The Journal of Portfolio Management

INVITED EDITORIAL COMMENT

Avoiding a Billion Dollar Federal Financial Technology Rat Hole

DAVID LEINWEBER

The Flash Crash, and an ongoing series of mini-crashes, some upward, have made many of us nervous. Some buy-side investors are using "electronic counter measure" trading systems, a sort of high-frequency (HFT) radar.

The heads of the SEC and CFTC often point out that they are running an IT museum. They have photographic evidence to prove it—the highest-tech background that *The New York Times* (on September 21, 2010) could find for a photo of Gregg Berman, the SEC's point man on the Flash, was a corner with five PCs, a Bloomberg, a printer, a fax, and three TVs on the wall with several large clocks.

A better measure of the inadequacy of the current mélange of IT antiquities is that the SEC/CFTC report on the May 6th crash was released on September 30, 2010. Taking nearly five months to analyze the wildest ever five minutes of market data is unacceptable. CFTC Chair Gensler specifically blamed the delay on the "enormous" effort to collect and analyze data.

What an enormous mess it is. In the SEC's request for comments on a new, improved consolidated audit trail system (CATS) is a long and sorry tale of bad software replacing bad paper systems with names like OATS, OTS, and COATS. The SEC explicitly blames the incompatibility of these Federal/FINRA IT legacies for the sorry state of our ability to know what is happening in markets and to make the case for an expensive IT upgrade—a proposed $4 billion to start, and $2 billion a year to operate, funded by trader and investor fees. A better name for CATS might be Andy Lo's "Capital Markets Safety Board." Aircraft have a much higher IT level than our capital markets.

Meanwhile, other federal agencies take proper interest in this from their perspectives. The emerging Office of Financial Research, charged with understanding systemic risk, has a clear role in the risks of systems. The long, appointed term of the director and lower legacy problems are encouraging.

The cyber world is dangerous. For many, the first thought during the Flash Crash was that it was a cyber attack. As far as we know, it wasn't, but it could have been. IARPA, the newly established intelligence research agency, is concerned. Discouragingly, the idea of security is not prominent in CATS discussions so maybe we'll get two systems—one for enforcement and the like, and one for cyber defense. Don't laugh. The Army and the Navy used incompatible radios for decades.

There is a long, sad history of large scale federal information technology flameouts. Tapes buried in the tundra. Billions of dollars have gone into the bit bucket. We need to make sure we get this one right.

WHY THIS CAN BE DONE RIGHT

As it turns out, this one has a better chance to be *done right* because large pieces of it have been *done before*. Largely unbeknownst to the financial information world, another zone of the federal government has been investing heavily in "big data" for science since the 1960s. These enormously capable supercomputers are relatively rare, and few financial people seem to have much sense of them.

Just looking at the size of the market data problem is instructive. SEC Chairperson Mary Schapiro estimated the

flow rate of the data stream: "We need… capability to receive something on the order of 20 terabytes of data in a month."

This is unbelievably small beans in the scientific supercomputing world of big data. Some experiments produce petabytes (1,000 terabytes) *per second*. Astronomy and earth-looking sensors produce similar flows millions of times larger than markets.

Being able to analyze and understand big data is just as important as having a place to put it. Science data were once as much of a mess as the market data are today. Millions of files scattered in many places, in many formats.

JIM GRAY, BIG DATA WIZARD

Readers may recall the news from 2008 of the search at sea for missing computer scientist Jim Gray. Microsoft, where Gray worked, Google, and thousands of volunteers mounted the most computer-intensive search effort in history. Gray's sailboat was never found, but why were so many people looking for him? *The New York Times*' John Markoff described Gray as the most insightful and accomplished computer scientist in generations. The closest thing to a Nobel Prize in computer science is the Turing Award. Gray won it in 1998.

Gray observed that the scientific data world had become a mess of millions of scattered, inconsistent, and disorderly files. It took months or years to put enough together to understand, for example, astronomical events observed in different ways and places.

Imagine, for the sake of comparison, a flash in the sky (not the stock market) that we want to understand and look at in other ways using other telescopes, satellites, and histories (not futures or options or multi-venues). Before Jim Gray, doing this was the same giant tangle of data we see in markets.

ONE BIG ELECTRONIC EYE

Because of Gray's ideas and efforts, all of the major telescopes on the planet and around it are now effectively one big electronic eye. The combined size of the data collected 24/7 is orders of magnitude larger than financial flows. *This can be done right*.

Unlike the financial world, astronomy and other sciences have been able to bring order from the information chaos by following Gray's principles and examples.

There are far too many gems in Gray's Fourth Paradigm opus on data collection, curation, analysis, modeling, exploration, and visualization to do them all justice here. I picked just one— Jim Gray's recipe for designing a database for a given discipline is that "it must be able to answer the key 20 questions that […we] want to ask of it."

KEY QUESTIONS

Regulators need to clearly spell out the right questions that we need future analysis systems to answer. Then, let the technology come from that, not from what's on the shelf. The following are some questions to get the process started:

Enforcement. Can we spot a market manipulator who works in microseconds, working similar scams across markets and different securities?

Systemic structural risk. How can we know if the complex interactions between market centers are a source of systemic risk due to unanticipated interactions between those systems when they are operating as designed?

Systemic implementation risk. The previous question, but recognizing that markets are built on real computers with delays, crashes, races, slowdowns, and all the ailments and errors that occur in real plugged-into-the wall IT machinery.

Policy. Can we simulate, analyze, model, and visualize what would happen if we make changes in the rules? We must avoid unintended consequences.

Financial cyber attack. The worst call the heads of the SEC/OFR/CFTC could get is: "Are our markets under attack?" If that happened, test probes would certainly precede it. Could we know if that was happening in time to take any action?

A business-as-usual update of fragmented, problematic systems is a recipe for disaster. I heard a "send lawyers, guns, and money" ringtone. Substitute "accountants" for "guns" and I fear we have a description of a likely approach to tackling the challenges of 21st century financial information. This is one of the most important computer science challenges we face.

The proposed billions that will need to be spent on this come from us. We should try to repeat the successes that science has shown with the biggest data problems. This counts as "disruptive innovation." Federal agencies, FINRA, and the supercomputing scientists have lots to talk about. They should get acquainted.

David Leinweber *is co-founder of the Lawrence Berkeley National Laboratory Computational Research Division's Center for Innovative Financial Technology in Berkeley, CA.*