August 9, 2010

Elizabeth M. Murphy, Secretary
Securities and Exchange Commission
100 F Street, NE
Washington, DC 20549–1090

Re: File No. S7–11–10 – Consolidated Audit Trail

Dear Ms. Murphy:

We appreciate the opportunity to comment on the Securities and Exchange Commission's proposed rule for a Consolidated Audit Trail. Rather than commenting on the central question of whether a Consolidated Audit Trail should be established, we focus on one particular, important, aspect – whether "SROs or the central repository [should] be allowed to make the data available to third parties, such as for academic research" (Federal Register Vol. 75, No. 109, p. 32581). As academic researchers with expertise in financial market regulation and information technology (IT)[1], and with a specific interest in how research in IT and other academic disciplines can be used to improve financial market regulation, we view issues about the availability of this data as highly important. While we recognize problems associated with making the data available for research (noted below), we urge the Commission to support a broad-based and open program of research that will improve the effectiveness of financial market regulation.

As context, let us start by saying that we firmly support the Commission's mission – "to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation." To do this, the SEC (and other regulatory organizations) should enlist the best minds in helping to develop regulatory approaches that take into account the complex and challenging environment cited by the Commission in its proposed rule. Further, we believe that academic researchers can add significantly to this work and can be important assets in furthering the work of the Commission and other regulatory bodies.

Access to real-world data can help research immensely. Theories about how markets work and how regulation and related IT systems can make them safer or more robust can be tested empirically to see if they really work. This argues for making the data widely available for academic research. But such availability poses important problems – access to raw, recent audit trail data can potentially expose the identities and proprietary business strategies of individuals and corporations. If not done carefully, this could result not only in the loss of anonymity and privacy, but could potentially increase resistance to auditing, raise questions about how such data might be used in the event of litigation, and possibly undermine the security of the markets in general. However, this does not necessarily mean that the Commission should close the door to third party access.

In the proposed rule the Commission asks specific questions pertaining to data availability, among them: whether certain data elements are more sensitive than others, whether a time

[1] We use the term "information technology" to refer broadly to computer science, information science, informatics, and related disciplines.

lag would help, and how controlled access and confidentiality could work. We comment below on these.

Data element sensitivity. The data in a consolidated audit trail is inherently sensitive. The data contains not only sensitive information about individual traders, brokers, and securities, but also—implicitly—proprietary strategies and business practices. While researchers have extensive experience with "de-identifying" data to protect individual identities[2], the proprietary strategies would be hard to protect. If researchers were given access to de-identified data, even if for the sole purpose of developing improved approaches to regulation, much would be learned collaterally about the proprietary algorithms and trading strategies of these anonymous financial actors. Plausibly, the application of machine learning and data mining technologies could be used to reverse-engineer strategies that might have taken millions of dollars and years of investment to develop. On the other hand, if data elements that might be susceptible to such reverse engineering were excluded from the data set, or if the data were aggregated to protect proprietary strategies and individual identities, it is much less likely that new ideas for improving regulation would emerge from the research.

Time lag. Publishing the data after a time lag would reduce the risk of revealing private information, but would also limit the usefulness of the research. Given the fast-moving evolution of both the trading infrastructure and the way it is used, research should be based on the most current data to have the greatest regulatory impact.

Confidentiality and limited availability. One approach would be to make the data available to limited groups of researchers under strict controls and confidentiality agreements. However, even this level of availability increases the risk of accidental release or misuse of data. At the same time, such an approach limits the research capacity to only a few. Because of the many challenges and complexities of building safe and robust regulatory systems, it will be beneficial to have many minds working on the problem. An open program of research devoted to developing better approaches to regulation would be useful, something analogous to the open source movement in software development[3] or open challenges in cyber security research[4], where significant problems are posed and publicized, and all interested parties are invited to help solve them. In this spirit, we believe that the data should be made widely available to researchers and as much research as possible should be conducted in the open[5].

And so, we have conflicting objectives. We want to develop new and innovative approaches to thwart prohibited activities by making the most current and complete data available to a broad

---

2. Uzuner, Sibanda, Luo, and Szolovits, 2008; Uzuner, Solti, and Cadag, forthcoming
3. Feller and Fitzgerald, 2000; Wu and Lin, 2001
4. Wybourne, Austin, and Palmer, 2009; Sheldon, Peterson, Krings, Abercrombie, and Mili, 2009
5. Of course, if used for financial market regulation, this approach presents obvious challenges. Merely stating a problem might expose issues that might be used to circumvent regulation or hide fraudulent behavior. Yet in many policy areas of comparable national importance and sensitivity—cyber security, homeland security, healthcare, and education, among others—dual threads of research are pursued simultaneously, one conducted in the open and the other not publicly available. In some of these research areas, academic researchers have access to large public databases on which to base their investigations.

group of researchers. But we also want to protect proprietary strategies as well as individual privacy. What are we to do?

The preceding discussion has an inherent assumption that may or may not be true. That to conduct research that will develop better means of regulation requires analyzing information that runs the risk of exposing trade secrets. What if this were not the case, and we could release the data to researchers in a form that doesn't expose this sensitive information? The data could be used for one purpose, but not another. The result would be an approach that would allow safe release of the data to the research community and ipso facto might lead to approaches to surveillance that relied on less sensitive information.

While using data for one purpose while intrinsically preventing its use for another may sound like having your cake and eating it too, it may not be an impossible task. Similar challenges were put to cyber security researchers at the dawn of the Internet age, when an emerging electronic-commerce industry wanted customers to send credit card information over an insecure Internet without first having to send them an encryption key over that same insecure network. Academic researchers have developed sophisticated techniques for communicating information that enables them to do exactly that, helping create a secure infrastructure for today's robust digital economy. The ideas that underlie this technology have been used to solve similar challenges, such as proving that an object (such as a patentable idea) exists at a particular point in time, without giving away any hints about what that object might be. There is a large cadre of faculty and student researchers who approach such difficult types of problems every day. We would relish the prospect of seeing what such researchers would develop if they were tasked with the challenge of developing new algorithms for monitoring our financial markets, ones that didn't run the risk of violating the privacy or proprietary interests of individuals and organizations in the system.

As with any research, we have no a priori way of predicting whether it might be possible to develop an approach to regulation that didn't depend on all the data. However, our experience tells us that, even if a total solution remained elusive, it is very likely that an active program of open research into this area would lead to improved and more efficient approaches for protecting our markets from fraud and instability. In other important areas, such "grand challenge" approaches have helped galvanize the research community around a particular issue and have led to significant new insights that might be difficult to obtain in a more closed environment.

And so we encourage the Commission to support both open and closed programs of IT research aimed at improving financial market regulation. The specific example we cite—figuring out how best to share information without giving away proprietary trading strategies or traceable identities—is one of many promising IT research questions in this area. A dedicated program of IT research similar to the National Science Foundation's CyberTrust and Digital Government programs, focused on the particular problem of regulating our financial markets and exploring fundamental IT issues in financial market regulation, would be a very useful byproduct of the Commission's proposed rule. The "Office of Financial Research" created by the Dodd-Frank Wall Street Reform and Consumer Protection Act could be tasked with developing such a

program of research, either by itself or in conjunction with the Commission and/or the National Science Foundation.

The nation's financial markets are an incredible resource that undergirds our nation's economic prosperity and competitiveness. A regulatory system that safeguards investors and allows our nation to compete globally is key to our future. We want our best and brightest helping to develop effective approaches to regulation. Given the key role that information technologies play in these markets, inviting a broad array of researchers to conduct open academic research on the technical underpinnings of financial market regulation would be a very positive step in that direction.

We appreciate the opportunity to comment on the proposed rule and would be happy to provide additional information regarding these research opportunities as they apply to challenges in financial market regulation. Please note that these comments reflect our personal opinions, and, although we list our professional affiliations for identification purposes, they do not reflect an official position of the University at Albany, the State University of New York, or the Institute for Financial Market Regulation.

Sincerely,

Peter A. Bloniarz
Dean
College of Computing & Information
p.bloniarz@albany.edu

George Berg
Associate Professor and Chair
Department of Computer Science
berg@cs.albany.edu

Sandor P. Schuman
Affiliated Faculty
Department of Informatics
sschuman@albany.edu

University at Albany/ State University of New York
1400 Washington Ave., Albany, NY 12222

References

Wu, M. and Lin, Y. 2001. Open Source Software Development: An Overview. *Computer* 34, 6 (June), 33-38.

Feller, J. and Fitzgerald, B. 2000. A framework analysis of the open source software development paradigm. In *Proceedings of the Twenty First international Conference on information Systems* (December 10-13, 2000, Brisbane, Australia). International Conference on Information Systems. Association for Information Systems, Atlanta, GA, 58-69.

Wybourne, M. N., Austin, M. F., and Palmer, C. C. 2009. *National Cyber Security Research and Development Challenges.* Dartmouth, New Hampshire: Institute for Information Infrastructure Protection.

Sheldon, F., Peterson, G., Krings, A., Abercrombie, R., Mili, A., Eds. 2009. *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research* (April 13-15, 2009, Oak Ridge National Laboratory, Oak Ridge, Tennessee). New York: ACM.

Uzuner, Ö., Sibanda, T., Luo, Y., Szolovits, P. 2008. A De-identifier for Medical Discharge Summaries. *International Journal Artificial Intelligence in Medicine*. 42, 1 (January), 13-35.

Uzuner, Ö., Solti, I., Cadag, E. (forthcoming). Extracting Medication Information from Clinical Text. *Journal of the American Medical Informatics Association*.