December 6, 2019

Via Electronic Mail (rule-comments@sec.gov)

Ms. Vanessa Countryman
Secretary
U.S. Securities and Exchange Commission
100 F Street NE., Washington, DC 20549

Re: **Market Data and Market Access (File No. 4-729)**

Dear Ms. Countryman:

On behalf of Data Boiler Technologies, I am pleased to provide the U.S. Securities and Exchange Commission (SEC) with our comments regarding improvements to the current market data system and discuss way for Market Authorities to reconstruct and analyze order books amid the unsynchronized clock issue[1]. In the capacity of an entrepreneurial inventor of a suite patent pending solutions for trade processing and analysis, I am advocating for making **market data available securely in synchronized time**[2]. Our solutions also address the related **surveillance** challenges that long been identified by the International Organization of Securities Commissions (IOSCO) since August 2012 per CR12/2012[3].

1. 17 CFR §242.603(a) interpretation is incomplete and requires clarification or appropriate updates

First of all, in adopting Regulation NMS and per 70 FR 37567[4] in 2005, the Commission stated that "adopted Rule 603(a)[5] prohibits an SRO or broker-dealer from **transmitting** data to a vendor or user any sooner than it transmits the data to a Network processor." Then, in Order 67857[6] in 2012, the Commission stated that "exchanges have an obligation under Rule 603(a) to take reasonable steps to ensure—through system architecture, monitoring, or otherwise—that they **release data** relating to current best-priced quotations and trades through proprietary feeds no sooner than they release data to the Network Processor, including during periods of heavy trading."

As highlighted in TABB Group's market data revenue analysis[7], "**faster access**" isn't only about accelerating SIP's processing speed, but there are issues with the current **aggregation distance**.

"Transmitting or releasing data no sooner than to a Network processor (SIP)" only describes one of the aspects of "fair and reasonable" and "not unreasonably discriminatory" principles required by Reg. NMS. It

---

[1] https://www.linkedin.com/pulse/clock-synch-challenge-new-solution-kelvin-to/
[2] https://www.linkedin.com/pulse/market-data-available-securely-synchronized-time-kelvin-to/
[3] https://www.iosco.org/library/pubdocs/pdf/IOSCOPD389.pdf
[4] https://www.govinfo.gov/content/pkg/FR-2005-06-29/pdf/05-11802.pdf
[5] https://www.law.cornell.edu/cfr/text/17/242.603
[6] https://www.sec.gov/litigation/admin/2012/34-67857.pdf
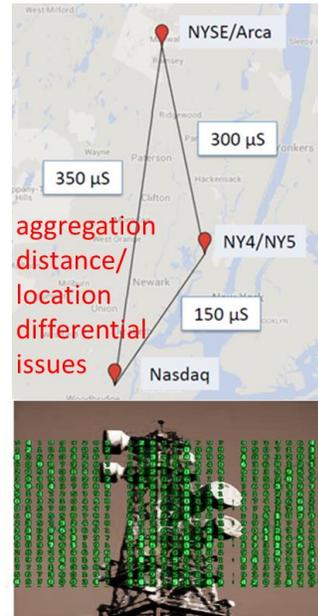[7] https://www.sec.gov/comments/4-729/4729-4559257-176198.pdf

omitted the fact that market data is highly valuable (it reflects the price discovery created by exchanges) and it requires **proper security protection**. Hence, the secured delivery (in-motion and at-rest)[8] and retrieval of data in a timely manner are equally important.

Speed and content differentials between the market data feeds provided by the SIPs and the proprietary products sold by the exchanges resulting in:

- *rent-seeking behaviors*
- *unhealthy competition for the fastest access*
- *impact to trade execution quality, and economic disincentive for non-HFT firms to maintain their trade activities*

Faster access isn't just about accelerating SIP's processing speed. It is also about the availability of proprietary feeds and SIP consolidated data securely in synchronized time.

So, instead of introducing greater competitive forces into the "dissemination" of core data that would result in multiple best-prices confusing the market, I recommend the SEC to consider **mandating the use of time-lock encryption** (see this for the general concept[9], and rest assure this is not another speed bump) for security protection of market data. It would allow proprietary feeds and SIP consolidated data to be "available"[10] securely in synchronized time.

2.   Context of the Problem; It's Broken, Fix it NOW

I applaud the joint statement[11] issued by Commissioner Peirce and Commissioner Roisman regarding the three possible paths that Exchanges can show how they would fulfill the "fair and reasonable" requirements per Reg. NMS. Yet, given the distinction between data transmission and data availability as mentioned earlier, I think the interpretations of 17 CFR §242.603(a)[5] is incomplete and requires clarification or appropriate updates.

Transmit ≠ Available — Incomplete – only one of the aspects of fair & reasonable

---

[8] http://www.databoiler.com/index_htm_files/DataBoilerInMotion.pdf
[9] https://people.csail.mit.edu/rivest/pubs/RSW96.pdf
[10] http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1126&context=wmblr
[11] https://www.sec.gov/news/public-statement/peirce-roisman-statement-101618

Rent-Seeking | Optimally restricts access to price info → unreasonably discriminatory

Asides, there is the problem of rent seeking behavior as pointed out by an empirical research – "Sale of Price Information by Exchanges: Does it Promote Price Discovery"[12] by Giovani Cespa and Thierry Foucault. The question is: what if Exchanges optimally restrict access to price information by charging a high fee so that only a fraction of speculators buy their proprietary products – would that constitutes as unreasonably, unfair and/or discriminatory?

Aggregation distance | Immutable latency because exchanges are immobilized

To level set, I want to acknowledge that Quantum computing[13] may address the aggregation distance/ location differential issues; however quantum computing may inadvertently cause new concerns, such as the lack of an audit trail. Given that, Market Authorities and the overall industry must recognize that, because of the immutable latency issue, there has to be some **gives and takes to optimize between processing speed and contents' richness for SIP**. Therefore, per my comment letter[14] to the SIP operating committee regarding the Odd Lots proposal, I encourage the industry to consider the following suggested priorities:

(i) First, address the speed differentials between the market data feeds provided by the SIPs and the proprietary products sold by the exchanges in order to get the biggest bang for the buck.

(ii) Second, demand for the depth-of-book information (a replication of the relative strengths in bid/ask price and steepness of the price curve in real-time), so at least the content would be a bit more compatible with the proprietary products sold by the exchanges, while minimizing drags of the SIP processing speed.

(iii) Third, pursuit market structure changes outside of the SIP that will make odd lots become true "outliers"[15] rather than the "norms", and/or ask for a "delayed" odd lot trades and quotations statistics, so that experienced market participants may use reverse-engineering methods to "figure-out" or "project" how these odd lots would play out in sequence. This "alternative", "compromise", or "trade-off" is based on the condition of Exchanges willingness to adopt point (i) – making market data available securely in synchronized time[2].

Per former Chair Mary Jo White's statement[16] in 2014, the objective here is: "deemphasize speed as a key to trading success" and let "trading venues have sufficient opportunity and flexibility to innovate successfully".

---

[12] https://pdfs.semanticscholar.org/b61b/597e0c4268eaec75fb744b4e1802c3beb8aa.pdf
[13] https://www.technologyreview.com/s/612421/us-china-quantum-arms-race/
[14] https://www.theice.com/publicdocs/DataBoiler_OddLots_Comments.pdf
[15] Instead of bluntly get rid of odd lots by changing the round lot size, one should ask what are the various reasons for the existence of odd lots. For example, the causes related to a lack of stock split, the segment's specific needs in moving blocks under the impending dynamics between lit and dark venues, etc.
[16] https://www.sec.gov/news/speech/2014-spch060514mjw

I personally attended the SEC- Fordham University event[17] in March 2019, and please allow me to restate these key questions raised by Chairman Clayton and Director Redfearn here:

"How can we help ensure that core data evolves along with the broader market ecosystem?

Speed:

SIPs is measurably slower than proprietary data feeds. The causes for this less timely data may include communications protocols, aggregation times, and geographical latencies associated with single point of consolidation. Regardless of the causes, concerns about latency differentials between the SIPs and the direct feeds are meaningful enough that staff will consider whether to recommend changes to ensure that core data is timely by today's standards.

Depth:

Brokers need to see depth-of-book quotes to handle the large orders of institutional investors. Market imbalance information is relevant past the top-of-book displayed quotes. During the early 2000s, when the markets converted to decimal trading, the extent of liquidity provided by top-of-book quotes in core data dropped significantly. This would likely also be the case if there was any potential reduction to the round lot size for higher priced stocks. As a result, staff will re-examine whether core data should be expanded to include liquidity beyond the top of each market's order book.

Fair and Efficient Access:

To what extent has the cost of timely access to competitive data and access affected the goal of fair and efficient access to markets at all venues, including the larger protected venues?"

Before jumping into my propose solutions, following table highlights the weaknesses of prevailing practices:

OLD Practices	Served Purpose	Shortfall
17 CFR §242 603(a)[5] interpretation	Transmitting or releasing data no sooner than to a Network processor (SIP)	70 FR 37567[4] and Order 67857[6] are ***Not*** equivalent to secured delivery and retrieval of data in a timely manner
Guidance[18] on SRO rule filings relating to Fees	Strong message to tell SROs that the SEC is watching market data price increases	No further exacerbating of market data price differences does ***Not*** ensure a fair and efficient access
Bring competition[19] to centralized processor infrastructure	Accelerate the dissemination of core data at possibly lower cost	***Not*** working, multiple NBBOs would confuse and further convolute the market; it won't curb rent-seeking behaviors, nor promote fairness

---

[17] https://www.sec.gov/news/speech/clayton-redfearn-equity-market-structure-2019
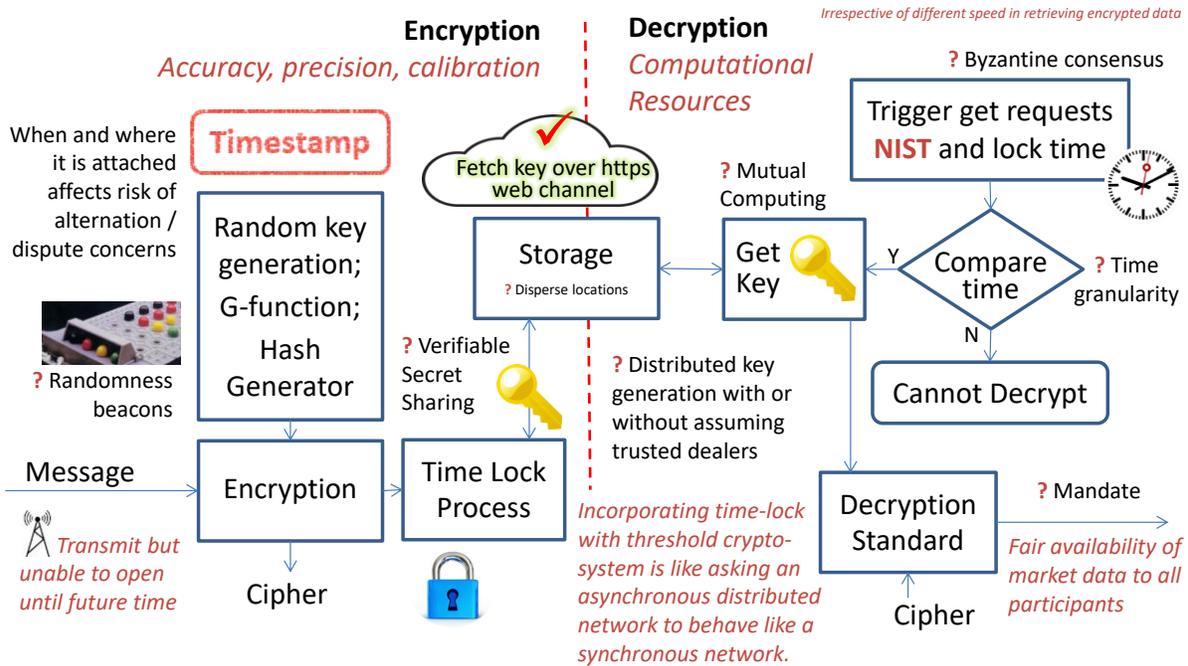[18] https://www.sec.gov/tm/staff-guidance-sro-rule-filings-fees
[19] https://www.sifma.org/wp-content/uploads/2018/10/File-No.-4-729-SIFMA-Comments-on-Roundtable-on-Market-Data-and-Market-Access-October-24-2018-002.pdf

3. Patent-Pending Solutions: Time-Lock Encryption, Data Compression and Delivery

Time-lock Encryption *Patent Pending*

Time-lock encryption is a method to encrypt data such that it can only be decrypted after a certain deadline has passed. The goal is to protect data from being decrypted prematurely. There are various ways to build time-lock encryption for different protection requirements. Please carefully pick and choose from followings:

☐ Randomness beacons?
☐ Verifiable secret sharing; disperse key storage locations; mutual computing to get key?
☐ Distributed key generation with or without assuming trusted dealers?
☐ Byzantine consensus?
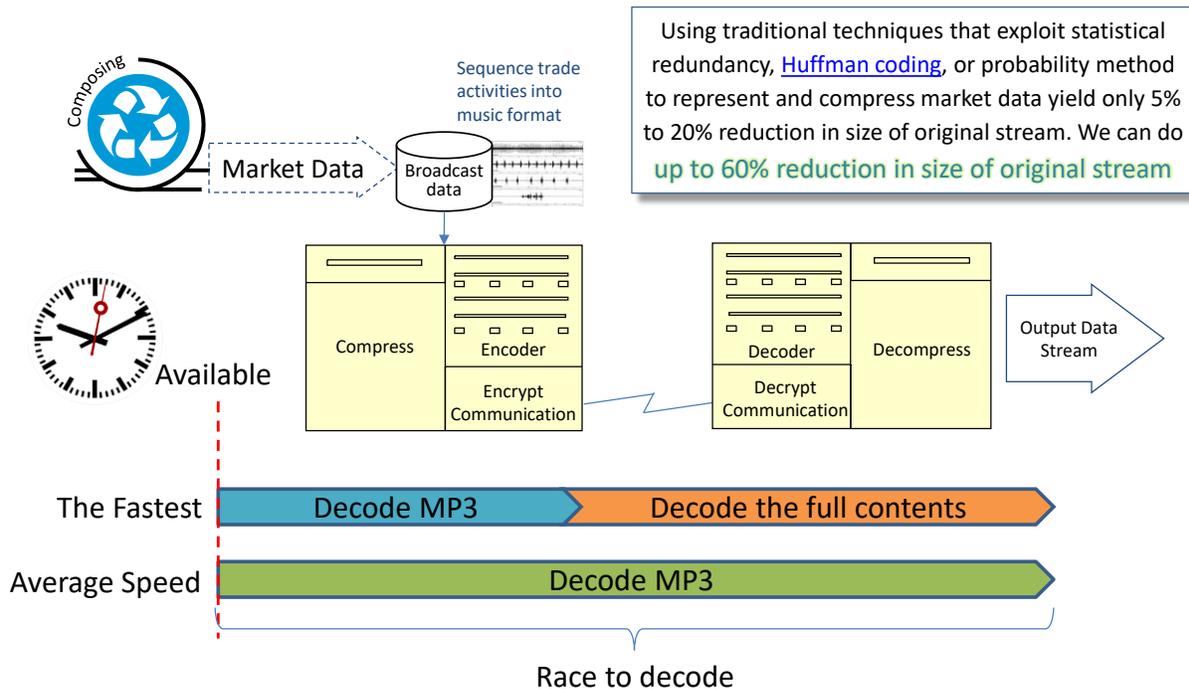☐ Time granularity requirement; mandate decryption standard?

The architecture needs **precise calibration of time** with an independent time aware atomic clock, such as the NIST[20]. Besides, we don't want to push the bottleneck to an arms-race of using high-performing computers[21] to decrypt data. Hence, **computational resources** and the type of data contents must also be considered in the design of a reliable encryption scheme.

---

[20] https://www.nist.gov/news-events/news/2018/11/nist-atomic-clocks-now-keep-time-well-enough-improve-models-earth
[21] https://insidehpc.com/hpc-basic-training/what-is-hpc/

212° DATA BOILER
TECHNOLOGIES, LLC

Data Compression & Delivery *Patent Pending*

Composing

Market Data

Sequence trade activities into music format

Broadcast data

Using traditional techniques that exploit statistical redundancy, Huffman coding, or probability method to represent and compress market data yield only 5% to 20% reduction in size of original stream. We can do **up to 60% reduction in size of original stream**

Available

Compress | Encoder | Encrypt Communication

Decoder | Decompress | Decrypt Communication

Output Data Stream

The Fastest — Decode MP3 — Decode the full contents

Average Speed — Decode MP3

Race to decode

In order to match up the race in decrypting/ decoding data, I looked for clues in the stories of MP3, Napster, and iPod to see how the audio sector is able to achieve optimization. I found that MP3 is indeed a lossy compression type[22], while human ears couldn't distinguish almost any difference from lossless music. Lossy methods yield a substantially **greater compression ratio** (60% or more of the original stream) as compared to traditional techniques (only 5-20%) that exploit statistical redundancy, Huffman coding[23], or probability method to represent and compress market data. My patent-pending methods reduce data storage and booster the efficiency in data distribution, while also enabling the **replicate the depth-of-book** information (relative strengths in bid/ask price and steepness of the price curve). I acknowledge that the buy-side also wants the SIP to include all the odd-lot[24] details amid some hidden cost for high priced stocks[25]. My response is: When we are in the midst of systemic reform, asking for too much or insisting on "complete" transparency, may indeed be detrimental to price discovery and the sustainable development of a healthy market. We do want to strike the appropriate balance in order to avoid a "**No fish can survive when the water is too clear**" situation. Given that, we'll preserve the richness of contents to the best we can, while making the tool fast, easy, secure and fit for the effective monitoring of trade activities.

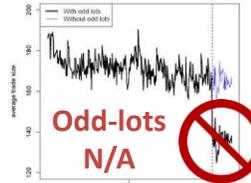[22] https://en.wikipedia.org/wiki/Lossy_compression
[23] https://en.wikipedia.org/wiki/Huffman_coding
[24] https://www.marketsmedia.com/flashback-friday-odd-lots-focus/
[25] https://www3.nd.edu/~scorwin/documents/OddLotPaperJOT_final.pdf

Depth-of-Book Replication *Patent Pending*

Odd-lots N/A

Users will figure out these outliers overtime using reverse engineering

In midst of systemic reform, should not ask for too much

Complete transparency may be detrimental to price discovery

No fish can survive when the water is too clear

Again, instead of bluntly get rid of odd lots by changing the round lot size, one should ask what are the various reasons for the existence of odd lots. For example, the causes related to a lack of stock split, the segment's specific needs in moving blocks under the impending dynamics between lit and dark venues, etc. Odd lot should be "outliers" rather than the "norms". To strike appropriate balance between dragging the speed of real-time SIP feed and content richness, I think a "delayed" odd lot trades and quotations statistics is reasonable compromise, because experienced market participants may use reverse-engineering methods to "figure-out" or "project" how these odd lots would play out in sequence. Following sum up the advantages over prevailing approaches:

Secure 🔒 — Protect data from being decrypted prematurely

Synch 🕐 — Accelerate process speed, savings from compression

Symmetry — Optimize contents' richness, balance diverse interests

BIG DATA | BIG PICTURE | BIG OPPORTUNITIES
We see big to continuously boil down the essential improvements until you achieve sustainable growth!
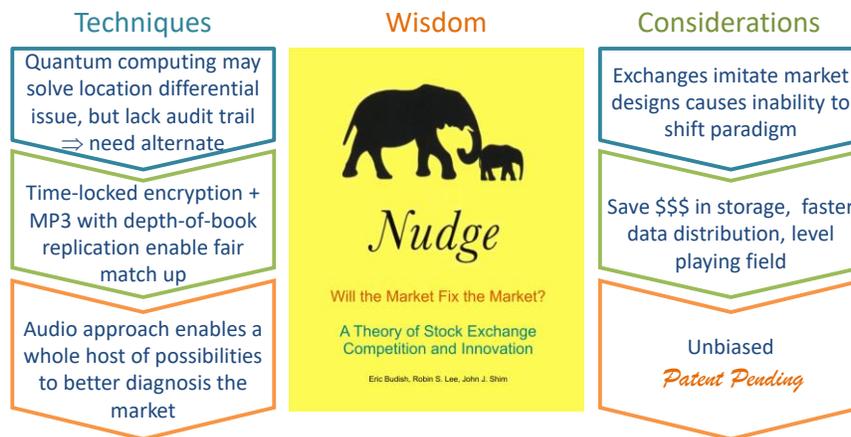☏ 617.237.6111   ✉ info@databoiler.com   databoiler.com

4. How the SEC can HELP

According to an empirical research – "**Will the Market fix the Market?** A Theory of Stock Exchange Competition and Innovation"[26] by Eric Budish, Robin S. Lee, and John J. Shim, it proven that "… If an exchange adopts a new market design that eliminates latency arbitrage, it would win share and earn economic rents. However, imitation by other exchanges would result in an equilibrium that resembles the status quo with competitive trading fees, but now without the rents from the speed race. This means that although the social returns to market design innovation are large, the private returns are much smaller and may be negative, especially for incumbents that derive rents from the status quo." It does require a little "nudging"[27], or the authors called it "regulatory push" to really fix this market data market access problem. Given that, **we urge the SEC to consider:**

(1) **Update interpretations of 17 CFR §242.603(a)**[5], so that it will cover not only the data transmittal aspect, but also emphasis on the synchronized availability of data between SIP and exchanges' proprietary products to satisfy the fair and reasonable, as well as non-discriminatory principles;

(2) Recognize that "real-time" market data is extremely valuable and it **must have proper time-lock security protection**, which the timing of permissible decryption must be regulated in accordance to an independent time aware atomic clock with commonly agreed time granularity, while considering the computing resources to decrypt SIP data cannot be unreasonably burdensome.

Grant Exclusivity

Making market data available securely in synchronized time[2] would align the start time of a "**drag race**". Yet, to deemphasize speed as a key to trading success[16], we have to think about "**curvy race tracks**" that place more emphasis on other skills, like agility to recognize patterns and resiliency to handle different volatile situations. On that I advocate for **democratization of technologies**, so more investors can trade like the Pros and fostering a healthy competition of discovering new patterns and trade strategies, instead of an arm race for the fastest speed.

---

[26] http://www.people.fas.harvard.edu/~robinlee/papers/ExchangeComp.pdf

[27] For behavioral economists, see this: https://en.wikipedia.org/wiki/Nudge_%28book%29, or in the sense of libertarian economists, this is the kind of rare situations where Sir Philip Haddon-Cave, the Master of "Positive non-interventionism", described as possible government intervention would yield necessary advantages.

5. Flash crash and market surveillance drivers propelled CAT as a mission critical project

As far as I know, the International Organization of Securities Commissions (IOSCO) has long identified the need of system, like the Consolidated Audit Trail (CAT), since August 2012 per CR12/2012[28]. In that consultation paper, there identified challenges to effective monitoring of markets and I'd like to highlight the followings:

- The lack of synchronized clocks among all of the entities that need to submit data;
- Ability to reconstruct and analyze order book(s) in the correct sequence, what tools are necessary to do so?

Since a thousand trades can occur in between the 50 +/- millisecond tolerance time allowed by CAT, the inexactitude in trade sequencing would cause analytic results that based on vector measurements/ visualized heat-map to be erroneous.

We at Data Boiler overcome this inherited problem of data imprecision, by applying my **patent-pending** "**music plagiarism detection**" method to achieve higher tolerance to the unsynchronized clock issue[1]. It is capable of recognizing patterns more quickly (up to 50 milliseconds top speed, as compared to taking hours or days or even months for trade review). Aside from the **accelerated speed** to decipher what's going on in the market, it has **fewer false-positives/ false-negatives** than the traditional techniques.

What is this about

New RiskTech that significantly increases the ability of market surveillance, sequence trade activities, reconstruct market events, monitor market behavior, conduct research, identify and investigate potential misconduct.

Better than measuring vectors graphically

Cross-Markets, Cross-Assets

Pattern Recognition

Context of the Problem

Too Fast — Risk defenses aren't match up with algorithm development

Dynamically changing — Market stress come suddenly, failures filled with surprises

Resources draining — Investigation is a pain, difficult to reveal what's going on

---

[28] https://www.iosco.org/library/pubdocs/pdf/IOSCOPD389.pdf

Following table summarizes the weakness of prevailing practices:

OLD Practices	Served Purpose	Shortfall
Risk Limits	Contain certain situations	*Not* able to catch intraday issues, and huge losses can be accumulated in a split second
Integrate risk controls into **product design**	Strong message to tell traders that they are being watched right from the get-go	*Not* able to detect possible bypass of controls that use synthetically created trades, breaches occur but remain hidden until problem gone haywire usually months later
Value-at-Risk (VaR) including other coherent risk measurements	Predict magnitude and probability of losses	*Not* able to tell when, *not* situational, *not* picking up insights from the field, VaR is too normalized/ over-fit model

The best talents will still fail if they aren't equipped properly to deal with these **sudden surprises**:

- The London Whale resurfaces: Bruno Iksil speaks out[29]
- CEO Blindsided as Bank Added to Risky Positions[30]
- Fined by SEC for Lapses That Caused Crashes[31]
- Flash crashes prod banks towards algo arms race[32]

Metrics/ KRIs aren't effective to deal with rapid issues proliferated by hidden problems and silos. We need an engineering approach to solve financial engineering problems. We need something fast, smart, and easy.

Fast — Real time (up to 50 milliseconds) *cross-over to apply concepts from music plagiarism detection* (patent pending)

Smart — Pattern recognition to curb abuses/ suspicious activities

Easy — Without the trouble to perfect the exactitude of clock synch

29 www.fnlondon.com/articles/the-london-whale-resurfaces-bruno-iksil-speaks-out-20170306
30 www.bloomberg.com/news/articles/2016-03-23/credit-suisse-ceo-blindsided-as-bank-added-to-risky-positions
31 www.bloomberg.com/news/articles/2016-09-26/bofa-fined-by-sec-for-lapses-that-caused-mini-flash-crashes
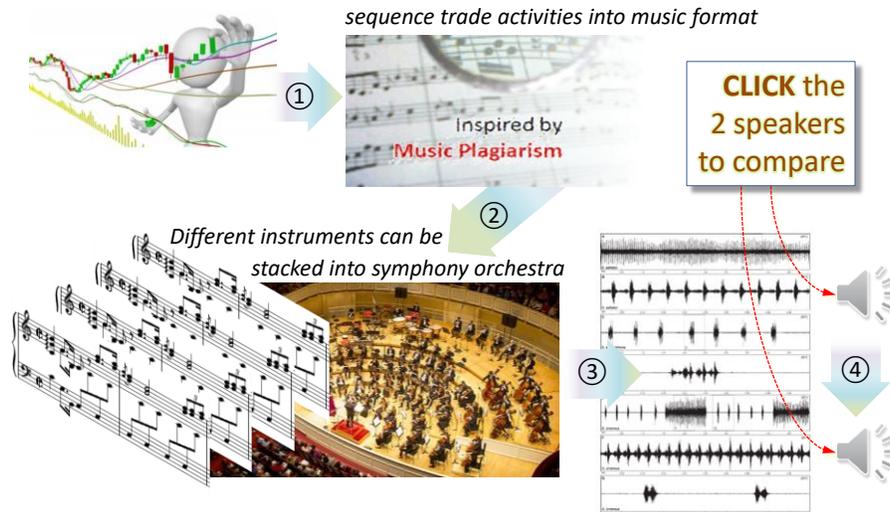32 www.reuters.com/article/us-britain-sterling-flashcrash-idUSKBN13X21N

6. Speeding Up Market Surveillance, Democratizing Technologies, and Level the Playing Field
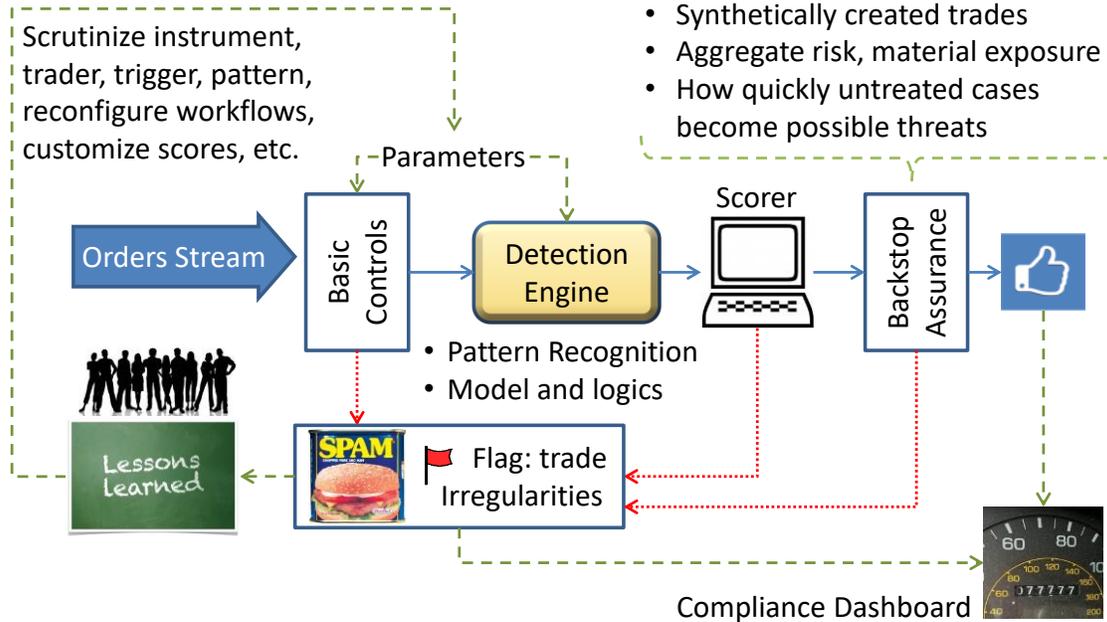
Capabilities of my suite of patent-pending inventions:

Abstract Concepts *Patent Pending*

sequence trade activities into music format

Inspired by Music Plagiarism

CLICK the 2 speakers to compare

Different instruments can be stacked into symphony orchestra

It makes implementation of preventive controls in real time possible; and there are other benefits such as the ease of trade reconstruction, order book replay simulation, backstop assurance, case management capabilities, crowd computing methods, and more.

Curb Gaming of Controls *Patent Pending*

Scrutinize instrument, trader, trigger, pattern, reconfigure workflows, customize scores, etc.

- Synthetically created trades
- Aggregate risk, material exposure
- How quickly untreated cases become possible threats

Parameters

Orders Stream → Basic Controls → Detection Engine → Scorer → Backstop Assurance → 👍

- Pattern Recognition
- Model and logics

Lessons Learned ← SPAM 🚩 Flag: trade Irregularities

Compliance Dashboard

Problem	Action
No pre-cautionary signal and the risk event is inexplicable even after the fact	Learn and be able to catch or avoid the issue next time
Know a threat exit, but incorrectly predict the magnitude.	Learn and continuously fine tune the algorithms with CEP lessons
Know a threat exit and the materiality of it, but override the controls for risky bets	Learn from past override implications to adjust and customize parameters; apply circuit-break to restrict override as appropriate
Know a problem exit and the seriousness, but nothing can be hedged (even one is willing to take a risky bet)	Simulate for best course of actions to minimize loss; prompt FINRA to use the halt authority as appropriate

There are known knowns, known unknowns, **unknown unknowns** about financial vulnerabilities and destabilizing threats. Much market research or monitoring is based on investigating known unknowns (i.e. develop a test hypothesis, then design experiments and hoping the results would be within range of known possibilities). They got stuck when seeing the unexpected.
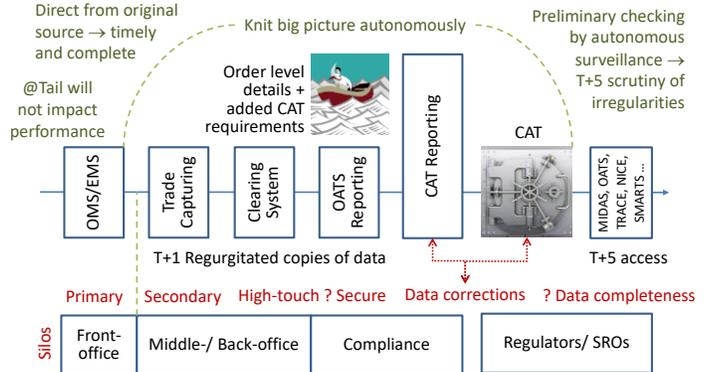
85% → 90% vs 98% → 99.9% incremental improvement
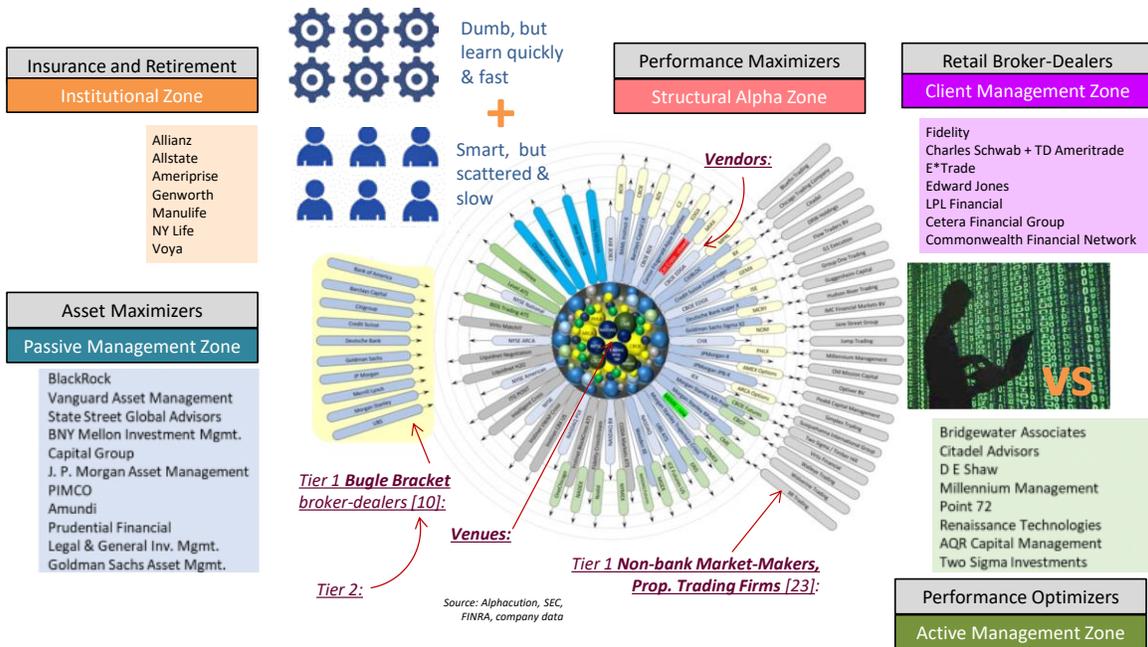
33% vs 95% error reductions

We focus on reducing the unknown unknowns by leveraging **crowd's wisdom and democratization of technologies**. To do that, we need supports from the SEC, FINRA, and the SROs to consider what I called "**H&R Block analogy for CAT**"[33]. I won't go over the details in here, but in short: gives the data submitters a comfort level if their transactions are "clean" or need to be further reviewed; and in the meanwhile, conducts preliminary analytics about suspicious activities.

Collecting Data - Easing Burden

Second, we are thinking about tools to help smaller firms and average investors to trade competitively against the Wall Street Titans that have the size, speed and other advantages. Trading technologies shouldn't be expensive and exclusive to the elite groups! Given that, if we can build a **utility model** where different market constituents can "economically"[34] use our tools to monitor, capture and create trade patterns to form a **community library of signals**. By then, we may be able to shift the paradigm from a "drag race" to rewarding talents who discover useful trade patterns (again, **reduce the unknown unknowns**) that benefit the overall markets.

Continuous Race

Source: Alphacution, SEC, FINRA, company data

---

[33] https://www.linkedin.com/pulse/hr-block-analogy-cat-combating-fraud-kelvin-to/

[34] We envisage multi-tiers licensing of our patent-pending technologies for different users, and possibly a FREE* community basic version if we can justify the overall costs and achieving critical mass in adopting the new tools.

I'll be happy to elaborate my suggestions further if we are to meet in-person. I hope your takeaways from this comment letter include:

- Use time-lock encryption to make market data available securely in synchronized time[2];
- Update interpretations of 17 CFR §242.603(a)[5] regarding the "data protection and availability" aspects;
- Prioritize on addressing the speed differentials issue, then consider the use of a replication of the Depth-of-Book curve and a "delayed" odd lot trades and quotations statistics;[14]
- Know that our patent-pending invention is the ONLY solution to solve the IOSCO[3] or CAT[33] surveillance challenges about analyzing order book(s) amid the unsynchronized clock issue[1];
- Way to deemphasize speed as a key to trading success[16] is by democratization of technologies – fostering a healthy competition of discovering new patterns and trade strategies instead of immerse in a drag race.

I hope our comments will be helpful to the SEC and benefiting the broader industry. Feel free to contact us with any questions. Thank you and we look forward to engage in any opportunities where our expertise might be required.

Sincerely,

Kelvin To

MSc Banking, MMGT, BSc
Founder and President
Data Boiler Technologies, LLC

CC: The Honorable Jay Clayton, Chairman
        The Honorable Robert J. Jackson Jr., Commissioner
        The Honorable Hester M. Peirce, Commissioner
        The Honorable Elad L. Roisman, Commissioner
        The Honorable Allison Herren Lee, Commissioner
        Mr. Brett Redfearn, Director, Division of Trading and Markets
        Ms. Andrea Orr, Counsel to the Director of Trading and Markets

This letter is also available at:
www.DataBoiler.com/index_htm_files/DataBoiler%20SEC%20MarketDataMarketAccess%20Comments.pdf