212° DATA BOILER TECHNOLOGIES, LLC

BIG DATA | BIG PICTURE | BIG OPPORTUNITIES

*Everybody talks about big data, but not anyone can be nominated for the Best Regulatory Compliance Solution!*

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

July 15, 2016

Mr. Brent J. Fields, Secretary
**Securities and Exchange Commission**
Email: rule-comments@sec.gov

Subject: **File Number 4-698**

Dear Mr. Fields:

On behalf of Data Boiler Technologies, I am pleased to provide the Securities and Exchange Commission (SEC) with our comments regarding the §613 Consolidated Audit Trail (CAT) project. Included within the "Respond to requests for comment" are detail opinions and suggestions to selected questions.

Based on our studies of the Joint Industry Plan; Notice of NMS Plan Governing the CAT, we feel there are several conditions hindering the CAT project success* (i.e. per SEC Chair Mary Jo White, "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct"):

| Issues | Suggestions |
|---|---|
| The plan lacks a dynamic analytical framework embedded in the design.<br>The T+5 Schedule for regulatory access is useless in terms of effective market surveillance in prevention of threats to the U.S. financial stability. | Huge loss can be accumulated within split-second. Market collapse does not take more than one day. Analysts need sensors, not encyclopedia. A good decision, made now and pursued aggressively, is substantially superior to a perfect decision made too late. Thus, the plan should mandate "real-time" (access to intra-day feeds) rather than the "8am following day" submission and the T+5 access schedule. |
| Major in the minors – overemphasis on storage, and not enough coverage of pattern recognition and/or systemic way to "red-flag" suspicious activities. | Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities. It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. |
| Stored data aren't openly shared to recognize its values to the fullness.<br>No mentioning of how to make certain delayed CAT data/ statistics available and transparent to the public for essential civilian oversight of the finance industry/ regulatory actions. | Instead of just letting the regulators be the primary user in analyzing CAT data, the plan should consider leveraging the crowd to unleash its powers. By granting certain limited access/ delayed data to the crowd (public disclosure), it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight. |

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

| | |
|---|---|
| A gigantic vault overemphasis on structure while there are other techniques, such as “chaos”, to comprehend the security defense. | (1) Segregation of duties: uses micro-tasking to farm out and distribute the work to various functional units so no one particular unit would have the full big picture. (2) Anti-reverse engineering: obfuscation, introduce randomness, separate and scramble. (3) Vulnerability scan |

We hope the above highlights and the enclosed comments will be helpful to advance the CAT project. Feel free to contact us for any questions, and/or shall our expertise be required. Thank you.

P.S. This letter and the enclosure are also available at:
www.DataBoiler.com/index_htm_files/DataBoiler%20CAT613%20Comments.pdf

Sincerely,

*Kelvin To*
Founder and President
**Data Boiler Technologies, LLC**

## Table of Contents

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

## Respond to requests for comment

**B. Central Repository**

**Question 40:**

No. The purpose of CAT should be about analytics. Analysts need sensors, not an encyclopedia in a "Central Repository". The question should NOT be framed from the prospective of "receiving, consolidating and retaining data", but one should ask "what are the vital few components to enable analytic works be done over the "captured intelligence" (audit trails)?" The plan ought to embed in its design a dynamic analytical framework. For example, having sensors directly conduct real-time analytics over streamed data where it was originated (see our respond to Q.59, Q.85, Q.87, Q.88, and Q.89 for an elaborated discussion). Hence, any irregular activities are spotted timely to become intelligence for further pursuit (i.e. research, and/or investigate collectively with other collected insights/ augmenting facts).

**Question 41:**

No. A good Data Loss Prevention/ Protection (DLP) infrastructure should incorporate, not only the access to stored data "At rest", but also consider treatments for data "In-motion", "In-use", as well as "Disposed". Some typical DLP methods include but not limited to the followings:

| Data | Typical protection methods, but not all |
|---|---|
| In Motion | Use SSL/TLS or IPsec and HAIPE encryption of the data transmitted over the network , Layer 2 "100Gbps" & "ESS" requirements |
| At Rest | Incorporate encryption by the Storage Area Network (SAN) Protect data at rest from access by a rogue host or from physical<br>*Note: virtual disk appears unencrypted to the server, and administrator with full access rights could see the unencrypted data even if s/he is not authorized to do so. Hence, n/a for DIU* |
| In Use | Because DIU has to be decrypted before it can be used, therefore tokenization is needed to encrypt more narrowly, or use database encryption to mitigate full exposure of data (i.e. restrict usage to specific range of data fields and/ or records) |
| Disposed | Use a degausser to thoroughly wiped device before it is discarded |

**Question 42:**

No, the plan should NOT mandate a particular data storage method. It'll make the structure too rigid and static, hindering the flexibility for future scalability. Therefore, we recommend the CAT development team to continuously learn from other industries to consider any innovative approaches.

Indeed, Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities (see related suggestions in our respond to Q.59). It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. We highly recommend CAT to consider using our patent pending invention.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**D. Data Reporting and Recording by Industry Members**

**Question 43:**

No. In order to "link information" to accurately reflect the lifecycle of an order, there requires a robust event sequencing method (suitable for subsequent patterns recognition and analytic processes), which we feel the plan lack sufficient coverage of the topic.

We applaud the efforts to capture both the trade execution information as well as order level details in CAT. Yet, we suggest a closer attention on how those data points would best be represented in a meaningful manner. In our opinions, trade order streams would best express in audio/ music notes form using our patent pending techniques, so it can be played out, reconstructed, orchestrated, and compared with ease and accuracy.

**Question 44:**

No, not until the system may robustly provide insights for the U.S. Suspicious Activity Reporting (SAR), as well as reconcilable and suitable for EU's Suspicious Transaction and Order Reporting (STOR).

**Question 45:**

We recommend the 'audio/ music note method' as stated in our respond to Q.43, while we discourage the plan to prescribe any single approach that may hinder scalability/ future system development.

**F. Times Stamps and Synchronization of Business Clocks**

**Questions 48, 49, and 50:**

We recognize the importance of clock synchronization and time stamp granularity, while we feel it is equally important to be practical. CAT's proposal is indeed synchronized with FINRA's proposal (Regulatory Notice 14-47): "the tolerance for computer clocks would be reduced to 50 milliseconds. The tolerance for mechanical time stamping devices would remain at one second (from the National Institute of Standards and Technology atomic clock)", which we agree.

Please refer to our responds to Q.101, Q.108, Q.114-117, 120, 122, 124-125 for elaborated discussion.

**G. Technical Specifications**

**Question 51:**

No. The list of items included in the Technical Specifications inappropriately constrained the design to be too rigidly follows a traditional SQL database design, when there can be better alternatives. The industry needs digitization, but not necessarily means "submitting" data in particular "format" and/or organizing the data in particular "structure".

Data analysts prefer flexibility to slice and dice data, so matters can be analyzing from multi-dimensional angles in discerning truths, or we refer it as "object-oriented analysis and design" (OOAD). We can adopt data in 'any digitized form' so long as it can be read by computers, rather than emphasis on Extract/ Transform/ Load (ETL) to convert data in particular format to feed certain SQL database's structure.

Modern technologies are capable of analyzing data beyond just text or structured data. Snapchat and other multimedia based apps are gaining popularity. Facebook just predict "the end of written word" (see related news).

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

There are many advantages in analyzing multimedia and/or voice recognized data than interpreting text or structured data alone. Multimedia describes messages in richer 'colors', whereas 'treated data' after the ETL process may discount a lot of useful signals. (Please also see Q.318 and Q.337 regarding standardized symbology)

Therefore, we encourage the plan to consider using "sensors" to capture digitized signals in real-time and in raw form so as to preserve the richness of signals, rather than requiring CAT-reporters to "submit" data in certain format and/or structure. CAT ought to embrace a forward looking design, in order to be flexible and scalable to cope with tomorrow's challenges.

**Question 56:**

Indeed we advocate for material amendments to the technical specification, because we see it as unsuitable to deal with modern day's surveillance challenges. See our respond to Q.51 for more info.

**Question 58:**

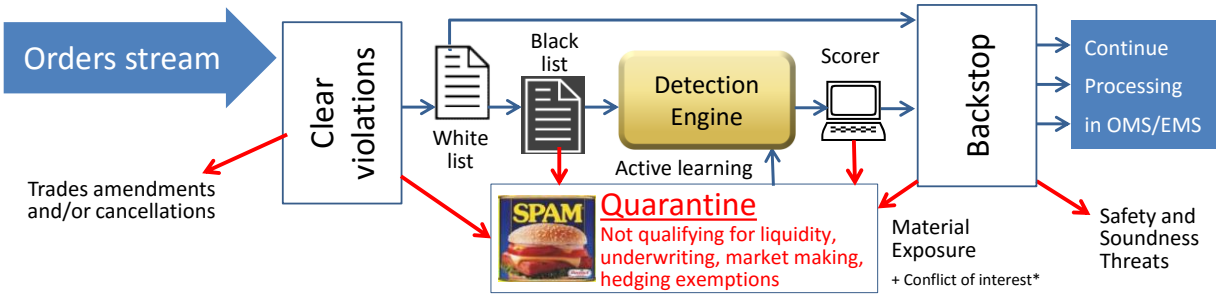
"The goal of next day recovery" and/or "48-hour recovery and restoration period" are nice to have. Consider the national exchange may have experienced down time in the past which management decided not to fall back to contingency mode and halt trading momentarily. If that's acceptable by the market, CAT's Business Continuity Plan/ Disaster Recovery (BCP/DR) should have requirements in relative to the national exchange, but not tighter because there are costs associated with it.

**H. Surveillance**

**Question 59:**

No, the "minimum functional and technical requirements" for the Plan Processor set forth in Appendix is NOT consistent with the creation of enhanced surveillance system. It lacks a dynamic analytical framework embedded in the design. Analysts need sensors, not encyclopedia in a "Central Repository". We recommend the use of sensors to perform real-time analytics over streamed data where it was originated. Hence, any irregular activities are spotted timely to become intelligence for further pursuit (i.e. research, and/or investigate collectively with other collected insights/ augmenting facts).

To facilitate the creation of enhanced surveillance system, we suggest the CAT NMS Plan to follow the approach of our patent pending invention to include the following features or components. Please see below diagram for a high-level illustration:



(a) Clear violations:

All clear violations (e.g. short selling for liquidity management trading desk, use of OTC derivatives for underwriting, etc.) are immediately treated as spam to block from further processing, or be flagged for further investigation. Then the orders stream goes through a comprehensive algorithm to distinguish any prohibited trades from the permitted hedging, market making, and underwriting activities. It automatically red-flags and/or quarantines transactions that are not in clear violation or legitimately clean. It preserves a full audit trail of all released approvals and incorporates a final QA check for backstop provision.

(b) White List:

A 'white list' in the algorithms specifies particular trade types and instruments that are precluded from prohibited trading activities. Let say repos or reverse repos for commercial banking transactions are allowed under Dodd-Frank Volcker Rule, it will bypass all other checking to go directly for a backstop final QA as a result. The backstop provision will examine repo transactions as if they may result in an effect of synthetic short sales for the appropriate quarantine, and prevention of other threats and/or material exposure.

(c) Black List:

As opposed to the 'white list', the 'black list' defines what isn't. Let's look at an example about market making. The algorithms should determine when and what inventory levels are 'inappropriate' for market makers. In other words, orders that are beyond the Dodd-Frank Volcker Rule reasonable expected near term demand (RENTD) limit and passively provide liquidity to the served markets need to be flagged for further investigation. So to establish the proper basis with valid assumptions for what is considered 'reasonable', one would make predictions on different liquidity and/or stress test scenarios. Market Risk analysts got to study the buying behaviors of clients, customers, and counterparties, different market scenarios, and which trade instruments to use. In short, advanced analytics can help develop and substantiate a 'reasonable' securities inventory plan.

(d) Detection Engine:

Moving on to the detection engine's filtering algorithm, it is basically a 'pattern recognition' tool used to quantify matters into a scoring model. For example, the detection engine performs an ABC analysis to look up and identify the value category for the trade order and adjusts the score accordingly to reflect the higher likelihood of that the trade should be flagged for suspicious activities. There are other parameters to consider, while the filtering mechanism depends on the sufficiency of signals picked up by many connected computers. This method implements in a low-latency system for real-time risk prevention has the advantage over human for its objectivity and consistency. More importantly, it is superfast and cheap, so it will save financial institutions from requiring "a lawyer, a compliance officer and a doctor … to detect traders' intents (see related news)".

(e) Backstop Quality Assurance (QA) Check:

Backstop is a final QA step serves like a catch-all clause that allows regulators to step in anytime if anything "may become a threat" to the U.S. financial stability. This shouldn't be treated merely as the

"spirit of the law" to tell them not to do anything stupid with taxpayers' and/or depositors' money. Instead this provision can be enforceable by considering the following practical implementation:

(i) Determine percentages of individual non-bank counterparty exposure and bank CET1/ Tier 1 capital which can be considered materially excessive following BCBS283 guideline: measuring and controlling large exposure;

(ii) Use a system to detect if permissible trades (e.g. repos, securities loans under the Volcker Rule) may synthetically be combined to become prohibited activities before execution;

(iii) Use a system to curb possible market manipulations.

In order to effectively curb any inappropriate use of derivatives and/or other exotic products that created through abusive use of financial engineering techniques, system to facilitate market surveillance must also adopt an engineering approach to address the problems as stated in (ii) and (iii). Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities. It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. We highly recommend CAT to consider the fit-for-purpose of our patent pending invention.

We also urge CAT design team to get familiar with these supervisory frameworks – OFR: Analyzing Threats to Financial Stability, and FSOC: Framework to Mitigate Systemic Risk.

(f) Workflow, dynamic upgrades, and others supporting features:
The system will need workflow processes to alert and escalate the suspicious activities to the relevant management of the financial institutions, trading venues, and related regulatory agencies in real-time. It will then document any released approvals (e.g. dismissal of any early warning signals and/or unflagging any suspicious trades), and change of course actions to the surveillance policies, procedures, and/or approaches.

The surveillance policies, procedures, and/or approaches resided in the system should not be static, but dynamically evolve through machine learning - "reinforcement model". This will enable continuous fine tuning of the system' parameters, to address any concerns about the consistent application of static formulas may potentially be reverse-engineered by rogue traders to bypass the system. The system's algorithms will evolve and benefit from the crowd collective intelligence. It's a machine that assimilates knowledge quickly from every move of its users (e.g. dismissal of any false positives or false negatives, and special scrutiny of particular nuances). The more users experience the system can accumulate, the better the detection algorithms will get.

If CAT reporters feel this utility model for CAT system design may expose their trading strategies to other participants in the network, there ought to be obfuscation techniques for necessary protection. Introduce randomness to resist pattern recognition, making it incompatible, separating and scrambling and/or aggregating rollup are effective mitigation methods. However, if in case there may be sufficient warning signal(s) of impending threat(s) to the U.S. financial stability, an emergency situation may be declared.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

During which, appropriate trade streams with complete transparency (unmasked) may feed into a higher-tier analytic platform, in order to facilitate a more in-depth study of the market dynamics by the regulators in monitoring any contagious effect/ systemic risks.

**Question 60:**

No, the online targeted query tool and user-defined direct queries and bulk extracts described in Sections 8.1 and 8.2 of Appendix D will NOT enable regulatory Staff to use the data. These query and/or bulk extract methods are "generic" and NOT fit for the purpose of financial market surveillance. Despite direct queries may be "user-defined", they AREN'T "engineering" approach to unveil whether trades may be synthetically created to bypass control.

Again, in order to effectively curb any inappropriate use of derivatives and/or other exotic products that created through abusive use of financial engineering techniques, system to facilitate market surveillance must also adopt an "engineering" approach to address the problems as stated in our respond to Q.59 – point (e)(ii) and (iii). Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities. It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. We highly recommend the CAT to consider using our patent pending invention.

**Question 61:**

No. There ought to be better obfuscation techniques for the necessary protection. Introduce randomness to resist pattern recognition, making it incompatible, separating and scrambling and/or aggregating rollup are effective mitigation methods that we suggest the plan to consider. Also, please refer to our suggestions in Q.150-155 regarding the use of trade pattern as "fingerprint" unique ID.

As per our respond to Q.59 - last paragraph, "However, if in case there may be sufficient warning signal(s) of impending threat(s) to the U.S. financial stability, an emergency situation may be declared. During which, appropriate trade streams with complete transparency (unmasked) may feed into a higher-tier analytic platform, in order to facilitate a more in-depth study of the market dynamics by the regulators in monitoring any contagious effect/ systemic risks."

**Question 62:**

No, the CAT user support functionality (as described in Section 10.2 of Appendix D) does NOT provide sufficient assistance to regulators in carrying out their regulatory functions. Again, in order to effectively curb any inappropriate use of derivatives and/or other exotic products that created through abusive use of financial engineering techniques, system to facilitate market surveillance must also adopt an "engineering" approach to address the complexity of today's financial markets' problems. See our respond to Q.59 for a more elaborated discussion.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**I. Information Security Program**

**Questions 63 and 64:**

The CAT information security program is largely off-the-mark to support CAT's intent purpose (i.e. "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct").

If CAT is designed with the analytic framework as suggested per our respond to Q.59, then regulatory users should theoretically be accessing CAT data only under the following circumstances:

- Query historical data for cases currently or previously under investigations/ prosecutions;
- Access to current trade activities information upon alert of early warning signals;
- Declared emergency when there may be sufficient warning signal(s) of impending threat(s) to the U.S. financial stability (see our respond to Q.59 – last paragraph for further details).

Other than the above scenarios, no one (Regulatory users, CAT-Reporters, Plan Processors, and Information Security Team included) should have access to unmasked "order level details" (i.e. transaction-by-transaction) information.

However, instead of just letting the regulators be the primary users in analyzing CAT data, the plan should consider leveraging the crowd to unleash its powers. By granting certain limited access/ delayed data to the crowd, it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight. Therefore, obfuscated CAT data on an "aggregated" level should be shared as public information.

Moreover, delayed CAT data pertaining to historical "prosecution" events (e.g. London Whale, LIBOR manipulation) and/or major financial crisis (e.g. LTCM, 2008 mortgage meltdown) should be scrambled and stored into database libraries, to be consumed by the public under registered access. It will facilitate the private sector (i.e. the academic and financial technology companies) to discover better approaches (as compared to sole effort by the regulators) to market surveillance. Hence, there will be continuous innovations (like there are FREE music libraries for anyone who can contribute to better music plagiarism detection) to help foster financial stability.

Given the above suggestions, the CAT information security program should consider a "multi-tiers" approach rather than one generic set of policies and procedures.

**(6) Financial Matters**

**Questions 65-67, 69-70, and 72:**

The funding and budget for CAT should benchmark in reference to the costs of the Dodd-Frank Volcker Rule compliance. In order to build a reasonably designed compliance program to effectively prevent Volcker Rule proprietary trading ban violations, we envisage a "utility model" similar to CAT that its cost will be shared by participating financial institutions. Given said that, the Office of Comptroller of Currency (OCC)'s had conducted an "analysis of 12 CFR Part 44" for the Volcker rule compliance (a salvaged copy may be downloaded at: https://perma.cc/BA7-R4PG).

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

The following numbers are extracted from "the OCC analysis of 12 CFR Part 44":

| RENTD Testing & Validtion | 7 large banks | 39 other banks | sub-total |
|---|---|---|---|
| 2014 | $70,900,000 | | $70,900,000 |
| 2015 | $70,900,000 | $31,700,000 | $102,600,000 |
| 2016 | $70,900,000 | $31,700,000 | $102,600,000 |
| 2017 | $70,900,000 | $31,700,000 | $102,600,000 |
| Total | $283,600,000 | $95,100,000 | $378,700,000 |
| Average | $40,514,286 | $2,438,462 | $8,232,609 |

Reasonable expected near-term demand (RENTD) is only one of the key components of the Volcker rule compliance, whilst algorithms to decipher trades as permissible versus prohibited should be the primary cost driver for a system that facilitate market surveillance. Therefore, the industry should expect an annual cost for CAT of approximate $2 million to 40 million per financial institution during the initial years.

We estimate the on-going costs for the CAT infrastructure (inclusive of BCP/ DR costs), to be about $28 million to $36 million annually assuming a low-latency platform running at about 50 millisecond speed.

Regarding the funding model for CAT on-going operations, we feel a toll charge based on "successful" quarantine or red-flag of suspicious trade messages would be a better alternative. If imposing toll charge on all trade messages that goes through CAT, it would essentially be like a "financial transaction tax" (see related news) that is unjustifiable and detrimental to the overall health of the U.S. capital markets.

Our recommended "successful red-flagging" approach will incentivize the CAT design team to come up with an effective algorithm to discern permissible versus suspicious/ prohibited activities. There should also be moving targets for number of trade messages expected to be quarantined/ red-flagged (in the range of at least 1 billion, to not over 12 billion in volume per annum) to ensure there won't be too many "false alarms" generated from the CAT algorithms. In return, the toll charge per "successful" quarantined/ red-flagged message should vary with its volume (in range of at least 0.5¢ to not over 5¢).

The price range and volume targets should be reviewed regularly on annual basis. The consideration factors for the review should include but not limited to: (1) the industry's capacity to review queue of "quarantined/ red-flagged" messages; (2) CAT algorithms' effectiveness; (3) price affordability per trading desk in average; (4) on-going operating and development costs for CAT. Given the diversified consideration factors for the price range and volume targets' review, the review and setting of standardize fee table would be appropriate to be done by an Independent Advisory Committee, rather than the Operating Committee.

In conclusion, our short answers to the various questions are – Q.65: NO, NOT appropriate and reasonable; Q.66: NO, NOT a fair allocation of CAT-related fees between Participants, other types of Execution Venues, and Industry Members. Q.67: NO, assessing fees based on market share and message traffic, as described in Sections 11.2 and 11.3 of the CAT NMS Plan, is NOT appropriate and reasonable; Q.69-70 and Q.72: NO, Operating Committee should provide information for the advisory board to consider, but NOT making decision that tolerates from a standardized fee table set by the Advisory Committee annually.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**(9) Plan Appendices**

**Questions 79:**

Using CAT as a replacement to OATS may just be "majoring in the minors". CAT should aim for facilitating a "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct", as well as enabling the identification of threats to the U.S. financial stability. Thus, CAT should be a boiling plate to holistically look into:

(a) Irregularities in trade orders (SAR, STOR);
(b) Excessive risk-taking and positioning;
(c) Capital adequacy, risk tolerance under stress scenarios;
(d) Market structure and liquidity issues (LCR, NSFR, leverages, PPNR);
(e) Risk convergence (e.g. a swift from market risk to credit risk, and vice versa);
(f) The migration of activities due to financial innovation and regulatory arbitrage;
(g) Clear and settlement pressure, market bubbles and any systemic risk concerns.

U.S. regulators have been collecting data and metric reports for the above matters separately (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) CAT presents an opportunity where all metrics reports may be consolidated to cross-tabulate into insights to better monitor and advance the development of the U.S. capital market.

Therefore, in addition to real-time monitoring of order level information, CAT should measure and scan the market, gauge the supply and demand of different instruments, venues, and the needs of investors in different segments. There ought to be continuous assessment of changing dynamics (e.g. maker-taker model versus trade execution quality, arbitrage versus market making, meaningful quoting, demand of liquidity timing, rebalancing, etc.). Other measurements include the appropriate accounting of instruments turnover rate, periodic review on reliability of securities supply, check for deterioration or losses in accrual loan portfolio, etc. All these measurements will help substantiate a meaningful scorecard metrics for reasonable activities in the market, and curbing any excessive risk taking. Please click here to see a related whitepaper for more information.

Besides, below is an extraction of Figure 2-6 from the Office of Financial Research (OFR)'s annual report that the agency has nicely laid out a list of risk indicators. We urge the CAT NMS Plan to consider.

Figure 2-6. Financial Stability Monitor Risk Classification and Examples of Indicators

| Risk | Definition | Indicators |
|---|---|---|
| Macro-economic | Evaluates risks that have the potential to affect financial stability through various macro channels such as growth, external balances, fiscal vulnerabilities, and confidence channels. | Financial conditions, output gap, sovereign debt levels and financing costs, foreign exchange reserves, current account balances, consumer and business confidence, inflation volatility, and inflation expectations |
| Market | Assesses the risk of destabilizing losses across key asset classes and investment strategies as a result of adverse movements in asset prices. | Duration, positioning, risk premiums valuations, and volatility |
| Credit | Measures the propensity of a counterparty to meet its financial obligations, and includes market-implied and balance-sheet measures of risk. | Corporate credit spreads, balance-sheet leverage, lending conditions, delinquencies, asset quality of households, corporates, banks, and nonbank financial institutions |
| Funding/ liquidity | Captures market liquidity, balance-sheet liquidity ratios, stress in funding markets, and the potential for vulnerabilities that arise from excessive leverage. | Broker-dealer inventories, turnover, volume, cash balances, dependence on wholesale funding, changes in short-term investor assets under management and tenors, foreign exchange basis swaps, short-term funding rates/spreads |
| Contagion | Measures the vulnerability of the financial system to sudden shocks that may spread as a result of interconnectedness. | Contingent claims analysis, conditional value at risk, systemic expected shortfall, distressed insurance premium, network analysis, cross-border exposures, sovereign-bank exposures, correlation risk |

Source: OFR analysis

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

Our short answer to Q.79 is: it should. At a minimum target, CAT should replace OATS on day-one of its implementation, while scope of CAT should be much broader as we have suggested above.

**Questions 80:**

No, there are not enough incentives for continuous CAT innovation and cost reductions by the Plan Processor and the Participants. As we have mentioned in section (6) Financial Matters, we recommend a "successful red-flagging" approach to incentivize the CAT design team to come up with an effective algorithm to discern permissible versus suspicious/ prohibited activities. There should also be moving targets for number of trade messages expected to be quarantined/ red-flagged (in the range of at least 1 billion, to not over 12 billion in volume per annum) to ensure there won't be too many "false alarms" / too little "early warning signals" generated from the CAT algorithms. In return, the toll charge per "successful" quarantined/ red-flagged message should vary with its volume (in range of at least 0.5¢ to not over 5¢).

We envisage that CAT algorithms will enable automated surveillance to help prevent financial crisis (which a crisis could cost the US economy more than $22 trillion per a study by the Government Accountability Office). We feel it will be appropriate to link CAT / surveillance budget with how effective the system can effectively mitigate threats to the U.S. financial stability. Therefore, we welcome competition in designing better surveillance algorithms to foster further innovations. Again, delayed CAT data pertaining to historical "prosecution" events (e.g. London Whale, LIBOR manipulation) and/or major financial crisis (e.g. LTCM, 2008 mortgage meltdown) should be scrambled and stored into database libraries, to be consumed by public under registered access. It will facilitate the private sector (i.e. the academic and financial technology companies) to discover better approaches (as compared to sole effort by the regulators) to market surveillance. Hence, there will be continuous innovations (like there are FREE music libraries for anyone who can contribute to better music plagiarism detection) to help foster financial stability.

In terms of CAT infrastructure cost, it should come down naturally according to Moore's law (processing power doubles every 18 months). CAT management should actively scout out for such technology advancements to consider possible cost reduction. Again, we suggest CAT to adopt our patent pending invention to convert/ sequence trade stream into music format for possible data storage cost savings and other benefits – ease of comparison with accuracy (pattern recognition speed up to 50 milliseconds).

**Questions 81:**

The plan lacks a dynamic analytical framework embedded in the design as we have mentioned in Q.59. CAT milestones need to be aligned with its objective (i.e. facilitate market surveillance), not for the shake of building a gigantic vault to store data. Hence the milestones set forth by the proposed CAT NMS Plan are unacceptable. As we have suggested in our respond to Q.79, CAT presents an opportunity where all the regulatory data submission and/or metrics reports (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) may be consolidated to cross-tabulate into insights for better market surveillance. We demand to see a better "roadmap" of how these data, metrics reports, and/or audit trails may be consolidated.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**B. Summary of Additional CAT NMS Plan Provisions – 1. Reporting Procedures**

**Question 84:**

No, data recording, reporting, and formatting procedures described in the CAT NMS Plan are inappropriate and unreasonable. Please see our respond to Q.43, Q.44, and Q.45 for an elaborated discussion.

**Question 85:**

Outreach, support, training, guidance and/or documentation are of minor relevancy to ensure that CAT Reporters are able to make data transmissions to the Central Repository. In order to major in the major, CAT should consider censoring (or tag information from) order management system (OMS)/ execution management system (EMS)/ portfolio management system (PMS), and other data sources that information was originated. Again, as we have mentioned in respond to Q.40, the plan ought to embed in its design a dynamic analytical framework (e.g. having sensors directly conduct real-time analytics over streamed data where it was originated). Hence, any irregular activities are spotted timely to become intelligence for further pursuit.

**Question 87:**

If CAT may adopt our suggestion to tag information as they were originated, there should not be worries about any updates and/or correction to CAT Reportable events. In big data, we prefer and believe the most original the data, the slighter the chance of introducing any noises and/or discounting the signals during the ETL process. According to Viktor Mayer-Schönberger and Kenneth Cukier in their book – Big Data: A Revolution That Will Transform How We Live, Work, and Think, "any particular reading may be incorrect (in a messy dataset), but the aggregate of many readings will provide a more comprehensive picture." Essentially, the messy whole can outperform exact, accurate subsets.

**Question 88:**

Why a more detailed schedule, with milestones, for CAT Reporters to adhere to in setting-up or configuring their systems to become CAT Data reporting compliant?! If CAT may adopt our suggestion per respond in Q.85 and Q.87 to tag information as data were originated in (or censoring) OMS/ EMS/ PMS, and other data sources, it won't be as invasive to the CAT Reporters to convert and submit data in particular standardized format (that it may introduce noises and/or discount the signals during the cumbersome ETL process).

**2. Timeliness of Data Reporting**

**Question 89:**

Analysts need sensors, not encyclopedia. A good decision, made now and pursued aggressively, is substantially superior to a perfect decision made too late. Thus, the plan should mandate "real-time" (intra-day feeds) rather than "8am Eastern Time the following day" submission.

Per our respond to Q.59 – point (e), in order to effectively curb any inappropriate use of derivatives and/or other exotic products that created through abusive use of financial engineering techniques, system to facilitate market surveillance must also adopt an engineering approach to address the problems as stated in (ii) and (iii). Data Boiler has a patent pending invention to convert/ sequence trade streams and using a financial engineering approach to recognize patterns. It'll enable ultra-fast analysis up to 50 milliseconds.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

Today, front office traders are using artificial intelligence (A.I.) and high frequency trading (HFT) algorithms. Regulators and the industry's middle-office risk and compliance team need to be equipped with real-time market surveillance in order to match-up with the front-office, or else huge losses in billions could happen in split second triggering another possible crisis. We can't turn back the clock, and CAT NMS plan must adopt a data timeliness standard at near real-time speed (i.e. 50 milliseconds in our opinion). We recommend the regulators to monitor a live feeding queue of irregular patterns/ suspicious activities that detected by the CAT filtering algorithms (see our respond to Q.59) using sensors captured "real-time" data. Please also see our respond to Q.194 and 195.

**Question 90:**

We believe that CAT Reporters will submit their reports at or about the same time, if an "unnecessary" deadline is set, such as 8:00 a.m. Eastern Time. Per our respond in Q.89, we advocate for "real-time censoring" of trade activities, to eliminate the need of data "submission" by a certain time.

Regarding spike in data submission volume at particular time frame, that should not be an "excessive" burden on the Plan Processor. If CAT infrastructure cannot even handle the "normal" day's spike in data volume, how would it be capable to handle the data deluge when the market goes haywire?!

Given said that, we are not undermining the challenge to make CAT infrastructure sturdy and scalable. One may be tempted by the petabyte processing power of Hadoop at less than a fraction of a percent cost to the upgrades quoted by certain "branded" vendors. The temptation is real because many have long been held hostile by these "branded" vendors who charges premium price for the necessary upgrades without giving the buyers any alternatives (i.e. locked-in because of compatibility issue). However, one should not jump to the conclusion in favoring a particular infrastructure approach.

To optimize Big Data's performance, creative methods are invented for specialized functions. Various NoSQL databases (column store, document store, key-value, graph DB, etc.) have different purposes. On top of that, there are the mix-and-matches of Flume/ SQOOP, HIVE/ PIG, B-Trees/ LSM, and more. There are relative advantages and compromises for each of these functions. Besides, how ready is CAT in letting go the structural schema of traditional database to gain efficiency in analyzing Big Data? If the CAT infrastructure team still has their mind anchored in the old-school approaches, such as traditional ETL (extract, transfer, and load) how would CAT move forward in the Big Data era (see Q.51's respond)?! Our point here is: it is likely going to be a mix-and-match approach for CAT.

The way to come up with the best mix indeed depends NOT on the hardware, but on the fundamental issue of "what problem CAT is trying to solve". Reference to our respond in Q.40, the plan ought to embed in its design a dynamic analytical framework. For example, having sensors directly conduct real-time analytics over streamed data where it was originated. Hence, any irregular activities are spotted timely to become intelligence for further pursuit. Instead of building a gigantic vault (see Q.81's respond), we suggest CAT to follow our suggestions in Q.59 so that CAT may effectively use a "divide and conquer" approach to deal with only the relevant signals at each "stage" (i.e. clear violation, white list/ black list, detection engine, and backstop QA).

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

In our opinion, there is no point in storing any data if users do not know when and where to use them. Our suggested "divide and conquer" approach will ensure "fit-for-purpose" that every data stored will be used effectively to facilitate the necessary surveillance goals, as well as minimizing resources waste.

**Question 91:**
The capacity is all depended on the CAT infrastructure design. If the technical specification stated in its current form, it would never be sufficient. If CAT may adopt our suggested "divide and conquer" approach to CAT infrastructure design (see Q.90's respond), then the capacity matter is much easier to manage because only meaningful signals are captured and stored at each stage, rather than storing the entire data deluge of everything.

**Question 92:**
1,800 CAT Reporters may sound a lot, but there are only a handful of major technology vendors (Calypso, Murex, Charles River, SunGard, Fidessa, Bloomberg, Markit, Aladdin, Advent, Eze, etc.) that most CAT Reporters are using their OMS/EMS/PMS (see Q.85). If CAT may consider censoring these systems directly, rather than having individual financial institutions submitting data, then the task will be easier to manage.

**3. Uniform Format**

**Questions 93-100:**
Please refer to our respond in Q.87, "the messy whole can outperform exact, accurate subsets" per Big Data: A Revolution That Will Transform How We Live, Work, and Think. It'll take "forever" to come up with a "golden" unified "single source of truth", and by the time a common standard is adhered, value of the data subsided to almost worthless in the context of market surveillance. Again, analysts need sensors, not encyclopedia. A good decision, made now and pursued aggressively, is substantially superior to a perfect decision made too late. Thus, our recommendations are:

- Preserve data in its most original form (the slighter the chance of introducing any noises and/or discounting the signals during the ETL process).
- Then convert/ sequence trade streams into "music format" for ease of comparison and storage advantages, as well as capable of facilitating real-time market surveillance (see Q.42).

Quality, reliability and accuracy, should be determined in relevant to the achievement of CAT's purpose (i.e. "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct"). Anything that deviates from such purpose cannot and should not be recognized as "quality", no matter how glamorous the CAT "Vault" / "Encyclopedia" may be. We feel that if CAT may adopt our patent pending invention (including a pattern recognition method that crossover from "music plagiarism detection" – see Q.59), it will help optimize the reliability and accuracy to extract sufficient insights in a timely manner to rightly serve the CAT's purpose.

**4. Clock Synchronization**

**Questions 101:**
Yes, a clock offset tolerance of 50 milliseconds is appropriate and reasonable.

**Questions 108:**

If certain categories of market participants can hold to a smaller clock offset tolerance of 50 milliseconds, then they should. Clock offset tolerance speed requirement should go hand-in-hand with how fast a market participant is allowed to conduct their HFT activities. In terms of sequencing of reportable events, it's all about "onset detection", or in other words: the likelihood of irregular activities detect within a time interval. We are aware that "NASDAQ is making investments in the Securities Information Processor that will dramatically increase its speed from 225 milliseconds a decade ago to 500 microseconds today, and soon to 50 microseconds". While there will always room for improvements, in the meanwhile, we feel the 50 milliseconds' clock offset tolerance is appropriate and reasonable for the time being.

Note: this requirement of 50 milliseconds clock offset tolerance should be reviewed regularly if it may introduce too much noise and/or overly distorted signals for market surveillance and manipulation detection purposes.

**5. Time Stamp Granularity**

**Questions 114-117, 120, 122, 124-125:**

We feel it is okay to record and report the time of each Reportable Event using time stamps that reflects current industry standards (i.e. at least the millisecond for electronic reportable event). Time stamp granularity requirement should go hand-in-hand with how fast a market participant is allowed to conduct their HFT activities. If certain categories of market participants can originate, modify, cancel, route, execute trade, and/or allocate an order in substantially less than one millisecond, then they should record and report the time of each reportable event using time stamps reflecting their sub-millisecond or microsecond processing capability.

Note: No one particular reportable event is more time-sensitive than the others from the perspective of market surveillance.

In terms of sequencing of reportable events, it's all about "onset detection", or in other words: the likelihood of irregular activities detect within a time interval. We are aware that "NASDAQ is making investments in the Securities Information Processor that will dramatically increase its speed from 225 milliseconds a decade ago to 500 microseconds today, and soon to 50 microseconds". While there will always room for improvements, in the meanwhile, we feel the one millisecond time stamp granularity standard for record and report the time of each electronic reportable event deems sufficiently precise for analytic purpose, including enabling the regulators' ability to reliably and accurately sequence events that occur in different execution venues.

Note: this requirement of one millisecond time stamp granularity for electronic reportable events should be reviewed regularly if it may introduce too much noise and/or overly distorted signals for market surveillance and manipulation detection purposes.

We acknowledge the SROs note in the Exemption Request that recording and reporting "manual order events" with a time stamp granularity of at least one second would result in little additional benefit, but we disagree with the "false sense of precision" part. CAT has to set a particular time granularity standard for non-electronic reportable events. Should one-minute or one-second be the appropriate time stamp granularity for "manual order events", it depends if it may introduce too much noise and/or overly distorted signals for market surveillance and manipulation detection purposes. In order to be practical, we suggest moving forward the CAT project with a more

relaxing time stamp granularity than the one-second for "manual order events" at the beginning. CAT may tighten the requirement/ standard during subsequent review (at least on annual basis).

After all, any time stamp granularity requirements must be enforceable by the CAT processor, or else the standard would be meaningless. In conclusion, our short answers to the various questions are – Q.114: Yes for electronic, No for non-electronic; Q.115: Yes, it is precise enough; Q.116: Yes, enable; Q.117 No particular reportable event is more time sensitive from the perspective of market surveillance; Q.120 Yes, the one-second time stamp granularity standard is appropriate and reasonable; Q.122: We partially agree with the SROs note to grant exemption for manual order events; Q.124: Yes; Q.125: CAT processor's ability to enforce the one-second time stamp granularity requirement.

**6. CAT-Reporter-ID**

**Questions 128-133:**
We believe that allowing the Existing Identifier Approach would be more efficient and cost-effective than the Rule 613 approach of requiring a CAT-Reporter-ID to be reported for each order and reportable event in accordance with Rule 613(c)(7). Regardless of which approach, CAT should tag related trade pattern with each identifiable trading desk and trader. The tag along trade pattern would become the unique "fingerprint" to hunt down rogue traders if they may hop from one trading desk or entity to another. Therefore, implementation of the Existing Identifier Approach would not mean transferring costs from CAT Reporters to the Central Repository.

CAT's purpose is to enable "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct", thus tagging unique trade pattern and assigning unique identifier are part of CAT's responsibilities.

Our short answers to the various questions are – Q.128: Yes, we believe the Existing Identifier Approach is more efficient and cost-effective; Q.129: No it shouldn't affect accuracy; Q.130: No, it shouldn't affect accessibility; Q.131: No, it shouldn't affect timeliness; Q.132: No, it shouldn't affect the security and confidentiality; Q.133: The challenges to link all SRO-Assigned Market Participant Identifiers to the appropriate CAT-Reporter-IDs are real, but that's part of CAT's responsibilities.

**Questions 134:**
We believe the CAT NMS Plan should mandate that Industry Members provide their LEIs, in condition of such information being already captured at respective systems where CAT may install appropriate "sensors" to capture the information (see Q.40, Q.51, Q.85, Q.87, Q.88, and Q.89).

**7. Customer-ID**

**Questions 135-142:**
We believe that allowing broker-dealers to report a Firm Designated ID to the Central Repository is more efficient and cost-effective than the Rule 613 approach of requiring broker-dealers to report a unique Customer-ID upon original receipt or origination of an order. We believe implementation of the Customer Information Approach would not mean transferring costs from individual broker-dealers to the Central Repository.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

Regardless of which approach, CAT should tag related trade pattern with each identifiable customer and counterparties. The tag along trade pattern would become the unique "fingerprint" (unique ID) to hunt down if a market maker may conveniently flip between categorizing a customer as counterparties (to escape fiduciary responsibility), and vice versa (e.g. relatively easier to qualify for the Volcker Rule exemption). Also, this is the only way CAT NMS Plan may provide sufficient safeguards or enforceable policies to assure that the same Firm Designated ID would not be used for multiple customers.

Besides, it is possible for a customer to concurrently be dealing with the same or another division of the same financial institution as counterparty. However, possible "conflict of interests" (a form of violation) may arise in such situation. Thus, tagging trade patterns as "fingerprint" (unique ID) to a customer and/or counterparty is part of CAT's responsibilities because it helps facilitate an exam by the regulators to check for consistency that supports their prosecution works. Again, CAT's purpose is to enable "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct".

Our short answers to the various questions are – Q.135: Yes, we believe allowing broker-dealers to report a Firm Designated ID to the Central Repository is more efficient and cost-effective; Q.136: No, insufficient details; Q.137: No it shouldn't affect accuracy; Q.138: No, it shouldn't affect accessibility; Q.139: No, it shouldn't affect timeliness; Q.140: It will be difficult for the Central Repository to utilize a Firm Designated ID for each account (e.g. to efficiently, reliably and accurately link orders and Reportable Events to a customer), but it is essential and part of CAT's responsibilities to tag trade pattern as "fingerprint" (unique ID); Q.141: No, insufficient details; Q.142: No, there are insufficient safeguards or policies.

**Questions 143:**
We believe the CAT NMS Plan should mandate that broker-dealers to provide LEI as part of the information used by the Plan Processor, but under a condition – such LEI information should be ready available at respective systems where CAT may install appropriate "sensors" to capture the data, rather than incurring additional efforts by broker-dealers to submit LEI data for CAT (see Q.40, Q.51, Q.85, Q.87, Q.88, and Q.89).

**Questions 144:**
No, it shouldn't affect the security and confidentiality. However, how regulators may use reverse-engineering approach to cross tabulate with other intelligence to decipher masked-PII confidential customer information is indeed outside scope of CAT's responsibilities in our opinion.

**Questions 145-148:**
If CAT may adopt our suggestion to tag trade pattern as "fingerprint" (see our respond to Q.135 - unique ID), then it doesn't matter to have the initial set of Customer Account Information and Customer Identifying Information being updated thereafter. We believe this approach is more efficient and cost-effective than the Rule 613 approach in kick starting the CAT launch date. It will always be difficult for the Central Repository to ingest the Customer Account Information and Customer Identifying information, and any updates thereafter. However, it is essential and part of CAT's responsibilities to tag trade pattern as "fingerprint" (unique ID) to efficiently, reliably and accurately link orders and Reportable Events to a Customer.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Questions 149:**

No, it shouldn't affect the security and confidentiality. However, how regulators may use reverse-engineering approach to cross tabulate with other intelligence to decipher masked-PII confidential customer information is indeed outside scope of CAT's responsibilities in our opinion.

**Questions 150-155:**

Yes, instead of using Customer's name, address, date of birth, and ITIN/SSN, CAT may consider masking these data points plus adopting our suggestion to tag trade pattern as "fingerprint" (see our respond to Q.135 - unique ID). Such approach will improve the completeness of the CAT audit trail and offers better confidentiality protection. Data quality and the related reliability of CAT audit trail will depend on the effectiveness of pattern recognition techniques to uniquely identify customers/ counterparties with minimum false positives / false negatives. Accessibility and timeliness shouldn't be affected and it offers better confidentiality protection.

Just relying on any "masked" data other than PII to identify a Customer wouldn't be a more efficient and cost-effective way to identify Customers, but it will, if combine with our suggested approach to tag trade pattern as "fingerprint". It will always be difficult to efficiently, reliably and accurately link orders and Reportable Events to a Customer, but that's part of CAT's responsibilities.

**Questions 156-160:**

We feel the "Account effective date" is reasonable and good enough for analytic purpose. It deems more efficient and cost-effective than requiring the Rule 613 approach. Rule 613 approach adds little benefits to require "the date the Customer's account was opened" when the difference between the two dates are generally insignificant. Again, per our respond to Q.87, the messy whole essentially can outperform exact, accurate subsets. No, it won't affect the quality, accuracy, completeness, accessibility or timeliness of the CAT data. The Plan Processor's ability to link a Customer's account with the Customer would be indifference if CAT adopts our suggestion to tag trade pattern as "fingerprint".

**Questions 161:**

The modification and cancellation instruction are of equal importance to other CAT Reportable events. Indeed, the person giving the modification or cancellation instruction is vital information for market surveillance purpose (e.g. hunt down who may be manipulating the market with spoofing). Besides, the market is vulnerable to attack, such as the 'flash crash', from anyone and anywhere in the World. CAT audit trail shouldn't leave out detail information about who initiated a modification or cancellation of an order. Therefore, we are strongly against the said exemption, because it will merely provide analysts with a "drop down list or check box" of – order initiated by either: Customer / a Broker-dealer / Exchange. The exemption will be a huge detriment to the effectiveness of market surveillance.

**8. Order Allocation Information**

**Questions 162:**

An allocation report that includes the Firm Designated ID when an execution is allocated in whole or part is undeniably useful information for analytic purpose. However, the availability of the report may not be at timely as real-time censoring of system where an execution is allocated. We also foresee challenges to link account and

subaccount information to which an execution is allocated. Therefore, the said exemption should be re-considered. It is not about the efficient and cost-effective concern, but the exemption is not a good alternative (in the context of facilitating market surveillance) in lieu of the requirement in Rule 613.

**Questions 163-167:**
Yes, it would affect the completeness of CAT data, as well as the accessibility and timeliness of allocation information. No, it should not affect the security and confidentiality of CAT data. We feel the Allocation Report Approach described by the SROs is not feasible (see Q.162).

**9. Options Market Maker Quotes**

**Questions 168-170:**
Per our study of Cost Survey Report on CAT Reporting of Options Quotes by Market Makers, we acknowledged that the "quote-driven" listed options markets are different from the equity stock markets that are typically "order-driven". The alternative approach of permitting Options Exchanges to report Options Market Maker quotes to the Central Repository may "appear to be sensible" because the market is reliant on intermediaries. However, it is detrimental to achieving the objective of capturing a "complete audit trails" of all the market activities. Inconsistency in data collection methods between of options market and others may lead to risk of failure similar to the Citi Incomplete Blue Sheet Data case. Another risk of the "alternative" approach is: it overly discounted/ distorted signals for market surveillance and manipulation detection purposes.

Per our suggestion in Q.40, Q.85, Q.87, Q.88, and Q.89, by picking up "audit trails" where CAT Reporters have "left behind" (tagging information as data were originated) at any digitized mediums where market makers may initiate/ withdraw their quotes, this won't be invasive at all to the CAT Reporters (indifference in cost savings). Given said that, CAT will have to accept the fact that today's options market makers quotes are filled with nuances, such as: a single message may contain multiple instruments and multiple types of requests; withdrawals are not explicitly linked to a specific quote message. CAT is not meant to fix all the market structure problems, nor forcing "quote-driven" markets to convert into "order-driven" markets.

The advantages of our suggested "tagging/ censoring" approach are:

- Consistency – market makers' order/ quote should "consistently" be censored regardless of options market or others;
- Completeness – no worries of receiving piecemeal information from different trading desks (options market versus others) using different data submission methods;
- Timeliness – "quote sent time" would already be captured, and no worries of reconciling in-fight conditions between new quotes and withdrawal requests due to data transmission delays and other interruptions, such as unexpected congestion in data traffic.

To deal with the mentioned nuances in the "quote-driven" listed options markets, CAT may encourage and incentivize the private sector to come up with creative "engineering" approach to "sort out" things such as: a single message may contain multiple instruments and multiple types of requests; withdrawals are not explicitly linked to a specific quote message. The techniques would likely involve "optimization" and cross-tabulation of data

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

about the market imbalance situation. Indeed, "censoring" gives regulators the complete picture to reconstruct all events, despite certain information are not make transparent in the Exchanges for listed options markets.

**11. Regulatory Access**

**Questions 190-195:**

No, we think there are insufficient details on how a regulator would access, use and analyze CAT data (see our suggestions in Q.59). The "Functionality of the CAT System" Section is definitely insufficient to address all regulators' end-user requirements. Per our respond to Q.81, CAT presents an opportunity where all the regulatory data submission and/or metrics reports (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) may be consolidated to cross-tabulate into insights for better market surveillance. We demand to see a better "roadmap" of how these data, metrics reports, and/or audit trails may be consolidated.

Reference to our respond to Q.40, CAT NMS plan ought to embed in its design a dynamic analytical framework. For example, having (i.e. research, and/or investigate collectively with other collected insights/ augmenting facts).

Permitting regulators to analyze sets of data within the CAT using applications or programs selected by the Commission, is NOT the same as our suggested approach to "embed in CAT design a dynamic analytical framework" per Q.40 and Q.59's responds. Our expectation is to have sensors directly conduct real-time analytics over streamed data where it was originated. Hence, any irregular activities are spotted timely to become intelligence for further pursuit, which is different from any after-the-fact investigations.

CAT should under ABSOLUTELY NO CIRCUMSTANCE (including BCP/DR) allow anyone the option to download the "entire" data sets, because this essentially opens a "backdoor" to significant security risk. No advantage would be big enough to justify such risk, because we consider anyone who possesses the entire CAT data sets may pose tremendous threat to the U.S. financial stability if they turn evil!

Per our respond in Q.60, the online targeted query tool and user-defined direct queries and bulk extracts methods will not enable regulatory staff to use the data. These query and/or bulk extract methods are "generic" and not fit-for-the-purpose of financial market surveillance. Despite direct queries may be "user-defined", they aren't "engineering" approach to unveil whether trades may be synthetically created to bypass control (see Q.59 – point (e)(ii) and (iii)).

The T+5 Schedule for regulatory access is useless in terms of effective market surveillance in prevention of threats to the U.S. financial stability. Huge loss can be accumulated within split-second. Market collapse does not take more than one day (also see our respond to Q.89). We advocate for "real-time" stream analytic. System to facilitate market surveillance must also adopt an "engineering" approach to address the problems (e.g. abusive use of financial engineering techniques to synthetically create trades/ derivatives to bypass controls) as we have stated in Q.59 – point (e)(ii) and (iii). Data Boiler has a patent pending invention to convert/ sequence trade streams and using a financial engineering approach to recognize patterns. It'll enable ultra-fast analysis up to 50 milliseconds.

Whereas the T+5 delayed CAT data would be appropriate to be consumed by the public under registered access, "if" they are properly being scrambled, aggregated, and/or partially released for selected historical "prosecution" events (e.g. London Whale, LIBOR manipulation) and/or major financial crisis (e.g. LTCM, 2008 mortgage

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

meltdown). Per our respond to Q.63 and Q.64, it will facilitate the private sector (i.e. the academic and financial technology companies) to discover better approaches (as compared to sole effort by the regulators) to market surveillance. Hence, there will be continuous innovations (like there are FREE music libraries for anyone who can contribute to better music plagiarism detection) to help foster financial stability.

**Questions 201:**
No, we don't think CAT NMS Plan has appropriately encourages or incentivizes the Participants and the Plan Processor to incorporate new technology and to innovate. Throughout the CAT NMS Plan or any related documentation, they reflect a traditional approach to build a data vault, rather than developing a modern ecosystem. The plan overemphasized on storage, which is majoring in the minors.

In our opinion, a modern ecosystem would censor any data exhausts and conduct stream analytics in real-time at the original data sources. The ecosystem would incorporate an analytic framework, like what we have suggested in Q.59, to allow noises to be filtered, and signals to be escalated and route to the respective decision makers for risk treatments at each stage of the analytic process.

We welcome this effort to solicit public comments, because it gives us an opportunity to showcase our patent pending inventions and other innovative technology ideas for the Participants and the Plan Processor to consider the fit-for-purpose. Nevertheless, to drive continuous innovations and ensure properly incorporate of new technologies, we recommend granting certain limited access/ delayed data to the crowd, it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight (see our respond to Q.63 and Q.64).

**Questions 202-203:**
No, per our respond in Q.60, the online targeted query tool and user-defined direct queries and bulk extracts methods will NOT enable regulatory staff to use the data. These query and/or bulk extract methods are "generic" and not fit-for-the-purpose of financial market surveillance. Despite direct queries may be "user-defined", they aren't "engineering" approach to unveil whether trades may be synthetically created to bypass controls.

**Questions 204:**
Logs of "who accessed to what" are generally important to the Plan processor and the Operating Committee in ensuring a smooth CAT operations, while regulatory access may related to confidential investigation that is too sensitive for the Plan Processor and Operating Committee to know. Thus, there should be system logics to alert Plan Processor any abusive usage of CAT (e.g. attempts to steal the entire CAT datasets – see Q.190-195), as well as sharing "aggregated" log information with the CAT Operating Committee.

**12. Security, Confidentiality, and Use of Data**

**Questions 206-211:**
We think there are rooms for improvement for the data security requirements set out in Appendix D. CAT information security program may want to include details as we have suggested in Q.41.

Besides, if CAT is going to embed with an analytical framework as we have suggested in Q.59, then forensic analysis of malware (e.g. look for evidence of code injection, check for signs of a rootkit, etc.) for in-memory usage

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

is necessary. Also, CAT's privacy and security programs should consider eradication (removal of data as soon as it has been transmitted/ used) and obfuscation (e.g. traffic flow security to mask patterns, de-identification) as appropriate.

To safeguards and prevent the misuse of CAT data by employees or agents of the Participants or other persons with access to the Central Repository, we suggest the followings: (1) Segregation of duties - uses micro-tasking to farm out and distribute the work to various functional units so no one particular unit would have the full big picture; (2) Vulnerability scan and audit of processes to identify system nuisances that can cause a security breach; (3) Precognitive fraud prevention. (See a related whitepaper)

Regarding the distinction between regulatory and non-regulatory staff, we think additional details may be needed if CAT can consider granting certain limited access/ delayed data to the public (academic, financial technology companies, industry associations and other related bipartisan organizations), it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight (see Q.63 and Q.64).

Regarding the distinction between the appropriate and inappropriate use of CAT data for commercial purposes, please refers to our respond in Q.212.

For data access and breach management provisions described in Appendix D of the CAT NMS Plan, we encourage CAT security policy to take a step back and consider the organization as a chain of capabilities. Who has the knowledge across most of these capabilities? Who are in control over most of the resources? Who can easily steal using their authorities, knowledge, and controlled resources? Middle and senior management of CAT indeed are most vulnerable. They face high temptations of wrong doing. The provisions may want to include additional details to address this issue.

In terms of who should have the responsibility to monitor for and prevent the misuse of CAT Data, this question will depend on the CAT management team dynamics and the person's ability instead of someone's title. From the perspective of RACI (a management framework), the information security team under leadership of the Chief Information Security Officer (CISO) should definitely have the monitoring "responsibility", while the CEO and/or Chief Compliance Officer would ultimately own the "accountability". During the monitoring process, there may be experts and other departments that the CISO should "consult" and/or "inform". A well-designed workflow and organization structure will help prevent the issue of "everybody owns nobody owns"!

**Question 212:**

It's not about what particular types of raw data would be appropriate or inappropriate to use for commercial purposes by Participants. The problem is: will participants use these data (regardless of it being a subset or its entire population) to develop algorithm(s) that can bypass surveillance controls. It will become threat(s) to the U.S. financial stability if they turn evil! "Introduce randomness" can be an effective way to resist pattern recognition and prevent the abusive situation.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Questions 213 and 214:**

PII should properly be safeguarded, while it will always be a race with any reverse-engineering approach to decipher who the person may be. Data obfuscation to mask, leave out certain details, or introduce randomness may help shield the person's identity, but nothing will be absolutely "bullet-proof". We believe the protection and security of PII in CAT is good enough, while using trade pattern as "fingerprint" (see our respond to Q.128-133 and Q.135-142) may supplement and/or replace the requirement to collect PII to the Central Repository.

**Question 215 and 229:**

For monitoring of threats, attacks, and anomalous activity, we recommend the Plan Processor to work with The National Cybersecurity & Communications Integration Centre (NCCIC).

**Questions 216-221:**

No, there may also be concerns for data-in-use and dispose (see our respond to Q.206-211). Also, besides the encrypted connection requirement, CAT security program should also consider: bare-metal cloud, size of counter for scalability, secure in-band / out-band management, certificate generation, hardware tempering controls, … as well as using eradication and obfuscation techniques where appropriate.

Memoranda of Understanding or Interconnection Security Agreements are "formality" on the paper that does little to no benefit to deter away hackers and/or prevent internal security breach. Using "chaos" to add complexity and use multi-factors authentications to comprehend the security controls may be more effective to discourage wrong doing.

NIST family of guidance and any other "industry best practices", are "nice-to-have" reference documents for the CAT security program. A static checklist or widely adopted industry standard may not necessary be the best security defense for CAT. No one or no reference guidebook should restrict the creativity to custom design a unique security program for CAT.

**Questions 222-224:**

We agree the Plan Processor should categorize, include labeling the confidentiality/ sensitivity of, data. However, why NIST SP 800-53, why not ISO-27001 or other types of third party risk assessment?! The emphasis shouldn't be favoring on a particular prescribed standard. We are okay to consider the Federal Information Processing Standard ("FIPS") 199 or NIST SP 800-60 as a baseline framework, but the key is: CAT needs independence privacy and security assessment at regular intervals. The assessment will include: vulnerability scan and identifying system nuisances that can cause or already caused privacy and security issues.

**Question 226:**

Independent assessment in the form of audit (3rd line of defense) should be carried out at least annually, while it is an on-going process to have privacy and security frameworks build into the design of CAT "right from the get-go".

**Questions 227-228:**

The requirements for conducting ad-hoc penetration testing (vulnerability scan) by a reputable third-party deems reasonable. Application security code audit may touch appoint sensitive codes of CAT, thus it'll be better NOT to allow a single third-party to perform audit over all CAT codes, as well as additional measures to ensure the code

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

auditors won't turn evil in using his/her access to CAT codes to develop algorithms that may become threats to the U.S. financial stability (self-compromise the security or confidentiality of CAT data).

**Questions 230:**
Verifying CAT reporters connect to the repository is definitely needed, but there are insufficient details about: (1) how to prevent ingestion of junk data intentionally or inadvertently by CAT reporters and/or their logins were compromised by hackers using social engineering techniques; (2) where there may be any single point of failure possibilities caused by system nuisances.

**Questions 231:**
Instead of a direct answer, we only have a sided comment here: too much detail to determine whether a regulator's queries are shielded from the Plan Processor indeed may attract hackers' interest.

**Questions 232:**
We believe CAT reporters would already have annual audit of their data security, there is no point to carry out separate review unless there is indication of weak controls at a particular firm.

**IV. Economic Analysis - D. Baseline**

**Questions 246 and 247:**
The fragmented nature of current data sources does pose significant challenges to regulators seeking complete data. Thus, we do not believe the HYPE that there will ever be any single source of truth or golden standard of all the complete and accurate data.

In big data, we prefer and believe the most original the data, the slighter the chance of introducing any noises and/or discounting the signals during the ETL process. According to Viktor Mayer-Schönberger and Kenneth Cukier in their book – Big Data: A Revolution That Will Transform How We Live, Work, and Think, "any particular reading may be incorrect (in a messy dataset), but the aggregate of many readings will provide a more comprehensive picture." Essentially, the messy whole can outperform exact, accurate subsets.

**E. Benefits**

**Question 258:**
The first biggest uncertainty of CAT is: whether the plan will adopt our suggestion to incorporate a "real-time" analytical framework (see our responds to Q.40 and Q.59) in the CAT design.

The second biggest uncertainty to CAT implementation is posed by: the future decision pertaining to the "per message based" funding model versus our suggested "successful red-flag" method.

These two uncertainties will dramatically affect the assessment of potential benefits. If CAT (in its current form) be implemented without a proper analytical framework, we envisage there will be absolutely ZERO BENEFIT to be realized out of building a gigantic vault that its data won't be effectively used for real-time market surveillance purpose. If imposing toll charge on all trade messages that goes through CAT, it would essentially be like a "financial transaction tax" that is unjustifiable and detrimental to the overall health of the U.S. capital markets (see our responds to Q.65-67, 69-70 and 72).

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Questions 259-260:**

We applaud the efforts to capture both the trade execution information as well as order level details in CAT, so "the inclusion of the data fields" part is welcoming. In fact, we disagree with "the centralized data source" part because we do not believe the HYPE that there will ever be any single source of truth or golden standard of all the complete and accurate data.

If CAT's objective is: "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct", then there should be roadmap of how collected data and metrics (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) may be consolidated to cross-tabulate into insights for better market surveillance (see our respond to Q.79). It is disappointing to know CAT have concluded that certain OATS/ EBS information is not included (e.g., OATS data fields that allow off-exchange transactions to be matched to their corresponding trade reports at trade reporting facilities and certain EBS elements).

**Question 263:**

The answer to this question depends on how "accuracy" is defined. If "accuracy" means transforming data into standardize format (disregard of how invasive it will be to the CAT-Reporters for all the ETL and the data submission requirements), then the answer is "yes". If "accuracy" means preserving signals and avoiding possible introduction of any noises and/or discounting the signals during the ETL and data submission processes (as oppose to our non-invasive suggestion of using "sensors" to pick up "audit trails" where CAT Reporters have "left behind" at the original sources – see our responds to Q.85, Q.87, and Q.88), then our answer would be "NO".

**Question 273:**

Yes, but not enough. We feel the Plan Processor may be able to develop expertise in linking data more efficiently than the regulatory staff members from each entity could on their own because of technical specialization. However, the technical experts resided at the Plan Processor will always need to collaborate with regulatory staff members and the industry to understand the financial "engineering" aspect of why certain data should/ shouldn't be linked (or be analyzed together/ separate).

As a side note, the word "link" indeed seems old school, like a traditional RDBMS SQL "JOIN" syntax. Although we recognize the usefulness of SQL JOIN function, but modern technologies such as semantic/ schema-less database and neural networks may offer alternate ways to model and analyze data. We want CAT NMS Plan to aware of these technology innovations and explore the possibilities in using them where they may be appropriated.

**Question 275:**

There are flaws to the approaches of CAT Customer and Reporter Identifiers, thus it has little benefit to improve the accuracy of information. See our suggestions in Q.128-133 and Q.135-142.

**Question 278:**

We believe "direct access" would help reduce the number of ad-hoc data requests. Whilst the question shouldn't be about sufficient capacity and functionality provided by the Plan. The core issue is: there is no point in keeping any data for the shake of keeping it. We see the plan lacks a dynamic analytical framework embedded in the design. The T+5 Schedule for regulatory access is useless in terms of effective market surveillance in prevention of

threats to the U.S. financial stability. It discourages the end-users to effectively use the system to facilitate the monitoring market behavior and identifying misconduct in "real-time".

**Question 279:**
Afraid the service level agreements (SLA) for "8am Eastern Time the following day" submission requirement and the T+5 Schedule for regulatory access would become service level "arguments" between the regulators and the industry. The CAT requirements aren't matching up with today's surveillance's need of "real-time". In case of a flash crash or major financial crisis, no one would expect the regulators to wait so many days before they can get hold on the essential data?! Although CAT offers the regulators on-demand query of delayed data that saves them multiple trips to request data from the financial institutions on normal days, in fact that does not necessary mean timeliness improvement. Again, we advocate for "real-time" censoring of data at the original sources (see Q.263).

**Question 280:**
A reasonably designed surveillance systems for modern day's needs require "real-time" intelligence instead of an "encyclopedia" of delayed data locked in a secured vault. Also, system to facilitate market surveillance must also adopt an "engineering" approach to address the problems (e.g. abusive use of financial engineering techniques to synthetically create trades/ derivatives to bypass controls) as we have stated in Q.59 – point (e)(ii) and (iii). Therefore, we do not believe CAT NMS Plan (in its current form) will facilitate the ability of each national securities exchange and national securities association to comply with the requirement in Rule 613(f).

Besides, market surveillance and/or industry oversight shouldn't be just relying on the hands of the national securities exchange and national securities association. We encourage the plan to consider leveraging the crowd to unleash its powers (see Q.201). By granting certain limited access/ delayed data to the crowd, it'll enable more creative approaches to market surveillance, foster industry collaboration (between regulators, exchanges, the industry, and the private sector), as well as augmenting the regulatory efforts for a more holistic industry oversight (through a surveillance network or utility model like what we have suggested in Q.59).

**Question 281:**
If CAT is not going to perform any "red-flagging" of suspicious activities (see our suggested approach in Q.59) and there lacks incentive (see our suggested "successful red-flagging" funding model in Q.65-67, 69-70, and 72) for the CAT to come up with an effective algorithm to discern permissible versus suspicious/ prohibited activities, then how would it be possible to facilitate the ability of regulators to conduct risk-based examinations! Quality of submitting CAT data is different from (or there's no proven correlation with) an entity's riskiness in their trading activities. Whereas the amount of "red-flagged" trade messages will be an effective gauges / basis for the regulators to rely on in conducting risk-based examinations.

**Question 282:**
It will slightly improve the efficiency of regulators' enforcement activities because CAT will save them multiple trips to request data from the financial institutions. However the benefit is minimal because it won't help the regulators to identify misconduct and/or recognize pattern of market manipulations in real-time. Huge loss can be accumulated within split-second. Market collapse does not take more than one day. Again, we recommend CAT to consider adopting our suggestions in Q.59.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Question 284:**
Overall, VERY MARGINAL IMPROVEMENTS but not a lot, that makes CAT NMS Plan (in its current form) not justify. To effectively and efficiently deter violative behaviors, CAT should consider our suggestions, particularly in Q.59.

**Question 285:**
There appears to be no analytic framework embedded in the design of CAT, thus there is ZERO improvement to reduce the percentage of activities that generate false positives (i.e., detection of behaviors that are not violative) and/or reduce the percentage of activities that are false negatives (i.e., not detecting behaviors that are violative).

Although the Commission's assessment may favorably argue for improvements contributed by CAT, assuming the quality of surveillance output is improved by the quality of data inputs. However, the definition of "quality" shouldn't be about the HYPE of a standardized golden source of data locked in a "vault", but the timeliness, relevancy and completeness of data in the eyes of the users/ analysts. CAT NMS Plan (in its current form) does not satisfy our quality expectation. Analysts need sensors, not encyclopedia. A good decision, made now and pursued aggressively, is substantially superior to a perfect decision made too late. Thus, the plan should mandate censoring of "real-time" information (rather than "8am Eastern Time the following day" submission and the T+5 regulatory access), as well as consider our suggestions in Q.59.

**Question 286:**
There are BIG GAPS between our expectation and the Commission's "overly optimistic" assessment of the economic effects of the improvements to surveillance, examinations and enforcements from the CAT NMS Plan. This is mainly due to the plan lacks an analytical framework embedded in the design. See our responds to Q.259-260, 263, 273, 275, 278, 279, 280, 281, 282, 284, and 285 for detail explanations.

**Question 287:**
We feel the improvements in efficiency and effectiveness of regulators conducting analysis and reconstruction of market events are MINIMAL, or not sufficient. This is mainly because the plan is majoring in the minors (i.e. overemphasis on storage, and not enough coverage of pattern recognition).

**Question 288:**
No, we think CAT NMS Plan (in its form) would do LITTLE help to facilitate market analysis and research. The plan should mandate "real-time" (access to intra-day feeds) rather than the "8am Eastern Time the following day" submission and the T+5 access schedule, in order to be useful for anyone to have a better understanding the dynamics of today's trading markets (e.g. flash crash and crisis could happen in lighten speed – milliseconds if not microseconds).

**Question 289:**
No other features of the CAT NMS Plan have caught our attention to consider them unique. Even the custom query functions deem generic, because they aren't financial engineering approach to curb issue as mentioned in our respond to Q.59(e)(ii) and (iii). Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities. It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. We highly

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

recommend CAT to consider using our patent pending invention, until then we'll see a possible increase in the benefits of CAT NMS Plan.

**Question 290:**
No, the provisions about future upgrades are infrastructure related, rather than "quality" improvements in the sense of timely insights to alert regulators of irregularities and/or better signals detection. Again, per our responds in Q.93-100, anything that deviates from the purpose of facilitating market surveillance cannot and should not be recognized as "quality", no matter how glamorous the CAT "Vault" / "Encyclopedia" may be.

**Question 291:**
No, there are insufficient incentives to reduce reporting errors. The approach to require data submission indeed is a setup for failure to the CAT Reporters because of its invasiveness and cumbersome ETL requirements. Instead of punishing the industry with fines (similar to the Citi Incomplete Blue Sheet Data case), why not consider our suggestion in Q.40, Q.59, Q.85, Q.87, Q.88, and Q.89? By picking up "audit trails" where CAT Reporters have "left behind" (tagging information as data were originated) at any digitized mediums through the use of "sensors", this won't be invasive and it'll preserve the complete audit trails without risks of discounting signals or introducing noises during ETL process.

**Question 293:**
YES, we ABSOLUTELY AGREE that communication of data feed delays for public consumption is BENEFICIAL to the operation and effectiveness of the CAT. Market surveillance and/or industry oversight shouldn't be just relying on the hands of the national securities exchange and national securities association complying with the requirement in Rule 613(f). CAT NMS Plan should definitely leverage the crowd to unleash its powers.

By granting certain limited access/ delayed data to the crowd (public disclosure), it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight. Please also see Q.201 and Q.280 regarding possible crowd access benefits.

There should be no additional cost to obfuscated/ aggregated data for public consumption, because such functionalities should already be included and built-in to cater for regulatory access as well. We estimate the costs to enhance the CAT's information security program to support public disclosure using a "multi-tiers" access approach as we have suggested in Q.63 and 64 to be less than half a million. For this minimal cost, it is definitely worth it. Besides, CAT as an "industry utility" is obligated to address the question of civilian oversight of the finance industry/ regulatory actions.

**Question 300:**
Please refer to Q.258 for the two biggest uncertainties (real-time and funding model) or factors relating to the operation and administration of the Plan that will dramatically affect the assessment of potential benefits. We appreciate the Commission's consideration of these critical factors in determining whether to approve (or in our opinion, to revisit the key fundamentals) the Plan.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

## F. Costs

**Question 301:**

No, the Commission's assessment of the potential cost per Table 6 is largely off the mark. Calculation issue is one thing, while the bigger issue is: the cost of building a "Vault" (CAT in its current form) is very different from having an analytical framework embedded in the CAT design to facilitate market surveillance.

From a financial number perspective, $40+ million in CAT implementation cost and $100+ million in CAT ongoing cost may "appear reasonable", but the devils are in the details. We cannot follow why there maybe any savings in surveillance cost by participants ($87.7 million a year for CAT versus $147.2 million for Current). Also, participants' data reporting cost is grossly underestimated.

We feel "the total implementation cost estimate of $17.9 million" is not too far off, while we expect more money for the consultants ($770,000 in current budget) at the initial stage than the full-time employee costs ($10.3 million) for operational, technical/development, and compliance-type functions. The "rip-off" is indeed related to the CAT "ongoing data reporting cost" – why should participants pay $720,000 annually to legal and consulting costs after CAT goes live, why shouldn't things be automated instead of incurring another $7.3 million annually in full-time employee costs for the participants?! This $14.7 million a year in total for the participants does not include the "per message" toll charge in CAT funding model. We think the actual cost is probably going to be more than the estimate of $14.7 million a year, but regardless, this cost represents how "invasive" the CAT NMS Plan (in its current form) will have an effect on the participants' daily operations. This cost would have been saved if CAT adopts our suggestion to use "sensors" in picking up "audit trails" where CAT Reporters have "left behind" at any digitized mediums (see Q.40, Q.59, Q.85, Q.87, Q.88, and Q.89).

Given CAT (in its current form) lacks analytical framework embedded in the design (i.e. does not have a systemic way to "red-flag" suspicious activities) we cannot imagine how surveillance cost may be reduced by more than 40% (1 – ($87.7 million / $147.2 million)) while CAT improving the effectiveness of surveillance?! By the way, we want to remind everyone that the CAT $87.7 million a year in surveillance cost does not include the "per message" toll charge in CAT funding model.

"Participants could realize efficiencies from having data standardized and centrally hosted that could allow them to handle fewer ad hoc data requests" is merely HYPE relates to the process of "after-the-fact" investigations. It has nothing to do with an effective surveillance program for real-time monitoring of market behavior and identifying misconduct. Also, the calculation seems "artificially' making "CAT implementation cost" of $41.1 million + "CAT ongoing" cost of $102.4 million closely approximate the current costs of $154.1 million, whilst the most significant cost – the per message "toll charge" in CAT funding model is omitted.

**Question 305:**

If the Plan (in its current form) is approved, Participants' data reporting costs will significantly increase, whilst surveillance costs will definitely not decrease.

Reference to our respond in Q.285, although the Commission's assessment may favorably argue for improvements contributed by CAT, assuming the quality of surveillance output is improved by the quality of data inputs. However, the definition of "quality" shouldn't be about the HYPE of a standardized golden source of data locked in

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

a "vault". Whatever "golden standard" that CAT may strike to achieve, it has no correlation with the surveillance costs. They aren't improving the timeliness and relevancy of intelligence for market surveillance purpose. CAT NMS Plan (in its current form) does not satisfy our quality expectation.

See our respond to Q.301 for elaborated explanation.

**Question 308:**
$50,000 to $180,000 per year sound reasonable, but why uses "service bureau" when CAT can adopt our suggestion to use "sensors" in picking up "audit trails" where CAT Reporters have "left behind" at any digitized mediums (see Q.85, Q.87, Q.88).

**Question 310:**
Please refer to our respond in Q.335.

**Question 318:**
Whether it is under approach 1 to submit data in existing format then converting data into uniform format in a second step, or ask for data in a mandatory format per approach 2, the cost should intuitively be like: having the expense incur from the left pocket versus the right pocket (i.e. indifference or of no significant difference). The only logical explanation we can think of is: CAT will have to use an expensive Enterprise Service Bus (ESB) to accept data in multiple formats under approach 1, while approach 2 does not. The ESB cost is assumed to be borne by CAT Reporters.

We can't imagine if CAT may ask the entire industry to consider replacing ticker symbols for the shake of capturing audit trails?! CAT shouldn't be, and has no authority to, change or invade the existing market practices/ processes. After all, we do not believe the HYPE that there will ever be a "golden" unified "single source of truth". This is because, by the time a common standard is adhered, value of the data subsided to almost worthless in the context of market surveillance.

**Question 323:**
Timing to retire duplicative reporting systems should be "now or never". We expect CAT to have milestone target of sun-setting OATS on Day 1 when CAT goes live.

**Question 325:**
Asking the question of "what costs that broker-dealers would face in accomplishing a period of duplicative reporting" is like: assuming that it is totally fine to be invasive in troubling the industry for unnecessary data submission. We despise that kind of attitude, and expect the broker-dealers community won't be happy about any CAT arrangements that give them no benefit.

**Question 327:**
No, we do not believe that the CAT NMS Plan (in its current form) would deliver additional cost savings from sources other than the retirement of duplicative reporting systems and a reduction in the amount of ad-hoc data requests to regulated entities.

If CAT may adopt our suggestions (particularly in our respond to Q.59), then it is possible to save the U.S. economy more than $22 trillion (see a study by the GAO) by avoiding the next crisis. What also needs to be changed is the regulatory exam and/or market surveillance practices. Instead of the current "passive" approach to investigate issues/ losses after-the-fact, we envisage the regulators to monitor queue of early warning signals and consider "proactive" risk prevention in "real-time". Reality is: huge loss can be accumulated within split-second. Market collapse does not take more than one day. If CAT is truly meant to enable effective market surveillance, then this is a "must have", not a "nice-to-have" functionality.

**Question 331:**
Please refer to our respond in Q.318.

**Question 332:**
Treating all costs related to development of the Plan as sunk costs may sound conservative. However, we think this is indeed a preferred approach based on our review of CAT's fit for its purpose (i.e. the cost of building a glamorous "vault" is of ZERO value to the purpose of facilitating market surveillance). See Q.258.

**Question 335:**
It is a significant cost-driver for the total industry (Outsourcers and the others) to provide customer information to the central repository. Keep in mind that customer information include PII confidential data, moving sensitive data require extreme precaution (including encryption, obfuscation, etc.) The more PII data that are "in-motion" and/or store at multiple systems, the bigger the risk and the higher the cost it will be.

**Question 336:**
We have suggested an alternate ways in which this data can be made available to regulators that would prove less costly to the industry and the investors per our respond in Q.40, Q.59, Q.85, Q.87, Q.88, Q.89, and Q.90.

We recognize that "censoring" may involve challenges to analyze unstructured data, but there are technology solutions, such as NoSQL document store/ semantic database to help. Also, per our respond in Q.51, data analysts prefer flexibility to slice and dice data, so matters can be analyzing from multi-dimensional angles in discerning truths, or we refer it as "object-oriented analysis and design" (OOAD). We can adopt data in 'any digitized form' so long as it can be read by computers, rather than emphasis on Extract/ Transform/ Load (ETL) to convert data in particular format to feed certain SQL database's structure.

The key advantage of our suggestion is: CAT may effectively use a "divide and conquer" approach to deal with only the relevant signals at each "stage" (i.e. clear violation, white list/ black list, detection engine, and backstop QA).

**Question 337:**
Please refer to our respond in Q.318 regarding the using of listing exchange symbology.

In addition, as we have mentioned in Q.51, modern technologies are capable of analyzing data beyond just text or structured data. There are many advantages in analyzing multimedia and/or voice recognized data than interpreting text or structured data alone. Multimedia describes messages in richer 'colors', whereas 'treated data' after the ETL process may discount a lot of useful signals. Our point here is: in the future, there may no longer be a need of "standardized" symbology, while traded product will better be identified through Blockchain and/or other

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

types of audit trails. In the meantime, in order to minimize cost and invasiveness to the industry, we suggest CAT to accept whatever existing tickers, CUSIPs, or any notations/ symbology "as-is".

**Question 341:**
There may be some savings from a reduction in the number (and ultimately the cost) of data requests as a result of regulators having direct access to CAT Data. In our opinion: about 5%, but definitely not over 10%. Per our respond in Q.301, we cannot imagine how surveillance cost may be reduced by more than 40% (1 – ($87.7 million / $147.2 million)) while CAT improving the effectiveness of surveillance?! The Commission's analysis of the potential cost savings in here is simply HYPE!!

**Question 346:**
ABSOLUTELY NOT, "Manual data entry option" exposes CAT to significant information security risk, such as: ingestion of junk data intentionally or inadvertently by CAT reporters and/or their logins were compromised by hackers using social engineering techniques (see Q.230). It also affects data accuracy and a whole host of problems related to updates and correction.

If CAT may adopt our suggestion (see Q.87) to tag information as they were originated, there should not be worries about any updates and/or correction to CAT Reportable events. In big data, we prefer and believe the most original the data, the slighter the chance of introducing any noises and/or discounting the signals during the ETL process. According to Viktor Mayer-Schönberger and Kenneth Cukier in their book – Big Data: A Revolution That Will Transform How We Live, Work, and Think, "any particular reading may be incorrect (in a messy dataset), but the aggregate of many readings will provide a more comprehensive picture." Essentially, the messy whole can outperform exact, accurate subsets.

### G. Efficiency, Competition, and Capital Formation

**Question 368:**
NO, we disagree with the Commission's analysis of the Plan's effects on the efficiency of market regulation and oversight. Per our respond in Q.301, we cannot imagine how surveillance cost may be reduced by more than 40% (1 – ($87.7 million / $147.2 million)) while CAT improving the effectiveness of surveillance?! "Participants could realize efficiencies from having data standardized and centrally hosted that could allow them to handle fewer ad hoc data requests" is merely HYPE that has little to no correlation with the efficiency of market regulation and oversight.

**Question 369:**
NO, we disagree with the Commission's analysis of the Plan's effects on market efficiency due to reductions in violative behavior (see Q.284). In order to effectively and efficiently deter violative behaviors, CAT should consider our suggestions, particularly in Q.59.

**Question 370:**
NO, we disagree with the Commission's analysis of the Plan's effect on efficiency related to reductions in ad hoc data requests from regulators. Per our respond in Q.341, we expect only marginal improvement, like 5%, but definitely not 40% savings as the Commission's analysis suggested.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Question 371:**

NO, we disagree with the Commission's analysis of the Plan's effect on efficiency due to reductions in duplicative reporting systems. Given OATS cannot even be sunset on Day 1 when CAT goes live, and we cannot see how other data collection/ metric reports may be consolidated through CAT, we are HIGHLY SKEPTICAL of what CAT (in its current form) will achieve.

On a separate note, we are aware of additional new data reporting requirements by the SEC (see related news on Bloomberg and Reuters). We urge the SEC to review any new data reporting requirements in conjunction with CAT implementation.

After all, with reference to our suggestion in Q.79, CAT presents an opportunity where all the regulatory data submission and/or metrics reports (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) may be consolidated to cross-tabulate into insights for better market surveillance. We demand to see a better "roadmap" of how these data, metrics reports, and/or audit trails may be consolidated.

**Question 374:**

"Commission preliminarily does not know whether Plan costs incurred by the industry are likely to be passed on to investors" is merely a diplomatic way to say: the costs are probable to be passed on to the investors, which we agree. "Competition in the market for broker-dealer services could mitigate some of these costs, but it may not minimize costs passed on to retail investors" in our frank interpretation means: most of the costs are expected to pass on to retail investors. In fact, we absolutely disagree and disbelieve that CAT (in its current form) will yield the "additional benefits they receive from the potential of a market that is more effectively regulated justify any additional costs they pay to access capital markets".

**Question 379:**

We believe that the CAT NMS Plan (in its current form) would "adversely" impact investor confidence, because the plan lacks connection with real-world problems (i.e. huge loss can be accumulated within split-second; market collapse does not take more than one day; abusive use of financial engineering techniques to synthetically create trades/ derivatives to bypass controls).

On the flip side, CAT presents an opportunity where all the regulatory data submission and/or metrics reports (e.g. CCAR/DFAST, TLAC, Volcker, Basel III, BCBS-283, etc.) may be consolidated to cross-tabulate into insights for better market surveillance. Unfortunately, these are missing in the current Plan.

In order to boost investor confidence, CAT must mandate "real-time" censoring of data and embed an analytical framework in its design. Again, we strongly recommend CAT to adopt our suggestions (see Q.59 in particular). It is possible to save the U.S. economy more than $22 trillion (see a study by the GAO) by avoiding the next crisis.

**H. Alternatives**

**b. Alternatives to the Approaches Permitted by the Exemption Order**

**Question 387:**

Please refer to our respond in Q.162.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Question 388:**

In our opinion, broker-dealers should and can track order allocation information, including many-to-many situation, like a material resources planning (MRP) system in manufacturing (which can be used in conjunction with a job costing system to become an enterprise resources planning (ERP) system).

**Question 389:**

Please refer to our respond in Q.135-142.

**Question 390:**

Please refer to our respond in Q.128-133.

**Question 391:**

Please refer to our respond in Q.134.

**c. Alternatives to Certain Specific Approaches in the CAT NMS Plan**

**Question 400:**

Please refer to our respond in Q.117.

**Question 404:**

No updates and/or correction will be necessary, if CAT may adopt our suggestion to tag information as they were originated. See our respond in Q.87 for explanation.

**Question 406, 407:**

Please refer to our respond in Q.66, 67.

**Question 422:**

Please refer to our respond in Q.318 and Q.337.

**Question 424:**

Please refer to our respond in Q.91.

**Question 425:**

We agree Daisy Chain is more favorable than a unique ID approach. Daisy Chain can be configured by connecting each component to another similar component, rather than require multi-interfaces to directly connect the system's core to each of the component being used. Although daily chain is easy and cheap, it has significant drawback – Daisy chain is suitable only for fixed functional units that cannot be removed.

In normal switchboard/ network topology operations, there may possibly be some components pulled out, and it will break the communication chain downstream. Fortunately, this issue can be resolved by considering a "tree" (hierarchical topology)/ "mesh" approach. "Tree"/ "Mesh", in essence, use sub-system that "self-heals" after a component unit is removed. By having separate "bypass switches" that close when components may be taken out, the system then close the loop to avoid data flow breaks. To further strengthen the "link", the system may use more than one device ("bypass switches") to act as controller, so the system may support more channels with more accurate timing that helps sequencing events.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

"Tree" and "Mesh" are more effective ways than Daisy Chain to cope with modern topology design requirements. In fact, it'll always be an art to choose between centralization ("tree") versus decentralization ("mesh"). There is no absolute right/ wrong answer, but we can help optimize between the cost of control and the most desirable "link" of orders.

### d. Alternatives to the Scope of Certain Specific Approaches in the CAT NMS Plan

**Question 435:**
As we have highlighted the issue about "customers" versus "counterparties" in our respond to Q.135-142, we think periodic refreshes of all customer information to the Central Repository is a bad idea. The only way CAT NMS Plan may provide sufficient safeguards or enforceable policies to maintain an accurate database of customer information is by tagging trade pattern to become unique "fingerprint" to hunt down if one may conveniently flip between the two.

Besides, it is possible for a customer to concurrently be dealing with the same or another division of the same financial institution as counterparty. However, possible "conflict of interests" (a form of violation) may arise in such situation. Thus, tagging trade patterns as "fingerprint" to a customer and/or counterparty is part of CAT's responsibilities because it helps facilitate an exam by the regulators to check for consistency that supports their prosecution works. Again, CAT's purpose is to enable "significantly increase the ability of industry oversight to conduct research, reconstruct market events, monitor market behavior, and identify and investigate misconduct".

### f. Alternatives Discussed in the CAT NMS Plan

**Question 438:**
Please refer to our respond in Q.42.

**Question 439:**
Please refer to our respond in Q.150-155, Q.213 and 214.

**Question 441:**
CAT should be system agnostic to accept the two (or more) data ingestion format approaches. FIX, SWIFT and any other formats that are recognized as "industry standards" are all good. The worst can happen is: when there is only one uniform format remains, because it'll be a monopoly. So be careful of what you wish for!

Translating different (including bespoke) formats into a uniform format for sequencing events, as well as facilitate pattern recognition/ data comparison for market surveillance purpose are part of CAT responsibilities. Indeed, Data Boiler has a patent pending invention to systemically "red-flag" suspicious trade activities (see related suggestions in our respond to Q.59). It includes a component to convert/ sequence trade streams that cross-over to apply concepts from music plagiarism detection. It'll enable ultra-fast analysis/ pattern recognition up to 50 milliseconds. Also, storing data in music format saves significant space, and it is easy to compare with accuracy. We highly recommend CAT to consider using our patent pending invention.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

**Question 444-446:**
We think CAT NMS Plan should NOT mandate a particular: development model, testing process, and quality assurance method. Appropriate management flexibilities/ discretions are needed.

**Question 447-448:**
Support and Help Desk costs could have been minimized or eliminated if CAT may adopt our suggestion to tag information as they were originated (see Q.40, Q.85, Q.87, Q.88, and Q.89), rather than the existing data collection and reporting ("submission") method. The whole process would be automated, rather than worrying about the user entitlement for the authorized submission/ reporting of data.

Regarding user creation and multi-role solutions, we advocate for granting certain limited access/ delayed data to the crowd (public disclosure), it'll enable more creative approaches to market surveillance, foster industry collaboration, as well as augmenting the regulatory efforts for a more holistic industry oversight. Please also see Q.201, Q.280, and Q.293 regarding possible crowd access benefits

**Question 449:**
We think having additional order events, such as the "results order event" and the "CAT feedback order event" are actually good idea from a "twisted" perspective. We envisage using them not as a separate "event type", but mixing in as a way to introduce randomness for the shake of improving information security control. "Chaos" somehow may be the best approach to comprehend the security defense ☺

**Question 450:**
The Plan should mandate the collection of SIP data in real-time, as opposed to through an end-of-day batch process. No additional processing support to deal with out-of- sequence or missing records would be required if CAT may embrace the big data theology of: the messy whole can outperform exact, accurate subsets (see Q.87 and Q.93-100). Allowing CAT Report to report order information on a next-day basis indeed is the BIGGEST FLAW.

Huge loss can be accumulated within split-second. Market collapse does not take more than one day. The T+5 Schedule for regulatory access is USELESS in terms of effective market surveillance in prevention of threats to the U.S. financial stability. Analysts need sensors, not encyclopedia. A good decision, made now and pursued aggressively, is substantially superior to a perfect decision made too late. Thus, the plan should mandate "real-time" (access to intra-day feeds) rather than the "8am following day" submission and the T+5 access schedule. If CAT (in its current form) be implemented without a proper analytical framework, we envisage there will be absolutely ZERO BENEFIT to be realized out of building a gigantic vault that its data won't be effectively used for real-time market surveillance purpose.

**I. Request for Comment on the Economic Analysis**

**Question 456:**
The probable unintended effect would be: "bureaucracy". See Q.279 for how SLAs may become "Service Level Arguments". Yet, the two uncertainties (real-time and funding model) as stated in Q.258, will dramatically affect the assessment of potential benefits.

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

If CAT (in its current form) be implemented without a proper analytical framework, we envisage there will be absolutely ZERO BENEFIT to be realized out of building a gigantic vault that its data won't be effectively used for real-time market surveillance purpose.

If imposing toll charge on all trade messages that goes through CAT, it would essentially be like a "financial transaction tax" that is unjustifiable and detrimental to the overall health of the U.S. capital markets (see our responds to Q.65-67, 69-70 and 72).

***** END *****