in the Era of Big Data

by

Peter Austin Moore, BA

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Public Affairs

The University of Texas at Austin

May 2015

Dedication

I would like to dedicate this paper to my family and friends, without whose support I would not have been able to even contemplate writing something like this.

Acknowledgements

I would like to acknowledge my co-workers at SIFMA for introducing me into this subject field and helping me make sense of it all.

Abstract

OATS, CAT, and CARDS:

Financial Regulation in the Era of Big Data

Peter Austin Moore, MPAff

The University of Texas at Austin, 2015

Supervisor: Kenneth Flamm

The explosion of data in the financial industry has led regulators to seek better ways to utilize big data analytics. This paper analyzes the inception and development of three major regulatory programs borne from market failures. These programs represent the promise of big data, but have had to withstand criticisms of their cost, effectiveness, and necessity. The focus is on the twin goals of these programs: to reconstruct the market and to detect market abuse; and how the promises have been met and criticisms have been replied to.

Table of Contents

List of Tables

List of Figures

List of Illustrations

EXECUTIVE SUMMARY

This paper explores the subject of big data through the lens of financial regulation, focusing primarily on two major programs utilizing big data analytics: the Order Audit Trail System (OATS) and the Consolidated Audit Trail (CAT). Further analysis is done on a third system still in its infancy but no less indicative of the things to come, the Comprehensive Automated Risk Data System (CARDS). Foremost, this paper helps to define what big data is, how it is being employed in financial regulation through OATS and CAT, what criticisms have been levied against these two programs, and to what extent can these criticisms be answered.

To answer these questions, we must examine the logic behind the necessity of such large quantities of information for market surveillance. Mostly, the necessity arises out the fact that the financial markets are themselves changing and continually expanding into the frontiers of information volume and processing speed. Regulators must adapt to these changing conditions if they are to succeed in their regulatory duties. Indeed in many ways the regulators require the data in order to understand how the markets are changing in order to react in an effective manner.

The necessity for data collection was borne of market failure. The NASDAQ price fixing scandal in 1996 and the flash crash of 2010 precipitated the creation of the two programs largely because of regulatory failures both preceding the events and in their analysis. Each failure in its own way provided reasonable necessity for more accurate and timely information in a cost effective manner. However, the findings from both can be boiled down into two general tasks for the surveillance mechanism, in whatever form it might be: (1) recreate market conditions; and (2) detect market abuses. These are the two

foci of big data as it arose in the form of OATS and CAT. The journey to fulfill these tasks, however, has not necessarily been a smooth one.

Almost from the outset, OATS and CAT came under scrutiny. Many of the comments received in 1996 as OATS began its request for comments phase were echoed nearly identically a decade and a half later as CAT began to take shape. Indeed it would take a full decade for the three phases of OATS to be completed. CAT is now in its fifth year since its recommendatory inception in 2010, and with no bidder chosen and a three year construction timeline, it is likely to take longer than OATS.

In seeking to analyze these two programs, I contend that OATS provides an analogous system to the proposed consolidated audit trail. Both were created to achieve very similar ends and under similar problematic circumstances. So how has OATS fulfilled its goal of recreating the market and detecting market abuses? Towards the goal of recreating markets, OATS has had a good track record of managing the influx of ever increasing data without equal increases in rejections or other issues. However, if OATS were sufficient in this regard, the flash crash analysis and the justification for CAT would not have existed. However, it is OATS' distinct limitations in scope that are exactly to blame for the creation of CAT. The current regulatory surveillance system, which takes 24 hours to detect an error and provides five days to correct it cannot withstand the new high-frequency trading systems creating terabytes of data in milliseconds. It is not farfetched, however, to recognize that as the scope was the downfall of OATS, even a consolidated audit trail may reach its limitations as markets change, expand, and grow.

On the second goal of detecting market abuses and failures, there have been some very notable and long term illicit activities by individuals and a recession under the watch of OATS. Since OATS' inception, there has not been a marked increase in disciplinary actions undertaken by the Financial Industry Regulatory Authority (FINRA). We can

infer from this that perhaps OATS is not in fact rooting out more regulatory infractions, even as the expenditures and staff of FINRA have continued to grow. We can note, however, that regulatory fees and penalties has increased as a proportion of the total revenues. However, a specific target of OATS, market manipulation through buying and selling large quantities of a specific security through various dealers in order to give the impression of activity and increase the price, also called churning, has resulted in a marked decline in the instances of churning to nearly a quarter of what they had been a decade ago.

The Securities and Exchange Commission (SEC), however, has seen an increase in their average yearly actions taken since OATS' inception, however specific activities that more heavily rely on OATS-like data, those of insider trading and market manipulation, have not seen as much of an increase. Market manipulation, however, shows signs that it is increasing, and the SEC has likewise put a larger emphasis and funding towards data analysis specifically towards the goal of detecting market abuse.

In many ways, from the analysis of these two goals, it remains to be seen if OATS will have a positive impact or if CAT can build upon it. We can infer though that eliminating the limitations of the current system will avoid the same issues that arose during the flash crash and the NASDAQ scandal before it. Further, as technology advances, implementing many of the reforms will be less costly and more feasible and already the estimated cost for CAT has dropped as the storage and processing costs have decreased in the market. It seems certain, however, that unless a viable alternative to reporting auditable information and data can be found, the need for these big data programs will only increase.

Three recommendations seem suitable to further and complete these goals. The first is that it might be suitable for the auditing process to create a tiered system that

focuses on heavy, frequent users and slower, more deliberate users separately. This might increase efficiency at both recreating the market and detecting market abuses specific to each niche. Secondly, creating a clear separate division for data use for regulation. This could increase the ability of regulators and the industry to leverage their respective technological abilities to find effective solutions. It would also allow regulators to provide better information about how effective data analysis is at detecting abuses and recreating the markets. Lastly, for the industry, contemplating a data standardization and centralization policy will prevent greater headache and expedite the system of collection, correction, and analysis on the regulatory side.

As our financial industry changes, so too will the necessity for regulators to change with it. Advances in technology and big data analysis are providing hope and visions of a future where simply collecting enough data will enable us to resolve the issues that continue to plague the financial system. However, the trials of this analytics are only in their nascent stage. And just as in 1996, it would have been hard to fathom the technological strides that have been made, in another twenty years the idea of collecting data at all may seem as passé as storing data on CD-ROMs does today.

CHAPTER 1: WHAT IS BIG DATA?

INTRODUCTION

"Big Data" has become a catch phrase for nearly every facet or major endeavor in predictive analysis using large data sets. Yet it has only been fairly recently that computational power has been sufficient and inexpensive enough to allow for meaningful insights to be gleaned from exceptionally large, so-called messy datasets. This isn't to say that collecting and using data is a new phenomenon. The mapping of trade winds and the introduction of calendars to create order out of seeming chaos were arrived at through collection of information, calculation, and ultimately deduction. Big data, however, involves advances in statistical methods coupled with data collected in quantities too vast for even modern computers to handle. But buried within this data lies a diamond in the proverbial rough. The promise of big data is that from collection and analysis, we can arrive at meaningful conclusions to broad questions, detect patterns, determine optimality, and predict the future. These promises are a means to the end of understanding nature; therefore, the philosophy goes, if we can feasibly collect it and analyze it, we should.

There are myriad issues and promises involved in using big data to an end as nebulous as understanding chaos. Generally speaking, I posit that they can be broken up into two elements: the technical and the human. The technical issues are found in the sheer size and complexity of the data, the methods of collection, reliability, processing time and power, and the costs involved with such endeavors. The drive to overcome these problems, however, have led to some of the greatest technological innovations of the past century. The human issues arise in the realm of personal freedom and security, abuse of data, and simple fact that conclusions drawn from the data have real world implications. As the benefits of big data are realized, more people enter into our digital society, and

there is a growing understanding of the underlying technology, the mystery and corresponding foreboding will dissipate as the analysis becomes more democratic.

THE TECHNICAL ELEMENT

As mentioned previously, computational power and storage capacity have exponentially grown since the dawn of the computer age. Computations that would have taken years can now be done in minutes. Storing the data, once a massive endeavor, is now a fairly simple practice. Even with these advances, however, the technological boundaries of what is possible often hampers meaningful analysis at large scales. In the future, it is likely that what we consider now "too large and complex" will seem as laughable as the computational power of the first space shuttle. The scale of the data produced is truly immense. In 2011, the global digital community was producing over 1.2 zettabytes (1 zettabyte is 2^{70} bytes) of data. That figure is expected to double every two years and by 2020 reach an estimated 35 zettabytes annually or 5 exabytes every day (See Illustration 1). Finally, the advent of "the cloud" and distributed servers and computing has buttressed this growth and pushed the boundaries of what can be analyzed and stored at a diminishing cost.

Ultimately, the analysis of data is predicated on the logic and scale of the processors that can handle this data. Central Processing Unit (CPU) processing power has, by and large, conformed to Moore's law, which stipulates the ever two years there will be a doubling of the number of transistors on an integrated circuit (see Illustration 2). This has largely been coupled with a 20-25% annual increase in the CPU processing power increase (See Illustration 3). Petaflop speeds, processing data at over a 10^{15} floating-point operations per second, are now fairly common among supercomputers.[1]

[1] Top500.org, "Top 500 The List," n.d., http://www.top500.org/.

However, access to their capabilities and the operating costs remain issues. No synopsis of this topic would be complete without an allusion to the perpetually five years away and pseudo-holy grail of quantum computing, which has the promise to redefine computation as we know it.

Collecting data itself can often prove to be an exceedingly tedious or difficult process. The format of the data collected, which will come either structured or unstructured, also plays a role in its eventual usefulness. Passive data collection, for example those from sensors, make the process easier. However, upkeep for the sensors and maintaining their functionality can often prove difficult. More active data collection, from visual inspection or collection, is often a labor intensive project subject to time and cost constraints. Both of these methods of collection can be structured, meaning they can have a set amount of fields and expected content. This makes it easy to analyze as operations on each data point can return an expected value. Unstructured data, however, does not conform to a number of fields or content. An example of this is a video. We can expect certain things: a length of time, a number of frames, and pixel content and resolution for each frame. But we cannot necessarily easily determine what the video is about or predict the content of the next video.

Finally, the resources available to put towards the data analysis end are limited. While most can argue the necessity of their cause, it is likely that they will be competing for the resources to conduct their analyses. Access to supercomputer and advanced computing is often limited to experts who understand how to utilize their power fully. While the field of data science, machine learning, and computing have grown tremendously, there is still room for growth. Large strides continue to be made as well in attempts to automate the analysis especially of unstructured data in order to glean useful information.

THE HUMAN ELEMENT

The human component of big data is a more difficult thing to measure. The dystopian futures portrayed in popular culture generally follow a similar vein: overzealous collection of personal information with the intent to control humanity in some form or another. The promise of big data, to some, represents a step towards this dystopia. Thus there is a philosophical opposition to this type of, especially government, collection. However, there are many additional agents and companies that make it their business to collect information about personal preference and their scope is often much larger than that of the government.

Security, especially of the personally identifiable information, is a growing concern in this field as well. Access control, integrity, and confidentiality are cornerstones of the cybersecurity policies that govern the large amounts of data that reflect the lives and choices of people. The internet itself was largely created as a tool for transferring data to be used for intellectual purposes. What cyberspace has become, however, is far beyond that. The opening of the internet has created equal amounts of opportunity and risk for losing a large amount of information that, in the wrong hands, present large issues.

Data itself is not an incorruptible medium. While many data and social scientists often concede or try to accommodate for unintentional data manipulation, the conclusions they reveal can sometimes be oversimplified for dissemination without regard for this process. The collection process itself can present many issues, which are said to be lessened simply by increasing the sample size. In a world where information is released at near fire-hose amounts and speeds, differentiating between the lies, damned lies, and statistics has become an even greater endeavor. Ultimately, data can be processed technically, but analyzing and interpreting it will always fall squarely on our shoulders.

None of this should dissuade anyone of the relevance or importance of this analysis and dissemination of data. While open data and greater access may not be a panacea to all of our social ills, empowering the public to fact-check and search for the meaning themselves will undoubtedly have a similar effect to the invention of the printing press.

CONCLUSION

Big data has firmly entered into the lexicon of those involved in public policy. The hope is that through data collection we can craft more effective legislation that carries out the directive of the government to represent their citizenry. This will have the added benefit of reducing the cost of programs by both reducing waste and redundant operations by automating where possible. As the remainder of this paper will hopefully ascertain, we are currently employing and seeking to expand the capabilities of big data in a way that can be measured and verified. This is important not only in order to make sure data is not collected for data's sake, but to ensure the goals of the data collection are being met.

We cannot forget the human element, however. While big data experts and evangelists, including myself, often quickly harp on the technical achievements and promise, the answer is only as good as the questions we ask. In the quest to create an unbiased truth, it is important to respect the fact that the analysis is a uniquely human endeavor and the implications will have real world consequences. It is estimated that bad or poor quality data accounts for a loss of $600 million annually to businesses and "performing data quality best practices can boost a company's revenue by 66%."[2] Those

[2] Chad Luckie, "'Big Data' Facts and Statistics That Will Shock You," *Fathom*, accessed March 18, 2015, http://www.fathomdelivers.com/blog/analytics-and-big-data/big-data-facts-and-statistics-that-will-shock-you/.

losses and that saving at least provide a good economic rationale into exploring the promise of big data.

CHAPTER 2: BIG DATA IN FINANCIAL REGULATION

INTRODUCTION

Why do financial regulators need big data, and why focus on financial regulation at all? It is true that there are many policy areas are currently employing big data, however financial regulation is one of the many policy realms that most interestingly and pertinently combines the human and technical promises and pitfalls of big data analysis and can be analyzed for how true to life those conclusions are in practice. The financial industry has a long history of data analysis, although the collection and analysis of such large data financial sets, especially for regulatory purposes, grew alongside the computing power of the time. But the world of markets and their regulation are growing and changing in fascinating and surprising ways. It is interesting both the scale of the operation and indeed the impact that it will have on all of our lives. Failure to properly regulate the market system has more than once caused deep hardship across the globe. That being said, we are moving into a world where information and processing speed is reaching new heights. We need to prepare for this world and this new paradigm.

Digitizing the financial markets has led to an explosive growth in the volume of trades. To keep up with the ability of computers to trade at inhuman speeds required regulators to collect data at inhuman speeds. This requires a suitable connection to the data, a place to put the data, and the means to analyze the data.

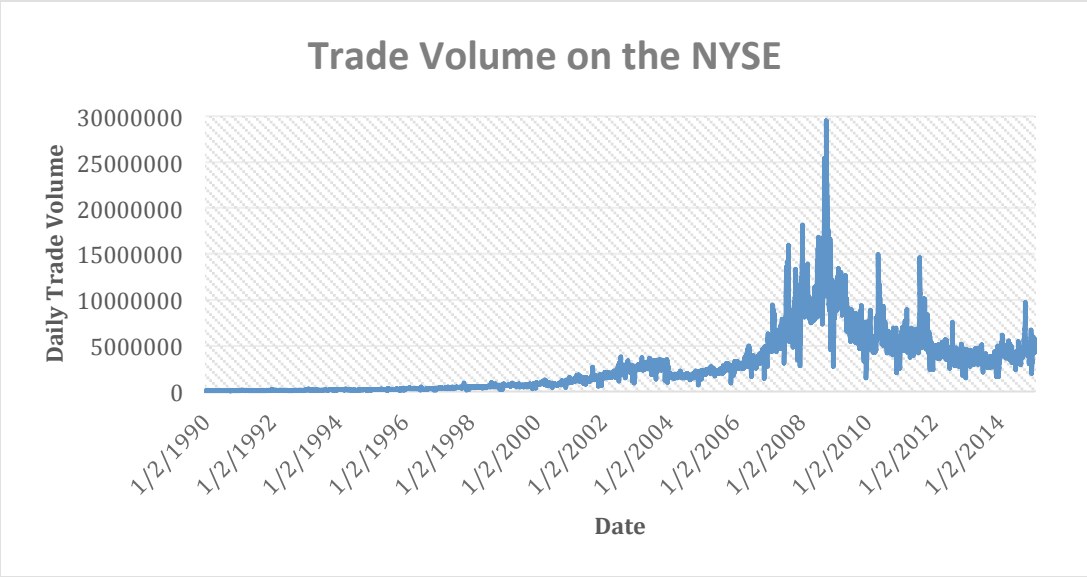


Figure 1.1 NYSE Trade Volume Data

The regulators' need for this data is primarily two-fold: (1) to be able to recreate market conditions; and, (2) detect and expose market abuses. Recreating market conditions allows the regulators to piece together the facts, which, often coupled with more thorough investigations, can lead to a determination of innocence or guilt or determination of a more systemic market problem.

More often than not, in exposing malfeasance regulators must request the information they need from the entity or person they are investigating. This has two particular drawbacks in that it puts a burden on the investigated party who has to spend time collecting and transporting the information and the burden on the investigators to then go through the information to reach meaningful conclusions. Both of these drawbacks also add to the time required to conduct and conclude an investigation. The difficulty of standardizing this information for completion and accuracy adds another layer to the trustworthiness of the data collected.

Recently, regulators have sought to use the power of greater market surveillance in order to cut back on both drawbacks and, hopefully, curtail the issues of timeliness, accuracy, and a general lack of collection coordination. Because the data would be

collected in its entirety, the requests made to businesses would not be one off, expensive inquiries but rather automated, efficient transfers of information. The idea being to collect everything all the time instead of collecting something some of the time. This would reduce the amount of time and energy required to collect the data. The investigators can then use aggregated, large datasets to detect anomalies from business as usual. Thus their investigations can be narrowed down before the problem becomes too immense and creates a proactive, rather than retroactive, system of regulation.

Major market events that may have been avoided through the use of better and more complete surveillance resulted in the creation to two specific programs that are the theme of this paper.

Financial Industry Regulators

Throughout this paper, there will be references to the regulating authorities of the financial industry, as well as specific regulation or programs created under these regulators. It is important to recognize their directives and the circumstances of their creation in order to understand their expansion into the world of big data. For the purposes of this paper, there are two major financial regulators that are of interest: (1) The Securities and Exchange Commission (SEC); and (2) The Financial Industry Regulatory Authority (FINRA). There are, of course, other federal, state, and local regulators, each with distinct rules and regulations to follow; however, their role in the collection and analysis of large amounts of financial data is less applicable.

SECURITIES AND EXCHANGE COMMISSION

The primary role of the Securities and Exchange Commission (SEC) is to "protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation."[3] The SEC is able to do this through two main laws.[4] Chronologically, the first is the Securities Act of 1933, which requires firms to register before offering market securities and to provide relevant and pertinent information to investors with the goal of preventing deceit, fraud, and misrepresentations.[5] Secondly, through the Securities Exchange Act of 1934, which, in addition to creating the SEC, more thoroughly defined the Commission's role in regulating and reporting activities on the various markets, such as NASDAQ and the New York Stock Exchange, primarily through cooperation with self-regulating organizations (SRO).[6] Additional laws include the Investment Advisers Act of 1940 and the Investment Company Act of 1940, which cover investor and company reporting regulations respectively. Newer laws, too, such as the Sarbanes-Oxley Act (2002), which "mandated a number of reforms to enhance corporate responsibility, enhance financial disclosures and combat corporate and accounting fraud" provided greater auditory oversight of firms[7]. The Dodd-Frank Act (2010) provided additional regulatory leverage for increasing transparency specifically in light of the 2008 financial crisis. Dodd-Frank was in some ways softened by the Jumpstart Our Business Startups (JOBS) Act (2012)[8]. Other rules proposed by the SRO must be reviewed and approved by the SEC prior to

---

[3] United States Securities and Exchange Commission, "The SEC - What We Do," *The Investors Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation*, accessed April 28, 2014, http://www.sec.gov/about/whatwedo.shtml.

[4] Matt Taibbi, "Why Didn't the SEC Catch Madoff? It Might Have Been Policy Not To," *Rolling Stone*, accessed April 29, 2014, http://www.rollingstone.com/politics/blogs/taibblog/why-didnt-the-sec-catch-madoff-it-might-have-been-policy-not-to-20130531.

[5] United States Securities and Exchange Commission, "The SEC - What We Do."

[6] Ibid.

[7] The Securities and Exchange Commission, "The Laws That Govern the Securities Industry," Government, (n.d.), http://www.sec.gov/about/laws.shtml.

[8] Ibid.

being adopted. Crucial to the SECs regulatory role, is its ability to bring civil enforcement action against individuals and companies for violations of securities laws.[9]

Since its inception the SEC has been involved in data collection. Supported by the 1934 Act, "companies with more than $10 million in assets whose securities are held by more than 500 owners must file annual and other periodic reports" in addition to "a variety of market participants to register with the Commission, including exchanges, brokers and dealers, transfer agents, and clearing agencies. Registration for these organizations involves filing disclosure documents that are updated on a regular basis."[10] This data is currently reported to the Electronic Data Gathering, Analysis, and Retrieval (EDGAR), which was created in 1984 to expand the ability of potential investors or clients to find information and inform their decisions.

Within the SEC is the Office of Compliance Inspections and Examinations (OCIE) and the Division of Enforcement are the dual bodies that carry out the enforcement and inspections of the regulatory purview of the SEC. However, the Division of Economic and Risk Analysis (DERA), created in 2009 and formerly known as the Division or Risk, Strategy, and Financial Innovation[11], is expressly tasked with integrating "financial economics and rigorous data analytics into the core mission of the SEC. The Division is involved across the entire range of SEC activities, including policy-making, rule-making, enforcement, and examination."[12] Since its inception, the budget for this division has more than tripled from $14 million in 2009 to over $40 million in

---

[9] United States Securities and Exchange Commission, "The SEC - What We Do."
[10] The Securities and Exchange Commission, "The Laws That Govern the Securities Industry."
[11] The Securities and Exchange Commission, "SEC Renames Division Focusing on Economic and Risk Analysis," Government, (June 6, 2013), http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171575272#.VP-KUWTF9kI.
[12] The Securities and Exchange Commission, "Division of Economic and Risk Analysis," n.d., http://www.sec.gov/dera.

2015. While this is still minute compared to the combined budgets of the OCIE and Enforcement, which together totaled just under $800 billion, this division is likely to continue its growth.[13]

FINANCIAL INDUSTRY REGULATORY AUTHORITY

The Financial Industry Regulatory Authority (FINRA) was formed in July 2007 out of the National Association of Securities Dealers (NASD) and regulators from the New York Stock Exchange (NYSE). NASD was founded in 1939 and worked with the SEC, recently formed from the Securities Exchange Act of 1934 in order to regulate its members and be subject to oversight from the SEC. As such NASD, and FINRA after it, was a self-regulating organization (SRO), which, as the name suggests, grants them the authority to create financial regulation and penalties for their members, however membership and control is comprised of firms within the financial industry. In this way, the SRO can in theory respond to the best interests of the financial institutions while being checked by the governance of the SEC.

In 1971, NASD formed NASDAQ[14], the National Association of Securities Dealers Automated Quotations, which would rely on computers for their trading process and remains the second largest financial market in the world behind the NYSE.[15] Along the way, there have been a few regulatory bumps for NASDAQ, some of which played a role in the creation of programs introduced in this paper. By 2001 and in the wake of the dot-com bubble, NASD had divested from NASDAQ and assumed a strictly regulatory role in the financial industry.

---

[13] The Securities and Exchange Commission, *FY 2016 Congressional Budget Justification*, February 2, 2105, http://www.sec.gov/about/reports/secfy16congbudgjust.pdf.
[14] John Allen, "FINRA| Broker Fraud Attorney," *Broker Fraud | Broker Fraud Attorney | Broker Misconduct*, accessed March 18, 2015, http://www.mybrokerfraud.com/faq/faq-general-finra/.
[15] World Federation of Exchanges, "Latest Statistics - Monthly Reports," March 2015, http://www.world-exchanges.org/statistics/monthly-reports.

On March 19, 2007, NASD filed a proposed rule change which would incorporate the NYSE Regulation, Inc. This came after the two organizations had announced in November of 2006 their plans to consolidate. In coming together, FINRA would accept responsibility "for regulatory oversight of all securities firms that do business with the public; professional training, testing and licensing of registered persons; arbitration and mediation; market regulation by contract for The NASDAQ Stock Market, Inc., the American Stock Exchange LLC, and the International Securities Exchange, LLC; and industry utilities, such as Trade Reporting Facilities and other over-the-counter operations."[16] On July 26, 2007, the SEC agreed to the merger. This substantially increased the regulatory purview of FINRA and further consolidated regulation within the financial industry.

With both the increase in market transactions and the increase in regulatory oversight, data and technology have remained integral parts of FINRA's regulatory toolbox. Today, FINRA spends over $30 million of its budget on the computers and systems needed to collect and analyze this data. With additional advances, FINRA has also begun the process of moving data to the cloud, which presents its own challenges and benefits.[17]

Financial Market Data Collection and Audit Trails

The financial markets are, in essence, large repositories of data. It is important that this data be as accurate and timely as possible as each day over six billion shares are

[16] The Securities and Exchange Commission, "Release No. 34-56145; File No. SR-NASD-2007-023," July 26, 2007, 3, http://www.sec.gov/rules/sro/nasd/2007/34-56145.pdf.
[17] Herbert Lash, "Wall St Watchdog Moves to Cloud, Big Data, to Boost Capabilities," *Reuters*, June 20, 2014, http://www.reuters.com/article/2014/06/20/finra-technology-idUSL2N0P105020140620.

traded across the financial markets worth an estimated $279 billion.[18] Trust in this system depends on the fidelity of the data collected by these financial market institutions. Indeed, the livelihoods and futures of large numbers of people in the country depend on knowing who did what and when. Market data by necessity then must be highly accurate. Many private enterprises use this market data in order to draw conclusions, often bred from experience or intuition, and predict the movements of the market. The financial markets are an oft-used system of judging the general health of the economy as a whole. However, with so much money at stake, there is a constant threat of cheating or defrauding for the purposes of financial gain. While the SEC and SROs are tasked with ensuring overall market stability, they are also tasked with routing out those who break the rules.

The SEC by its own admission lists three particular uses for audit trails and other sources of market data:

(1) inform its priorities for examinations of broker dealers, investment advisers and SROs;

(2) supplement the data and information it collects during those examinations; and,

(3) determine the nature and scope of any potential misconduct the examinations identify.[19]

---

[18] Sam Mamudi and Whitney Kisling, "Volatility Spurs Volume Boon in U.S. Stocks With Trading Up 11%," *Bloomberg*, February 14, 2014, http://www.bloomberg.com/news/2014-02-14/volatility-spurs-volume-boon-in-u-s-stocks-with-trading-up-11-.html.

[19] The Securities and Exchange Commission, "Release No. 34-67457; File No. S7-11-10," October 1, 2012, 22, http://www.sec.gov/rules/final/2012/34-67457.pdf.

LIST OF FINANCIAL ISSUES

Financial regulators can use market data to reveal anomalies, deconstruct the market in order to determine ownership, and look for signs of impropriety. This evidence can be collected for later regulatory action by the SEC and/or FINRA. These regulatory actions can include fines, limitations of business, or outright barring from the financial industry. The actions are compiled in monthly and annual reports, which can be easily found on each organizations website.

In 2014, FINRA enumerated a list of issues that they would focus on throughout the year.[20] These issues can be found in list 2.1 FINRA Issue Priorities below.

Table 2.1 FINRA Issue Priorities

Business Conduct Priorities

- Suitability
- Recidivist Brokers
- Conflicts of Interest
- Cybersecurity
- Qualified Plan Rollovers
- Initial Public Offering Market
- General Solicitation and Advertising of Private Placements
- Due Diligence and Suitability of Private Placements
- Offerings of Securities through Private Placements
- Anti-Money Laundering (AML)
- Municipal Advisors
- Crowdfunding Portals
- Senior Investors

Financial and Operational Priorities

- Funding and Liquidity Risk
- Risk Control Documentation and Assessment
- Accuracy of Firm's Financial Statements and Net Capital
- Auditor Independence

Market Regulation Priorities

- Algorithmic Trading and Trading Systems
- Audit Trail Integrity
- Best Execution of Equities, Options and Fixed Income Securities

Fraud Priorities

- Microcap Fraud
- Insider Trading

Of these priorities listed, however, the data that is able to be passively collected limits in some ways the scope of what can be investigated to those practices that are easily accessible. Specifically, the market regulation priorities, fraud priorities, and financial

[20] ACA Compliance Group, "FINRA 2014 Annual Regulatory and Examination Priorities," January 9, 2014, http://www.acacompliancegroup.com/news/compliance-alert/finra-2014-annual-regulatory-and-examination-priorities.

and operational priorities are the most currently data intensive priorities from this list. However, as the collection methods have improved and expanded, other areas, such as suitability can be analyzed.

Interestingly, the SEC also uses this data "to improve its understanding of how markets operate and evolve, including with respect to the development of new trading practices, the reconstruction of atypical or novel market events, and the implications of new markets or market rules."[21] This is an important point, as the markets continue to change drastically as technology and other practices continue to influence them. Some issues have persisted since 1934, while others, such as algorithmic trading, present new and novel methods of playing the stock market game.

THE ORDER AUDIT TRAIL (OATS)

The Order Audit Trail (OATS) was created in 1996 as part of litigation by the SEC against FINRA (then known as National Association of Securities Dealers) and other regulators for failing to suitable monitor the markets, in particular NASDAQ. OATS is fundamentally an auditing system for the financial market, encapsulating the who, what, where, and when of transactions from daily reporting from financial industry firms.

OATS reports can be made for single or multiple orders on the financial market. In total, the information collected on the OATS report includes 75 separate pieces of information. However, the OATS report specifically looks at:

- Customer or firm order,
- Date and time of receipt,
- Order ID,

[21] The Securities and Exchange Commission, "The Consolidated Audit Trail," 22.

- Terms of the order (i.e. buy, sell, sell short, security, price, shares, account type and handling instructions),
- If the order was received manually or electronically,
- If the order was routed manually or electronically,
- Where the order was routed for execution,
- Any modifications to the order including the date and time of any modifications,
- Execution information including partial executions, price, date, time and capacity in which the firm acted in the trade.[22]

Events Leading Up to OATS

The NASDAQ Price Collusion Scandal

The roots of the NASDAQ price collusion scandal's eventual uncovering can be traced to an article written in 1994 by William Christie and Paul Schultz entitled: "Why do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" As the title suggests, the authors noticed a distinct pattern in how quotes, and by extension, their spreads, were made on NASDAQ. Their conclusion invokes game theory to posit the idea that tacit collusion among the market makers is an equilibrium of the market structure at that time and indeed state that they are "unable to envision any scenario in which 40 to 60 dealers who are competing for order flow would simultaneously and consistently avoid using odd-eighths quotes without an implicit agreement to post quotes only on the even price fractions." However, their report concludes that this data does "not provide direct

[22] Jeff Van Blarcom, "The Order Audit Trail System / OATS - Series 55," *Investopedia*, accessed March 1, 2015, http://www.investopedia.com/study-guide/series-55/commissions-and-trade-complaints/order-audit-trail-system-oats/.

evidence of tacit collusion."[23] The evidence required to definitely prove this collusion would have to come from the SEC.

The SEC did indeed conduct an investigation and concluded in 1996 that there was reasonable evidence to support the tacit collusion on the NASDAQ market and that NASD had not done enough to combat this issue. The report, which the SEC landed particularly hard on FINRA's predecessor, NASD, resulted in the SEC sueing NASD for failing to regulate NASDAQ and remarked that:

> In the course of the investigation, the Commission staff encountered significant difficulties reconstructing activity in the Nasdaq market. Broker-dealer order tickets, among the most fundamental of records, were too often unavailable or inconvenient to retrieve. Time stamping was often unreliable for the purposes of determining compliance with applicable rules, such as the firm quote rule and limit order protection rules.[24]

It seems all the more surprising given the fact that it took two years from the publication of the article regarding the oddities of the NASDAQ market and the investigation to conclude. However, the SEC in some ways blamed itself for failing to demand greater oversight from NASD.

Creation of OATS

The creation of OATS begins in 1996 with the settlement between NASD and the SEC to resolve the myriad issues found in the preceding price collusion scandal. For the following two years, working with the industry, the rules and form of OATS were

---

[23] William G. Christie and Paul H. Schultz, "Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?," *The Journal of Finance*, December 1994, Vol. 49, No. 5 edition.

[24] The Securities and Exchange Commission, *Appendix To Report Pursuant To Section 21(a) of the Securities Exchange Act of 1934 Regarding the NASD and the NASDAQ Market*, August 1996, https://www.sec.gov/litigation/investreport/nd21a-appx.txt.

determined and in March of 1998 the rules were approved by the SEC giving them the force of law. Throughout this two year process, requests for comment produced a dizzying array of responses and feedback from the industry.

The synchronization of clocks, a pertinent finding from the SEC's report, was to be completed first through a determination of what standard to use followed by general adoption. The final date that this process was to be completed was August of 1998 for computer clocks and July 1999 for manual clocks. The standard that was chosen was the National Institute of Standards and Technology (NIST) clock. At the time, the specification required reporting down to the nearest second.

The implementation of OATS reporting was broken down into three distinct phases in order to allow for compliance and because of general feasibility. Phase one involved reporting from those firms already utilizing certain electronic orders from market makers, large firms already heavily involved in the markets, and the trades over the electronic communications network. This reporting was to be completed by March of 1999 and by August of the same year phase two, that of all electronic orders being reported, would be completed. Phase three was the inclusion of all manual orders. While this was scripted to be completed by July of 2000, it wasn't until six years later that phase three was fully completed (*see Important Dates)*.

Particularly in response to the aforementioned market and regulatory failure, OATS was presented with two main objectives:

(1) provide an accurate, time-sequenced record of orders and transactions, and

(2) provide for market-wide synchronization of clocks used in connection with the audit trail.[25]

---

[25] The Securities and Exchange Commission and NASD, "Release No. 34-39729; File No. SR-NASD-97-56 - NASD Rulemaking: Various Orders Relating to the Creation of an Order Audit Trail System," March 6, 1998, http://www.sec.gov/rules/sro/nd9756o.htm.

These two projects are the basis for which OATS can be judged for this success or failure given much larger scope of FINRA's surveillance. Market data collection before OATS was plagued with an inaccurate clock system and a general lack of standardization. "The OATS Rules currently impose obligations on FINRA member firms to record in electronic form and report to FINRA on a daily basis certain information about orders originated, received, transmitted, modified, canceled or executed by firms relating to OTC equity securities and equity securities listed and traded on The Nasdaq Stock Market, Inc. (NASDAQ)."[26]

OATS was "designed to prevent fraudulent and manipulative acts and practices" and to "protect investors and the public interest." This remains part of the fundamental goal of OATS beyond simply ensuring clock synchronization or reporting standards. The data collected by OATS "would provide a substantially enhanced body of information regarding orders and transactions that would improve the NASDR's ability to conduct surveillance and investigations of member firms for violations of Association rules."[27]

THE CONSOLIDATED AUDIT TRAIL (CAT)

Events Leading up to CAT

One of the leading justifications for OATS was to create "an audit trail sufficient to enable the Association to reconstruct markets promptly, conduct efficient surveillance and enforce its rules."[28] While this is a worthy goal, and for a period OATS remained

---

[26] The Financial Industry Regulatory Authority, "FINRA Manual - Notices - 2011 - 11-03 FINRA Expands the Order Audit Trail System to All NMS Stocks; Effective Date: July 11, 2011," 2011, http://finra.complinet.com/en/display/display_main.html?rbid=2403&element_id=9945.
[27] The Securities and Exchange Commission, "Self-Regulatory Organizations; Notice of Filing of Proposed Rule Change by National Association of Securities Dealers, Inc. Relating to the Creation of New Rules 6900 Through 6970 or an Audit Trail System Owned and Operated by the National Association of Securities Dealers, Inc.," *Federal Register*, August 28, 1997, Vol. 62, No. 172 edition.
[28] Ibid.

sufficient to meet this goal, it has recently become clear that OATS and other market surveillance programs, are unable to keep pace with today's markets. High-frequency Trading (HFT), algorithmic trading, and the connectedness of the markets themselves have fundamentally changed the speed and dynamic nature of the stock markets and ushered in a new era of financial regulation.

The national market system is not what it once was and the days of simple market surveillance have ended. The ease with which large volumes can be traded within and across markets led the president of FINRA to remark that regulators do not currently have a consistent convention in place to identify a market participant across different markets and no easy way to determine which desk within a firm is responsible for a trade.[29] He further lamented that current audit trails are limited in their ability to identify trading activity that results from direct market access and specific trading strategies. This means that the audit trails themselves have become insufficient for successfully meeting their goals.

The Flash Crash

On May 6th, 2010 the Dow Jones Industrial Average (DJIA) experienced the biggest one-day point decline in its history[30]. Over the course of three minutes, from 2:41 pm to 2:44 pm, the DJIA lost 600 points, about 6%, and in almost the same amount of time regained it. Shares in Accenture fell from about $40 to just over $0.01, and many other stocks who were either directly following or were The report for what happened on this three-minute period wouldn't be completed until September of the same year, four full months later. The brevity and severity of this fluctuation in the market was deemed

[29] Financial Industry Regulatory Authority, "2009 Year in Review," 2010, http://www.finra.org/sites/default/files/Corporate/p121646.pdf.
[30] John Carney, "What Really Caused the Flash Crash," *CNBC*, accessed February 8, 2015, http://www.cnbc.com/id/42919462.

the "flash crash," and continues to give pause on the ability of the markets to withstand the brave new technological world.

The SEC report squarely puts the large trade of a particular stock, the e-mini, in particular the decision to sell a large quantity, as the root cause of the flash crash. This was exacerbated by certain traders on the market, which were programmed to respond in a certain manner to this sell off. Because the e-mini was supposed to track or mirror the movements of the S&P 500, a general indictor for the market as a whole, its sudden decline caused a knock on effect to other stocks which comprised or also mirrored that market. The result was that stocks in markets other than the DJIA saw marked decreases in very short time periods as well. There is some dispute, however, that sell-off was the root cause of the flash crash and instead reflects a more systemic market issue.[31]

The buzzword that was features prominently in this report is that of HFT and the high frequency traders who inhabit this space. On this particular day in May, 16 high frequency traders, a mere one tenth of one percent of the total number of traders for the day, accounted for nearly 30% of the trades.[32] This is not unexpected, as a similar level of trades by similarly few traders were seen in the days and weeks preceding the flash crash.

The report eventually came to two main conclusions of relevance for this paper[33]: (1) regulators were not able to provide timely monitoring of the financial markets; and, (2) "the events of May 6 clearly demonstrate the importance of data in today's world of fully-automated trading strategies and systems. This is further complicated by the many

---

[31] Philip Stafford, "Flash Crash Explanation Questioned," *Financial Times*, April 3, 2013, http://www.ft.com/intl/cms/s/0/52e2e7e0-9c5e-11e2-9a4b-00144feabdc0.html#axzz3V26HCHrh.

[32] The Securities and Exchange Commission and The Commodity Futures Trading Commission, *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, September 30, 2010, http://www.cftc.gov/ucm/groups/public/@otherif/documents/ifdocs/staff-findings050610.pdf.

[33] There were many other, more financially focused and technical conclusions reached as a result of this report. They are listed in greater detail in the SEC-CFTC report cited below.

sources of data that must be aggregated in order to form a complete picture of the markets upon which decisions to trade can be based. Varied data conventions, differing methods of communication, the sheer volume of quotes, orders, and trades produced each second, and even inherent time lags based on the laws of physics add yet more complexity."[34] In essence, the regulators were unable to gather data sufficiently or effectively in order to reconstruct the market in order to determine the cause of the flash crash.

High-Frequency Trading

The story of CAT would not be complete without a brief aside into a style of trading known as high-frequency trading (HFT). The idea of high frequency trading arises from recent technological advances that allow computers to trade at exceptionally large volumes and in exceedingly small timeframes. In some cases, the trade may be ordered and canceled in the same instant.[35] The SEC itself defines HFT as "HFTs are proprietary trading firms that use high speed systems to monitor market data and submit large numbers of orders to the markets. HFTs utilize quantitative and algorithmic methodologies to maximize the speed of their market access and trading strategies."[36]

Algorithmic trading, as the name suggests, allows the computer to make trading decisions independently of a human and significantly more rapidly and is the method that allows HFT to exist. The speed is increased through a process of colocation, where the servers handling the trades for the HFT firm are located close to, or sometimes even within the same building as, the exchanges server. This proximity allows the trades to

---

[34] The Securities and Exchange Commission and The Commodity Futures Trading Commission, *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.

[35] Austin Gerig, *High-Frequency Trading Synchronizes Prices in Financial Markets*, n.d., http://www.sec.gov/dera/staff-papers/working-papers/dera-wp-hft-synchronizes.pdf.

[36] The Securities and Exchange Commission and The Commodity Futures Trading Commission, *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.

take place in magnitudes of time faster, however the orders are still handled in the same system as every other trade.

One particular struggle with HFT and algorithmic trading is the simple fact that there is a chance, however, slight, that the incorrect "decision" may be made causing greater loss on the market in a shorter period of time. In short, the volatility of the markets could be impacted by these systems. The code and algorithms used for these platforms is still created by humans and there are certainly historical examples of HFT failure. How regulators should regulate this style of trading is the basis for continuing discussion. Following the flash crash, the SEC has worked with the markets to look at possible kill switches, which would be triggered if a certain amount is lost over a certain period of time.

The debate will likely continue into whether HFT is "good" or "bad" for the market, but until it is deemed illegal or becomes unprofitable, it is likely here to stay. The SEC released a white paper detail the ways in which HFT may be beneficial in creating more accurate and less difference between the buy price and sell price, thereby arriving at a more true value. On the other hand, these proprietary data feeds have been viewed as a substantial, almost illegal, competitive advantage. This is because the information these data feeds receive is much timelier, often with millisecond delays, as opposed to what regular, non-proprietary data feeds receive which may be in the order of tens of seconds.

Creation of CAT

On May 26, 2010, twenty days following the flash crash, the SEC proposed rule 613, which called for the creation of a consolidated audit trail (CAT) - Release No. 34-62174; File No. S7-11-10. One of the primary cited reason for the creation was that the "proposed consolidated audit trail, if implemented, would have significantly enhanced the

Commission's ability to quickly reconstruct and analyze the severe market disruption that occurred on May 6, 2010. If approved and implemented, the proposal also would enhance the Commission's ability to similarly respond to future severe market events." Simply, the interconnectedness and speed of the markets could no longer be ignored.

The proposed consolidated audit trail is intended to[37]:

- Provide regulators direct and timely access to uniform consolidated order and execution information for all orders in National Market System (NMS) securities from all market participants across all markets.
- Enable SROs to better fulfill their regulatory responsibilities to oversee their markets and their members.
- Enable the SEC to better carry out its oversight of the NMS for securities and to perform rapid and accurate market analysis.

Among the other considerations were the various audit trails that already existed within the SRO and market frameworks, including the current Electronic Blue Sheet model employed the SEC itself. As discussed previously, OATS had been in place for over a decade. Additionally, the New York Stock Exchange had created the Order Tracking System (OTS) in 1999, and a year later several other options markets had adopted the Consolidated Options Audit Trail system (COATS). However, each of these audit systems required varying amounts and types of information, different reporting processes, and, as a result, could not easily be aggregated. CAT would substantially increase the regulatory oversight where the aforementioned audit trails had failed.

More succinctly, the creation of CAT followed directly the twin goals of market surveillance and auditing in general: (1) "A consolidated audit trail will increase the data

[37] The Securities and Exchange Commission, "SEC Proposes Consolidated Audit Trail System to Better Track Market Trades," Government, (May 26, 2010), http://www.sec.gov/news/press/2010/2010-86.htm.

available to regulators investigating illegal activities such as insider trading and market manipulation." [38] Meaning, this data can be used to detect market abuse that is otherwise hidden in the tangle of data currently collected. And that CAT will (2) "significantly improve the ability to reconstruct broad-based market events in an accurate and timely manner." [39]

In addition to the two goals of detecting abuse and recreating market conditions, the SEC hoped that CAT would "significantly increase the ability of regulators to monitor overall market structure and assess how SEC rules are affecting the markets, and will reduce the regulatory data production burdens on SROs and broker-dealers by reducing the number of ad hoc requests from regulators presently."[40] This entailed that regulators would have greater access to information without require an equal increase in the methods they already use for gathering information.

[38] The Securities and Exchange Commission, "SEC Approves New Rule Requiring Consolidated Audit Trail to Monitor and Analyze Trading Activity," July 11, 2012, http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171483188#.VO0AGyvF98E.
[39] Ibid.
[40] Ibid.

CHAPTER 3: BIG DATA ISSUES RAISED WITH OATS AND CAT

INTRODUCTION

As with most financial regulation, and regulation in general, there were initial issues raised with many facets of the both the OATS and CAT programs. Many of these issues were questions about the cost and effectiveness of the systems, particularly in how the SRO would be able to collect and analyze such large amounts of data in a suitable timeframe to carry out their regulatory duties. The sheer scale of the data collection was daunting to say the least and presented an initial concern. Data retention issues were raised especially in regard to how data errors would be efficiently identified and processed. Further, in retaining the data, how would security be assured became a growing issue.

Additional arguments have been made regarding the necessity of such systems given that other, perhaps more effective, solutions exist. Notably, the data the uncovered the NASDAQ scandal that preceding the creation of OATS was in fact readily available, and in some ways, blindingly obvious, to those looking at it. Data analysis was required, but the data used in the report was not proprietary nor difficult to obtain. At root for many of the late 20th and early 21st century's biggest scandals, too, was whistleblowing or industry calls to investigate as opposed to data collection and analysis.

"Electronic reporting requirements continue to have the most adverse effect on capital expenditures of the firm. Such initiatives as OATS, OTS, TRACE, MSRB, Reg SHO and other reporting requirements continue to be introduced or amended at a record pace. The programming costs on these types of initiatives have grown over the last two years. In addition, many firms have utilized outside vendors for software solutions to

meet these deadlines. The costs associated to these software solutions can and have had a definite impact on compliance-related capital expenditures."[41]

EFFECTIVENESS

The effectiveness of the OATS and CAT program can be measured using three primary criteria with multiple subcategories comprising each. The first of these criteria is data collection. Technically, this criteria is a measure of how effectively the data can be ingested given the size and time requirements placed upon the system. The human factor arises in the fact that over 1,000 financial institutions will eventually report to CAT, each of these institutions will have people maintaining, collecting, and reporting this data. Even in the most robust and automated situation, there is always the chance that information will be improperly collected or maintained.

The second criteria is the retention of the data once it is inside the system. Since this data will contain information that is personal or can be seen as useful, its security is an important consideration. The ability also to correct data that has been rejected remains a topic. Access to the data by the members reporting to it, as is currently the practice with OATS, is also a metric by which the data projects will be measured.

Lastly, the analysis of the data once it is in the system is nearly the entire reason the project exists in the first place. The goal is to reconstruct markets much faster than before and with the same or greater detail. Actions taken by the SEC and FINRA should experience a decrease in time or at least some variation in frequency.

---

[41] Stephen J Nelson, "Commentary: The Promise of a Consolidated Order Audit Trail," *Traders Magazine Online News*, September 16, 2010, http://www.tradersmagazine.com/news/consolidated-order-audit-trail-sec-nasdaq-nyse-106337-1.html.

Data Collection

The collection of data is the first step in the analysis project and can present a myriad of issues that primarily effect the initial stage of the big data project. Firstly, setting up the database and creating the connections to the reporters presents the first large challenge. OATS has been shifted to the cloud in order to obtain efficiencies not found in their current systems and is expected to save FINRA $10 to $20 million annually.[42] In terms of CAT, there were varying bids that utilized cloud or distributed databases in order to cut costs to creating dedicated connections and databases in order to solve the problems unique to CAT. The most pressing issue for both has been the clock synchronization, with OATS currently requiring second level data, but CAT migrating to the more granular 50 millisecond level.

Quantity

The amount of data collected under the proposed CAT system is immense. The requirements for the CAT system stipulate that the system must be able to handle over 5 terabytes of data per day, at a minimum. Within five years, the database was expected to grow to well over 21 petabytes of data.[43] The system must be able to process at least one terabyte a day from a variety of sources and at a variety of times in order to adequately fulfill its duties. One particular bidder for the CAT project stated that the current big data systems that are employed are not good enough for this project, even as they handle record amounts of data.

The physical space required to store that data is immense. Currently, the NYSE data center in Mahwah, New Jersey occupies 400,000 square feet, about 60,000 for actual

[42] Lash, "Wall St Watchdog Moves to Cloud, Big Data, to Boost Capabilities."
[43] Financial Industry Regulatory Authority, "Summary of the Consolidated Audit Trail Initiative," January 2015, http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p571933.pdf.

data storage, and has the capacity to use 28 MW of electricity and cost around $500 million to build.[44] The NSA, meanwhile, constructed for $1.2 billion[45] its data center in Utah. The expansion of the cloud, generally an interconnected web of large data centers, has cut the cost and made the necessity for a standalone data center less appealing.

Clock Synchronization

In stock market trading, it is said that timing is everything. Indeed knowing when an event took place is becoming a major factor in the financial markets and, as alluded to earlier, has given rise to a style of trading for which every millisecond matters. This subsection's subject, timing, requires clocks to be synchronized to within a certain bounds of a standard.

In the creation of OATS, the SEC laid out that "the proposal would require member firms to synchronize their business clocks and continually to keep them synchronized with a specific time designated by the Association."[46] This was easier said than done. Writing in the Federal Register, NASD remarked that "members and electronic communication networks and service bureaus use a variety of methods for synchronizing business clocks" and as such "accuracy of recorded times may vary significantly among member firms."[47]

---

[44] Data Center Knowledge, "Closer Look: NYSE Euronext's NJ Data Center," *Data Center Knowledge*, accessed March 19, 2015, http://www.datacenterknowledge.com/closer-look-nyse-euronexts-nj-data-center/.

[45] Kashmir Hill, "Blueprints Of NSA's Ridiculously Expensive Data Center In Utah Suggest It Holds Less Info Than Thought," *Forbes*, accessed March 19, 2015, http://www.forbes.com/sites/kashmirhill/2013/07/24/blueprints-of-nsa-data-center-in-utah-suggest-its-storage-capacity-is-less-impressive-than-thought/.

[46] The Securities and Exchange Commission and NASD, "Release No. 34-39729; File No. SR-NASD-97-56 - NASD Rulemaking: Various Orders Relating to the Creation of an Order Audit Trail System."

[47] The Securities and Exchange Commission, "Self-Regulatory Organizations; Notice of Filing of Proposed Rule Change by National Association of Securities Dealers, Inc. Relating to the Creation of New Rules 6900 Through 6970 or an Audit Trail System Owned and Operated by the National Association of Securities Dealers, Inc."

For CAT the issue of clock synchronization exposes how divergent trading timescales have become. The rise of high frequency trading and proprietary data platforms gives ample advantage to those who can afford to pay to know. At the time of the flash crash, the NYSE created its consolidated quotes with a 10 second delay, reaching a 40 second lag at one point. For the proprietary feeds, that delay was 8 milliseconds (0.008 seconds).[48] By many measures, the price we see is already ancient.

The concern arises, then, from accurately acquiring the data while not causing increased burdens on those that are not operating at millisecond intervals. The solution is for CAT to operate that 50millisecond offset. According to one survey by the Financial Information Forum[49], nearly 40% of their respondents are not currently reporting at the 50 millisecond level. Further, synchronizing the clocks to such a minute level would require GPS links or Precision Time Protocol (PTP) upgrades from current Network Time Protocols (NTP) that are fairly expensive for many firms. The survey concluded that in order to achieve a 50 millisecond offset, their respondents would have to invest nearly $13 million initially to obtain that level of granularity and that annual costs would increase 31% to maintain it.[50]

Data Retention

Often the collection of data, especially with automated systems, is not necessarily the major issue. Instead, the major costs can come from storing and maintaining the dataset, keeping the data secure while also allowing useful analysis of the data, and allowing for corrections when data has been incorrectly entered.

---

[48] The Securities and Exchange Commission and The Commodity Futures Trading Commission, *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.
[49] Manisha Kimmel, "FIF Clock Offset Survey Preliminary Report," February 16, 2015, http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p602479.pdf.
[50] Ibid.

Security

There is little doubt that the information collected by CAT and OATS are exceptionally significant. In the wrongs hands, or perhaps better said, in the *right* hands, this data offers a treasure trove of information regarding trading strategies, general market trends, and person-specific information. Security, then, has consistently been high on the list of requirements for both systems. While fear, uncertainty, and doubt tend to prevail and cloud discussions of this topic, as with any connected system, the right precautions need to be taken in order to insure information security.

To date, OATS has not had a reported serious security breach. However, OATS remains a relatively small target compared to the financial market and institutions that report to it. CAT, however, may present a target that is much more comparable, though no less challenging to penetrate. The threat vectors to the system are numerous. They include not only the physical security of the database, but also the data in transit, a risk that can be mitigated through encryption and other devices. However, encrypting the data will add an additional layer of cost and trouble for the reporters and further the data will have to be decrypted in order to be analyzed. With such large quantities of data, this may become more than a trifling issue.

Ultimately, these security of the data will likely follow the established C.I.A. triad of cybersecurity: Confidentiality, Integrity, and Accessibility. In short, the data will have remain confidential with strict controls on who can see what data. The data must retain its integrity, meaning that it cannot be changed or altered except by someone authorized to do so. And finally, the data must be accessible, meaning that it is available for use for as much time as possible. Nearly every member of this project from FINRA and the SEC to the financial industry members themselves have been quick to take information security

seriously. The process of balancing the risks and costs will likely continue and be better informed as more attention is paid to adopting stricter cybersecurity practices.

Correction

As many computer users can attest, working with data big or small is often a tedious task. It is a near certainty that there will be errors within the data that will require correction from their source. Finding and correcting that data can be challenging and as the data set grows, using more automated systems, such as machine learning for pattern detection and outliers, becomes a necessity. Once an issues is found, it then needs to be corrected, which means soliciting additional data from the source and additional research.

Currently, the OATS program requires 24 hours for errors to be detected, another up to 5 business days for the error to be corrected. CAT is seeking to decrease this to three days. Indeed, some of the bidders contend that given today's technology, error detection should be done instantly and correction swiftly. The five day delay, however, means that regulators will be operating with incomplete data and will be analyzing the market in many ways as it existed five days prior. As the second chapter should have made clear, this may not be a very efficient way of measuring markets that are operating at millisecond speeds.

Part of this system too relies on a good feedback loop between the processor and the reporter. OATS currently implements a reporter portal, which allows the users to maintain their supervision of the reports they have submitted and look for rejections and anomalies. Many of the bidders for CAT stated goals for a similar process that would look for particular patterns and create a dashboard so that if a certain increase in rejects is occurring a certain time or from a certain place, the firm can look into it and solve the issue at the source.

Data Analysis

Given the size and complexity of the data collected, many issues were raised with how the regulators would be able to monitor, collect, correct, and then analyze the data to produce useful results in acceptable timeframes. The technical issues of big data are evident in how the various bidders for CAT presented their prospective programs. The who's who of the database creation, management, and search were present. The bidders critiqued and expounded upon the virtues of various styles of collecting and analyzing the market data that is collected. Some bidders suggested that stand alone systems are required for such an immense and large scale project. A greater majority offered proprietary solutions based upon existing enterprise-level data management programs that were currently in place.[51]

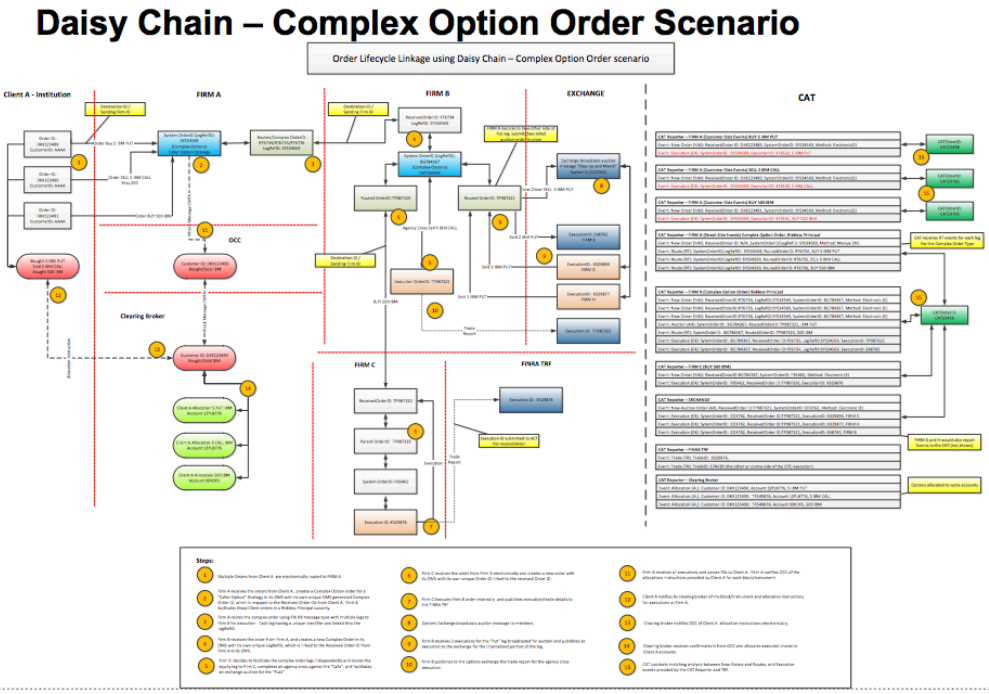
System	Strengths	Weaknesses
Relational DBMS	Complex, flexible schema. ACID properties.	Poor scalability
Hadoop/MapReduce + HDFS/GFS	Highly scalable Organize many compute nodes	Slow response times Dedicated storage/compute ratios
HBASE/Cassandra/BigTable	Indexing added without the scalability issues of RDBMS	Generic approach adds large storage overhead
Apache SPARK/Google Dremel/ BigQuery/Redshift	Scalability of Hadoop with better response times	Optimized for specific analyses which don't map well to CAT (security and optimization)

Table 3.1 Big Data Systems

Further complicating matters in the analysis is getting a precise idea of the order tracking. Currently OATS employs a method known as the "daisy chain," which as its name suggests looks for where an order appears to begin and end and matches both with other information presented by the various reporters to CAT. Eventually, this will lead to

[51] Thysis Technologies, "Proposed CAT Solution: SIFMA CAT Bidder Summit," July 29, 2014, http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/thesys_technologies_cat_presentation_sifma_summit_final_posted_072914.pdf?n=06244.

an originator and end point that can be traced back through the complete cycle. This cycle can be incredible complex and can be easily broken by a failure to report or an incorrect identifier to link the events. Below is an image presented by one of the bidders that highlights this complexity in a visual manner.



[52]

Figure 3.1 Daisy Chain Linkage System

COST

[52] Sungard, "Consolidated Audit Trail Solution Overview: SIFMA Bidders Summit," July 29, 2014, http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/sungard-cat-presentation-sifma-bidders-summit-july292014.pdf?n=66627.

Table 1: Current Costs: Large Respondents Summary (49 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$310,000	9.56	$3,800,000	$180,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$1,000	0.13	$52,000	$1,000
Maximum	$6,000,000	190.00	$76,300,000	$6,000,000
Count of Zero Responses	31	25	25	36
Count of Blank Responses	0	0	0	0

Table 2: Current Costs: Small Respondents Summary (118 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$130,000	2.36	$950,000	$130,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$1,000	0.15	$60,000	$1,000
Maximum	$14,000,000	68.00	$27,300,000	$6,500,000
Count of Zero Responses	96	89	89	93
Count of Blank Responses	0	0	0	0

Table 3.2 Current Cost of OATS Compliance

As with most new regulation, there will be costs involved as firms must implement their controls in order to become compliant and on the side of the regulators to ensure compliance. OATS and CAT both experienced considerable pushback on the costs versus the benefits they seem to confer. Compliance with OATS in particular has given many pause about the subsequent expansion with CAT. Specifically, the fines levied for not complying are seen as strictly punitive and unconstructive.[53]

While data on the total implementation costs of OATS are difficult to discern, one industry survey and report into OATS' expansion in the National Market System was completed in 11 months at a cost of nearly $50 million. This does not include the additional costs to FINRA of expanding their operations. According to the cost benefit study conducted for the CAT NMS plan, however, industry respondents reported an

[53] Dan Jamieson, "Brokers Charge FINRA with Becoming OATS Mill," *Investment News*, April 24, 2011.

"average cost of $4,290,000, and small firms reported an average cost of $1,210,000 for current reporting costs."[54] However, OATS specific data reveals that OATS respondents pay considerably more than their non-OATS counterparts at both small and large firm levels. Large OATS reporting firms reported an average of "$8,320,000 while large non-OATS respondents estimated an average cost equal to $1,324,600. Small OATS reporters estimated an average cost equal to $3,500,000 while small non-OATS respondents estimated an average cost equal to $433,800."[55] Thus an expansion of OATS might produce additional burdens on an already comparatively overburdened reporting base.

Table 7: Approach 1 Implementation Costs: Large Respondents Summary (49 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$580,000	11.00	$4,400,000	$72,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$5,000	0.02	$8,000	$1,000
Maximum	$10,000,000	142.00	$57,000,000	$2,000,000
Count of Zero Responses	28	27	27	41
Count of Blank Responses	0	0	0	0

Table 8: Approach 1 Implementation Costs: Small Respondents Summary (118 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$5,200	1.17	$470,000	$76,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$1,000	0.20	$80,000	$1,000
Maximum	$500,000	20.00	$8,000,000	$8,000,000
Count of Zero Responses	95	94	94	95
Count of Blank Responses	2	0	0	1

Table 3.3 Approach I Implementation Costs

[54] Participants of NMS Governing Plan, "National Market System Plan Governing the Consolidated Audit Trail Pursuant to Rule 613 of Regulation NMS under the Securities and Exchange Act of 1934," February 27, 2015, Appendix C, http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p602500.pdf.
[55] Ibid.

The costs of cost-benefit report for CAT are broken up into two different approaches to ingesting data into the system. The first relies on firms submitting data based upon their own current data formats and standards, while the second would require data to be in a specific format like that of OATS. Both offer various savings and costs to OATS and non-OATS reporting firms both large and small.

Table 13: Approach 1 Maintenance Costs: Large Respondents Summary (49 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$210,000	8.54	$3,400,000	$52,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$2,000	0.02	$8,000	$1,000
Maximum	$5,200,000	152.00	$61,000,000	$1,000,000
Count of Zero Responses	28	27	27	41
Count of Blank Responses	1	0	0	0

Table 14: Approach 1 Maintenance Costs: Small Respondents Summary (118 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$1,600	1.12	$450,000	$24,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$500	0.15	$60,000	$500
Maximum	$120,000	18.00	$7,200,000	$1,500,000
Count of Zero Responses	96	93	93	96
Count of Blank Responses	0	0	0	0

Table 3.4 Approach I Maintenance Costs

The report concluded that "total average (median) costs for Approach 1 Implementation are estimated to be $5,052,000 ($0) for large firms, and $551,200 ($0) for small firms."[56] For OATS reporting firms the average is slated to be nearly double, approximately $3,000,000 higher, than that of non-OATS reporting large firms, while small OATS reporting firms are expected to pay nearly 10 times the amount of small

[56] Ibid.

non-reporting firms. Largely, the small firms will have to upgrade their hardware and employ more people to comply with the new regulation.

The maintenance costs for the first approach averages $3,662,000 for large firms and $475,600 for small firms. Between the large firm OATS and non-OATS reporters, the maintenance costs are over one million dollars higher for OATS reporters than the non-OATS reporters. A bulk of this difference come from employee costs. For small firms, the cost is significantly greater for OATS reporters, at nearly $1.5 million versus only $121,200 for non-OATS reporters. Like the large firms, the large difference in costs arise from employees as well as much larger hardware and outsources costs relative to the non-OATS reporting firms.

Table 19: Approach 2 Implementation Costs: Large Respondents Summary (49 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$570,000	10.15	$4,100,000	$68,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$5,000	0.02	$8,000	$1,000
Maximum	$10,000,000	116.00	$46,600,000	$2,000,000
Count of Zero Responses	28	28	28	41
Count of Blank Responses	0	0	0	0

Table 20: Approach 2 Implementation Costs: Small Respondents Summary (118 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$5,000	1.08	$440,000	$16,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$1,000	1.00	$400,000	$1,000
Maximum	$500,000	20.00	$8,000,000	$1,000,000
Count of Zero Responses	98	96	96	97
Count of Blank Responses	1	0	0	1

Table 3.5 Approach II Implementation Costs

Meanwhile, the average for the second approach, that of OATS like standard data ingesting, was lower with a “total average costs for Approach 2 Implementation are

estimated to be $4,738,000 for large firms, and $461,000 for small firms."[57] Specified between the OATS and non-OATS reporters, however, the cost will decrease for both, although the gap between the costs are nearly $200,000 wider for this approach, with the savings made by non-OATS reporters.

Table 25: Approach 2 Maintenance Costs: Large Respondents Summary (49 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$200,000	7.27	$2,900,000	$48,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$2,000	0.00	$0	$1,000
Maximum	$5,200,000	102.00	$40,900,000	$1,000,000
Count of Zero Responses	28	28	28	41
Count of Blank Responses	1	0	0	0

Table 26: Approach 2 Maintenance Costs: Small Respondents Summary (118 Firms)				
	Hardware / Software	FTE Counts	FTE Costs	Third Party / Outsourcing
Average	$1,500	1.06	$430,000	$10,000
Median	$0	0.00	$0	$0
Minimum	$0	0.00	$0	$0
Minimum (non-zero)	$500	1.00	$400,000	$500
Maximum	$100,000	18.00	$7,000,000	$1,000,000
Count of Zero Responses	97	94	94	93
Count of Blank Responses	2	0	0	5

Table 3.6 Approach II Maintenance Costs

The maintenance costs for the second approach, that of ingesting through an OATS like data standardization, average costs for Approach 2 Maintenance are estimated to be $3,148,000 for large firms, and $441,500 for small firms. The large firms differed in their maintenance costs by over $2 million dollars, with OATS reporters reporting a larger cost. For small firms, as with the first approach, the OATS reporters can expect to pay nearly ten times the amount required of non-OATS reporters for maintenance costs.

[57] Ibid.

From these reports we can conclude that most of the rise and the largest differences in costs for OATS and CAT are borne by small firms regardless of the approach decided upon for CAT. Current OATS reporters, who will more than likely report to the CAT system as well, stand to pay a considerable share larger than current non-OATS reporting firms, even though non-reporting firms will still face a greater increase in costs as a result of this program.

INDUSTRY CONCERNS

Naturally, the financial industry was quick to post comments during the request-for-comment (RFC) period as many of the proposed rules would have direct effects on their business operations.[58] Most of the comments for both OATS and CAT recognized what the programs were attempting to achieve were admirable, but questioned some of the methods and scope. Some of the concerns that echoed in both the OATS and CAT comment letters involved the anti-competitiveness of the programs, in that the burden to comply would be much greater on smaller firms than on larger ones. Additionally, the duplicity of systems currently employed and leveraging existing systems for the same effect were prevalent in these comments.

Anti-Competitiveness

The demands for OATS and CAT systems have raised anti-competitive concerns because they will be unequally burdensome for financial institutions. Specifically, smaller institutions that perhaps do not have the resources to devote to advanced information technology or data collection and reporting will face a higher cost to comply with the new audit systems. It may be, as well, that these small institutions are not even

[58] The Securities Industry and Financial Markets Association, *Industry Recommendations for the Creation of a Consolidated Audit Trail*, March 28, 2013.

able to comply with the reporting system given their particular methods of trading, collecting data, or the ability to report it. As a result, certain provisions and exemptions were made in the roll out of OATS and have already begun for CAT, even though reporting has yet to begin.

Duplicative Systems

One of the main omissions of the CAT policy was the lack of a timeframe to sunset the current collection programs such as OATS. As CAT is intended in many ways to replace, standardize, and expand the current systems being used, it makes sense that these other programs would wind down. Especially worrisome is the fact that the similar systems will require varying degrees of reporting, metrics, data, timelines, and even timescales. Further, the reporters will likely have to work with different parts of the compliance agency for very similar purposes.

CHAPTER 4: BIG DATA ISSUE RESPONSES AND ANALYSIS

INTRODUCTION

This chapter will build on and respond to the previous chapter's questions into the necessity of big data programs. The purpose is to glean useful information regarding how greater surveillance and analytics is useful in a "real world" example. While CAT remains a proposal at this stage, OATS has been working in various capacities since 1998. This has not been a continuous motion, however. OATS has been expanded, changed, and altered throughout this time to respond to changing economic and political conditions.

The responses by regulators to the criticisms and issues raised on the previous chapter are a starting point for looking at the analysis of CAT and OATS. Throughout the comment and adoption periods for each program, the SEC and FINRA have been quick to note the differences and discrepancies for the two in addition to the overall effect that audit trail systems have on the agency and SRO mission. These responses provide the logical basis for the big data projects and overtly hint at the panacea they represent. Further, we can use the responses to better inform our metrics in analyzing the systems.

This chapter moves on first with an argument about why OATS can be used as an indicator for the CAT project and why the two are analogous. This is important in establishing a basis for proceeding to the analysis of OATS' effectiveness in relation to the promise of CAT. From here, we can look at how OATS has met or failed to meet the metrics for analysis.

Finally, this chapter concludes with a discussion of the Comprehensive Automated Risk Data System (CARDS), which stands to increase the ability of regulators to determine a more difficult regulator target: that of risk and suitability.

REGULATOR RESPONSES

In seeking to understand the effects that CAT may have on financial regulation as a big data program, it is important to return to the source of CAT and the regulator's responses to the issues of financial regulation and OATS. It cannot be understated that the sole purpose that billions of dollars, from both taxpayers and the private sector, is spent on financial regulation is to create a fair and free market, provide security by enforcing rules, and increase trust. In that endeavor, there is latitude for compromise, especially when a regulation becomes too burdensome or unfairly affects a particular market participant. That latitude ends, however, when not implementing a regulation or program might result in a continuation of a status quo that led to poor consequences. This continual balancing act is the political reality that CAT and OATS were born into. Regulators thus had provide concrete reasons why implementing these programs would support their overall mission.

Disparate Systems

One of the major justifications and criticisms for the introduction for CAT was that the markets were regulated under different disparate systems.[59] One of the first questions to arise in the face of CAT was how adding yet another large systemic reporting mechanism would help. Yet it was precisely the fact that there are so many systems: OATS, COATS, OTS, etc.[60] that trying to combine the data and analyzing it is not realistic. Steve Cohen, an investigator at the SEC, remarked in 2012, two years following the flash crash, that the Commission had to "cobble together data from disparate systems, each incomplete, inaccurate, inaccessible and untimely in their own

[59] Nelson, "Commentary."
[60] For a more comprehensive list, please see "Other Reporting Systems" illustration.

way."[61] These systems are incompatible in their scope and data created a system where gathering the data was a costly and time consuming endeavor.

As noted previously, the different audit systems put different degrees of burden on the markets themselves, which in turn could affect whether or not a firm would choose to enter that market. In this sense, there was an incentive to have the least burdensome audit system while still complying with regulation instead of more comprehensive requirements. Therefore, merging these systems would eliminate that form of competition and provide a standardization that would apply across the markets.

Inadequate Data

In addition to the disparate systems producing varied and untimely data, the data itself lacked meaningful information. The lack of access to customer information, for instance, has "hindered and slowed the division's investigations, hiding potentially manipulative activity."[62] Further, the errors and corrections for the data put into the audit trails was an area of concern. The timeline for OATS of 24 hours to detect an anomaly or issue, followed by the five days it could take to receive a correction, is not adequate for a market that can drastically change in minutes. Implementing CAT has the distinct possibility and goal of modernizing and speeding up the investigation and analysis process. This is suggested to make it easier for firms to report their data without having the regulators unnecessarily focusing on getting the data as opposed to the greater issues of market malpractice and failure.

---

[61] Steven SloanNina Mehta, "SEC Votes to Require Consolidated Audit Trail for Markets," *Bloomberg.com*, accessed March 1, 2015, http://www.bloomberg.com/news/articles/2012-07-11/sec-votes-to-require-consolidated-audit-trail-for-markets.
[62] Ibid.

Increasing Capability

The regulator responses have not all been focused on the negative aspects of data collection, analysis and regulation. In the yearly address, the SEC stated that "over the last several years, OCIE has made significant enhancements in data analytics that enable us to efficiently and effectively analyze the data to which we have access. We will use these capabilities to focus on registrants and firms that appear to be potentially engaged in fraudulent and/or other potential illegal activity."[63] As greater funds and personnel are allocated with specific tasks of analyzing and reporting on the data received, the usefulness of the data will increase.

ANALYSIS OF OATS EFFECTIVENESS AS AN INDICATOR FOR CAT

It is said of history that it doesn't often repeat exactly, but it most assuredly rhymes. Likewise, the CAT project was borne of similar tumultuous roots and to achieve similar notable ends as OATS. The substantive causes and justifications for each program remain relatively unchanged: (1) reconstruct the market; (2) detect market abuses. Consistently listed among the criticisms of the CAT program is the fact that OATS, in essence, is already supposed to complete the task of CAT. It stands to reason, then, that simply expanding or adapting the OATS program would be a cost effective and easy method for solving the issues that prompted the creation of CAT. But how has OATS met or failed to achieve the goals and how can it be measured as a big data system? I will look at how OATS has measured up to the goals of reconstructing the market and detecting market abuses.

---

[63] The Securities and Exchange Commission and The Office of Compliance Inspections and Examinations, "National Exam Program Examination Priorities for 2015," 2015, http://www.sec.gov/about/offices/ocie/national-examination-program-priorities-2015.pdf.

EFFECTIVENESS

It would appear from the outset that OATS has indeed failed to keep pace with the technology and dynamic nature of the market – embodied by the flash crash of 2010. Similarly, there have been spectacular and very notable market failures and abuses in the years following the adoption of OATS that might have, in part, been handled with adequate data analysis. In order to evaluate the effectiveness of OATS, this section reviews the ability of OATS to recreate markets using the metric of data collection and aggregation. A metric of the activities and actions opened and closed by regulators over time can be used as a metric for how much more efficiently or effectively regulators have utilized these new technologies, all else being equal. The cost of the regulation, already partially discussed in chapter three, can additionally be viewed and examined through the lens of the resources devoted to regulating and OATS in particular over the period of time since OATS' inception. Finally, the time to completion of regulatory actions, a key selling point for this new technology, can be determined using new and historical data.

Recreating Markets

One of the major roles for OATS is to recreate market conditions at specific points or durations of time. This must be done for a variety of reasons, but the fact of the matter is that in order to achieve the recreation results, accurate, timely data is required. This has proved challenging in many respects but hinges primarily on the issue of data collection. The penalties for failing to comply with the data reporting can be considerably

harsh.[64] Sometimes failure to report is tagged on to previously uncovered illicit activities.[65]

Data Collection

While the arguments seem persuasive that the sheer scale and size of the data collected by OATS would be prohibitive to the project's success, recent years have seen large increase in the amount and frequency of data collected by FINRA. Figure 1 below shows data collected by FINRA regarding the OATS program and provides a good overview of the preceding decade in reporting.

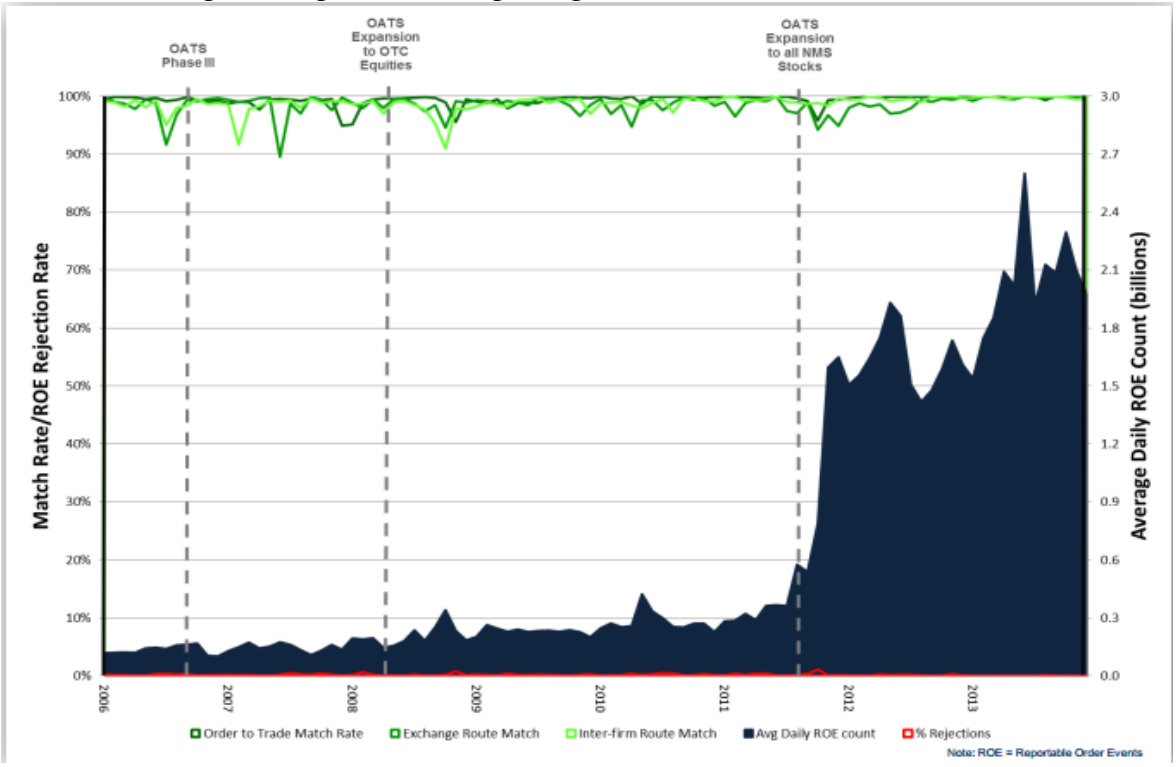


Figure 4.1 OATS Reporting Statistics[66]

[64] Financial Industry Regulatory Authority, "FINRA FINES 3 FIRMS A TOTAL OF $1.6M FOR OATS VIOLATIONS," *Bloomberg*, May 15, 2008, http://www.bloomberg.com/apps/news?pid=newsarchive&sid=awcbFR97xMCc.

[65] Marlene Y. Satter, "SEC, FINRA Enforcement: Money Manager Charged in Free-Riding Scheme," *ThinkAdvisor*, September 5, 2013, http://www.thinkadvisor.com/2013/09/05/sec-finra-enforcement-money-manager-charged-in-fre.

Indeed it is clear to see that although the reportable order events have increased substantially since late 2011 when OATS program was expanded to all National Market System stocks, the match rate has remained near 100% and rejection rate near 0%.

In 2010, it could not be said that OATS itself failed to recreate the market following the flash crash that prompted CAT. In many ways, it was the failure of regulatory oversight in general, however reporting across disparate systems, markets, and at varying degrees of specificity made the regulators job of recreating the market impractical in a short period of time. The issue then was that the right type, frequency, and place of data being collected was insufficient, rather than an overall lack of data collection in general.

Detecting Abuse

Detecting market abuse and illicit activities is one of the core and consistent themes for adopting greater surveillance. These activities range from insider trading to more complicated front running and churning, which involves a broker dealer trading with a frequency that is not in line with a particular investors goals in order to obtain a higher commission.[67] Data can be used to both provide evidence and detect these crimes. We would expect, however, that with the increase in data surveillance more of such activities would be found. Operating under the assumption that there are not fewer illicit actors today than before, we can begin our analysis of OATS in detecting market abuse.

Concrete data is difficult to come by for how many investigations were solely based upon the analysis of market data. Many actions arose from whistle blowers and

---

[66] Financial Industry Regulatory Authority, "Consolidated Audit Trail," July 29, 2014, http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/thesys_technologies_cat_presentation_sifma_summit_final_posted_072914.pdf?n=06244.

[67] The Securities and Exchange Commission, "What Is Churning?," Government, *Churning*, (n.d.), http://www.sec.gov/answers/churning.htm.

industry insiders themselves, which has been the goal and a major push for regulators seeking greater involvement from their regulated counterparts. Further, the types of illegal activities has shifted during the implementation period of OATS. However, in ascertaining the detection of market abuse, we can start by asking simply if the introduction and expansion of OATS has coincided with an increase in detection of abuse.

Data from NASD is limited[68], however, we can get disciplinary action data as far back as 2003, well before OATS phase III was entirely completed in late 2006. Data prior to 2003 is sporadic, as it was not consistently reported in the annual reports. Prior to and including 2006, there were an average of 1,353 disciplinary actions of all types filed per year. During these four years, NASD resolved an average of 1,288 actions. However, after 2006, FINRA the year average action declined to 1,335 with case resolutions also declining to an average of 1,182. It should be noted, however, that the period after 2007 also witnessed the fewest number of actions taken (in 2008) and the most (in 2012). This coincides with both extreme market activity, as with the great recession in 2008, and the expansion of FINRA into the national market system in late 2011. However, it does not appear that OATS or the data collection apparatus itself has effectively increased disciplinary actions as a whole.

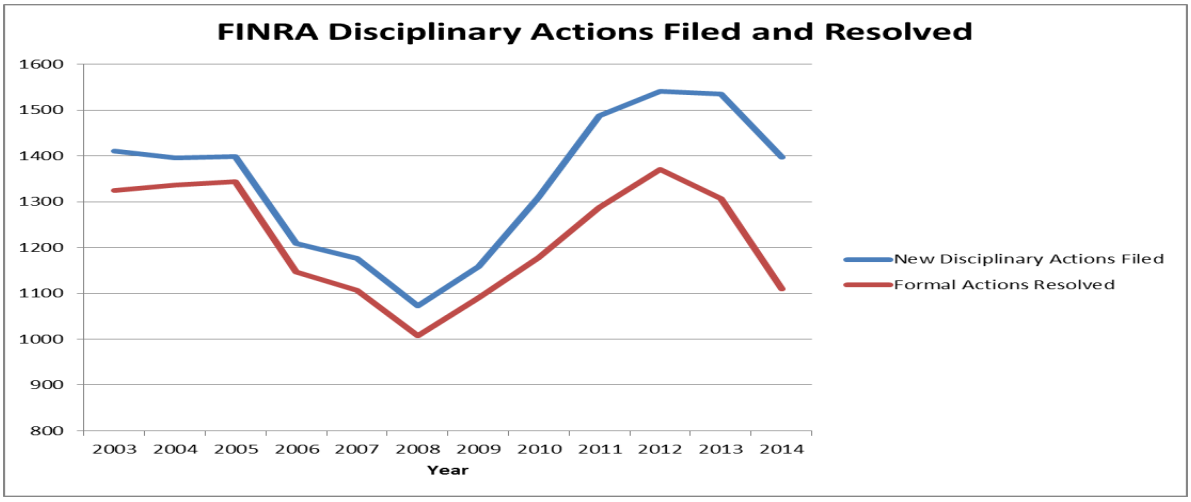
Table 4.1 Number of Disciplinary Actions Filed and Resolved Per Year

Year	2003	2004	2005	2006	2007	2008
New Disciplinary Actions Filed	1410	1396	1399	1209	1177	1073
Formal Actions Resolved	1324	1336	1344	1147	1107	1007

[68] Please see Appendix A for a complete listing and explanation of historical data sources.

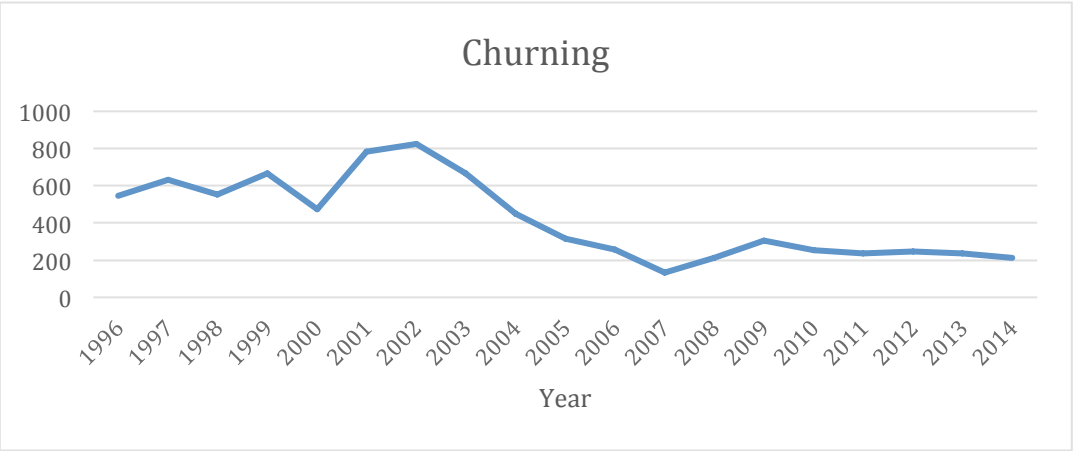
Year	2009	2010	2011	2012	2013	2014
New Disciplinary Actions Filed	1158	1,310	1,488	1,541	1,535	1,397
Formal Actions Resolved	1090	1,178	1,287	1,370	1,307	1,110

Figure 4.1 FINRA Disciplinary Actions Filed and Resolved



This data however accounts for all the various disciplinary actions that FINRA can take against individuals and firms under their jurisdiction, including those that OATS is not intended to cover. It should be noted too that the limit for reporting an action is $1,000, and some fines for improperly reporting to OATS fall below this threshold. Therefore, there may be under reporting of the true regulatory cost of OATS throughout this data.

Figure 4.2 Incidents of Churning



As part of FINRA's duties as an arbiter for cases of financial malfeasance, data is collected on specific areas that arise, including churning. Over the past decade it has become clear that current churning claims have decrease to a quarter of their height in 2002. Churning can be easily detected through the data captured, for instance, by OATS because each trade can be linked and the frequencies can be exposed. Such a sharp decrease is even more startling with respect to FINRA's expansion into other markets as the primary regulator. Indeed levels have not climbed precipitously since its inception in 2007.

Table 4.2 Incidents of Churning Per Year

Year	1996	1997	1998	1999	2000	2001	2002
Incidents of Churning	545	633	552	665	473	784	824

Year	2003	2004	2005	2006	2007	2008	2009
Incidents of Churning	449	315	257	133	212	306	254

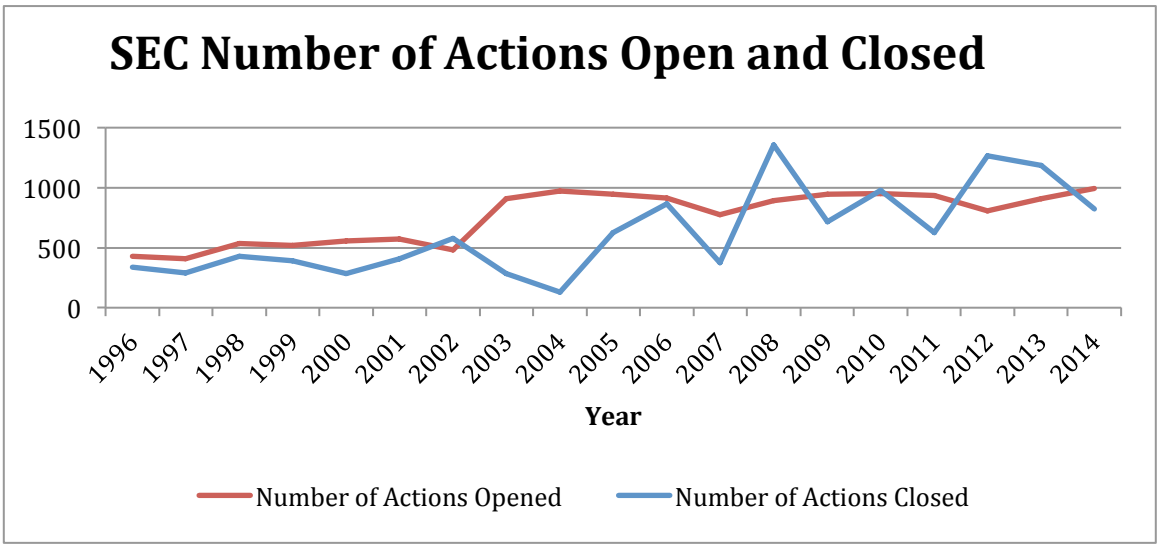
Year	2011	2012	2013	2014
Incidents of Churning	236	245	238	212

The SEC has a longer trail of data for their market abuse detection. Going back to 1996 with the inception of OATS and over the decade until its completion, the SEC was averaging 658 cases per year. Of these cases, an average of 420 would be resolved per year. However, since 2006, the SEC has averaged 900 actions opened versus an average of 915 closed. This shows a marked increase in the number of actions opened and closed by the SEC. However, using the same time periods as the actions filed for FINRA, we find that from 2003 to 2006 there were an average of 936 cases opened but only an average of 476 cases closed. 2008 witnessed an abnormally large number of closed cases relative to their opened cases with over double the cases closed from 2007.

Table 4.3 Number of Actions Opened and Closed per Year

Year	1996	1997	1998	1999	2000
Number of Actions Opened	426	408	536	520	558
Number of Actions Closed	340	289	430	393	284
Year	2001	2002	2003	2004	2005
Number of Actions Opened	570	479	910	973	947
Number of Actions Closed	409	578	283	129	625
Year	2006	2007	2008	2009	2010
Number of Actions Opened	914	776	890	944	952
Number of Actions Closed	868	374	1355	716	975
Year	2011	2012	2013	2014	
Number of Actions Opened	933	806	908	995	
Number of Actions Closed	628	1263	1187	822	

Figure 4.3 Number of Actions Opened and Closed



As with the FINRA data, the total number of opened actions encompasses a wide variety of investigations including ones perhaps not specifically suited to data analysis. To drill down deeper into the data, we can look at two specific actions that can be affected by greater data access, insider trading and market manipulation. During the period from 1996 to 2006 when OATS reached its maturity, we find that the SEC averaged 33 market manipulation cases and 47 insider trading cases. Since 2007, that number has risen to an average of 44 market manipulation cases and 51 insider trading cases per year. While this cannot necessarily be considered a vast increase in detection of market manipulation or insider trading, especially market manipulation actions seems to be increasing. Nevertheless, the SEC continues to tout the abilities of big data to root out insider trading[69] and recently conducted its first action against a HFT firm for illegal market practices.[70]

---

[69] Kevin Cirilli, "SEC chief hails 'big data' in insider trading bust," Text, *TheHill*, (October 14, 2014), http://thehill.com/policy/finance/220698-sec-director-big-data-clamps-down-on-insider-trading.

[70] Matthew Rossi et al., "US SEC Brings First Enforcement Action For Market Manipulation Through High-Frequency Trading," *Mayer Brown*, October 23, 2014, http://www.mayerbrown.com/US-SEC-

Cost

One of the criticisms of OATS has been the cost of adoption versus the benefits reaped. This has largely escalated into divergent views. One in which compliance is the main action target, not the market failures or illegal activities the regulators purport to be trying to uncover. The second is the one largely agreed upon that in order to make accurate decisions, the data must be timely, precise, and complete, which necessitates spending the necessary amount of money.

Table 4.4 FINRA Regulatory Revenue, Total Revenue, and Expenses Per Year

Year	1996	1997	1998	1999	2000	2001	2002
Regulatory Revenue (millions)	$47.90	$58.10	$71.46	$132.86	$140.10	$150.90	$233.58
Total Revenue (millions)	$792.00	$935.00	$1,073.00	$1,672.00	$2,037.00	$2,058.00	$1,612.00
Expenses (millions)	$667.00	$832.00	$975.00	$1,358.00	$1,688.00	$1,935.00	$1,453.00

Year	2003	2004	2005	2006	2007	2008	2009
Regulatory Revenue	$225.70	$279.20	$224.90	$260.70	$393.90	$498.50	$427.00
Total Revenue	$1,322.00	$1,322.00	$1,356.00	$688.00	$826.00	$885.00	$834.00
Expenses	$1,424.00	$1,249.00	$1,230.00	$962.00	$927.00	$1,061.00	$967.00

Year	2010	2011	2012	2013
Regulatory Revenue	$464.67	$428.70	$419.60	$421.30

Brings-First-Enforcement-Action-For-Market-Manipulation-Through-High-Frequency-Trading-10-23-2014/.

Total Revenue	$921.00	$926.00	$906.00	$915.00
Expenses	$1,022.00	$1,047.00	$1,024.00	$1,015.00

We can see from table 4.3 above that since 1996 FINRA has steadily accumulated more revenue from regulatory actions, fines, and dues from members in order to carry out these regulations. In fact, over this time regulatory revenue has gone from a relatively small percentage of the overall revenue stream, 6% in 1996, to just under half (46%) in 2013. Further the percentage of the expenditures devoted to technology (computer operations) has decreased from 8% of budget in 1996 to a mere 3% in 2013. From this data is does not appear terribly clear if the greater access to data has led to any corresponding cost reduction for regulation of the industry.

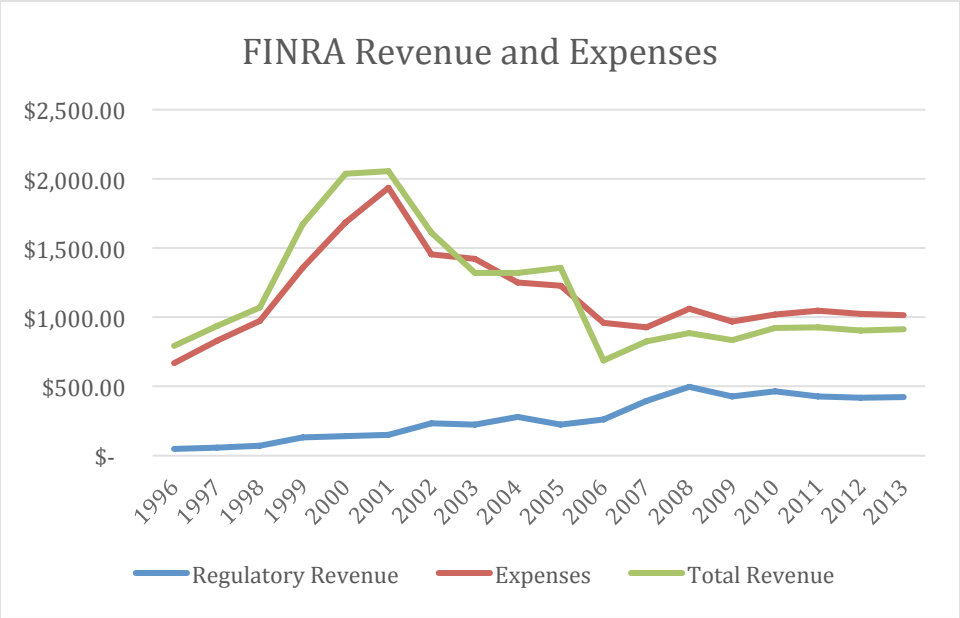


Figure 4.4 FINRA Revenue and Expenses

Information Technology is known for its costliness from both the capital and labor aspects. Investing in technology requires either purchasing the required hardware and software to complete regulatory requirements in-house or paying a third party. Along with that cost comes the necessity to hire people who must understand the technical and regulatory aspects of the program. Coupled with the fact that industry members, through their payments to the SRO, are in effect paying twice for these new regulations. It must be noted, however, that these programs are largely justified by real events that preceded them as described in their creation narratives and neither OATS nor CAT arose purely for the sake of collecting more data.

In response to criticisms against the OATS program being used as a "cash cow" by punishing firms for not complying with an intentionally difficult, tedious, and technical system, FINRA retorts that the regulatory action for reporting compliance account for only 7.5% of the total disciplinary actions and a mere 3% of the fines levied.[71] Further the SEC is working to expand its own data processing to ensure good quality data without simply punishing firms. However, the SEC does not seem to be slowing their investigation of reporting failures.[72]

Time to Completion

The time of analysis is one of the main selling points for data collection and analysis.[73] However, data on timing has only recently become available to supplement the overall amount of time it takes to investigate and enforce an action. It is also difficult for

---

[71] Dan Jamieson, "Brokers Charge FINRA with Becoming OATS Mill."

[72] Patrick Hunnius, "The SEC – Now Killing Many Birds with One 'Big Data' Stone: 5 Takeaways | Insights | DLA Piper Global Law Firm," *DLA Piper*, accessed March 19, 2015, https://www.dlapiper.com/en/us/insights/publications/2014/09/sec-many-birds/.

[73] Susan F. Axelrod, *Remarks from the PLI Seminar on Broker-Dealer Regulation and Enforcement 2012*, October 24, 2014, http://www.finra.org/newsroom/speeches/102412-remarks-pli-seminar-broker-dealer-regulation-and-enforcement-2012.

the time to be measured as there are variety of factors that may cause investigations or arbitration cases to take longer or shorter amounts of time. However, we would expect some decrease in the length of the investigation if the increase in data utilization and collection was similarly increasing.

To measure this, the SEC has collected and reported data on different metrics and "performance goals" over the years. For time to completion, performance goal 1.3.2 evaluates the need "to balance the need for complete, effective and fair investigations with the need to file enforcement actions in as timely a manner as possible."[74] This balance is measured in a percentage of enforcement actions filed with two years of the opening of the investigation. Their goals in general range from 65%-70%. Since 2011, the SEC has also begun collecting the actual average number of months between opening the investigation and commencing the enforcement action.

Table 4.5 Percentage of First Enforcement Actions Filed Within Two Years Per Year

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Percentage	62%	69%	65%	64%	64%	62%	70%	67%	61%	63%	58%

[74] The Securities and Exchange Commission, *FY 2013 Annual Performance Report*, 2015, http://www.sec.gov/about/reports/sec-fy2013-annual-performance-report.pdf.

TABLE 1.10

PERFORMANCE GOAL 2.3.3
Average months between opening a matter under inquiry or an investigation and commencing an enforcement action

Description: This metric captures the average number of months between the opening of an investigation and the filing of the first enforcement action arising out of that investigation. If the investigation was preceded by a matter under inquiry, the metric draws on the date of opening of the matter inquiry. In conducting investigations, the enforcement program continually strives to balance the need for complete, effective, and fair investigation with the need to file enforcement actions in as timely a manner as possible. While not all investigations result in the filing of enforcement actions, this metric provides information concerning the pace of investigations that do lead to such actions and supplements the previous goal, which measures the percentage of first enforcement actions filed within two years.

Fiscal Year	FY 2009	FY 2010	FY 2011	FY 2012	FY 2013	FY 2014 Plan	FY 2014 Actual	FY 2014 Results
Months	N/A	N/A	22	21	21	20	21	Not met

Responsible Division/Office: Division of Enforcement

Data Source: HUB case management and tracking system for the Division of Enforcement

Plan for Improving Program Performance: To achieve its goal of 20 months on this metric, Enforcement will continue to look for ways to manage investigations effectively in order to promote speed and efficiency while maintaining an appropriate degree of thoroughness and completeness. Enforcement will strive to improve through effective management of cases, as well as by leveraging various processes and initiatives designed to promote efficiencies in investigations, such as technology, training, and regular case assessments. Enforcement leadership also will encourage appropriate use of tools such as subpoena enforcement actions in order to ensure that investigations proceed on an appropriate timeframe.

Figure 4.5 Average Months Between Opening and Commencing Enforcement

From these two tables we can see that over the past decade, the percentage of enforcement actions starting within two years has not substantially increased. In fact, over the last few years, the percentage has seen a notable decrease. Some factors, such as a decrease of staff or funding do not seem to account for this. It may be that the scope of the investigations themselves. The SEC notes that in trying to increase this number, they are working on "further leveraging technology to expedite investigative activities." The average monthly difference, with only three years of data, gives a good bench mark going forward but is not enough information to prove any technological impact.

THE COMPREHENSIVE AUTOMATED RISK DATA SYSTEM

The future of big data is bright. But with this explosion of data, comes questions that begged to be answered following the thought "what if we got more?" Specifically related to the topic of this paper is the Comprehensive Automated Risk Data System (CARDS). Risk has long been an aptly economic term and its measurement remains a

central subject for economists, politicians, and businessmen alike. CARDS offers a look what the future of big data in the sphere of financial regulation might look like.

CARDS in essence will collect not only information similar to OATS and CAT but also data at the account level. This is in order to determine suitability, which is the expected risk that a certain account would take given certain factors such as the amount of money in the account, the type of account, and others. If a particular broker dealer is taking too many risks with what is supposed to be a less risky account, it can be said that what that broker dealer is doing is not suitable. In 2014, FINRA arbitrated 1,326 cases of unsuitability, so having the data required to uncover these schemes is likely to benefit investors greatly.[75]

The issues with CARDS come in the form of an overstep of surveillance into what is a more personal metric along with a superfluous duplication of CAT. Collecting this information would involve reporting much information about particular investors, much of it personal information, and then running analyses that haven't yet be proven. The determination of suitability is often one between the individual and the broker dealer who has the conversation about goals, risks, and opportunities. The big data challenges then are immense, as the question being analyzed is not one that is easily solvable and, as noted, often requires arbitration.

However, the data collected by CARDS does not need to identify the specific person attached to an account but instead needs only the details to run the analysis. This data is currently required during investigations and does not constitute a change in policy for firms. The data can be analyzed based on certain metrics and using baselines created

[75] Financial Industry Regulatory Authority, "Dispute Resolution Statistics," February 2015, http://www.finra.org/arbitration-and-mediation/dispute-resolution-statistics.

by examining similar accounts. Implementing CARDS thus presents a great opportunity to increase oversight and, done correctly, do it in a safe and effective manner.

CHAPTER 5: RECOMMENDATIONS AND THE FUTURE OF BIG DATA

INTRODUCTION

It seems more and more clear that the convergence of big data systems in the sphere of financial regulation will be messier before it gets clearer. While the CAT program has moved into its bidder stage, FINRA has already commented that if CAT does not collect the same data as OATS, they will "have to continue with OATS collection." This is, rightfully, seen as a major barrier to creating less costly and efficient big data programs. The cost of these programs continues to be quite large, although it pales in comparison the expenditures made by financial institutions seeking to capitalize on new technologies and methods of operating on the market.

It may also be that by the time CAT is implemented, FINRA will have already completed a system under OATS that in essence covers the areas that CAT is expected to cover. As of this writing, FINRA will have surveillance of over 99% of all market activities through consolidation of various market reporting mechanisms. That does not entail, however, that the analysis of the data it is receiving is greatly impacting either the cost or length of time for investigations and regulatory actions. For that, advances and technology, information sciences, and, simply, time will have to tell.

RECOMMENDATIONS

With the hopes of adding to this conversation regarding OATS, CAT, and CARDS, I would add some recommendations going forward, which may help with additional systems in the future. Specifically, I have three recommendations, two for the regulators and one for the industry itself in order to more effectively utilize the data analysis and technology advances that have arisen in the last few years. The first is a tiered auditing system that accounts for the fact that trading systems have diverged

irreconcilably. Secondly, creating a specific division to handle this data and its analysis will enable a more robust and agile technology group that can assist more thoroughly the mission of the financial regulators. For the industry, it is coming time for data, specifically its standardization and centralization, to play a prominent role in nearly every facet of the business. Completing the task of organizing an industry standard now will likely avoid greater headaches and growing pains in the future.

Tiered Auditing

The idea of tiered auditing recognizes the fact that advances in technology have in effect created two divergent systems with competing goals and timeframes. We see a system where the average holding period for a stock, once measured in years, is down to 5 days today, and continues decreasing.[76] High-frequency trading and other methods for quickly turning around a stock are operating in the same field as long term investments and regular broker dealers. It seems somewhat inane that the reporting requirements should likewise be the same.

This difference is apparent too in the goals of the audit trail systems. In recreating the markets, those operating at millisecond frequencies can provide greater granularity for smaller time frames, while those that are not engaged in that are less engaged in the increase in volume will fill in the picture. Requiring these HFT firms to report at millisecond levels that they are already operate at will likely not increase their costs or require radical updates to their technology. Conversely, requiring firms to report at levels that they are not consistently operating at will require more upgrades. This will also increase the speed of correcting by focusing on systems that are relatively faster and slower more effectively and specifically.

---

[76] Sam Ro, "Stock Market Investors Have Become Absurdly Impatient," *Business Insider*, accessed March 21, 2015, http://www.businessinsider.com/stock-investor-holding-period-2012-8.

In the realm of detecting abuse, algorithmic trading that operates at inhuman levels can be reviewed for specific patterns that indicate an abusive process that it is programmed to do. At non-HFT levels, the abuses are likely in line with current, conscious decisions to break established rules such as insider trading. This would further allow the regulators to focus on HFT-specific issues in one area and the more traditional trading issues in another.

Separate Division

In order to effectively handle and analyze this data, CAT and future audit trails that pull large amounts of data should be a specific entity within the regulatory framework. As long as human beings run the companies and sit at the computer terminals, examinations and inspections will continue to play a major role in regulating and ensuring the fairness of the markets. However, data analysis is a growing part of the regulatory equation and will likely continue as they search for needles in an ever increasing data haystack. Regulators, specifically SRO's, must ensure that those who seek to defraud or abuse the market do not feel safe in their obscurity.

A separate division will also enable the regulators and industry to have a closer connection in the data collection, use, and storage. This will only be a part of the conversation that SRO's have as the mediators of the markets and the public governance. This connection will also enable the regulators to make faster changes in response to changes in the market and various regulations regarding personal information and security. This division can also more actively seek adoption of cost cutting technologies.

This division would also play a closer role with regulators seeking information about their own cases in an effective manner. This division can also report on how many actions were initiated specifically as a result of data analysis and the costs that this

analysis is incurring. This increases transparency in the regulatory sphere and could provide a clearer path to increasing efficiency in reporting and processing.

Data Standardization and Centralization

The transition to a more effective regulatory method that capitalizes on advances in data analytics will require the industry to take a hard look at their own data standards practices. While an industry wide standardization policy might be infeasible due to varying policies of collection and reporting, it would be worth looking into how certain data standards could be used to reduce the costs of reporting and regulation. Many firms already employ data officers or have positions regarding data management, but specific rules regarding how reportable data in handled might decrease the pain across the industry of centralizing and standardizing their data. The philosophical underpinning of seeking standards in data for reporting now is that it will be increasingly difficult to do later.

CONCLUSION

The dreams of big data scientists to draw meaningful analysis from large volumes of data are quickly becoming reality as computational power increases while storage costs decrease. However, there are still many hurdles to be overcome before the age of big data can be fully realized. The variety of systems, the sheer quantity of data, and the specific challenges of financial regulation make the programs of OATS and CAT particularly interesting in this field. While the cost-benefit of these programs seems to suggest that, while good intentioned, they have not entirely lived up to their promise, the era of data analysis is at hand and is not likely to slow down any time soon. It is certainly the case that the financial industry has no intention of slowing down its investment in processing speed and power. If we truly wish to see regulators that are able to piece the

markets back together after a crisis, to identify market malpractice, or to avoid both of these events entirely, then we must be prepared to equip them with the necessary strategy and technology to do so.

Appendix A: Historical Data

SEC

Historical data for the SEC can be located on their reporting page, including financial and annual information.[77] Each year of data comes from the respective annual report, however the exact page or location varies, as the reports do not follow identical layouts and orders. Case data is often found as an appendix to the annual report, while the financial data is often within the main body.

FINRA/NASD

As NASD no longer exists, historical data had to be obtained using the way back machine curated by the Internet Archive.[78] FINRA maintains a statistics page[79] that includes data for various years, however this has a limited functionality for historical data. The data presented in this papers correspond to the year when the data is available on the particular page, as the way back machine does not always capture every subpage on the website. The way back machine does include images of the NASD statistics page, which closely mirrors that of the current day FINRA page.[80] For older FINRA pages, the way back machine also has older versions archived at various dates.[81]

---

[77] The Securities and Exchange Commission, "Reports," n.d., https://www.sec.gov/about/secreports.shtml.
[78] Internet Archive, "Way Back Machine," n.d., http://archive.org/web/.
[79] The Financial Industry Regulatory Authority, "FINRA Statistics," n.d., https://www.finra.org/newsroom/statistics.
[80] The Internet Archive, "Dispute Resolution Statistics," accessed August 3, 3004, http://web.archive.org/web/20040803010657/http://www.nasdadr.com/statistics.asp.
[81] The Internet Archive, "FINRA Way Back Machine," n.d., http://web.archive.org/web/20150423193422*/http://www.finra.org/.

Table A.1 January 2013 OATS Reporting Statistics – Number of Firms by Record Volume[82]

Total number of firms reporting in January 2013	1,051
Number of firms reporting 100 million or more records	33
Number of firms reporting between 3 million and 99,999,999 records	52
Number of firms reporting between 100,000 and 2,999,999 records	89
Number of firms reporting between 10,000 and 99,999 records	169
Number of firms reporting 9,999 or less records	708

[82] Financial Industry Regulatory Authority, "Consolidated Audit Trail."

Table A.2 Important Dates (OATS) [83]

FINRA vs SEC Settlement (calling for creation of OATS)	August 1996
OATS rules approved by SEC	March 1998
Phase I: Electronic Communications Network and electronic orders of market makers reported to OATS	March 1999
Phase II: All electronic orders are reported to OATS	August 1999
Phase III: All manual orders must be reported to OATS	July 2000 (*actual* July 2006)
All computer clocks must be synchronized (manual clocks synchronized)	August 1998 (July 1999)
Regulation National Market System adopted	June 2005 (*actual* October 2007)
OATS is expanded to all OTC equity securities	June 2007
SEC Approves rule allowing FINRA to collect all NMS stocks	November 2010
FINRA, through OATS, begins collecting all NMS stocks	October 2011
FINRA expands market surveillance to include BATS, increasing cross market surveillance to 99% of all listed equities	February 2014
If possible, OATS trade data must be reported in milliseconds	September 2014

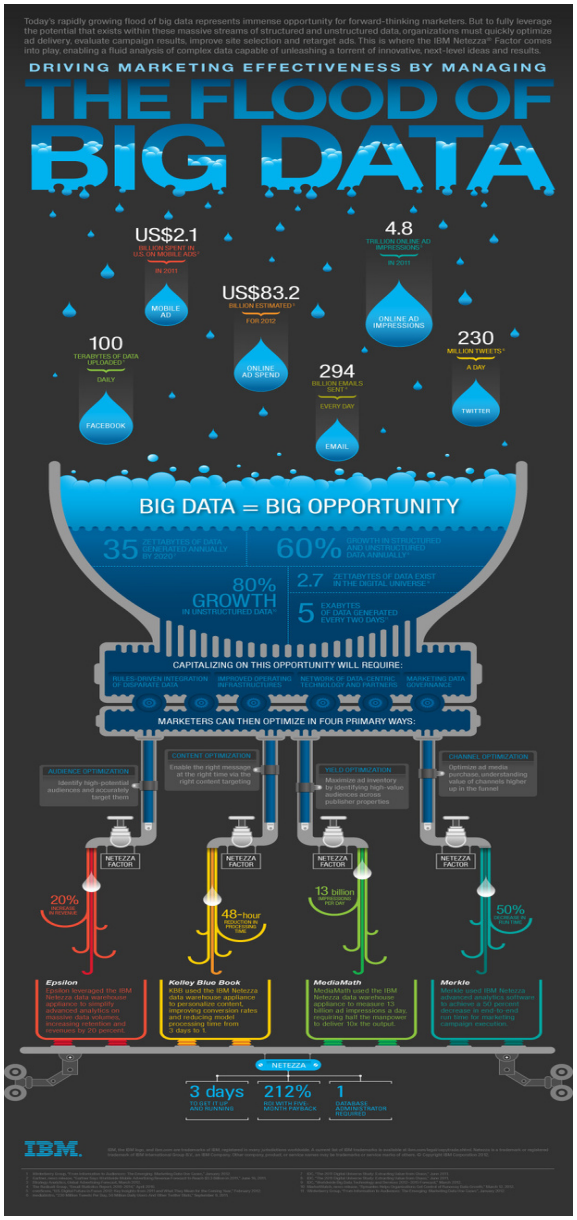
[83] Financial Industry Regulatory Authority, "OATS," n.d., https://www.finra.org/industry/oats.

Table A.3 Important Dates (CAT) [84]

SEC Adopts Rule 613 Calling for Creation of CAT	July 11, 2012
Flash Crash	May 6, 2010
CAT Request for Proposal	February 26, 2013
Bidder Shortlist produced	July 1, 2014
Selection Process Plan Approved	February 14, 2014
Implementation	
CAT Processor Selected by NMS Plan Participants	Within two months after effectiveness of the approved NMS Plan
Business Clock Synchronization for SROs and Broker-dealers	Within four months after effectiveness of the approved NMS Plan
SROs begin submitting data to the central repository	Within one year after effectiveness of the approved NMS Plan
SROs must implement enhanced surveillance using CAT data	Within 14 months after effectiveness of the approved NMS Plan
SRO members, except small members, must begin submitting data to the central repository	Within two years after effectiveness of the approved NMS Plan
Small SRO members must begin submitting data to the central repository	Within three years after effectiveness of the approved NMS Plan

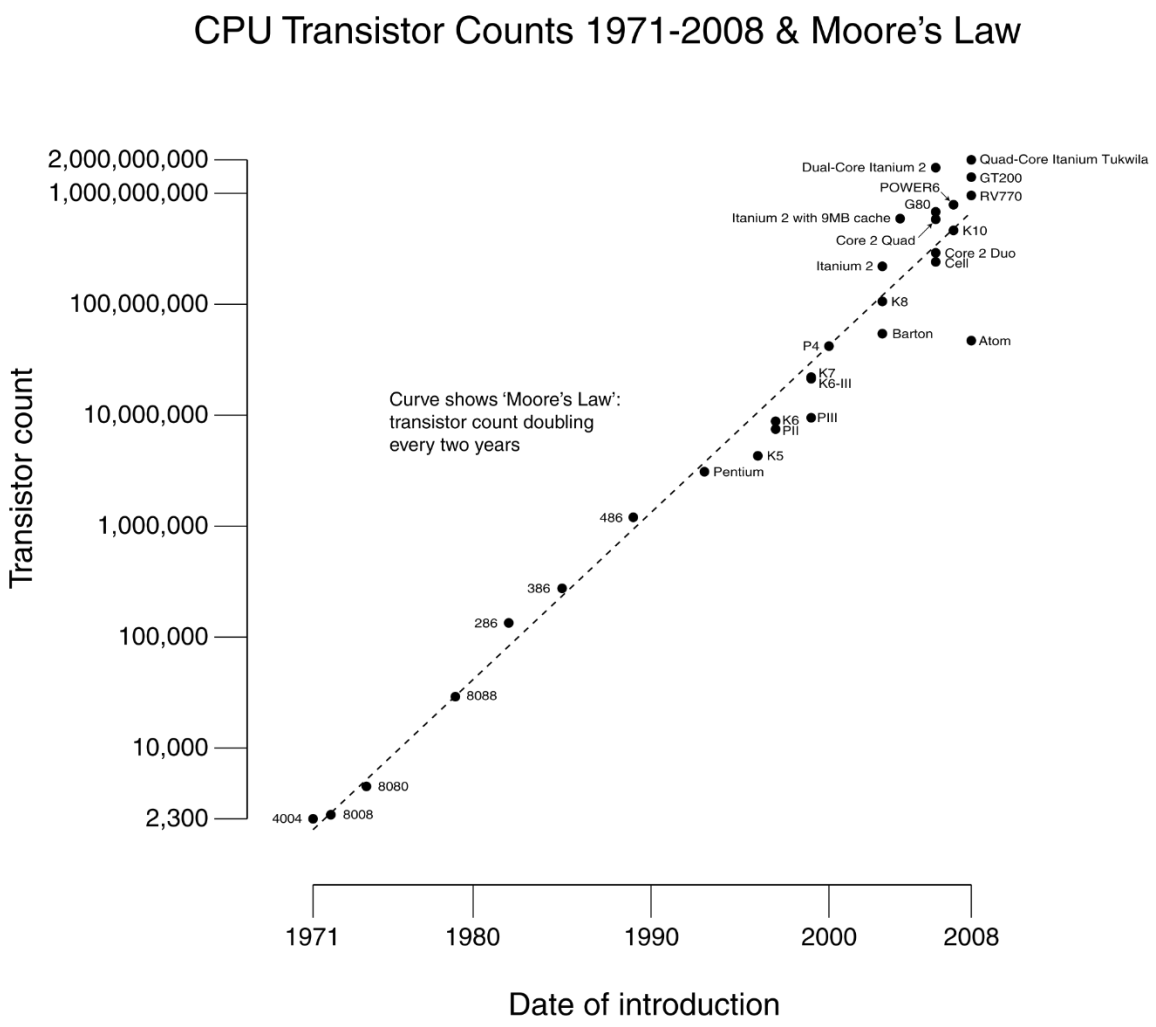
[84] Participants of NMS Governing Plan, "National Market System Plan Governing the Consolidated Audit Trail Pursuant to Rule 613 of Regulation NMS under the Securities and Exchange Act of 1934," 6.

Illustration A.1 IBM Big Data Infographic[85]



[85] Financial Industry Regulatory Authority, "OATS," n.d., https://www.finra.org/industry/oats. data-facts-and-statistics-that-will-shock-you/","author":[{"family":"Chad Luckie","given":""}],"accessed":{"date-parts":[["2015",3,18]]}}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} Chad Luckie, "'Big Data' Facts and Statistics That Will Shock You."

Illustration A.2 Number of Transistors[86]



[86] Wikipedia, "CPU Transistor Counts 197102008 & Moore's Law," accessed February 2, 2015, http://upload.wikimedia.org/wikipedia/commons/thumb/0/00/Transistor_Count_and_Moore's_Law_-_2008.svg/2000px-Transistor_Count_and_Moore's_Law_-_2008.svg.png.

Illustration A.3 Processing Power[87]

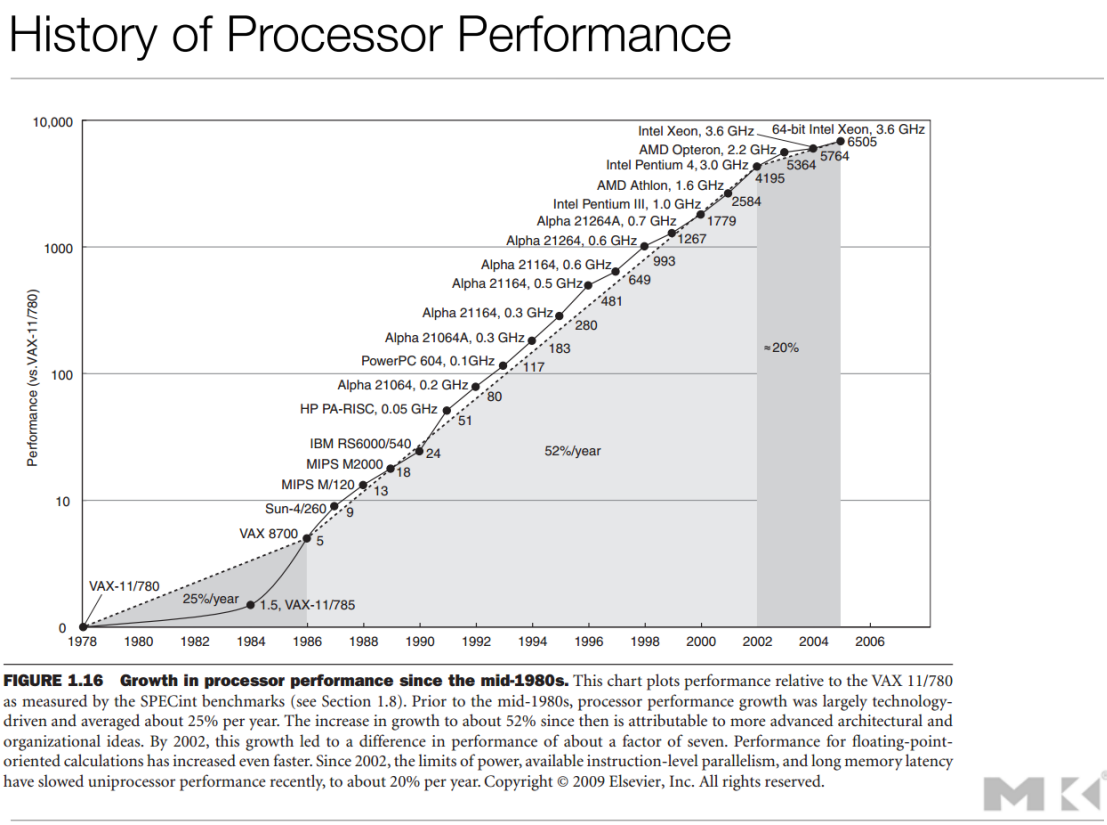
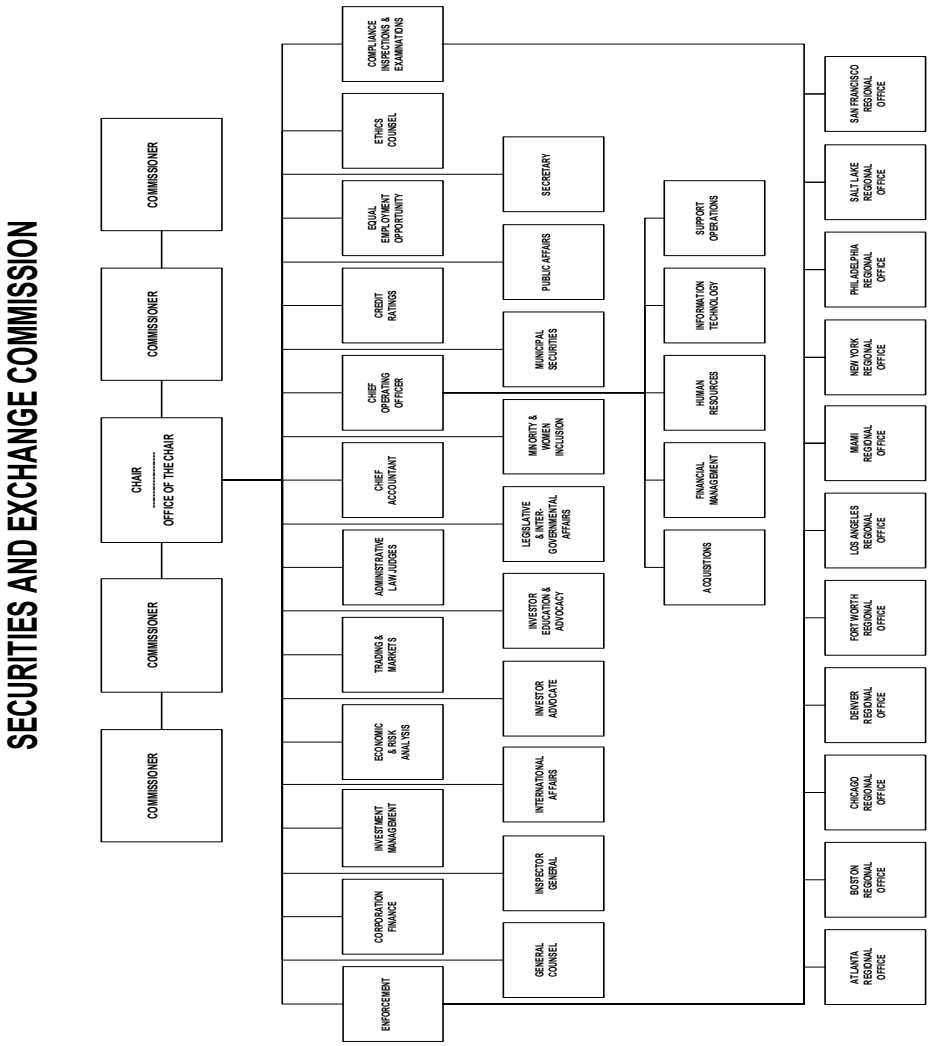


FIGURE 1.16 Growth in processor performance since the mid-1980s. This chart plots performance relative to the VAX 11/780 as measured by the SPECint benchmarks (see Section 1.8). Prior to the mid-1980s, processor performance growth was largely technology-driven and averaged about 25% per year. The increase in growth to about 52% since then is attributable to more advanced architectural and organizational ideas. By 2002, this growth led to a difference in performance of about a factor of seven. Performance for floating-point-oriented calculations has increased even faster. Since 2002, the limits of power, available instruction-level parallelism, and long memory latency have slowed uniprocessor performance recently, to about 20% per year.

[87] Elsevier, Inc, *History of Processor Performance*, April 24, 2012, http://www.cs.columbia.edu/~sedwards/classes/2012/3827-spring/advanced-arch-2011.pdf.

Illustration A.4 SEC Organizational Chart[88]



[88] The Securities and Exchange Commission, *FY 2013 Annual Performance Report*.

Table A.4 Other Reporting Systems and Rules[89]

Overview of Other Reporting Systems and Rules

Reporting System / Rule	Nature of Rule	References
Order Audit Trail System (OATS) Owner: FINRA Rule: FINRA Rules 7410 - 7470	An integrated audit trail of order, and trade information for NMS securities. FINRA uses this audit trail system to recreate events in the lifecycle of orders and to more completely monitor the trading practices of member firms. FINRA member firms are required to develop a means for electronically capturing and reporting to OATS specific data elements related to the handling or execution of orders, including recording all times of these events in hours, minutes, and seconds, and to synchronize their business clocks.	FINRA Rule 7400 Order Audit Trail System OATS Reporting Technical Specifications OATS Reportable Securities
Electronic Blue Sheets (EBS) Owner: SEC Rule: SEC Rule 17a-25	Requires brokers and dealers to submit electronically to the SEC, upon request, information on customer and firm securities trading, including order execution time. Designed to improve the Commission's capacity to analyze electronic submissions of transaction information, thereby facilitating Commission enforcement investigations and other trading reconstructions.	Final Rule: Electronic Submission of Securities Transaction Information by Exchange Members, Brokers, and Dealers EBS Submission Specification
Equity Cleared Reports Owner: NSCC	This report is generated on a daily basis by the SROs and is provided to the NSCC in a database accessible by the SEC, and shows the number of trades and daily volume of all equity securities in which transactions took place, sorted by clearing member.	
Large Trader Reporting Owner: SEC Rule: SEC Rule 13h-1	Assists the SEC in both identifying and obtaining trading information on market Participants that conduct a substantial amount of trading activity, as measured by volume or market value, in the NMS Securities. Rule 613 requires broker-dealers to maintain and report	Final Rule: Large Trader Reporting

[89] The Securities Industry and Financial Markets Association, *Industry Recommendations for the Creation of a Consolidated Audit Trail*, 66.

Reporting System / Rule	Nature of Rule	References
	data that is largely identical to the information covered by the Commission's Electronic Blue Sheets (EBS) system [29]– the system the SEC currently uses to collect transaction data from broker-dealers. LTR provides a source of data to support investigative and enforcement activities, and helps to reconstruct trading activity following periods of unusual market volatility, and to analyze significant market events for regulatory purposes.	
Consolidated Options Audit Trail System (COATS) Owner: FINRA, Options Exchanges Rule:	A consolidated audit trail that enables the Options Exchanges to reconstruct markets promptly, effectively surveil them and enforce order handling, firm quote, trading reporting and other rules. Requires that each order, change to an order, or cancellation of an order transmitted to the exchange be "systematized," in a format approved by the exchange, either before it is sent to the exchange or contemporaneously upon receipt on the floor of the exchange, and prior to representation of the order.	
NYSE Rule 410B	Transactions effected in NYSE listed securities by members and member organizations, which are not reported to the Consolidated Tape must be electronically reported to NYSE including date of transaction, customer name, address(es), branch office number, registered representative number, whether order was solicited or unsolicited, date account opened and employer name and the tax identification number(s).	Rule 410A. Automated Submission of Trading Data
PHLX 1022 Owner: Philadelphia Stock Exchange	Specialists or Registered Options Traders must report orders for the purchase or sale of securities underlying any stock or Exchange Traded Fund Share options contract traded on the exchange, including securities convertible into or exchangeable for such underlying securities regardless of whether the Specialists or Registered Options Trader makes a market for the related option.	
CBOE 8.9 Owner: Chicago Board Options Exchange	Clearing firms, with respect to transactions to be cleared in accounts of market makers, must report executed orders for the purchase or sale of positions in securities underlying options	

Reporting System / Rule	Nature of Rule	References
	traded on the Exchange, including securities convertible or exchangeable into such securities, regardless of whether the market maker makes a market for the related option. In addition, clearing firms must also report market maker executions and positions with respect to securities traded on the Exchange.	
Large Options Position Reporting (LOPR)[30] Owner: FINRA Rule: FINRA Rule 2360(b)(5)	Requires member firms to file reports for each account in which a member has an interest, each account of a partner, officer, director, or employee of the member; and of each customer, non-member broker, or non-member dealer that has an aggregate position of 200 or more options contracts (whether long or short) on the same side of the market covering the same underlying security or index.	FINRA Rule 2360 Reference Guide for LOPR Firms
Rule 4560 – Short Interest Reporting Owner: FINRA	Member firms are required to report total short positions in all customer and proprietary firm accounts in all equity securities to FINRA on a bi-monthly basis.	Short Interest Reporting

Table A.5 OATS Reporting Types[90]

Order Audit Trail System (OATS) – Overview

Currently, eight report types are submitted by broker-dealers to OATS.

	OATS Report Type	Description
1	New Order Report	Records the receipt of an order from a customer, another member or non-member, or orders originated within a member firm
2	Route Report	Records the routing of an order to another member, ECN, non-member, or national securities exchange
3	Desk Report	Records the receipt of a full or partial transmittal to a desk or department within the member firm
4	Execution Report	Records the full or partial execution of an order
5	Cancel Report	Records the full or partial cancellation of an order
6	Cancel/Replace Report	Records the modification of an order
7	Combined Order/Execution Report	Records the receipt of an order from a customer, another member or non-member, or an order originated within a member firm and full execution on the same day
8	Combined Order/Route Report	Records the receipt of an order from a customer, another member or non-member, or an order originated within a member firm and full route of the order on the same day to another member, ECN, non-member or exchange

[90] Financial Industry Regulatory Authority, "Consolidated Audit Trail."

BIBLIOGRAPHY

ACA Compliance Group. "FINRA 2014 Annual Regulatory and Examination Priorities," January 9, 2014. http://www.acacompliancegroup.com/news/compliance-alert/finra-2014-annual-regulatory-and-examination-priorities.

Chad Luckie. "'Big Data' Facts and Statistics That Will Shock You." *Fathom*. Accessed March 18, 2015. http://www.fathomdelivers.com/blog/analytics-and-big-data/big-data-facts-and-statistics-that-will-shock-you/.

Christie, William G., and Paul H. Schultz. "Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" *The Journal of Finance*, December 1994, Vol. 49, No. 5 edition.

Cirilli, Kevin. "SEC chief hails 'big data' in insider trading bust." Text. *TheHill*, October 14, 2014. http://thehill.com/policy/finance/220698-sec-director-big-data-clamps-down-on-insider-trading.

Dan Jamieson. "Brokers Charge FINRA with Becoming OATS Mill." *Investment News*, April 24, 2011.

Data Center Knowledge. "Closer Look: NYSE Euronext's NJ Data Center." *Data Center Knowledge*. Accessed March 19, 2015. http://www.datacenterknowledge.com/closer-look-nyse-euronexts-nj-data-center/.

Elsevier, Inc. *History of Processor Performance*, April 24, 2012. http://www.cs.columbia.edu/~sedwards/classes/2012/3827-spring/advanced-arch-2011.pdf.

Financial Industry Regulatory Authority. "2009 Year in Review," 2010. http://www.finra.org/sites/default/files/Corporate/p121646.pdf.

———. "Consolidated Audit Trail." July 29, 2014. http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/thesys_technologies_cat_presentation_sifma_summit_final_posted_072914.pdf?n=06244.

———. "Dispute Resolution Statistics," February 2015. http://www.finra.org/arbitration-and-mediation/dispute-resolution-statistics.

———. "FINRA FINES 3 FIRMS A TOTAL OF $1.6M FOR OATS VIOLATIONS." *Bloomberg*, May 15, 2008. http://www.bloomberg.com/apps/news?pid=newsarchive&sid=awcbFR97xMCc.

———. "OATS," n.d. https://www.finra.org/industry/oats.

———. "Summary of the Consolidated Audit Trail Initiative." January 2015. http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p571933.pdf.

Gerig, Austin. *High-Frequency Trading Synchronizes Prices in Financial Markets*, n.d. http://www.sec.gov/dera/staff-papers/working-papers/dera-wp-hft-synchronizes.pdf.

Internet Archive. "Way Back Machine," n.d. http://archive.org/web/.

Jeff Van Blarcom. “The Order Audit Trail System / OATS - Series 55.” *Investopedia*. Accessed March 1, 2015. http://www.investopedia.com/study-guide/series-55/commissions-and-trade-complaints/order-audit-trail-system-oats/.

John Allen. “FINRA| Broker Fraud Attorney.” *Broker Fraud | Broker Fraud Attorney | Broker Misconduct*. Accessed March 18, 2015. http://www.mybrokerfraud.com/faq/faq-general-finra/.

John Carney. “What Really Caused the Flash Crash.” *CNBC*. Accessed February 8, 2015. http://www.cnbc.com/id/42919462.

Kashmir Hill. “Blueprints Of NSA’s Ridiculously Expensive Data Center In Utah Suggest It Holds Less Info Than Thought.” *Forbes*. Accessed March 19, 2015. http://www.forbes.com/sites/kashmirhill/2013/07/24/blueprints-of-nsa-data-center-in-utah-suggest-its-storage-capacity-is-less-impressive-than-thought/.

Lash, Herbert. “Wall St Watchdog Moves to Cloud, Big Data, to Boost Capabilities.” *Reuters*. June 20, 2014. http://www.reuters.com/article/2014/06/20/finra-technology-idUSL2N0P105020140620.

Mamudi, Sam, and Whitney Kisling. “Volatility Spurs Volume Boon in U.S. Stocks With Trading Up 11%.” *Bloomberg*, February 14, 2014. http://www.bloomberg.com/news/2014-02-14/volatility-spurs-volume-boon-in-u-s-stocks-with-trading-up-11-.html.

Manisha Kimmel. “FIF Clock Offset Survey Preliminary Report.” February 16, 2015. http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p602479.pdf.

Marlene Y. Satter. “SEC, FINRA Enforcement: Money Manager Charged in Free-Riding Scheme.” *ThinkAdvisor*, September 5, 2013. http://www.thinkadvisor.com/2013/09/05/sec-finra-enforcement-money-manager-charged-in-fre.

Matthew Rossi, J. Gregory Deis, Jerome J. Roche, and Kathleen M. Przywara. “US SEC Brings First Enforcement Action For Market Manipulation Through High-Frequency Trading.” *Mayer Brown*, October 23, 2014. http://www.mayerbrown.com/US-SEC-Brings-First-Enforcement-Action-For-Market-Manipulation-Through-High-Frequency-Trading-10-23-2014/.

Mehta, Steven SloanNina. “SEC Votes to Require Consolidated Audit Trail for Markets.” *Bloomberg.com*. Accessed March 1, 2015. http://www.bloomberg.com/news/articles/2012-07-11/sec-votes-to-require-consolidated-audit-trail-for-markets.

Nelson, Stephen J. “Commentary: The Promise of a Consolidated Order Audit Trail.” *Traders Magazine Online News*, September 16, 2010. http://www.tradersmagazine.com/news/consolidated-order-audit-trail-sec-nasdaq-nyse-106337-1.html.

New York Stock Exchange. “NYSE Historical Transaction Data,” n.d. https://www.nyse.com/data/transactions-statistics-data-library/.

Participants of NMS Governing Plan. “National Market System Plan Governing the Consolidated Audit Trail Pursuant to Rule 613 of Regulation NMS under the

Securities and Exchange Act of 1934," February 27, 2015. http://catnmsplan.com/web/groups/catnms/@catnms/documents/appsupportdocs/p602500.pdf.

Patrick Hunnius. "The SEC – Now Killing Many Birds with One 'Big Data' Stone: 5 Takeaways | Insights | DLA Piper Global Law Firm." *DLA Piper*. Accessed March 19, 2015. https://www.dlapiper.com/en/us/insights/publications/2014/09/sec-many-birds/.

Sam Ro. "Stock Market Investors Have Become Absurdly Impatient." *Business Insider*. Accessed March 21, 2015. http://www.businessinsider.com/stock-investor-holding-period-2012-8.

Stafford, Philip. "Flash Crash Explanation Questioned." *Financial Times*, April 3, 2013. http://www.ft.com/intl/cms/s/0/52e2e7e0-9c5e-11e2-9a4b-00144feabdc0.html#axzz3V26HCHrh.

Sungard. "Consolidated Audit Trail Solution Overview: SIFMA Bidders Summit." July 29, 2014. http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/sungard-cat-presentation-sifma-bidders-summit-july292014.pdf?n=66627.

Susan F. Axelrod. *Remarks from the PLI Seminar on Broker-Dealer Regulation and Enforcement 2012*, October 24, 2014. http://www.finra.org/newsroom/speeches/102412-remarks-pli-seminar-broker-dealer-regulation-and-enforcement-2012.

Taibbi, Matt. "Why Didn't the SEC Catch Madoff? It Might Have Been Policy Not To." *Rolling Stone*. Accessed April 29, 2014. http://www.rollingstone.com/politics/blogs/taibblog/why-didnt-the-sec-catch-madoff-it-might-have-been-policy-not-to-20130531.

The Financial Industry Regulatory Authority. "FINRA Manual - Notices - 2011 - 11-03 FINRA Expands the Order Audit Trail System to All NMS Stocks; Effective Date: July 11, 2011," 2011. http://finra.complinet.com/en/display/display_main.html?rbid=2403&element_id=9945.

———. "FINRA Statistics," n.d. https://www.finra.org/newsroom/statistics.

The Internet Archive. "Dispute Resolution Statistics." Accessed August 3, 3004. http://web.archive.org/web/20040803010657/http://www.nasdadr.com/statistics.asp.

———. "FINRA Way Back Machine," n.d. http://web.archive.org/web/20150423193422*/http://www.finra.org/.

The Securities and Exchange Commission. *Appendix To Report Pursuant To Section 21(a) of the Securities Exchange Act of 1934 Regarding the NASD and the NASDAQ Market*, August 1996. https://www.sec.gov/litigation/investreport/nd21a-appx.txt.

———. "Division of Economic and Risk Analysis," n.d. http://www.sec.gov/dera.

———. *FY 2013 Annual Performance Report*, 2015. http://www.sec.gov/about/reports/sec-fy2013-annual-performance-report.pdf.
———. *FY 2016 Congressional Budget Justification*, February 2, 2105. http://www.sec.gov/about/reports/secfy16congbudgjust.pdf.
———. "Release No. 34-56145; File No. SR-NASD-2007-023," July 26, 2007. http://www.sec.gov/rules/sro/nasd/2007/34-56145.pdf.
———. "Release No. 34-67457; File No. S7-11-10," October 1, 2012. http://www.sec.gov/rules/final/2012/34-67457.pdf.
———. "Reports," n.d. https://www.sec.gov/about/secreports.shtml.
———. "SEC Approves New Rule Requiring Consolidated Audit Trail to Monitor and Analyze Trading Activity," July 11, 2012. http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171483188#.VO0AGyvF98E.
———. "SEC Proposes Consolidated Audit Trail System to Better Track Market Trades." Government, May 26, 2010. http://www.sec.gov/news/press/2010/2010-86.htm.
———. "SEC Renames Division Focusing on Economic and Risk Analysis." Government, June 6, 2013. http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171575272#.VP-KUWTF9kI.
———. "Self-Regulatory Organizations; Notice of Filing of Proposed Rule Change by National Association of Securities Dealers, Inc. Relating to the Creation of New Rules 6900 Through 6970 or an Audit Trail System Owned and Operated by the National Association of Securities Dealers, Inc." *Federal Register*, August 28, 1997, Vol. 62, No. 172 edition.
———. "The Laws That Govern the Securities Industry." Government, n.d. http://www.sec.gov/about/laws.shtml.
———. "What Is Churning?" Government. *Churning*, n.d. http://www.sec.gov/answers/churning.htm.
The Securities and Exchange Commission, and NASD. "Release No. 34-39729; File No. SR-NASD-97-56 - NASD Rulemaking: Various Orders Relating to the Creation of an Order Audit Trail System," March 6, 1998. http://www.sec.gov/rules/sro/nd9756o.htm.
The Securities and Exchange Commission, and The Commodity Futures Trading Commission. *Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, September 30, 2010. http://www.cftc.gov/ucm/groups/public/@otherif/documents/ifdocs/staff-findings050610.pdf.
The Securities and Exchange Commission, and The Office of Compliance Inspections and Examinations. "National Exam Program Examination Priorities for 2015," 2015. http://www.sec.gov/about/offices/ocie/national-examination-program-priorities-2015.pdf.

The Securities Industry and Financial Markets Association. *Industry Recommendations for the Creation of a Consolidated Audit Trail*, March 28, 2013.

Thysis Technologies. "Proposed CAT Solution: SIFMA CAT Bidder Summit." July 29, 2014. http://www.sifma.org/uploadedfiles/issues/legal,_compliance_and_administration/consolidated_audit_trail_(cat)/thesys_technologies_cat_presentation_sifma_summit_final_posted_072914.pdf?n=06244.

Top500.org. "Top 500 The List," n.d. http://www.top500.org/.

United States Securities and Exchange Commission. "The SEC - What We Do." *The Investors Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation*. Accessed April 28, 2014. http://www.sec.gov/about/whatwedo.shtml.

Wikipedia. "CPU Transistor Counts 197102008 & Moore's Law." Accessed February 2, 2015. http://upload.wikimedia.org/wikipedia/commons/thumb/0/00/Transistor_Count_and_Moore's_Law_-_2008.svg/2000px-Transistor_Count_and_Moore's_Law_-_2008.svg.png.

World Federation of Exchanges. "Latest Statistics - Monthly Reports," March 2015. http://www.world-exchanges.org/statistics/monthly-reports.