

*Remarks of Richard Berner, Director, Office of Financial Research,
at a joint conference on Big Data in Finance hosted by the OFR
and the Center on Finance, Law, and Policy at the University of Michigan
October 27, 2016, Ann Arbor, Michigan*

Thank you, Michael, for that kind introduction, for your collaboration, and for the center's support for this conference. Thanks also to our collaborators from here at the University of Michigan on this event — the Institute for Data Science, the Ross School of Business, the College of Engineering, and the Law School. I would also like to thank our sponsors — the Smith Richardson Foundation and the Omidyar Network — and our speakers and presenters for joining us and making this event possible.

Most of all, I want to thank you — our conference participants — for being here to focus on a fundamental element of the new world order: Big Data.

This is our second joint interdisciplinary conference, bringing together data scientists, economists, lawyers, mathematicians, and others to share ideas, scientific findings, and alternative perspectives on opportunities and problems facing our financial system.

Big Data

Big Data captures one of the defining traits of our era, and likely the future.

Some prosaic examples: We pull out our smart phones to recall who played Kramer in the TV show Seinfeld or to navigate around traffic jams. Our cars have dozens of computers that feed data to us and to the manufacturer, and will soon drive themselves.

In virtually every aspect of our personal and professional lives, our thirst for information, for decision making based on detailed information, for convenience, and for speed have fueled the demand for data, while technology has raced to supply them.

The value of data explodes when they can be compared and linked with other data from a variety of sources, especially when they are highly granular. Small wonder that enterprises and governments alike view data as valuable assets. So do criminals and other bad actors.

However, making sense of, and managing, the torrent of data create a tsunami of challenges. We use visualization and other techniques to understand the patterns in large sets of data, but here we've only scratched the surface.

Some of the emerging techniques, like artificial intelligence, raise the kinds of moral questions that genetic engineering has raised for decades. And the data deluge also raises serious questions about data privacy, confidentiality, ownership, appropriate access, security, management, stewardship, integrity, analytics, and retention. It's hardly surprising that the private and official sectors have elevated chief data officers to senior positions.

In my remarks, I'll try to offer a few answers to those questions in the context of our work, and to frame the discussion over the next two days. And of course, I will challenge you, the people in this room, to help us find even better answers.

Data and Financial Stability

Although we have access to vast stores of data, we still struggle to separate the information from the noise. What data do we truly need and how can we best use them? And how do we act on the recognition that all the data in the world are no help without being harnessed, organized, and understood?

Marrying the data with analytics clearly creates value. That's why IBM purchased the Weather Company's assets last year; its Watson computer can sift through troves of weather data and make better predictions. Forecasting weather has improved markedly over the decades in large part because the availability of good weather data has exploded — and the ability to analyze massive amounts of data exploded as well.

Let me give you context for thinking about data needs for our work in financial stability analysis.

Some people propose that the OFR should be a financial weather service, poring over troves of data and identifying patterns of financial storm signals to predict a gathering crisis.

That's an exciting metaphor, but even with great data and tools to analyze them, I don't think we can predict — much less prevent — the next financial crisis. Instead, we seek to make the financial system more resilient to shocks by helping to identify and analyze vulnerabilities that can morph into systemwide threats.

To do that, we also work to improve the quality and accessibility of the data we have, identify and fill gaps in the data landscape, and develop appropriate tools to analyze them. That will vastly enhance our ability to look around corners — and in the shadows — for building threats.

An Evolving Financial System and Regulatory Framework

Our financial system and the regulatory framework governing it have evolved rapidly, paralleling the revolution in data and technology. We have evolved from a financial system that was largely domestic, where threats may have been confined to particular sectors like banking, and specialized regulators conducted oversight of institutions and markets.

The financial crisis exposed gaps in our understanding of the financial system and in data to measure financial activity. The crisis also underscored the strong need for financial policymakers and regulators to collaborate across jurisdictions and regulatory silos.

Before the crisis, the International Organization for Securities Commissions, or IOSCO, an association of the world's securities regulators, and the Financial Stability Forum (now the Financial Stability Board) began to promote reform in international financial regulation.

Domestically, we saw enhanced coordination through the informal President's Working Group on Financial Markets and the Federal Financial Institutions Examination Council, a body of federal financial regulators that sets standards for examining financial institutions.

Our Dodd-Frank Act responded to a growing recognition that financial activity and regulation are now interconnected, globally and jurisdictionally. To break down barriers to collaboration in

our regulatory infrastructure, Congress created two complimentary institutions to identify and respond to threats to U.S. financial stability, wherever they emerge: The Financial Stability Oversight Council — or FSOC — and the Office of Financial Research.

The post-crisis framework for global coordination also improved. The G20 began a high-level political impetus to enact reforms, the Financial Stability Board gained legal status in Switzerland, and IOSCO joined forces with the Basel Committee on Payments and Market Infrastructures on projects such as harmonizing swaps data reporting, setting standards for the governance of central counterparties, and coordinating global standards for cyber security.

To me, the most notable aspect of these emergent organizations and affiliations is their interdisciplinary nature. Economists no longer work only with other economists, or lawyers only with other lawyers. Each must get out of their comfort zones and work closely from the beginning of a project to the end with the other and, more broadly, in a team that includes data scientists and information technologists — or their solutions fall flat.

The OFR's Interdisciplinary Approach

At the OFR, we are keenly focused on this point.

Late last year, we adopted a programmatic approach to our work, which identifies core areas of concentration that align our priorities to our mission. We are initially focusing on eight core areas. Some relate to institutions and markets — central counterparties, market structure, and financial institutions, while others involve tools — monitors and stress testing.

The final three programs focus on the scope, quality, and accessibility of data — the topic of this conference.

Our programmatic approach is interdisciplinary by design. A senior staffer with relevant expertise leads each of our program teams. That person might be an economist, market analyst, policy expert, or data scientist. Each team is made up of researchers, data experts, lawyers, and technologists. In addition, the teams include external affairs specialists who help us align our priorities with stakeholders' needs, and communicate our work and our findings.

This retooling of the way we work — by convening centers of interdisciplinary coordination — is already paying off. For example, our U.S. Money Fund Monitor is an interactive visualization tool to display highly granular data collected by the Securities and Exchange Commission (SEC). Both policymakers, who are analyzing the effects of “Brexit” and the impact of the SEC's new fund rules on U.S. markets, and the news media, have cited its utility.

In designing the monitor, analysts who are experts in these markets worked with lawyers who negotiated data rights, technologists who built the user-friendly tool, and public affairs specialists who helped figure out how to effectively communicate the most valuable information to our stakeholders. I invite you to visit our website to use this tool for yourself.

Data Scope, Quality, and Accessibility

Our OFR data programs echo the three themes of this conference and ask three basic questions about data:

1. Do the data have the necessary scope? That is, are the data comprehensive and, at the same time, granular? And where are the key gaps in the data?
2. Are the data of good quality? Are the data fit for purpose, and capable of providing actionable information, either alone or in combination with other data?
3. Finally, are the data accessible? Are they available to decision makers for well-informed and timely decisions?

Data Scope

Regarding data scope, I will start by making an important point: More data are not necessarily the answer. We must have the right data. That might mean using existing regulatory, commercial, or public collections. It could also mean that some data are not doing the intended job and so the collections no longer make sense. If the financial system has evolved and moved on, so should our data collections.

Granular data are essential for our work. That's because, like policymakers and risk managers, we are in the business of assessing tail risks. Looking at medians and means is helpful for sizing a market or an institution, but risk assessment requires analyzing the whole distribution. Granular data and their analysis help us gauge risks related to particular activities, and to concentration, interconnectedness, complexity, financial innovation, and the migration of financial activity.

So, granular data are critical for us to update our Financial Stability Monitor, which assesses vulnerabilities in the financial system based on five functional areas of risk: (1) macroeconomic, (2) market, (3) credit, (4) funding and liquidity, and (5) contagion.

If we see a consequential data gap, we consider filling it. For example, data describing bilateral repurchase agreements and securities lending were scant in the run up to the financial crisis, and they still are. To understand how best to fill those gaps, the OFR, the Federal Reserve System, and the SEC together recently completed voluntary pilot surveys.

We reported results of the pilots to FSOC and the public. Guided by the pilots, we are pursuing a permanent data collection for repo transactions. These data will help us better monitor a \$1.8 trillion component of the \$4.4 trillion securities financing markets — one that amplified the financial crisis through runs and asset fire sales.

Under our data scope program, we also consider what other datasets exist on the servers of our sister agencies that are necessary for better stability monitoring. We work closely with fellow regulators to figure out who has what. The results are filed in the Interagency Data Inventory, a catalog of metadata — data about the data — acquired by financial regulators. We update the inventory annually.

We also collaborate with industry, market utilities, and other data providers to see if the data we seek may already exist. In fact, the statute requires that we check whether data exist before

launching any collection. We want to be sure that any new data collection minimizes the burden on firms providing the data, while maximizing benefits.

For our repo and sec lending pilots, we worked directly with the firms to develop the data template — a shining example of government and industry working together to solve problems. Following these best practices in data collection also aligns the data with the risks and aligns industry interests with ours.

Data Quality

Our second data-related program — data quality — focuses on standardizing and harmonizing data to make them useful. An example is our Legal Entity Identifier (LEI) program. The LEI is like a bar code for precisely identifying parties to financial transactions.

Although industry hungered for such a standard, the LEI did not exist before the crisis, so the industry, regulators, and policymakers were practically unable to link datasets or even figure out “who is whom” and “who owns whom” in our financial system.

Under OFR leadership, the LEI system now exists and almost 500,000 legal entities from almost 200 countries have LEIs for reporting and other uses. The system is now rolling out the ability to reveal the ownership structures of companies, and thus how firms are exposed to one another.

The next step is implementation of global standards for instrument identifiers, which will help us understand “who owns what” — and “who owns the risk” through financial instruments.

These critical interdisciplinary building blocks help assure data quality. To realize the full benefits of the LEI system, we continue to call on regulators to require the use of the LEI in regulatory reporting.

Data Accessibility

Our data accessibility program starts from an obvious premise: What good is any dataset if you can't get it and use it when you need it?

A major challenge is to achieve a balance between securing confidential data and making data appropriately available to stakeholders, including policymakers, regulators, markets, and the public. This program aims at finding that balance.

Trust and verification are crucial for sharing data. Data providers such as financial firms, domestic regulators, and foreign authorities are reluctant to share data without trusting that (1) the need for confidentiality is recognized; and (2) once shared, the data will not be breached or carelessly shared further. Verification, even of trusted parties, helps build that trust.

Reputations are at stake and any regulator, including the OFR, recognizes that it must protect confidential data, or prospective data providers will be reluctant to cooperate in the future. At the OFR, we have been highly successful at gathering data voluntarily from other regulators, market infrastructures, and firms.

We have dozens of memorandums of understanding, or MOUs, that reflect common understandings of the importance of strong information security regimes, agreement on what data must be secured at what level of security, and other process-oriented clauses dealing with court subpoenas and Freedom of Information Act requests.

We have found this approach fruitful. In fact, the OFR has been leading an interagency working group to develop best practices for data sharing. The group is working on a common vocabulary for identifying data, definitions of information security levels, and model language for MOUs.

This project is particularly exciting because, for the first time, we have created a community of financial regulatory lawyers specializing in data sharing agreements and memorandums. I believe this interagency partnership will greatly speed the creation of MOUs — and lead to greater familiarity and trust.

The OFR's data collection rulemaking and subpoena authority are also critical for our work. We intend to conduct a rulemaking on repo markets soon. Of course, a rulemaking is superfluous if the desired data already exist elsewhere, either at a regulator or at a firm.

A subpoena is a great tool to have in the toolkit. It enhances our power to persuade. Someone recently said to me that you can learn a lot more from a subpoena than you can from a regression analysis. Of course, this tool must be used judiciously. A subpoena carries costs — to the reputation of the organization and through the sometimes time-consuming process of judicial enforcement.

So far, we have chosen to pursue the cooperative approach to data sharing. This approach is not perfect because the process takes persistence and time, and the data, once obtained, may not fit their intended purpose. Moreover, the provider of the data may impose limits on further sharing the data, making use of the data for public or regulatory reporting challenging.

International data sharing can be even more challenging because of the absence of a common overseer and legal framework. In that environment, MOUs also advance the game. We have one with the Bank of England, and markets regulators and law enforcement entities have long relied on informal MOUs and international “soft law” to gain cooperation.

At last year's conference here in Ann Arbor, we heard of many promising technologies that might help us solve the trust problems that can impede data sharing. For example, computing techniques may be able to mask counterparty data, but still reveal concentrations of a particular counterparty or network. As these technologies mature, they might help us solve a problem such as combining U.S. data on swaps positions with those of European regulated entities, without revealing the names of the firms themselves.

As an economist by training, even one supported by a cadre of technologists, lawyers, and data scientists, I won't presume to enumerate the possibilities that exist in these other domains. But you can, and I hope you will.

I hope our discussions here can help us imagine how to develop ways to use modern data and information technology science to collect data efficiently, improve data quality, and make data appropriately accessible to those who need them.

Thank you again for your engagement here. I would be happy to take some questions.