Comments of the World Privacy Forum regarding Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning

Via regulations.gov and via email to 2021-RFI-AI@cfpb.gov

Thank you for the opportunity to comment on Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning, https://www.federalregister.gov/documents/2021/03/31/2021-06607/request-for-information-and-comment-on-financial-institutions-use-of-artificial-intelligence#footnote-5-p16839. The World Privacy Forum (WPF) is a non-profit public interest research group that focuses on consumer data privacy issues, including those relating to emerging technologies, identity, data brokers, AI, health, and other topics. WPF is a non-political, non-partisan organization. WPF focuses on privacy and data governance, and is one of the only NGOs that focuses on objective research so as to produce fact-based consumer data privacy work. Our research, testimony, consumer education, and other materials are available on our webpage, www.worldprivacyforum.org.

We have been researching and working on issues related to AI and ML in the financial sector for over a decade, including testimony before Congress on this issue area. WPF was part of the AI Experts Group that crafted the OECD Principles on AI, now ratified by the US, and now WPF is engaged with the OECD on implementation of the principles.

The RFI has come at an important time in AI and ML development. In 2020, Patrice Alexander Ficklin, Tom Pahl and Paul Watkins wrote that:

> AI has the potential to expand credit access by enabling lenders to evaluate the creditworthiness of some of the millions of consumers who are unscorable using traditional underwriting techniques. These technologies typically involve the use of models that allow lenders to evaluate more information about credit applicants. Consideration of such information may lead to more efficient credit decisions and potentially lower the cost of credit. On the other hand, AI may create or amplify risks, including risks of unlawful discrimination, lack of transparency, and privacy concerns. Bias in the source data or model construction can also lead to inaccurate predictions. In considering AI or other

technologies, the Bureau is committed to helping spur innovation consistent with consumer protections.[1]

WPF agrees with this assessment. In this document, we focus on case studies that we see as having high impact on consumers and privacy, and case studies that point to solutions that could help mitigate the risks. Specifically, these comments focus on:

- Credit scoring models: Beneficial examples of uses of alternative data
- Credit scoring models (Regulated versus unregulated credit scores case study)
- Real time AI-based financial ecosystems and their regulation (FINRA case study)

I. Credit Scoring models: Beneficial examples of uses of alternative data in the thin file context

Some beneficial examples in the AI /ML scoring context are found in the area of "thin file" consumer scoring products. These types of credit scores are well understood in the marketplace. Typically called "alternative credit scores," thin file credit scores are almost always brought in as regulated scores under the FCRA.[2] Alternative credit scores typically use a small alternative data set to calculate thin file scores. Utility payments, rent payments, phone bill payments, and other types of steady payments can be used as predictors for credit risk for people who may not have purchased a home, a car, and may not have an extensive credit history for a variety of reasons.

Exemplars include the FICO UltraFICO, https://www.fico.com/en/products/ultrafico-score and ID Analytics use of alternative credit data https://www.idanalytics.com/solutions-services/credit-risk-solutions/alternative-credit-data/, particularly the Credit Optics Full Spectrum https://www.idanalytics.com/solutions-services/credit-risk-solutions/. These products utilize *alternative data* to provide credit score analysis, and at last check, the companies consider the products to be regulated under the FCRA.

---

[1] Patrice Alexander Ficklin, Tom Pahl, Paul Watkins. Innovation spotlight: Providing adverse action notices when using AI/ML models, CFPB. July 7, 2020. https://www.consumerfinance.gov/about-us/blog/innovation-spotlight-providing-adverse-action-notices-when-using-ai-ml-models/

[2] It is important to disambiguate thin file or alternative credit scores from with *aggregate credit scores*. Companies building *aggregate credit scores* typically do not see these models as regulated under the FCRA, because aggregate scores apply to households, not individuals. This is a loophole in the FCRA, as the FCRA only applies to individuals. Aggregate credit scores that are created at a household level are not regulated, but they nevertheless might be applied to individuals by companies seeking an unregulated predictive score. Aggregate credit scores can use hundreds and up to more than a thousand factors, and can be quite accurate. In short, aggregate credit scores can act as an unregulated proxy for the traditional credit scores originally regulated under the FCRA. This is in contrast to thin file, alternative credit scores, which are regulated scores that can be beneficial to previously unscored consumers or consumers with minimal credit histories.

We acknowledge the benefits of thin file credit scores. However, we remain concerned about the data being utilized for the scoring. We note that there needs to be clear regulation around what data types may be used in the analysis, when, and with what consumer consent. Otherwise, the lack of clarity around what number and types of factors considered appropriate for use in thin file scores could allow unfairness and bias to creep in. We note that many risk scores that exist in the grey area of the FCRA regarding risk scoring similarly lack clear regulatory guidance. In the absence of such guidance, the opportunity for the market to go off course is substantial.

II. Credit scoring models: Regulated versus unregulated models

Credit scores as defined under the Fair Credit Reporting Act (FCRA) and applied to individuals are regulated under the FCRA. However, since at least 2013, the World Privacy Forum[3] has identified and documented "unregulated credit scoring" or "non-FCRA credit scores." These types of scores are not considered by industry to be regulated under the FCRA for three reasons:

1. Non-FCRA credit scores are scored to a household or a small group of households. (Regulated FCRA credit scores are scored to an individual).
2. Non-FCRA credit scores can use elements of traditional credit report data, or they may use third party data, or both, plus additional data types.[4]
3. Non-FCRA credit scores are purportedly used for "marketing purposes" and not for eligibility and other purposes as defined and regulated under the FCRA.

In 2013, WPF testified before Congress about non-FCRA or unregulated credit scores, warning that they were problematic and could create consumer harm. In 2014, we wrote a report called *The Scoring of America* that more fully documented the non-FCRA credit scores.[5] In June 2019, we testified before the Senate Banking Committee about the risks that unregulated credit scores pose to consumers. In our research, we have found that unregulated credit scores are now widespread and are being used on data broker lists and in electronic data append services. We are concerned that the use of unregulated credit scores is poised to create substantial, widespread consumer harm as the use of these scores becomes an entrenched business practice.

Before continuing, we want to differentiate unregulated credit scores from "alternative credit scores." Alternative credit scores are regulated credit scores that use limited third party data to score "thin file" consumers. Alternative credit scores are applied to individuals, and are subject

---

[3] The World Privacy Forum is a non-profit public interest research and consumer education group. Our focus is on conducting research in the public interest regarding privacy, with a focus on health, finance, AI / machine learning, data brokers, identity and biometrics, privacy governance, and other topics. See: https://www.worldprivacyforum.org.

[4] We note that the RFI defined alternative data. WPF is adopting this definition for our comments. The definition is as follows: "For the purposes of this RFI, alternative data means information not typically found in the consumer's credit files of the nationwide consumer reporting agencies or customarily provided by consumers as part of applications for credit."

[5] Pam Dixon and Robert Gellman, The Scoring of America, World Privacy Forum. April 2, 2014. https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/

to the FCRA.[6] We are not referring whatsoever to these kinds of alternative credit scores when we describe "non-FCRA or unregulated" consumer scores.

The unregulated credit scores we are concerned about are credit scores that score and determine risk like a regulated credit score, however, instead of being analyzed initially at the individual level, the scores are analyzed geographically, to one or more households, thus evading the application of the FCRA. Because the FCRA applies only to individuals, not to households, household level credit scores or Zip +4 credit scores are seen at this time as not being subject to the FCRA.

However, we note that even if the analysis is conducted at the household or Zip +4 level, the final analysis can still be applied to individuals, thus acting for all intents and purposes as a regulated credit score. For example, some consumer reporting agencies utilize APIs that allow businesses to look up an address or last name and retrieve the Zip+4 credit score associated with that address, and at the end of the day, the individuals residing ther.

Even though some of the companies selling the scores are careful to state that they can only be used for non-FCRA purposes, it does not mean the scores are actually used in that way. Words on a website do not take away the deeper problems involved with using these scores — including for marketing purposes.

The non-FCRA credit scores go by quite a few different names in the marketplace, including aggregate credit scores, aggregated credit statistics, household scores, Zip+ 4 credit scores, Geo Scores,Geo Credit, modeled credit statistics, or modeled credit scores. There are more names for these kinds of scores, but for clarity, in these comments we are using the terms "aggregate credit score," "unregulated credit score," or "non-FCRA credit score" to refer to these types of credit scores.

When WPF first researched unregulated credit scores, the use of these scores was limited. However, the use of unregulated credit scores has expanded greatly. There are more scores being offered now; for example, all of the major credit bureaus offer an unregulated credit score, plus additional companies also offer variations of these types of scores. Unregulated credit scores have been added to large data broker mailing lists, and are being used "as a service" in electronic data appends at scale. In other words, millions of people can be scored at the household or Zip + 4 level now. Prospect lists can be electronically appended with unregulated credit scores for "marketing uses." Businesses or others can "pull a household credit report" or "Zip +4 credit report" for any person whose address they know.

The unregulated scores have also become much more accurate. As machine learning and other analytic techniques have advanced profoundly, so too, have the sophistication of the unregulated credit scores. Just as the advances in machine learning has transformed accuracy rates in other areas or scoring (biometrics), so too have these advances transformed the accuracy of household-level models, which now boast equivalency and use similar scoring ranges as traditional, regulated credit scores.

---

[6] There are numerous regulated alternative credit scores. For example, PRBC is a Credit Reporting Agency that is registered with the CFPB and it operates under the FCRA. It uses uses "non-traditional information" to score credit, including utility, mobile phone, and other bills. Alternative credit scores are regulated, and are not at all the same as "non-FCRA" household-level or zip +4 unregulated credit scores.

Equifax offers **Aggregated Fico Scores**.[7] This is Equifax's version of an unregulated credit score. Equifax states:

> "Aggregated FICO® Scores offer firms an aggregated, micro-neighborhooded form of the FICO® Score to enhance marketing applications. This enables credit grantors, insurance companies, and other firms to utilize an aggregated version of the industry accepted credit risk assessment measure for non-FCRA marketing purposes.
> Based on the most widely used and accepted risk score in the marketplace, Aggregated FICO® Scores can be used by credit grantors, insurance companies, and other firms for marketing campaigns throughout the customer lifecycle, including to enhance prospecting, segmentation, campaign planning, targeting, and offer development.
>
> Key Benefits
> - Better determines market size and share of total outstanding credit behavior through detailed credit categories
> - Enables credit grantors, insurance companies, and other firms to utilize an aggregated, modeled version of the industry accepted FICO® Score to enhance offline and online marketing
> - Developed solely for non-FCRA marketing purposes
> - Aggregated, modeled scores designed to protect consumer privacy
> - Can be applied immediately as a standalone solution or combined with our other marketing products such as CreditStyles Pro"

Transunion offers **Transunion Audiences**. This is what the company calls a "summary level view of credit profiles at a geographic (Zip +4) level. This is Transunion's version of an unregulated credit score, and the scoring is offered as a service.

"TransUnion audiences are sourced from anonymized, aggregated consumer credit data, delivering valuable credit behavior intelligence. Built from TransUnion's consumer database consisting of more than 230 million U.S. records, aggregated credit data provides a summary level-view of credit profiles at a geographic (ZIP+4) level. TransUnion audiences target the consumers most likely to have the financial ability to qualify and respond."[8]

Analytics IQ offers a **GeoCreditIQ** product,[9] which is its version of an unregulated consumer score. Analytics IQ states that:

> "Credit related data, even summarized at a geographic level, should always come directly from the source – U.S. based credit bureaus. That is the approach AnalyticsIQ takes to create the foundation of our GeoCreditIQ data. By working directly with the bureaus, our GeoCreditIQ data is extremely accurate and predictive. With GeoCreditIQ marketers get the best of both worlds. The data correlates highly to actual credit scores, however, it less restrictive and very powerful in everyday marketing activities."[10]

Experian offers its **Premier Aggregated Credit Statistics** score. The "The Premier Aggregated Credit Statistics product is derived from the credit profiles of more than 220 million credit-

---

[7] Equifax Aggregated Fico Scores. https://www.equifax.com/business/aggregated-fico-scores/

[8] Transunion Audiences, https://www.transunion.com/solution/digital-marketing-audience-segmentation. The quoted material is from

[9] https://analytics-iq.com/what-we-do/ . For a better description, see https://analytics-iq.com/downloads/analyticsiq-productsheet-geocreditiq.pdf.

[10] Analytics IQ GeoCreditIQ brochure, https://analytics-iq.com/downloads/analyticsiq-productsheet-geocreditiq.pdf.

active consumers and averaged at the ZIP Code level."[11] Experian states that this score is "Beneficial to virtually any industry, including debt collections, education, government, financial services, capital markets and data analytics."[12] Experian states that customers can "Get unprecedented insight into the credit health of neighborhoods across the United States." And it also states that it can be used for debt collections, which typically is applied at an individual level. It has used its data to score the top 25 neighborhoods with the most mortgage debt, for example.[13] Experian's zip code credit score is offered as a service.

WPF has found numerous applications and uses of the non-FCRA credit scores, or unregulated credit scores in the market. The uses have spread dramatically from when we first began researching the uses of these scores. Marketers now offer the ability to add unregulated credit scores to marketing lists as a standard addition or "data append," as well as conduct digital list enhancements as part of a more modern data append technique. Data analysts and brokers are also layering multiple data sets with the unregulated "Zip +4" credit scores to create household-level scores. Unregulated scores can be produced at a neighborhood, census block, Zip +4 or the household level, and can be applied to an identifiable address or the name of an individual living there. Unregulated credit scores can be applied to any individual living in that household or Zip +4 area.

For example, Experian's Zip +4 credit score allows an analysis of its median Zip +4 credit score to an individual name level. This can be seen, for example, in ActiveProspect. ActiveProspect uses the Experian Zip +4 Median Credit Score,[14] noting: "ActiveProspect partners with **Experian** to offer theirZIP+4 Credit Score service directly through LeadConduit." On its website, ActiveProspect states:

> "Experian's "Median ScoreX PLUS Score" service allows you to obtain median credit score data on a ZIP+4 area based on a valid mailing address.
>
> The median credit score for the immediate surrounding area can translate into a block, a building, or a single organization.
>
> Experian assigns a descriptive label to the provided lead, based on the bracketing of credit scores:
>
> **Super Prime:** 740+
> **Prime:** 680 – 739
> **Near Prime:** 620 – 679
> **Non-Prime:** 550 – 619
> **High Risk:** <550
>
> For example, if the median credit score for a ZIP+4 address is 800, the lead will be assigned a bracket of **Super Prime."**

The "fidelity of the score" analysis from Equifax data is not just focused on the neighborhood level. It can yield information about last names and single addresses. In this way, Zip +4 credit

---

[11] Experian Premier Aggregated Credit Statistics. Available at: https://www.experian.com/consumer-information/premier-aggregated-credit-statistics.html.

[12] Supra 7.

[13] Experian Blog Post, Zip Codes with the Highest Mortgage Debt, July 22, 2019. https://www.experian.com/blogs/ask-experian/research/zip-codes-with-the-highest-mortgage-debt/

[14] ActiveProspect https://support.activeprospect.com/hc/en-us/articles/205106225-Experian-ZIP-4-Median-Credit-Score

scores (unregulated credit scores) can be applied at an individual level. See Appendix A for the documentation of this data flow.

Numerous companies now offer mailing lists sorted by unregulated credit scores. Hedges Company offers an automotive mailing list with what it calls "inferred credit scores":

> "... AUTOMOTIVE MAILING LISTS WITH INFERRED CREDIT SCORE
> Reach current vehicle owners with a mailing list built on inferred or modeled credit score. These mailing lists can be used as filters for the in-market mailing lists above, or they can be used on their own.
> These automotive mailing lists are households whose purchase history, debt and household income profile can correlate with categories of credit scores. It is important to note these are implied or modeled credit scores, not actual FICO credit scores as listed with Experian, Equifax or TransUnion."[15]

Environics Analytics[16] has a detailed offering of geographic-based credit characteristics called "Neighborhood View."

DMDatabases.com sells an unregulated credit score, it calls it a "modeled credit" list.[17] Its website describes the modeled credit scores: "...First, actual credit data from the three credit bureaus is collected for all households at the block level. Next, we combine the credit bureau data with mortgage deed data, tax assessor's data, past buying history, and hundreds of demographic and lifestyle variables from our master consumer database. The end result is a modeled credit score list that has similar FICO credit scores. The credit scores range from 450 to 850."

NextMark sells a data broker list of "Summarized Credit Scores Fico Like Mailing List."[18] The data card states: "Summarized Credit Scores are used to help our clients target segments of the population at varying levels of credit worthiness. It is carefully built upon the historic financial transaction data of hundred of millions of consumers, aggregated at the zip+4 level." The data card has further recommendations for use:

> **"Recommendations for Banking, Insurance and Automotive Industries:**
> Overlay summarized credit scores on your database to determine credit worthy, or subprime for special finance offers.
>
> Recommendations for mortgage industry:
> Subprime Program: Identify consumers with debt and credit challenges: Choose summarized credit FICO-like ranges of less than 600, specific loan dates and loan amounts or LTV...."

The Dataman Group has "Modeled Credit Score Prospect Lists."[19] The lists include a profitability score, and uses layers of data to score at the household level.

---

[15] Hedges Company Consumer Automotive Mailing Lists https://hedgescompany.com/consumer-automotive-mailing-lists/conquest-inmarket-mailing-lists/

[16] https://www.environicsanalytics.com

[17] DMDatabases https://dmdatabases.com/databases/specialty-lists/modeled-credit-score-direct-mail-email-list

[18] https://lists.nextmark.com/market;jsessionid=624D63468C12F73E52082D474F1C49C9?page=order/online/datacard&id=281247

[19] Dataman Group, Modeled Credit Score Lists, https://www.datamangroup.com/modeled-credit-score-lists/

> "This new ConsumerView Profitability Score list select helps identify households likely to pay their debts and ranks households by profitability, allowing marketers to target the best prospects based on:
> - Profitability
> - Approval Rates
> - Response Rates
>
> The scores align very closely to bonafide Credit Scoring – and with this file – no pre-approval is needed!
>
> The ConsumerView Profitability Score combines a robust scoring model that offers high levels of refinement for selecting the most profitable prospects combined with our top-notch Consumer Database. This gives you greater precision in predicting, identifying and targeting prospects at the Household Level."

These are just a few exemplars of the ways in which unregulated credit scores are being used today.

A. Third party data does not equal unregulated

Even though unregulated credit scores use third party data, this use does not automatically mean the scores are unregulated. The alternative credit scores such as those offered by PRBC are regulated credit scores. Alternative data is considered regulated just as if it were credit bureau data. This creates a strong basis for determining that it is not just the use of traditional credit bureau data that causes the applicability of the FCRA to a score.

Using third party data does not constitute a condition under which a score does not fall under FCRA regulation.

B. **Household-level scores may still be applied to individuals**

Even though companies and credit bureaus creating and using unregulated versions of credit scores make great efforts to explain that the scores are aggregated to a household level data, or census block-level data, or zip +4 data, it does not mean that the data will not be used as a proxy for a credit score of an individual living at that address.

If a credit score is applied to an individual at any point in its lifecycle, then the credit score, even if it is an aggregate, zip +4 modeled score, still must be regulated under the FCRA because it is being applied to an individual.

We stress that as long as a person's home address is known, then a Zip + 4 credit score can be applied to that person as an individual. Additionally, any person who gives a general zip code at a point of purchase, for example, could be scored in near real-time and decisions can be made about that person as an individual based on the zip code of the neighborhood they live in. In this way, too, unregulated credit scores may be applicable to individuals.

C. Credit scores must only be pulled for purposes strictly defined in the FCRA; they cannot be used for general marketing purposes

It is established policy, and law, that credit scores cannot be used for general marketing purposes except in situations expressly defined by the FCRA. Given that unregulated credit scores are accurate proxies for regulated credit scores, and that they are being applied to individuals, the use of unregulated credit reports violates established law and public policy about uses of credit scores. If credit scores were meant to be used for expansive marketing purposes, then the FCRA would permit such uses. In the materials written by unregulated score creators, including the credit bureaus, numerous financial sector, insurance sector,

collections, and many other "marketing" uses are anticipated for the unregulated credit scores. These uses far exceed uses deemed acceptable in the FCRA.

Because scores can and are being applied to individual consumers, the FCRA still applies. Because the FCRA applies, the general marketing uses of these scores is an unfair practice. It is unknown how or if the scores are being used in eligibility circumstances as defined by the FCRA, but given the expansion in the use of unregulated scores, it is probable that such uses have occurred.

We do not see mechanisms that have been made available for making the uses of these scores transparent to consumers. We do not see any effort by credit bureaus to allow consumers to see their Zip + 4 credit scores, nor household scores, nor reveal who has requested their unregulated credit score. We do not see any mechanisms for consumers to correct errors in their unregulated scores, or to prevent other abuses the FCRA and FDCPA were designed to address. We do not know how or if the credit bureaus are affirmatively tracking and monitoring the uses of unregulated credit scores, and we are greatly concerned that these scores may also be easily used both applied at an individual level and used for eligibility purposes. We do not see the credit bureaus and others reporting publicly their technological proof of compliance with the FCRA regarding the unregulated credit scores.

Unfortunately, consumers are not able to avoid the harms involved with unregulated credit scoring. The lists and databases of millions of consumers appended with their unregulated credit scores occur without consumers' knowledge or ability to correct the data. Financial, educational, employment, and other opportunities based on a person's unregulated Zip+4 or household credit score may have profound impacts on individuals, but they will not be able to learn of this.

Unregulated credit scores are a violation are a violation of the FCRA when the scores are applied to individuals, even in marketing situations, because a credit score cannot be used for general marketing purposes.

III. The Rise of Real Time Data and the Need for Real Time Data Governance in AI/ML ecosystems

The global health crisis caused by the COVID-19 virus has accelerated both digitization and utilization of AI/ML. It has also set a spotlight on the increasing use of and need for data that flows in real time or near real time through complex data ecosystems. The combination of AI/ML plus real time data flows is something that simply has not been grappled with yet by regulators. However, of the sectors, the financial sector is well-positioned to utilize existing exemplars to determine regulatory templates. We see FINRA as an extremely important exemplar to utilize in this regard.

FINRA, the Financial Industry Regulatory Authority, is a self-regulatory organization under the Securities and Exchange Act ('34 Act). It is authorized by the U.S. Congress to issue rules under Section 15A(b)(6) of the '34 Act in order to "…prevent fraudulent and manipulative acts and practices, to promote just and equitable principles of trade, and, in general, to protect investors and the public interest and Section 15A(b)(9) of the Act."

FINRA is a key exemplar of real-time information governance. It has moved from a legacy, paper-based reporting system to an AI-enhanced, real-time system that processes an average of 66.7 billion up to 100 billion market transactions every day. Along the way, FINRA has developed core competencies in both technology and multiple aspects of governance that could provide important lessons for the "real-time" world that is rapidly approaching. By looking at FINRA's evolution, it is possible to glean how other information systems might be able to evolve into real-time systems with appropriate, fair, robust, and democratic information governance and privacy guardrails.

In the past, FINRA produced periodic summarized reports to support its mission. This was fine, and entirely appropriate for a paper-based economy and era. From the 1930s when the modern U.S. securities law framework was established through to the present, regulators such as the Securities and Exchange Commission and SROs such as the New York Stock Exchange and the National Association of Securities Dealers (whose SRO powers were eventually transferred to FINRA) had no choice but to rely on periodic reporting from regulated entities as their primary source of information. Staff members of regulated entities spent huge amounts of time boiling down vast quantities of raw data into highly simplified, abstract form for reporting. Then staff members of regulators tried to develop an accurate understanding of the complex reality summarized in the reporting forms through a combination of analysis of the reporting forms and selective audits. These paper-based reporting and regulatory processes were normal and appropriate and used throughout the American economy and world for most of the 20th century.

The computerization of American financial markets was driven in the late 1960s and 1970s by the "paperwork crunch" on Wall Street. As trading volumes increased, paper-based clearing and settlement systems became overloaded, making it impossible to settle all of one day's transactions before the start of the next trading day. The first response to the paperwork crunch was to close markets earlier, which was obviously not a solution that appealed to either financial firms or their clients.

By the end of the 1970s, clearing and settlement systems were running on mainframe computers and American banks, brokerage firms and insurance companies were world leaders in the computerization of their back-office systems. The regulatory financial reporting obligations of these firms were met through a combination of reports generated by mainframe computer systems and information collected and summarized by staff members. These reporting and regulatory oversight processes were based on point-in-time, low-resolution snapshots of the business operations of regulated entities. Regulators could see the equivalent of the tip of an iceberg and were forced to guess the characteristics of the submerged portion of the iceberg. The executives running regulated entities were in much the same position.

In his book, “Seeing Like a State,” Harvard political science Professor James Scott articulated the challenges that modern regulators face when forced to make decisions on the basis of the kind of highly compressed summaries of complex realities found in periodic reporting by regulated entities. The regulator can literally “see” only what is presented in the summary, and on the basis of that kind such summaries, make educated guesses about where to look more closely for evidence of violations of law.

Following the Stock Market Crash of 1987, regulators began working with regulated entities to better understand the operation of their computer systems and to integrate the functioning of those computer systems more directly into their regulatory oversight activities. As regulators gained greater direct access to the information begin generated by the information systems operated by regulated entities, they gradually were able to “see” something closer to what the executives of regulated entities could see.

By the 2000s, financial market regulators such as the SEC and FINRA were developing the capacity to collect and analyze raw data feeds directly from regulated entities. This brings us to today, where FINRA is using the availability of increased technological capacity to acquire real-time transaction data regarding TRACE - eligible securities (Trade Reporting and Compliance Engine). Instead of receiving periodic reports, those subscribing to FINRA’s TRACE reporting system now have firehoses of real time data to manage and analyze.

In the FINRA real-time environment, regulators now have to develop their own capacity to analyze these data feeds and draw their own inferences from them, which requires huge investments in computing capacity and staff with relevant subject matter expertise. After these systems are fully operational, then in theory regulators should be able to “see” whatever data executives at regulated entities can “see.” The starting point of the dialogue between regulators and regulated entities can focus on comparing the results of the regulators’ analyses and the regulated entities’ analyses of the same raw data generated by the regulated entities’ computer systems.

FINRA's TRACE reporting system was developed specifically to assist with this process. To meet its primary mission, FINRA will need to continue to ensure that the kinds of compliance problems they look for, such as concealed shell companies, achieve maximum benefits from the data volume and velocity “real time” affords. “Real time” does not automatically equal “better” unless foundational work has been done to ensure that the data has been properly tagged and organized to facilitate compliance reporting and response. For example, compliance alerts in real-time systems are typically based on some form of trigger. Various kinds of data tags and identifiers are particularly important to construct properly to fulfill this task. With proper triggers in place,

real-time data firehoses can be purposefully and reliably analyzed at scale and at speed in order to create accurate real-time governance feedback.

The ability of regulators to request real-time data from regulated entities and to engage in real-time analysis of that data for evidence of compliance or violations of the law by the regulated entities represents the beginning of a new era of "real-time governance." In a real-time governance system, regulators should be able to respond almost as quickly as regulated entities to evidence of a risk of noncompliance. The expansion of real-time governance in the U.S. and around the world promises a fundamental breakthrough in risk management: citizens should be able to enjoy the best quality goods and services and the benefits of rapid technological innovation while at the same time also being provided better protection from risks.

In order to lay a foundation for continuous improvement of real-time governance systems, regulators and regulated entities will need to collaborate to increase the standardization of data formats. Back in the 1970s, when each financial service firm was installing its own mainframe computer, it was not uncommon for each firm to acquire custom-developed, bespoke software application. Standards were developed for transaction data so that first could sent and receive order and execution information from exchanges and other firms quickly and accurately, but there was no need to standardize other parts of the firms' computer systems.

By the 2000s, the result was significant diversity across firms in the way that some of the information relevant to their reporting obligations was generated and stored. Limited standardization of data formats and software architectures across regulated entities increases the challenges to regulators to move to real-time governance because of their need to compare compliance-related behaviors across different firms with different computer systems.

Lack of standardization of data formats hampered regulators' ability to respond to the 2008 collapse of Lehman Brothers and the 2010 Flash Crash. Regulators' efforts to track down the course of large volumes of computer-generated orders were hampered by the difficulty of comparing data generated by different firms. One problem in particular had to do with lack of standardization in how customers that were "legal persons" (e.g., corporations), were identified. The same corporation's name might be entered into different firm computers differently due to the use of non-standard abbreviations or even typographical errors. The lack of global standards for identifying common ownership of financial accounts by business entities quickly and accurately was hampering tax and anti-money laundering regulatory efforts as well.

In 2011, the Depository Trust & Clearing Corporation (DTCC) and the Society for Worldwide Financial Telecommunications (SWIFT) launched a collaborative, global standard-setting effort that led to the creation of the "Global Legal Entity

Identifier" standard. This standard has been endorsed by the Financial Stability Board and the G20 and designated as International Organization for Standardization ISO standard 17442. Some jurisdictions outside the United States have begun mandating the use of LEI numbers in certain financial service markets in order to increase the effectiveness of regulatory oversight processes (e.g., EU Markets in Financial Instruments Directive known as MiFID II).

Any legal entity anywhere in the world can obtain quickly, easily and cheaply a globally unique 20 digit LEI number from the LEI issuer of their choice, and be confident that it will be accepted by regulators and counterparties around the world for compliance purposes. The LEI Regulatory Oversight Council and the Global Legal Identifier Foundation (GLEIF) jointly administer the LEI system. This includes the oversight of a global network LEI issuers that compete with each other to issue LEI numbers to entities; providing the Global LEI Index, an open, searchable database of LEI numbers, and monitoring emerging technologies and updating the standard as needed to accommodate them.

The LEI ROC and GLEIF provide a clear example of the kind of transparent, accountable and inclusive governance processes that are needed to insure that real-time governance serves the public and is not captured by industry or leveraged by owners of proprietary technologies. The LEI ROC and GLEIF operate in all global markets simultaneously to reduce compliance burdens on regulated entities, amplify the effectiveness of national and global regulators' efforts to protect the public and are completely transparent to end users.

But the public, the regulators that represent the public interest, and private firms cannot enjoy any of those benefits of real-time governance without a very large, one-time investment by the private sector in business process reengineering. That is because all private enterprises today have some system for identifying themselves to their counterparties and keeping track of their counterparties that was developed before the global legal entity identifier standard was developed. The problem from a software programming perspective is similar to the Y2K problem at the end of the 1990s: software programs that only allocated two digits for storing information about years had to be modified to accommodate four digit years in order to insure that the year 2000 was not interpreted by the software as 1900 instead. In a similar manner, all business software systems will have to make a one-time change to adopt GLEI and phase out whatever other system they were using. Depending on how a firm's computer system is organized, this may require undertaking a long, slow, difficult process to achieve what appears to be a simple and obvious outcome to anyone not familiar with the challenges of business processing reengineering.

With regard to the ability of FINRA or any other regulator working with real-time data feeds to fulfill their public service mission through real-time governance processes, increasing standardization of data formats is an essential part of the process of increasing the accuracy of regulators' ability to predict the behavior

of investors, regulated entities and markets generally. The kind of predictions that the use of big data and artificial intelligence make possible are statistical inferences about the probability of different outcomes. The use of data analytics would permit a regulator to estimate the probably that certain data revealed a violation of the law.

Using real-time data flows and real-time governance processes in this way permits regulators to engage in provable, fact-based, and "risk based" regulation. This would permit regulators to adjust dynamically and in real-time their allocation of scarce enforcement resources to those situations where they would create the most value for the public. They could use real-time governance mechanisms to identify those situations where the regulator believes the probability of a violation of the law occurring is the highest and the risk of harm to the public as a result of that violation is the highest, and concentrate their resources there.

The migration by regulators to real-time governance in effect levels the playing field with regard to what the executives of regulated entities know and what regulators know. In addition, regulators gain deeper insight into the behavior of markets generally because unlike the executives of regulated entities who can see in detail only their own firms' internal operations, regulators will be able to learn from comparing detailed, accurate information about operations of all regulated entities.

As regulators give up the 20th century system of regulation based on information contained in point-in-time, low resolution snapshots of the behavior of regulated entities and move to real-time governance instead, regulators will be able to use whatever resources they have more effectively, the public will be better protected and regulated entities will benefit from greater predictability and consistency of regulatory enforcement actions.

It is difficult to overstate the potential significance of the move from 20th century command and control bureaucratic regulatory processes to real-time governance process not just in financial services but in every sector of the American economy and across global markets. In the 19th century, governments could only act as a "night watchman state" because of their limited capacity to regulate the economy. By the 20th century, the modern regulatory state had come into being and could act to protect the public from tainted food, poisonous medicines and lethal workplaces. The Administrative Procedure Act of 1946 (APA) was enacted to insure that the power of the modern regulatory state was exercised in a manner consistent with the rule of law.

The fundamental advances in accountability and effectiveness ushered in by the APA such as notice and comment rulemaking cannot meet the challenge of insuring that regulatory power exercised through real time governance processes also conforms to the rule of law. In order to lay a statutory foundation

for the transparent, accountable and inclusive exercise of regulatory power through real-time governance processes, a fundamentally new approach to regulation is required. FINRA's processes and controls should be utilized as a model and case study by financial sector regulators to understand how the processes are working on the ground. The approaches will need to be adapted to other businesses, and likely much smaller entities.

One potential for adapting regulations for real time governance that would be congruent with the APA would explicitly authorize regulators to leverage voluntary, consensus standards developed by private standard-setting organizations that have committed to observing due process. Public-private collaborations between federal regulators and private sector standard developing organizations (SDOs) have been taking place for decades with the framework of Office of Management and Budget Circular A-119 governing Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities, which had its most recent update in 2016. This is one potential solution for addressing the need for fine-tuning and contextualizing AI/ML systems, particularly for systems operating in whole or in part with real time data flows and analysis.

Real-time financial sector analysis is no longer a single-jurisdiction endeavor. It requires multi-level cooperative efforts. The example of the Global LEI standard demonstrates that the use of a legislative interface through which regulators and private standard-setting organizations can collaborate to achieve real-time governance that serves the public can work any context, not just information privacy law. It also demonstrates that the transparency, accountability and inclusiveness of real-time governance can be supported by cooperative efforts with global standard-setting organizations as well as American standard setting organizations.

How these cooperative efforts are accomplished requires careful and methodical decision making and planning -- private organizations and the public sector both need to be fully committed to insuring the fundamental fairness of their own processes. FINRA's system gives us a view into the implications of the world to come, and the depth of its new technical and policy requirements.

Thank you for the opportunity to comment.

Respectfully submitted,

Pam Dixon
Executive Director, World Privacy Forum