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

CAT: the "A through Z" Security and Privacy Requirements

When we compare the current Consolidated Audit Trail (CAT) design with our "A-Z" requirements listed in below table, we see **significant deficiencies and ineffective controls in CAT that require immediate attention**. Is that the reason why the CAT operating committee seems to hesitate to respond to each of our 26 suggestions?[1] These principle based requirements are better than the Enhanced Data Security proposal[2] which only ask FINRA CAT LLC to comply with outdated security guidance – the revision 4 of SP800-53 by the NIST.[3] To address the flaws in CAT's outdated design and to **resolve CAT's challenges, it will take not just cooperation and collaboration, but development and deployment** efforts.

#	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.
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'[4] = 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[5] without crossing the line of privacy[6] 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.

#	Suggested Clauses (continue)	Rationale/ Justifications
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.
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)'.
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.
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).

*By **Kelvin To**, Founder and President of Data Boiler Technologies*

At Data Boiler, we see big to continuously boil down the essential improvements that fit for your purpose. Between my patented inventions and the wealth of experience of my partner, Peter Martyn, we are about finding rare but high-impact values in controversial matters, straight talk of control flaws, leading innovation and change, creation of viable paths toward sustainable development and economic growth.

[1] https://www.sec.gov/comments/4-698/4698-8573527-230862.pdf

[2] https://www.sec.gov/rules/proposed/2020/34-89632.pdf

[3] 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/ A.I. 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).

[4] 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.

[5] https://people.eecs.berkeley.edu/~jfc/'mender/IEEESP02.pdf

[6] https://www.fdic.gov/regulations/examinations/financialprivacy/handbook/