212° DATA BOILER TECHNOLOGIES, LLC

BIG DATA | BIG PICTURE | BIG OPPORTUNITIES

We see big to continuously boil down the essential improvements until you achieve sustainable growth!

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

January 27, 2021

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

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

Re: Proposed Amendments to the National Market System Plan Governing the Consolidated Audit Trail to add industry-standard Limitation of Liability Provisions to the Reporter Agreement and Reporting Agent Agreement + Charles River Associates' economic analysis of the liability issues presented by a potential CAT data breach[1]

File No. 4-698 (Release #: 34-90826)

Dear Ms. Countryman:

On behalf of Data Boiler Technologies, I am pleased to provide the U.S. Securities and Exchange Commission (SEC) with our comments on this release concerning Limitation of Liability for Stakeholders of Consolidated Audit Trail (CAT) system. We applaud the SEC and the CAT processor – Financial Industry Regulatory Authority (FINRA) for recognizing the importance of information security[2] over CAT data and commissioned the Charles River Associates' Economic Analysis (CRAEA). In our humble opinion, **CAT is in a dilemma** that its **original design as a golden-source or a "gigantic data-vault" is out-of-date** and unfit for modern surveillance and cybersecurity/ privacy protection requirements. Per Appendix B of the SEC's proposal, authors of the CRAEA did point out the danger of a "centralized database being a prime **target for hackers' attacks** … retail investors' information are particularly vulnerable in the event of unauthorized access and used." There are better ways to enhance market surveillance to mitigate and **prevent Flash Crashes**[3] without any unnecessary movement of data in-and-out of the data vault or data source, hence reducing cybersecurity and privacy risks.

Before we discuss our alternative suggestions, we want to make clear that (1) we **despise "kicking the can down the road" on CAT**; (2) we like to reiterate the **civic concerns** per our comments to the SEC on November 30, 2020.[4] We disagree with the authors of the **CRAEA** because their **three types of breaches scenarios are insufficient to represent the potential damages** to our country's economy and national security in case of breach. Their estimates of "greater than $100 million damage" or 95% percentile loss may **misguide policy makers into falsely believing the risks may possibly be accepted when it should not**. The magnitude of damage if CAT may ever be infiltrated by foreign enemies, or internally compromised, could potential cause a major downfall to the U.S. capital market which trade in trillion dollars daily. The **CRAEA failed to account for scenario, such as the Edward Snowden case**[5] where information from CIA systems got exposed to WikiLeaks. The CRAEA also neglected the scenarios, such as the **2015-2015 SWIFT banking hack**,[6] where hackers used stolen information of a foreign central bank to initiate the scam/ scandal to theft on the Federal Reserve Bank of New York; or **Market Chaos** such as the **GameStop phenomenon**[7]. We can go on-and-on with additional scenarios about exploitations or abuse of CAT. In any case, the SEC's proposed standard Limitation of Liability Provisions (LLP) to the Reporter Agreement and Reporting Agent Agreement is **inconsistent with the Exchange Act** because these threats could escalate into National Security issues which are outside the jurisdiction of the SEC.

[1] https://www.sec.gov/rules/sro/nms/2020/34-90826.pdf
[2] https://www.sec.gov/rules/sro/nms/2020/34-90096.pdf
[3] https://youtu.be/dIq16lZBnDY
[4] https://www.sec.gov/comments/s7-10-20/s71020-8068693-225956.pdf
[5] https://en.wikipedia.org/wiki/Edward_Snowden
[6] https://en.wikipedia.org/wiki/2015%E2%80%932016_SWIFT_banking_hack
[7] https://www.bloomberg.com/news/videos/2021-01-27/gamestop-rally-to-push-some-hedge-funds-to-bankruptcy-gartman-video

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

The **CRAEA did not address** the civic concerns about **massive government surveillance**. According to M.I.T. professor Gary Marx's statements in this Stanford University's study[8], *"…most people in our society would object to this solution, not because they wish to commit any wrongdoings, but because it is invasive and prone to abuse … fails to take into consideration a number of important issues when collecting personally identifiable data or recordings … such practices create an archive of information that is vulnerable to abuse by trusted insiders … In addition, allowing surreptitious surveillance of one form, even limited in scope and for a particular contingency, encourages government to expand such surveillance programs in the future. It is our view that the danger of a 'slippery slope' scenario cannot be dismissed as paranoia … When data is collected, whether such data remains used for its stated purpose after its collection has been called into question… even when two databases of information are created for specific, distinct purposes, in a phenomenon known as 'function creep' they could be combined with one another to form a third with a purpose for which the first two were not built… This non-uniqueness and immutability of information provides great potential for abuse…"*

Please be reminded that the **Fourth Amendment right to be free of unwarranted search or seizure**, recognized by the Supreme Court as protecting a general right to privacy. No-one wants his/her data be used by regulator(s) to develop policies that potentially may discriminative against him/her. **Suspicion of crime or anticipation of market turmoil should begin with some basis or require 'search warrant'** before permissible surveillance on information that would otherwise be considered as private. For civic concerns, the defined **purposes of accessing CAT should be much narrower** than the broadly defined "regulatory purposes". Using tax filing to the Internal Revenue Service (IRS) as an illustrating analogy, the IRS asks for income information, but would not ask for the complete customer and supplier lists and detail transactions unless the party is being summoned in court. Therefore, we argue that there should be **no access to CAT** for 'market surveillance' purpose **prior to identifying symptoms of irregularity that are substantiated** by data at SIPs and/or analytical procedures at SROs/ the SEC.

Within the current permissible rules, we think it is okay for Regulators to demand Broker/ Dealers to provide better Suspicious Activity Report (SAR) and/ or order improvements of trade controls to fulfill essential compliance requirements. We also think the SEC has rights to adopt the "A-Z" clauses that we suggested in ***Table 1*** of our November 30, 2020 comments and shown below, as part of the minimum requirements for principle based rules rather than making specific reference to revision 4 of SP800-53 by the NIST.[9]

#	Suggested Clauses	Rationale/ Justifications
A	CAT should minimize 'data-in-motion' whenever and wherever possible;	The more frequent the transmittal of data in-and-out and within CAT, the more vulnerable it is.
B	Whenever and wherever the data is consumed or 'in-use', it has to serve 'defined purpose(s)' and be at a 'secured environment';	Civic concern of massive government surveillance. 'Data-in-use' is more vulnerable than 'at-rest'. The more users/ devices access to data, the greater the risk hackers may alter/ add/ insert/ use the data abusively.

[8] https://cs.stanford.edu/people/eroberts/cs181/projects/ethics-of-surveillance/ethics.html

[9] NIST's CISP revision 4 of SP800-53 has been superseded by revision 5 since September 2020. Also, NIST's recommended best practices alongside other Cybersecurity and Privacy protection standards/ guidelines, such as ISO/IEC 27001 and 27032, Gramm-Leach-Bliley Act §6801, and FINRA's cybersecurity rules and guidance, etc. may continue to have updates and new added contents. We have multiple concerns if CISP is referencing to a particular NIST publication, including: (1) potential of complying with the bear minimal requirements rather than pursuing the best practices; (2) new emerging cyber threats that the corresponding mitigation method(s) have yet to be incorporated in newer standard – i.e. the in-between time awaiting to adopt new policy; (3) non-synchronize with international rules, such as the EU's General Data Protection Regulation (GDPR).

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

#	Suggested Clauses (continue)	Rationale/ Justifications
C	The appropriate eradication or removal of data as soon as data has been transmitted or used to avoid 'function creep';	Omission or incomplete or untimely eradication would introduce opportunities for hackers.
D	No usage or possession outside of 'defined purposes';	'Function creep' = abuse of CAT related tech or data.
E	When data is 'at-rest', it must be stored at designated 'secured environments';	Data-vault, data-lake, and 'golden source of data' are indeed targets attracting hackers to treasure hunt.
F	'Secured environments' must be segregated in accordance to 'sensitivity' of stored data;	Minimize vulnerability to specific range of data fields and/or records.
G	If data is considered 'sensitive', it must be obfuscated at all times ('at-rest'/ 'in-motion') except when it is 'in-use'; whenever 'alternate' surveillance methods are available, CAT users should refrain from querying 'sensitive' data.	Personal identifiable information (PII) or any data similar to that nature is deemed sensitive. If there is a way(s) to enable surveillance intelligence[10] without crossing the line of privacy[11] hazard, CAT must adopt.
H	'Defined purposes' are limited to 'market surveillance', 'specific case investigation' and/or 'rule enforcement' only;	Again, the Civic concern as stated in "B". No-one wants his/her data be used by regulator(s) to develop policies that potentially may discriminative against him/her.
I	If using metadata can achieve the 'defined purpose', CAT should by all mean avoid collecting or creating repetitive copies of raw data;	Prevent information leakage. Somehow metadata is more useful than raw data, especially when raw data is inherited with imperfect quality (50±ms tolerance).
J	If using 'integrated' data can achieve the 'defined purpose', CAT should avoid collecting data at lower domain;	Roll-up aggregation is another technique similar to masking or obfuscation that helps prevent leakage.
K	All data trajectory must be mapped, assessed, and monitored;	Scrutinize any Repurpose or Reuse or Recycle of data.
L	All users' entitlement in accessing CAT or its data must be duly authorized and maintained without delay;	Share access is a common threat, and lapsed entitlement introduces opportunities for hackers.
M	No access to CAT before a 'defined purpose' is identified and a secured connection is established;	Access entitlement does not mean there is no usage limit on CAT. Gateway and proxies need appropriate inspection to deter unsecure connection to CAT.
N	All user activities must be logged timely in the system;	For scrutinization of any abnormal activities.
O	CAT functionalities and 'data-in-use' should be segregated based on 'defined purpose(s)' of specific user group(s);	Restrict the usage to specific range of data fields and/or records that fits the 'defined purpose(s)'.
P	Whenever possible, apply analytic techniques closest to the original source of data rather than making redundant copies of data;	Redundant copies of data affect data quality and expose the information to higher chance of unauthorized access.
Q	Use of 'predefined automated analytical steps' instead of ad-hoc data query wherever possible;	'Predefined automated analytical steps' require proper testing and authorization by Operating Committee.
R	Volume and frequency of ad-hoc data queries for 'specific case investigation' or 'rule enforcement' purpose is limited;	E.g. to < 0.001% of daily order volume of the targeted broker-dealer with suspicious activity per-query per-user per-day; < 0.01% in aggregate every two weeks.

[10] https://people.eecs.berkeley.edu/~jfc/'mender/IEEESP02.pdf
[11] https://www.fdic.gov/regulations/examinations/financialprivacy/handbook/

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

#	Suggested Clauses (continue)	Rationale/ Justifications
S	No access to CAT for 'market surveillance' purpose prior to identifying symptoms of irregularity that are substantiated by data at SIPs and/or analytical procedures at SROs/ the SEC;	Again, the Civic concern as stated in "B". Suspicion of crime or anticipation of market turmoil should begin with some basis or require 'search warrant' before permissible surveillance on information that would otherwise be considered as private.
T	Bulk data extraction is generally prohibited, except during 'market crash' with special authorization from the SEC;	Where 'market crash' period may refer to Limit Up-Limit Down trigger or exchange halt scenarios.
U	Database server infrastructure and configuration should prioritize 'consistency' and 'partition tolerance' over 'availability', and CAT system should be in compliant with Atomicity, Consistency, Isolation, and Durability (ACID).	The controversy is that CAT as a surveillance tool is supposed to prioritize 'availability' over the two other attributes. Real-time or velocity of data serves to provide a higher values than veracity of data during a 'market crash'. The T+5 access defeats CAT purpose.
V	Data loss protection (DLP) infrastructure must include proper steps for effective and efficient data disposal;	Retaining more data than necessary is a liability. Record retention must be enforced diligently.
W	Audit logs (including user activities, network performance and other system gauges for automated threat detection) must be readily available for exam upon request;	The timelier the review, the higher the chance to salvage a loss situation.
X	Abnormality to CAT or its data or connectivity, or breach of control must be reported in timely manner;	Give the reviewers the authority to provide non-bias and timely report of problems to the upmost Seniors.
Y	Any control compromised must be diligently rectified; independent assessment to recommend interim actions;	Avoid 'bandage' or temporary fix, or a fix in one area may inadvertently cause vulnerability in other area(s).
Z	Must actively observe, adopt and pursuit relevant information security and privacy best practices.	Continuous improvement, ensure forward looking (e.g. today's encryption will be obsoleted upon quantum).

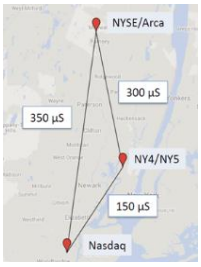
While critical to many aspects of the CAT project since our initial comments in July 2016[12] (many concerns still remain as of today – e.g. "T+5" regulatory access being too late, missing Futures data, etc.), we understand there is no turning back in the CAT project. **Again, CAT is in a dilemma** that its original design as a gigantic vault is out-of-date. It overemphasized on structure rather than embedding a dynamic analytical framework in the design, which unnecessarily redundant copies of data as it affects data quality and exposes the information to higher chance of unauthorized access. User Defined Direct Query and bulk extraction increase the vulnerability of data being misused for impermissible purposes. **Cybersecurity and privacy risks must NOT be accepted in any way or form** because there is no expressed consent from people having a retirement or investment account to share their information with CAT. So, **the only viable option is to reduce risks by significantly reduce the storage, transmission and usage of data to-and-from and in CAT**, other than evidence or symptoms of prosecutable crime.

In view of the dilemma we mentioned earlier while achieving the purposes of CAT, we favor real-time analytical platform (RTAP) to help conduct automated surveillance at high "velocity" efficiency rather than "subjective" to user defined queries or bulk data extraction. With regard to **big data, a timely early warning to facilitate analysis and good decisions is substantially superior than perfecting whatever 'golden-source' of data**. Automated checking of trade irregularities according to certain "defined purposes" would improve "objectiveness" of the surveillance scan.

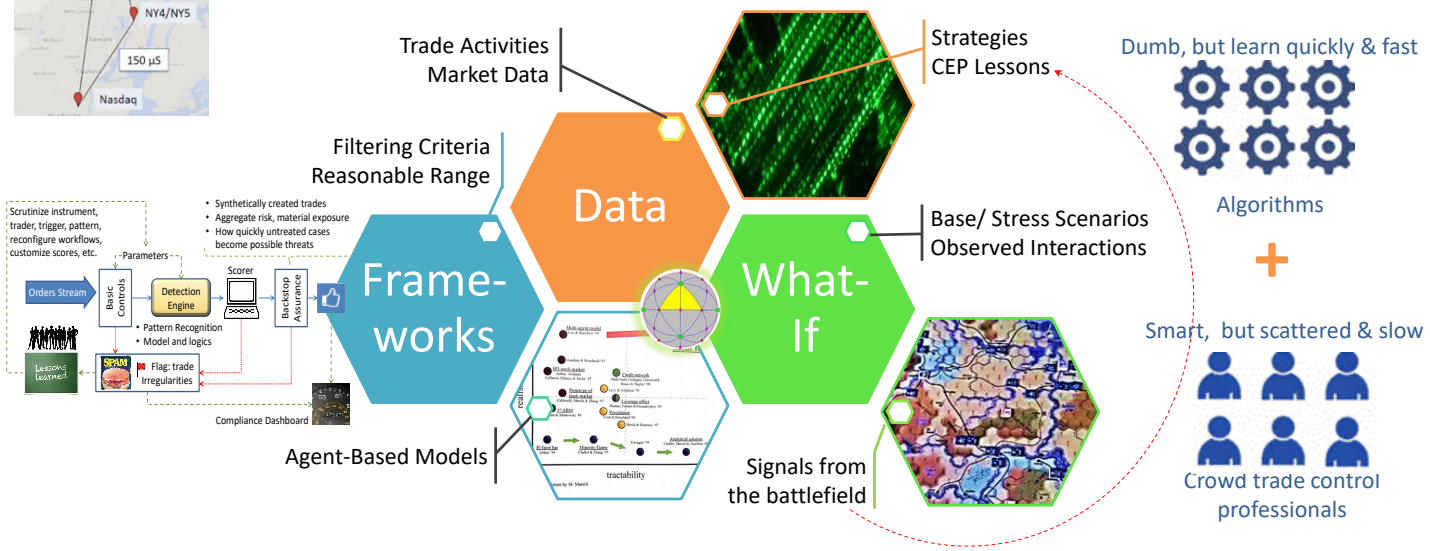
[12] https://www.sec.gov/comments/4-698/4698-4.pdf

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

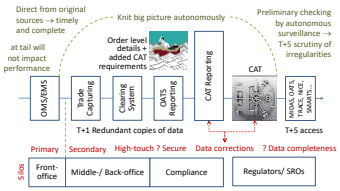
To augment the deficiencies of CAT, please refer to our counter suggestions in ***Figure 2*** of our November 30, 2020 comments, and as illustrated below:



1. If SIPs will expand core data to include depth-of-book, odd-lots, auction, etc. (i.e. 'order level' details), then there is no need to overlap this publicly available information at CAT.
2. Cross tabulate early warning signals from mining of SIP data with intelligence gathered from SROs' surveillance systems (i.e. onset signal detection from better data (captured data at SIPs and SROs are at nanoseconds rather than CAT data with 50± milliseconds tolerance).



3. Enabling the crowd with an open platform consists of: (a) a community library of basic trade patterns, similar to a music editor tool for the crowd to modify basic patterns into new trade strategies to discover alpha (or the unknown unknowns), a simulation tool where new strategies will be back tested, and a reward scheme if back-testing results shown useful insights into reducing unknown unknowns.
4. By leveraging crowd intelligence in step 3 (learn from the field that is closest to the market and in real time) and combining with surveillance findings gathered from steps 1 and 2, it would substantiate enough supporting arguments for identified 'symptoms' of market or trade irregularities.



5. Then, whatever residual data that are not otherwise available at SIPs or SROs would be collected through CAT, which substantially reduces CAT storage (avoid unnecessary regurgitation or repetitive copies of data) to only those data that are essential for law enforcement or prosecution investigation related to specific cases.
6. Through pre-defined automated pre-screening of data that broker-dealers are about to submit to CAT (i.e. targeted to look for specific 'red-flag(s)'), if no 'red-flag' is found relevant to previously identified symptom(s), then trade activities may be certified 'clean' and that batch of data is no longer required to be stored at CAT.
7. With substantiated 'symptoms' and identified 'red-flag(s)', retrieve or subpoena evidence by querying relevant data field/ records would be permissible.

We believe that the **broker-dealer community would welcome a "clean-scan"** on data exhaust from their order management systems **(OMS)** or execution management systems **(EMS) than the burden of data submission** for CAT and **filing SAR**. After the scan they can be provided with a percentage indicator that the broker/ dealer' trade activity may be "**certified clean**" or subjected to the SEC/ FINRA/ SROs exams. This method is indeed drawing a real-life analogy from the Internal Revenue Services (IRS)'s 'My Free Taxes' initiative.[13] The noteworthy fact is: designated private tax filing firms concurrently analyze the data for and on-behalf of the IRS. Allowing the IRS to only **focus on those high-risk candidates for**

[13] https://www.myfreetaxes.com/

☏ 617.237.6111 ✉ info@databoiler.com databoiler.com

scrutinized exams, as a majority of good citizens can handle their annual tax return with ease. Please see this article[14] for how the analogy can be applied in context of CAT and market surveillance.

Again, analyzing the data directly at the original source avoids unnecessary making of redundant copies of data. By reducing the amount of 'data-in-motion'[15] it will make CAT much more **secure**, **effective** (OMS/EMS capture trade orders at nanoseconds rather than CAT data with 50± milliseconds tolerance[16]), and **efficient** (T+0). To effectively mitigate privacy and security risks without creating bureaucracy, do keep in mind the following three management fundamentals: (i) segregation of duties[17], (ii) keep clean with high incentives (e.g. whistleblower award), and (iii) precognitive prevention by reducing the amount of unknown unknowns[18], such as flash crash.

We envisage a crowd model to **reduce unknown unknowns while enhance security of CAT**. The benefits of our suggested approach are: (a) dramatically reduce CAT footprint or data storage and traffic by avoiding unnecessary redundant copies of data and minimize 'data-in-motion'; (b) confine access to CAT data to 'targeted search' of relevant data that fits the 'defined purposes'; and (c) better intelligence for market monitoring by enabling and rewarding the crowd for identifying early warning signals to potential flash crash or other trade irregularities.

Civic concerns about **massive government surveillance** should be taken seriously. We disagree with the CRAEA because the three scenarios being insufficient and the estimates of damage grossly undermine the National Security threats it may cause. **Limit liabilities on CAT processor, users, and data submitters would be detrimental to rights of ordinary investors** and inconsistent with the Exchange Act. The US markets should be opened for any investors to freely trade or invest, as long as the trade activities are not suspicious nor violate any rules. Our suggested approach would **strike appropriate balance** between augmenting the missing components of CAT (or IOSCO – CR12/2012[19]) for effective market surveillance and preserving CAT to conduct its essential functions **without compromise on cybersecurity and privacy protection**. This is a "win-win" for everyone in charging forward of CAT that we hope to get bipartisan support. Feel free to contact us with any questions. Thank you and we look forward to engage in any opportunities where our expertise might be required.

Sincerely,

Kelvin To

Founder and President

Data Boiler Technologies, LLC

Former member of Financial Services Roundtable – BITS (Banking Policy Institute) information security committee

CC: The Honorable Allison Herren Lee, Acting Chair
The Honorable Hester M. Peirce, Commissioner
The Honorable Elad L. Roisman, Commissioner
The Honorable Caroline A. Crenshaw, Commissioner

This letter is also available at:
https://www.DataBoiler.com/index_htm_files/DataBoiler%20SEC%20CAT%20Limitation%20Liability.pdf

[14] https://www.linkedin.com/pulse/hr-block-analogy-cat-combating-fraud-kelvin-to/
[15] https://www.databoiler.com/index_htm_files/DataBoilerInMotion.pdf
[16] https://tabbforum.com/opinions/is-clock-synch-the-cats-fatal-flaw/
[17] https://www.linkedin.com/pulse/big-data-privacy-security-control-kelvin-to/
[18] https://www.pmi.org/learning/library/characterizing-unknown-unknowns-6077
[19] https://www.iosco.org/library/pubdocs/pdf/IOSCOPD389.pdf