HOSTED BY THE FEDERAL OFFICE OF FINANCIAL RESEARCH AND THE CENTER ON FINANCE, LAW, AND POLICY

BIG DATA IN FINANCE

OFR OFFICE OF FINANCIAL RESEARCH

CENTER ON FINANCE, LAW, AND POLICY
UNIVERSITY OF MICHIGAN

Panel 3: Big Questions About Big Data
Thursday, October 27, 2016 at 2:00 p.m.
University of Michigan Law School, Hutchins Hall 100

Moderator:
Michael S. Barr, Roy F. and Jean Humphrey Proffitt Professor of Law, University of Michigan Law School; Professor of Public Policy, University of Michigan Gerald R. Ford School of Public Policy; and Faculty Director, Center on Finance, Law, and Policy, University of Michigan

Panelists:
Thomas Brown, Partner, Global Banking and Payment Systems Practice, Paul Hastings

Kabir Kumar, Director for Policy and Ecosystem Building, Omidyar Network

Graham Steele, Minority Chief Counsel, U.S. Senate Committee on Banking, Housing, and Urban Affairs

Marisabel Torres, Senior Policy Analyst, Wealth-Building Policy Project, National Council of La Raza

Introduction

Big Data can create value for consumers and firms alike. Companies use it to glean insights about their customers and deliver new and complex products to market. Banks use it to improve fraud detection. Healthcare providers leverage it to improve patient treatment. Big data and the analytics and tools to process it drive benefits that can be felt across a range of sectors, in both large and small firms. At the same time, big data can raise big questions for society.

This panel will wrestle with several big questions raised by Big Data. Who should own and control big data? How should we think about privacy in the era of Big Data? How can Big Data be used to reduce discrimination, or conversely, what are the dangers that Big Data can exacerbate discrimination? How can policy makers and the public get access to the information they need to make or evaluate public policy, understand risks in the financial system, or explore tradeoffs in different public policy approaches?

NOT FOR CITATION OR QUOTATION. DISCUSSION DRAFT ONLY. Professor Barr thanks research assistants Amelia Bailey, Andrew Kang, Charlie McDonald and Steve Vnuk, who prepared this draft note. Note contents should not be attributed to the moderator or panelists.

Ownership of Consumer Financial Data

There is a fierce debate today over whether consumers should be empowered to own their own data, and thus to introduce new competition into the financial services space. Fintech companies assert that they should have the right, upon permission from consumers, to access bank data to upload information to apps that help consumers manage their finances, for example, while banks contend that such data must be protected and secured by the bank itself, rather than opening up such consumer information to fraud or abuse.

JP Morgan Chase CEO Jamie Dimon's annual letter to shareholders earlier this year included a discussion of the steps the bank is going to take to change the way consumer banking data is shared with third-parties, which includes many fintech companies. Dimon's proposed solution might involve Chase "pushing" consumer banking data to third parties based on consent received from consumers. Chase would control how the data is shared and how much of it is shared. Fintech companies and venture capitalists backing them, on the other hand, would prefer to keep the status quo or improve the way data is "pulled" from banks, with consumer consent, through open APIs or similar interventions. Some believe that regulators should intervene to ensure that banks don't block access to data, as a number of banks have reportedly done over the last year. Both sides claim to advocate for the consumer's interest: the banks on the grounds of security and privacy and the fintech industry on the grounds of access and product innovation.

Moving forward on this debate is not as straight forward as both sides claim. It involves both technical improvements on security, as well as technical and regulatory changes that may require open access API. In fact, open access API may seem premature to some in the absence of a legal and policy architecture that provides for security, privacy, consumer protection and consent. Further reforms will also be required to lower the costs of account switching, and to empower consumers to act on such reforms. One such measure is the CFPB's authority, which the Obama Administration included as section 1033 of the Dodd-Frank Act, to require banks (and others) to provide consumers with access to their account and usage data in standardized, machine-readable form.

Other reforms might include: payment systems reforms to move towards instant payments, with appropriate consumer protections and anti-money laundering detection and enforcement; and changing good funds availability rules so that deposits are instantly available (with appropriate anti-fraud protections), reducing the risk of overdraft and helping consumers better to match their income with expenses. Overdraft reforms can also further reduce the costs of account ownership and better enable consumers to take control of whether and when overdrafts occur.

Background

As mentioned above, banks have raised security and privacy concerns over allowing third party access and control over consumer financial data, while fintech companies have argued the banks' denial of access stifles financial services innovation to the detriment of consumers. In taking such positions, both banks and fintech companies claim to be advocating for consumers' best interests.

Banks' Concerns

Data Security & Privacy

Many third-party financial services providers use financial data aggregation to offer e-banking customers convenient intra-bank payment and account information services. For example, the Mint app aggregates financial data from a customer's different financial services accounts to present one coherent picture of that customer's finances. To access this financial data, third party personal financial management (PFM) services, like the Mint app, request the customer's e-banking access data, essentially asking the customer to hand over the keys to their bank account. The third party then uses the customer's access data (such as account password, pin number, etc.) to log in to the customer's online bank account the same way a customer accesses their account via their bank's website. Once inside, the third party then transfers the customer's account information to the PFM using a process called "screen scraping," whereby the information is copied and transferred into the customer' PFM account, aggregated according to the particular features chosen by the customer. While this may be convenient for consumers, encouraging accountholders to give their bank account passwords to third parties so that they may essentially copy and paste their financial data into a mobile phone app raises serious data security and privacy concerns.

Impersonation Issues

First, by logging into the using the customer's access data to log into their account, the third-party is essentially impersonating the customer. This impersonation approach is the same approach used in phishing attacks, and the customer's bank has no way of identifying whether the individual accessing the account is actually the customer, a third-party with the customer's permission, or some other party with criminal intent. Further, every time the customer refreshes their app or uses a certain feature, the third-party logs back into the customer account to scrape more data and deliver it to the platform. This creates problems for banks' monitoring of customer accounts for fraudulent activity, and several large banks have complained that their servers have been overloaded by the increased online account activity brought on by third-party data aggregating services.[1]

Loss of Control

Aside from creating difficulties for financial institutions' fraud monitoring and data privacy protection procedures, banks have warned consumers that "handing over your password to your bank account just isn't a good idea."[2] A working group at Lucerne University analogized the process of allowing third party access to consumer accounts to a customer walking into a travel agency to book a vacation "and then simply logging on the travel agent into your e-banking account and then leaving the shop."[3] In his letter to shareholders, Jamie Dimon criticized third-party data aggregators, alleging that many collect more data than necessary, hold it for longer than necessary, and often sell that information at the expense of consumer privacy. Dimon also warned that rogue employees of trusted financial intermediaries could use the customer's login information to steal their money or pass their information along to another person with criminal intent.

Consumer Protection

While some have criticized Dimon's comments as scare tactics,[4] most industry commentators agree that repeated copying and transferring of data by third-parties, or partners of third-parties, increases opportunities for information theft and phishing attacks. Banks have also pointed out that it's not clear whether a consumer who willingly provides confidential information to a third-party financial intermediary would be protected if that intermediary were hacked. A PFM data breach could also indirectly harm clients of small financial service professionals that use PFMs to aggregate sensitive client information. Further, data aggregators, including Yodlee, the aggregator that delivers information to Mint users, have come under fire for selling customer transaction data to hedge funds, investor research firms, and even banks.[5] Many PFM services contain terms that allow data aggregators to use the customer's financial data for commercial purposes. While consumers may think to look for these terms when dealing with their bank, the casual PFM app user may not. Banks argue that their duty to protect customer financial data requires that they control the flow of this data to third-parties, even where customers themselves have given the third party permission to access it.

Fintech Companies' Concerns

Security, Access Rights & Innovation

Most in the fintech industry agree that customer data security and privacy is of paramount importance. Instead, many argue that the banks' desire to restrict data aggregators from accessing a customer's full financial profile is incompatible with that customer's right to access their own financial data, which encompasses a right to provide that data to a third-party intermediary of their choosing.

Consumer Choice

First, fintech companies argue that while banks have a duty to protect the financial information of their customers, the information belongs to the customer and the customer has the right to access this data and provide it to third parties if they choose. In support of this position, advocates point to Section 1033 of the Dodd-Frank Act, which gives the CFPB authority to require banks to provide customers' financial information upon their request and in machine-readable form.[6] Many in the industry also believe that consumers should have the right to give any third party permission to access their personal financial data, regardless of whether the customer's bank approves of the customer's decision to share their information. Alternative financial services offered by fintech companies have filled a void in the financial services industry, and most of these services are complimentary to those provided by the banks. Fintech advocates argue that blocking or limiting third party access to customer accounts places restrictions on the customer's ability to control their own financial data and denies them access to beneficial new personal financial management tools.

Blocking Aggregators Is Not A Long-Term Solution to Security Risks

The data security concerns highlighted by the banks are most prevalent when a data aggregator uses impersonation techniques to access and screen scrape the customer's account. Consensual impersonation is also the primary cause of the fraud monitoring difficulties and server slow downs described by banks. The most effective way to eliminate customer impersonation is through multi-factor authentication, a security feature that banks have used to block third party intermediaries from accessing customer accounts.[7] While some commentators have characterized the banks' restrictions as an attempt to stave off competition from data aggregators, banks understand that blocking aggregators is not in their long-term interests. In fact, many have spent considerable resources to accommodate data aggregators, white-listing their IP addresses to avoid triggering fraud alerts and limiting access to low-traffic timeframes so as not to overwhelm web servers. Distinguishing between aggregators and nefarious impersonators has become more difficult with the rising number of PFM providers, some of which are startups with unproven security systems. In 2014, the Financial Services Information Sharing and Analysis Center (FS-ISAC), a financial sector working group, recommended banks and aggregators establish application programming interfaces (APIs) to allow aggregators direct access to customer information through a dedicated portal, substantially eliminating the need for impersonation and screen scraping.[8] While the response has been slow, some banks have taken steps towards API partnerships, whereby designated data aggregators would act as intermediaries between the bank and PFM services providers.[9]

APIs That Restrict Data Flow Will Not Eliminate Screen Scraping

Representatives of established data aggregators have argued that there is a need for aggregators to collect data through both APIs and screen scraping in order to provide the best customer experience. If the industry moved to one data exchange method, it could leave aggregators and their PFMs vulnerable to receipt of outdated or incomplete data from banks. Further, making the APIs accessible only to aggregators chosen by banks could stifle innovation in the industry by blocking startups from accessing to financial data. The proposal to "push" information to third parties has also fueled suspicion that banks will delay updates to their API data feed or refuse to provide full access to the customer's account information. As a result, data aggregators have refused to commit to abandoning screen scraping – a practice that, according to both sides, "nobody likes," – unless banks provide APIs that serve as legitimate alternatives to the full access offered by impersonation.[10]

When considering data portability, what are other countries doing?

<u>European Union</u>

Two major regulatory frameworks govern use of consumer financial data in the European Union: the Payment Services Directive (PSD2) and the General Data Protection Regulation (GDPR). Under the PSD2, banks must give licensed third party financial services providers access to customer financial data where the customer has given explicit consent to that access, and customers have a right to information on the extent of that access. Further, banks are barred from placing restrictions on third-party information access, and must treat payments made through third parties the same as payments made directly by the customer.[11] The European Banking Authority published a consultation paper on the draft Regulatory Technical Standards for PSD2, the language of which proposed the requirement that banks and third party payments providers apply "strong customer authentication" to ensure security of customer bank accounts.[12] While the EBA's standards are still in draft form, some commentators view the proposed standards as an attempt to strike a regulatory balance providing a technology neutral, flexible framework that leaves room for payments and security innovation before the regulation takes full effect in 2018.[13]

The GDPR applies to data controllers and data processors that operate within EU territory, as well as those outside the territory that offer goods or services to, or monitor the behavior of, EU data subjects.[14] Entities that perform large scale or highly sensitive data processing must designate a Data Protection Officers to oversee internal monitoring of data security and privacy programs. Third parties that perform data processing services for data controllers must also implement compliance measures and abide by substantially similar obligations as those imposed on data controllers. Additionally, data subjects must provide unambiguous, informed, and freely given consent to processing of their personal data. Such consent must be recorded, and opt-in/opt-out forms with sufficient information to support an informed decision by the data subject are permissible. The GDPR states that issues arise in regards to “freely given consent” where performance of a service is conditional upon consent to processing data that is not necessary to perform the service. Further, data subjects must be informed of their right to object to data processing for direct marketing. Data controllers must also provide clear notice to consumers at the time their data is obtained, disclosing information such as the length of time their data will be stored and their right to withdraw consent. Finally, the GDPR prohibits the transfer of personal data to countries outside of the European Economic Area and countries designated by the European Commission as providing adequate data protections. While the U.S. has not yet been designated as fully adequate, the E.C. designated approved the Privacy Shield framework for U.S.-E.U. personal data transfers in a commercial context.[15] Corporations outside adequate privacy areas can continue to engage in intra-organizational international data transfers by establishing Binding Corporate Rules, which are EU-approved internal polices that mirror EU data protection regulations and are binding on an the organization and its affiliates.[16]

The UK

The UK has been experimenting with various measures over the past three years to increase competition in the banking market through encouraging account switching. UK banks are supporting a platform that facilitates the moving of information (bill pay, direct deposit) from a prior account to a new account. The Competition and Markets Authority in the UK has recently called for “Open Banking” implementation by early 2018. Under the UK’s approach, consumers would be able to share their data with third parties using secured protocols. The UK is also undertaking other measures to increase account switching, including requiring reporting on bank quality of service and various “prompts” to encourage consumers to consider whether their service and costs are appropriate.

Information Privacy: Re-thinking the “Right to be Left Alone” in the age of Big Data

The concept of information privacy was born at the turn of the 20th century, inspired by concerns with “newspaperization” and the newly invented Kodak camera[17]. In “The Right to Privacy”, Samuel Warren and Louis Brandeis articulated the seminal idea that information privacy should be a legal right - “the right to be left alone”[18].

In the age of Big Data, never before have people seemed to care less about being “left alone”, at least in the sense that Warren and Brandeis might have imagined. With remarkable frequency, individuals voluntarily publish photos and personal information on social media networks in order to participate in an interconnected world[19]. Consumers willingly share behavioral data to companies, in order to enjoy better products and services.

But while the "right to be left alone" might seem like an outdated way to articulate privacy concerns today, policymakers and stakeholders should continue to heed the spirit of Warren and Brandeis' message: information privacy should be a means of individual empowerment. In this discussion, panelists will discuss what that empowerment should mean in the age of Big Data. What is the nature of our interest in information privacy in the age of Big Data? What is the best approach to fashioning a sensible privacy policy for Big Data? What is the best regulatory framework to achieve the goals of that policy?

In the age of Big Data, what is the nature of our interest in information privacy?

The nature of our interest in information privacy shifts continuously with changing technology and generational attitudes. To Warren and Brandeis, information privacy meant minimizing public disclosure of private information.[20] Today, the focus has shifted from restricting public disclosure of personal information to ensuring that individuals could control its access and flow. To quote Alan Westin, privacy is "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others."[21]

In the age of Big Data, the unprecedented flow of personal information and markedly different attitudes about sharing it necessitate another look at the nature of our interest in information privacy. Should it be viewed as a consumer protection interest, a personal dignity interest, a civil liberty interest, a commodity interest, all or none of the above?[22]

Some approaches to reform suggest that the time has come for information privacy rights to converge with property rights[23]. If consumers owned their data, market forces would drive policy protection. Consumers could negotiate with firms about the uses to which they are willing to have personal data put, and businesses would internalize the societal costs of personal data processing[24]. Data ownership could also empower greater consumer autonomy. For example, ownership of banking data could allow consumers to have better control over their access to value-adding financial products and portability over banking services[25].

What is the best approach to fashioning a sensible privacy policy for Big Data? Navigating tradeoffs between transparency, accountability and utility

Today's information privacy policy is fundamentally based on access control[26]. In other words, privacy protection involves providing individuals with control over how personal information is accessed[27]. This is evident in the fact that "notice and consent" is the dominant paradigm of online privacy protection – those who seek to use an individual's private information must provide a "notice", or a presentation of terms, and that individual must "consent" to those terms[28].

But with Big Data, the volume and variety of information flows greatly erode the effectiveness of access control. Notice fails to be useful when it would take the average person 181 hours per year to read every privacy policy applicable to her[29]. Consent to sharing seemingly innocuous "public" data fails to be informed when such data is used to extract sensitive information – for example, when Target guesses which of its customers are pregnant solely by tracking purchasing habits[30]. How should these weaknesses be addressed?

The path of least resistance would be to fortify access control by improving transparency. For example, streamlined, "layered" privacy notices might convey the most important elements in a clear and understandable way with optional links that provide more detail[31]. Or third party organizations could closely vet privacy policies on behalf of individuals and create a standard set of "privacy profiles" with which individuals can associate themselves, creating a marketplace for the negotiation of community standards[32]. But would increased transparency be enough?

Many believe it is not, and that the conceptual focus of privacy protection should shift from controlling the access and flow of information to controlling its use. Doing so would fundamentally shift the burden of privacy protection to the firms that use data. Through "privacy by design", data controllers can bake in privacy protections throughout the different phases of big data value chain[33]. For example, data can be sourced through "clean data" techniques (i.e. anonymization[34], de-identification and encryption), in the same way that the energy sector looks for clean energy sources Or more robust data retention and deletion policies can honor the "right to be forgotten", an increasingly contentious subject of debate in the European Union[35]. But use control is not without its own problems – where it obviates concerns with transparency, it creates new concerns of accountability. For example, it has been argued that the use-based approach taken by the FCRA produced systemic unaccountability, errors that caused people financial harm, and crimes[36].

More creative solutions involve re-imagining the notion of "privacy protection" altogether as a multi-dimensional concept rather than a monolithic one. For example, policy could work toward the "contextual integrity" of information flow, where finely calibrated systems of social norms or rules would govern the flow of personal information in distinct social contexts[37]. Or the use of "privacy substitutes" can facilitate more nuanced trade-offs between privacy and other interests.[38] Or perhaps the concept of "privacy" should be eschewed altogether in favor of robust principles of "information ethics"; firms who engage in consumer research would create small internal committees to ensure compliance with such principles[39].

With all these possible directions, what is the best way to fashion a sensible privacy policy that meets the needs of Big Data? In thinking about solutions, it is also imperative to consider the complex economic and social tradeoffs at work – after all, regulation is not a costless exercise[40].

What is the optimal regulatory framework for privacy protection laws?

Unlike the top-down approach to privacy regulation in the EU, the United States employs a sectoral approach that focuses on regulating specific risks of privacy harm in particular contexts, such as health care and credit[41]. For example, the 1970 Fair Credit Reporting Act (FCRA) protects consumers by providing specific rights to access and correct the information assembled by consumer reporting agencies[42]. The 1996 Health Insurance Portability and Accountability Act (HIPAA) addresses the use and disclosure of individuals' health information by specified "covered entities"[43]. More narrowly, the Children's Online Privacy Protection Act (COPPA) regulates the acquisition of information for marketing purposes about those under 13 years of age.

The sectoral approach places fewer broad rules on the use of data, allowing industry to be more innovative and flexible in its products and services, but it also leaves unregulated potential uses of information that fall between sectors[44]. For instance, it fails to regulate "data brokers" that do not fall within statutory definitions or that operate outside of particular terms of service[45]. Furthermore, those who are subject to the sectoral legislation can simply transfer information to third parties who are not[46].

Many believe that United States privacy law is a "patchwork," and that decades of self-regulation have left the fox in charge of the henhouse[47]. Others still believe in the efficacy of our current legal infrastructure[48]. Does the current regulatory framework make sense for Big Data? Can gaps that exist between sectoral laws simply be plugged, or should the U.S. fundamentally move toward a more comprehensive, top-down approach?

Data-driven Discrimination: Exercising Big Judgment[49] with Big Data

Data-driven discrimination is nothing new. For generations, the government, banks and service providers used geographic data to draw red lines on maps indicating areas in which they would not invest or provide services. But location served as a proxy for race, and "redlined" areas were dominated by racial minorities. Minorities in these redlined neighborhoods were systematically excluded from banking, healthcare, retail merchandise, and even groceries.

With rapid advances in data processing technology, the dark history of redlining should serve as a cautionary tale. With behavioral data and advanced algorithms, firms can draw granular distinctions between individuals and "score" them for a growing range of purposes[50]. Used correctly, these sophisticated scoring processes can scrub away inappropriate human biases or empower inclusion of marginalized groups. But they can also mask biases and perpetuate exclusion – in essence, drawing red lines around discrete groups of people.

This discussion will explore several questions. As a practical matter, how should we prevent algorithmically driven processes from intentionally or unintentionally harming marginalized or disadvantaged groups? As a normative matter, should we continue to encourage the development of a “scored society”[51]? Finally, what role can equal opportunity laws play in the prevention of harmful discrimination?

How do we prevent inequitable results from algorithmically driven processes?

Algorithmic scoring can facilitate new techniques in price discrimination and targeted advertising, through which firms can expand the size of their markets and maximize profits. These practices raise several major issues. Most obviously, algorithms can fuel faulty insights, which harm people by lumping them into the “wrong” group[52]. But even when they operate correctly, they run the risk of unfairly harming historically disadvantaged groups, whether intentional or not.

In theory, price discrimination and targeted advertising should especially benefit historically disadvantaged groups, because they are generally known to be more price-sensitive[53]. But there are many documented examples of the harmful impacts of price discrimination on disadvantaged groups, whether intended or not. Major companies have been known to track information based on physical location to display different online prices to different customers, but offer better deals to higher-income locations than lower-income ones because the poorer areas had fewer retail outlets competing with online stores[54]. Less ethical companies have used data to seek “vulnerable” prospects to exploit with scams and misleading offers[55]. In fact, evidence suggests that targeted advertising helped fuel the subprime mortgage crisis: it steered approximately 30,000 Black and Hispanic consumers into costly subprime mortgages during 2004 – 2009 and charged them higher fees than white borrowers[56].

How should policymakers deal with data-driven price discrimination in a way that maximizes its value but minimizes inequitable results? Should it focus on encouraging competition by improving disclosures rather than limiting differential pricing? Or should it go even further and limit the use of personalized pricing to offline settings?

Should we continue to encourage the development of a “scored society”?

Data and predictive algorithms have enabled firms to develop and use more prevalent and sophisticated “scoring” processes. Scores can range from traditional credit scores to newer “consumer evaluation or buying power scores” that are highly valuable to companies[57]. On the whole, these scoring processes can lead to more accurate insights and fuel the development of more sophisticated products and services. But as a normative matter, is it desirable to continue to encourage the development of a “scored society”?

Enhanced scoring processes can certainly empower greater inclusion. For example, decision-makers can use new kinds of data to determine who gets access to credit, and on what terms[58]. This “alternative data” can empower financially underserved consumers, whose traditional credit reports might not serve as the best indication of their creditworthiness[59]. At the same time, companies can underwrite financial products in instances where a full credit report is unavailable by using a much wider spectrum of alternative data[60].

Still, scoring with Big Data remains an ethically questionable process. First of all, “Raw data” is an oxymoron[61] – human biases, no matter how small, will taint each step of the Big Data process, from collection to interpretation. Less obviously, these processes will necessarily exclude those at Big Data’s margins – for instance, those who lack access to smartphones or are not “plugged in” to the system[62]. Finally, on a more fundamental level, over-reliance on Big Data and scoring might threaten individual autonomy itself. The vast volume and variety of data blurs the line between measurement and manipulation, between privacy and probability, and between free will and the dictatorship of data[63].

What role should equal opportunity laws play in the prevention of harmful discrimination?

A number of federal equal opportunity laws offer protection from data-driven discrimination based on protected characteristics such as race, color, sex or gender, religion, age, disability status, national origin, marital status, and genetic information[64]. To prove a violation of these laws, plaintiffs typically must show “disparate treatment” or “disparate impact.” Disparate treatment involves treating an applicant differently based on a protected characteristic[65]. Disparate impact occurs when a company employs facially neutral policies or practices that have a disproportionate adverse effect or impact on a protected class, unless those practices or policies further a legitimate business need that cannot reasonably be achieved by means that are less disparate in their impact[66].

Big Data and advances in data processing technology make it harder to detect possible violations of equal opportunity laws. On one hand, as predictive insights can be drawn from a wider variety of seemingly more “neutral” data inputs, it becomes harder to “smoke out” improper biases from harmful results[67]. A combination of seemingly innocuous inputs such as zip codes and buying patterns can serve as proxies that hide charged classifications such as race or age. On the other hand, harmful impacts become harder to detect and easier to justify in terms of legitimate business need.

Is it time to re-evaluate these equal opportunity laws to better protect against harmful discrimination from Big Data?

Big Data in Policy Making

In the lead-up to the financial crisis, the U.S. and global financial sectors were over-leveraged, short-funded, risky, and opaque. Shadow banking permitted institutions to avoid comprehensive supervision and capital requirements. Innovation outpaced the ability or willingness of private-and public-sector guardians to rein in risks. An asset bubble fed the system, until the market imploded in the fall of 2008. When the crisis hit, our society found itself ill-equipped to deal with the failure of leading financial firms. The financial crisis crushed the economy, cost millions of people their homes and their jobs, wiped out household savings, and shuttered businesses.

One of the challenges faced by regulators and market participants is data. In both the run up to the crisis and during the crisis itself, data on the holdings of financial firms were hard to access or understand. This meant that neither market participants nor regulators could accurately measure the scope or scale of the problems. Or, when they did receive access to data, they could not reasonably conclude that the data was accurate, if it was stored and transmitted in a format they could access at all. Analyzing the available data was also often beyond the ability of both participants and regulators, and available computational and analytic tools fell behind the needs of regulators. Moreover, both psychological and organizational behavior factors, as well as incentive structures within firms, reduced the willingness and ability of market participants and regulators to gain access to data, to understand the data, and to act on that understanding. Fortunately, there may be new sources of data and new analytical techniques available that could better inform market participants and regulators about risks in the financial system, and better enable them to act on such information. For example, the Office of Financial Research has developed a global legal entity identifier to track connections among institutions through deal-specific data collection.

Big data can dramatically improve our ability to understand risk in the financial system, but only if we overcome barriers to data standardization, data collection, data analytics, and the managerial constraints, incentive problems, organizational design and socio-technical complications.

References

[1] Robin Sidel, "Big Banks Lock Horns with Personal-Finance Web Portals," WALL ST. J., Nov. 4, 2015, http://www.wsj.com/articles/big-banks-lock-horns-with-personal-finance-web-portals-1446683450

[2] Tilos Demos, "Fintech Firm Plaid Raises $44 Million," WALL ST. J., Jun. 19, 2016, http://www.wsj.com/articles/fintech-firm-plaid-raises-44-million-1466377808.

[3] Sofia, "Should FinTech Startups Have Access to Banking Data?" LetsTalkPayments.com, Jun. 21, 2016, https://letstalkpayments.com/should-fintech-startups-have-access-to-banking-data/.

[4] Samantha Sharpe, "Jamie Dimon Doesn't Think You Are Smart Enough to Manage Your Own Financial Data," FORBES, Apr. 26, 2016, http://www.forbes.com/sites/samanthasharf/2016/04/26/jamie-dimon-doesnt-think-you-are-smart-enough-to-manage-your-own-financial-data/#2f2ef0c43ee8.

[5] Bradley Hope, "Provider of Personal Finance Tools Tracks Bank Cards, Sells Data to Investors," Aug. 6, 2015, http://www.wsj.com/articles/provider-of-personal-finance-tools-tracks-bank-cards-sells-data-to-investors-1438914620.

[6] 12 U.S.C. § 5533.

[7] Sidel, "Big Banks Lock Horns with Personal-Finance Web Portals."

[8] Mary Wisniewski, "Is It Tome to End Screen Scraping?" Nov. 7, 2014, AMERICANBANKER, http://www.americanbanker.com/news/bank-technology/is-it-time-to-end-screen-scraping-1071118-1.html

[9] Demos, "Fintech Firm Plaid Raises $44 Million."

[10] Id.

[11] European Commission, Press Release, "European Parliament Adopts PSD2" (Oct. 8, 2015), http://europa.eu/rapid/press-release_IP-15-5792_en.htm?locale=en; Angus McFayden, "Key Features of PSD2 and What They Mean for the Payments Industry," Pinsent Masons LLP, Jan. 26, 2015, http://www.out-law.com/en/articles/2015/january/key-features-of-psd2-and-what-they-mean-for-the-payments-industry/.

[12] European Banking Authority, Consultation Paper, EBA-CP-2016-11 (Aug. 12, 2016), http://www.eba.europa.eu/documents/10180/1548183/Consultation+Paper+on+draft+RTS+on+SCA+and+CSC+%28EBA-CP-2016-11%29.pdf.

[13] Chris Finney, "PSD2: 'Strong Customer Authentication' – What, When & How?" Cooley LLP, Aug. 19, 2016, http://www.lexology.com/library/detail.aspx?g=d0d14bd7-29cd-4876-9f87-ff2146fc0200.

[14] Antonio Martinez et. al., Allen & Overy LLP, The EU General Data Protection Regulation (2016), http://www.allenovery.com/SiteCollectionDocuments/Radical%20changes%20to%20European%20data%20protection%20legislation.pdf.

[15] U.S. Department of Commerce, Fact Sheet: EU-U.S. Privacy Shield Framework (July 12, 2016), https://www.commerce.gov/sites/commerce.gov/files/media/files/2016/fact_sheet-_eu-us_privacy_shield_7-16_sc_cmts.pdf.

[16] European Commission, "Data Protection: Overview on Binding Corporate Rules," (Feb. 8, 2016), http://ec.europa.eu/justice/data-protection/international-transfers/binding-corporate-rules/index_en.htm.

[17] Dorothy J. Glancy, *The Invention of the Right to Privacy*, 21 ARIZ. L. REV. 1, 8 (1979).

[18] Samuel Warren & Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890).

[19] *See* Mary Madden & Aaron Smith, *Reputation Management and Social Media, Pew Internet & American Life Project*, at http://www.pewinternet.org/Reports/2010/Reputation-Management.aspx.

[20] *See* Warren & Brandeis, *supra* note 2, at 196.

[21] ALAN F. WESTIN, PRIVACY AND FREEDOM 7 (1967).

[22] Pamela Samuelson, *Privacy as Intellectual Property*. 52 Stan. L. Rev. 1125, 1170 (1999).

[23] *Id*. at 1127.

[24] William McGevaran, *Revisiting the 2000 Stanford Symposium in Light of Big Data, in*, FUTURE OF PRIVACY FORUM, BIG DATA AND PRIVACY: MAKING ENDS MEET DIGEST 86 (Stan. Law Sch. Center for Internet and Society ed., 2013).

[25] Karl Antle, *The Looming Battle over Customer Data*, THE CLEARING HOUSE, BANKING PERSPECTIVE Q1 (2016).

[26] Lalana Kagal & Hal Abelson, *Access Control is an Inadequate Framework for Privacy Protection* 1 (MIT Computer Science and Artificial Intelligence Lab).

[27] *Id*.

[28] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, REPORT TO THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE, 38 (2014).

[29] Aleecia M. McDonald & Lorrie F. Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: A JOURNAL OF LAW AND POLICY 542, 560.

[30] Charles Duhigg, *How Companies Learn Your Secrets*, THE N.Y. TIMES MAGAZINE (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

[31] *See* Center for Information Policy Leadership, *Ten Steps to Develop a Multilayered Privacy Notice* (White Paper, 2007).

[32] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, *supra* note 12, at 40.

[33] *Report by the European Union Agency For Network And Information Security on "Privacy by Design in Big Data,"* 1, 5 (2015).

[34] In general, as the size and diversity of available data grows, the likelihood of being able to re-identify individuals (that is, re-associate their records with their names) grows substantially. See PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., EXEC. OFFICE OF THE PRESIDENT, *supra* note 12, at 38-39.

[35] *See* Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos, 2014 E.C.R. 317.

[36] Chris Jay Hoofnagle, *How the Fair Credit Reporting Act Regulates Big Data, in*, FUTURE OF PRIVACY FORUM, BIG DATA AND PRIVACY: MAKING ENDS MEET DIGEST 86 (Stan. Law Sch. Center for Internet and Society ed., 2013).

[37] HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY AND THE INTEGRITY OF SOCIAL LIFE, 3 (2010).

[38] Jonathan Mayer & Arvind Narayanan, *Privacy Substitutes: A Thought Experiment*, 66 STAN. L. REV. ONLINE 89, 90 (2013).

[39] Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97, 97 (2013).

[40] Adam Thierer, *A Framework for Benefit-Cost Analysis in Digital Privacy Debates*, 20 GEO. MASON L. REV. 1055, 1056 (2013).

[41] EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES, 17-18 (2014).

[42] *Id*.

[43] Susan Freiwald, *Managing the Muddled Mass of Big Data*, *in*, FUTURE OF PRIVACY FORUM, BIG DATA AND PRIVACY: MAKING ENDS MEET DIGEST 31 (Stan. Law Sch. Center for Internet and Society ed., 2013).

[44] *Id*.

[45] *Id*.

[46] *Id*.

[47] Calo, *supra* note 23, at 15.

[48] Deirdre Mulligan and Kenneth Bamberger argue, in part, that the emergence of the privacy professional has translated into better privacy on the ground than what you see on the books. *See id*.

[49] This term comes from: Shvetank Shah, Andrew Horne & Capella, Jaime, *Good Data Won't Guarantee Good Decisions*, HARV. BUSINESS REV. (2012).

[50] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*. 104 CAL. L. REV. 671, 719 (2016).

[51] *See* Danielle Keats Citron & Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 WASH. L. REV. 1 (2014).

[52] For example, Kevin Johnson, a condo owner and businessman, found that after returning from his honeymoon, his credit limit had been lowered from $10,800 to $3800. The change was not based on anything he had done but, according to a letter from the credit card company, he had shopped at stores. Newman, Nathan, *How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population*, 6 (Public Comment to the Federal Trade Commission, 2014).

[53] EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND DIFFERENTIAL PRICING, 4-5 (2015).

[54] Jennifer Valentino-Devries, Jeremy Singer-Vine & Ashkan Soltani, *Websites Vary Prices, Deals Based on Users*, WALL STREET JOURNAL, Dec. 23, 2012; http://on.wsj.com/Tj1W2V.

[55] The data broker industry even has a term called "sucker lists" for poor, old and less educated groups.

[56] Newman, *supra* note 36, at 6.

[57] Pam Dixon & Robert Gellman, *The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future* 27 (2014)

[58] Robinson + Yu, *Knowing the Score: New Data, Underwriting, and Marketing in the Consumer Credit Marketplace*, 2 (Ford Foundation, 2014).

[59] *Id*.

[60] *Id*.

[61] For a collection of essays regarding inherent biases in data, see "RAW DATA" IS AN OXYMORON, (Lisa Gitelman ed.) (2013).

[62] Jonas Lerman. *Big Data and its Exclusions*, 66 STAN, L. REV. ONLINE 55, 55 (2013).

[63] *See* VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK (2013).

[64] Such laws include the Equal Credit Opportunity Act ("ECOA"), Title VII of the Civil Rights Act of 1964, the Americans with Disabilities Act, the Age Discrimination in Employment Act, the Fair Housing Act, and the Genetic Information Nondiscrimination Act. *See* FED. TRADE COMMISSION, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION, iii (2016).

[65] *Id*.

[66] *Id*.

[67] Barocas & Selbst , *supra* note 34, at 723.