WORLD **PRIVACY** FORUM

Comments of the World Privacy Forum to the Consumer Financial Protection Bureau regarding Request for Information Regarding Data Brokers and Other Business Practices Involving the Collection and Sale of Consumer Information, Docket No. CFPB-2023-0020

Sent via regulations.gov with cc: to DataBrokersRFI_2023@cfpb.gov

Erie Meyer, Chief Technologist and Senior Advisor, Office of the Director
Davida Farrar, Counsel, Office of Consumer Populations
Request for Information Regarding Data Brokers
Consumer Financial Protection Bureau
1700 G Street NW
Washington DC 20552

15 July 2023

The World Privacy Forum is pleased to provide comments regarding the Consumer Financial Protection Bureau's *Request For Information (RFI) regarding Data Brokers,* 88 FR 16951, https://www.federalregister.gov/documents/2023/06/13/2023-12550/request-for-information-regarding-data-brokers-and-other-business-practices-involving-the-collection . The World Privacy Forum (WPF) is a nonprofit, non-partisan 501(c)(3) public interest research group.[1] WPF focuses on multiple aspects of privacy, with governance of complex data ecosystems being among our key areas of work (technical, legal, and policy). We have conducted and published extensive research for 20 years and counting, including original peer-reviewed data and technology governance research published at the highest levels,[2] among collaborative multi stakeholder work at the multilateral level.

Specific to the RFI regarding data brokers, WPF has conducted extensive past and current research and work specific to data brokers and data broker ecosystems. Our reports about data brokers include *The Scoring of America: How secret consumer scores threaten your privacy and your future,*[3] which was the first predictive analytics report analyzing data broker activity in regards to Artificial Intelligence and machine learning. *The Scoring of America* was

[1] World Privacy Forum, https://www.worldprivacyforum.org

[2] *A Failure to Do No Harm: India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the U.S.* Pam Dixon, Springer Nature, Health Technology. DOI 10.1007/s12553-017-0202-6. http://rdcu.be/tsWv. Open Access via Harvard-Based Technology Science: https://techscience.org/a/2017082901/.

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

cited by the Obama White House in its White House Big Data report[4] as well as by the FTC report on data brokers. Another WPF report, *Data Brokers and the Federal Government*,[5] led to positive change in practices regarding certain consumer practices. We have testified regarding data brokers before Congress four times, each time submitting substantive written testimony, and we testified and participated in the Vermont educational hearings process regarding data brokers which led to the nation's first data broker registry. WPF has additional substantive expertise in identity ecosystems, and WPF's executive director was named one of the leading global digital identity experts in 2021. Digital ID and data broker ecosystems are intertwined, something that has not been well-documented in the US Federal work on data brokers.

Beyond our own research and work, WPF co-chairs the data governance working group in the UN Statistical Commission's Global Task force, and participates in the World Health Organization as co-chair of the Research and Academia Network Constituency, and serves on a separate WHO data governance workgroup. WPF participated in the OECD's AI Network of Experts during the drafting of the OECD Recommendations on Artificial Intelligence and currently participates in the AIGO Working Party in three expert working groups, including the AI Foresight Group. You can find out more about WPF's work and see our reports, data visualizations, testimony, consumer guides, and comments at http://www.worldprivacyforum.org.

The CFPB RFI regarding data brokers is broad, and requests information about a broad array of topics in the data broker ecosystem. In considering what would be most useful, these comments outline, describe, analyze, and document the data broker activities at an ecosystem level regarding technical, legal, and policy components of the ecosystem, and what approaches could help mitigate the problems in the ecosystem. Much of the work being done today regarding data brokers does not encompass the ecosystem level structure and dynamics of the issue, and WPF has concerns that this will lead to piecemeal approaches that do not tackle the root challenges. WPF also has concerns that the data broker ecosystem is either at or in the process of passing through a watershed point beyond which mitigations will become less and less possible.

The conversation about data brokers in the US context does not appear to be fully aware of the past, nor even the present state of data broker ecosystems; and it is no wonder; the data broker ecosystem is stunningly complex. These comments attempt, with as much brevity as possible, a snapshot of this ecosystem and its evolution, concluding with how the risks this ecosystem poses might be mitigated.

To begin, these comments examine the exponential curve of the data broker ecosystem by comparing where data brokering was in the late 1980s and early 1990s to where it is today. There is very little documentation of the early data brokering ecosystem anymore, however, WPF has gathered key aspects of this data and presents it in these comments in brief. These comments also discuss the relationship of the Fair Credit Reporting Act to the data broker ecosystem. And finally, these comments briefly touch on the role of digital identity in the data broker ecosystem, a neglected aspect of regulation that needs attention. Taken together, the comments will provide a landscape view of the data broker ecosystem, at technical and policy

[4] *Big Data, Seizing Opportunities, Preserving Values*, Executive Office of the President of the United States (White House Big Data Report).

[5] Robert Gellman and Pam Dixon, *Data Brokers and the Federal Government: A new front i the battle for privacy opens*, 30 October 2013. https://www.worldprivacyforum.org/2013/10/report-data-brokers-and-the-federal-government-a-new-front-in-the-battle-for-privacy-opens/ .

levels, which is essential to understanding how to proceed forward toward mitigations and solutions.

I. Introduction

The modern privacy and governance challenges that data brokers pose to the American public are profound, and they operate in the rarified air of growth curves that are among the most difficult for human brains to process; that is, exponential curves. These are the kinds of curves that have befuddled even the smartest, most well-educated people, and have caused many to underestimate materially important matters such as the speed of climate change, how debt to income ratios work, and today, how data brokers are operating in today's complex and entangled data ecosystems in ways that affect everything from business processes to peoples' and groups of peoples' lives. Data brokers are not operating on a linear curve. They are operating on an exponential curve, and this has consequences for policy and for people.

Willis Ware, the computer scientist who famously worked with John von Neumann building an early computer at Princeton in the late 1940s to early 1950s, is credited with creating the field of computer security in 1970 with his landmark publication, *The Ware Report*. [6] Ware understood exponential curves, he understood computer ecosystems, and he understood privacy. Thanks to these gifts, he saw around a lot of corners. He wrote in the late 1960s: "The computer will touch men everywhere and in every way, almost on a minute-to-minute basis." He penned these words just before he chaired the the famous U.S. hearings [7] that provided the evidentiary basis for the "HEW Report," shorthand for a bedrock report on privacy which stated in full is *Records, Computers and the Rights of Citizens: Report of the HEW Advisory Committee on Automate Personal Data Systems*,[8] which in turn first articulated the Fair Information Practices, or FIPs.

At the time, there was a great deal of concern about the Social Security Number and its potential for abuse, and generally about the "automation of personal data record keeping operations" by the US government. Ware looked at the emerging world of networked computers, databases, unique personal identifiers like the SSN, and third party data use and saw specific risks which he and others at the time worked collaboratively to mitigate.

FIPS became a bedrock for early privacy law; the European Data Privacy Directive, EU 95/46, is in large part dependent on Ware's initial collaborative work on developing the Fair

[6] The 1967 Spring joint Computer Conference session organized by Willis Ware and the 1970 *Ware Report* are widely held by computer security practitioners and historians to have defined the field's origin. See: IEEE Annals of the History of Computing https://dl.acm.org/doi/10.1109/MAHC.2016.48 Willis H. Ware Papers , CBI 40, http://purl.umn.edu/41431; See also W.H. Ware, RAND and the Information Evolution: A History in Essays and Vignettes, RAND, 2008; www.rand.org/pubs/corporate_pubs/CP537.html .

[7] Hoofnagle, Chris Jay, *The Origin of Fair Information Practices: Archive of the Meetings of the Secretary's Advisory Committee on Automated Personal Data Systems (SACAPDS).* Archival text uploaded July 15, 2014. https://www.law.berkeley.edu/research/bclt/research/privacy-at-bclt/archive-of-the-meetings-of-the-secretarys-advisory-committee-on-automated-personal-data-systems-sacapds/

[8] *Records, Computers and the Rights of Citizens: Report of the HEW Advisory Committee on Automate Personal Data Systems*, DHEW Publication No. (OS) 73-94 (July 1973). https://www.justice.gov/opcl/docs/rec-com-rights.pdf

Information Practice principles (FIPS).[9] EU 95/46 was grounded in FIPs, something Ware was proud of.[10] This early EU data privacy law went on to form the backbone of the modernized General Data Protection Regulation (GDPR) in 2018. FIPs is also the structure upon which HIPAA in the US rests, among other privacy laws in the US and around the world. [11]

The data ecosystems Ware saw as being subject to exponential curves in his time have indeed proven to be exactly that; the ecosystems have grown in scope and complexity as he expected, and steep section of the exponential curves are becoming apparent.

Ware was among the first scientists to fully articulate computer and data risks beginning in the late 60s. These comments begin with Willis Ware because today, it will be necessary to think like Ware, but for our time, to see where we are now and to work to anticipate and mitigate what comes next. US policy makers have failed to reign in data broker activities, even when there is clear evidence of harms to people and groups of people resulting from data broker activities. We can and we must do better, or risk being locked into an unhealthy ecosystem. Data broker lock-in is a real possibility at this point, which these comments will explain.

The evidentiary hearings that led to the enactment of the Fair Credit Reporting Act curtailed certain abuses of credit reporting in the United States at that time. These discussions were part of a first, early push in US privacy regulations, which also generally included in the US the Privacy Act and the Family Educational Rights and Privacy Act (FERPA). It is becoming more apparent that the US has reached a point where something similar will need to occur for a changed data era.

II. Data broker ecosystems

Data brokers are not a shiny new topic. There have been extraordinary reports about data brokers and harms resulting from data brokers. Here, these comments begin with a seminal report by Chris Jay Hoofnagle, now a Berkeley Law Professor, who in 2003 published *Big Brother's Little Helpers.* This report stands as the first major modern reporting of data broker activities.[12] In this report, Professor Hoofnagle documents the myriad ways that the US government relies on data collected by third parties, data that has levels of accuracy that are non-transparent and questionable.

Surprised at the time that there was not a regulatory response to Professor Hoofnagle's report, WPF followed on Hoofnagle's work with a 2013 report, *Data Brokers and the Federal*

[9] Robert Gellman, *Fair Information Practices: A Basic History.* http://bobgellman.com/rg-docs/rg-FIPShistory.pdf. A brief introduction is here http:/www.worldprivacyforum.org/2008/01/report-a-brief-introduction-to-fair-information-practices/.

[10] *An Interview with Willis H. Ware, Oral History 356*, Conducted by Jeffrey R. Yost on 11 Auugust 2003, Santa Monica, California. University of Minnesota. https://conservancy.umn.edu/bitstream/handle/11299/107703/oh356ww.pdf .

[11] A current discussion among privacy and AI ethicists is if FIPs is enough for the changed AI -driven ecosystems. An early consensus is forming that the FIPs will not be sufficient, and will need be transmuted and built upon to adapt to the changes.

[12] Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and other commercial data brokers collect and package your data for law enforcement*, EPIC, 2003. https://lawcat.berkeley.edu/record/1118906

Government. This report analyzed the data broker purchasing activities of the Federal Government in light of the then-new OMB guidance regarding its Do Not Pay policy. The Treasury's implementation of the "Do Not Pay" portal included information from a commercial database called the Work Number, a database that was not a government-held database and was not subject to the Privacy Act. The WPF report concluded:

> The government must bring itself fully to heel in the area of privacy. If it is going to outsource its data needs to commercial data brokers, it needs to attach the privacy standards it would have been held to if it had collected the data itself. Outsourcing is not an excuse for evading privacy obligations.
>
> This report discusses new Office of Management and Budget (OMB) guidance for an initiative (Do Not Pay Initiative)[13] that on one hand provides for expanded use of commercial data brokers by federal agencies and on the other it establishes new privacy standards for the databases used in the Initiative. Although incomplete, its extension of privacy standards to commercial databases purchased by the federal government is groundbreaking. As such, this report recommends that OMB should expand its new guidance to cover all government data purchases, bartering, and exchanges from commercial data brokers and databases containing personal information. The problems created by unregulated government use of commercial data sources need to be seen clearly and addressed directly.
>
> If all federal government uses of commercial data brokers are not required to satisfy the new OMB guidelines at a minimum, then the very databases that are supposed to be used for society's benefit will be less accurate, timely, relevant, and complete, and can therefore cause unnecessary and avoidable harms such as garbled identities, blocking individuals from government benefits, and potential misclassification or even law enforcement actions against people due to errors in data. On a broader level, a lack of trust in the government's ability to properly protect fair information rights in a new digital era can be the expensive societal result."

Although Professor Hoofnagle wrote his report now 20 years ago, and although WPF published its report regarding government use of commercial data 10 years ago, the lessons articulated to the US government have yet to be digested and acted upon. In 2023, the Office of the Director of National Intelligence (ODNI) released and declassified a report discussing problems with the US use of commercial data brokers.[14] While this action was the right thing to do, it inadvertently documented the practices that are as of yet still not constrained by appropriate guardrails.

The ODNI report is helpful in several respects in determining the contours of the modern data broker ecosystem. The report, *The Office of the Director on National Intelligence Senior Advisory Group Panel on Commercially Available Information*, approved for release 5 June 2023, documents that the US government intelligence community purchased commercial

[13] OMB Memorandum M-12-11, Reducing Improper Payments through the "Do Not Pay List" (Apr. 12, 2012), available at http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12- 11_1.pdf.

[14] *The Office of the Director on National Intelligence Senior Advisory Group Panel on Commercially Available Information*, approved for release 5 June 2023. https://www.odni.gov/files/ODNI/documents/assessments/ODNI-Declassified-Report-on-CAI-January2022.pdf .

available information, which is described in the report as clearly providing intelligence value. The ODNI report states that commercially available information:

> "…clearly provides intelligence value, whether considered in isolation and/or in combination with other information, and whether reviewed by humans and/or by machines. It also raises significant issues related to privacy and civil liberties. The widespread availability of CAI regarding the activities of large numbers of individuals is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function."

This is a clear indication that data brokers' practices of selling data about consumers to the Federal government is not going to be stopping any time soon, not if national security interests in the US are finding that data useful. It is helpful that the ODNI markedly stated that they saw the data brokering raises significant issues relating to privacy and civil liberties, however, while the World Privacy Forum agrees with the ODNI that data broker activities are a rapidly growing and increasingly significant part of the information environment, it cannot be reiterated too many times that data brokering activities are operating on an exponential curve, where activities are doubling. This means that data brokering will get much, much more prevalent much more quickly than policy makers may realize. Because growth of this type is very difficult to manage as it reaches the upper curves, which is where data brokering is heading now, it is no longer enough to simply state what has been documented for 20 years now: that collection and sale of commercially available information about consumers in ODNI's words, "raises significant issues related to privacy and civil liberties."

WPF's analysis is that government use of data broker data, by ODNI and other national security and law enforcement entities, will require a different set of guardrails than CFPB is considering. For this reason, in these comments we set aside the discussion of this aspect of the data broker ecosystem in these comments save for mentioning one last item: which is that government purchase of data broker data in general provides substantive baseline funding for the data broker ecosystem as a whole. Without government support of data broker activities, the economic fundamentals of the data broker industry would be much weaker. Beyond law enforcement and national security purchases of data broker data, which deserve extensive fresh discussions, so too does purchases of consumer data from data brokers by other Federal agencies.

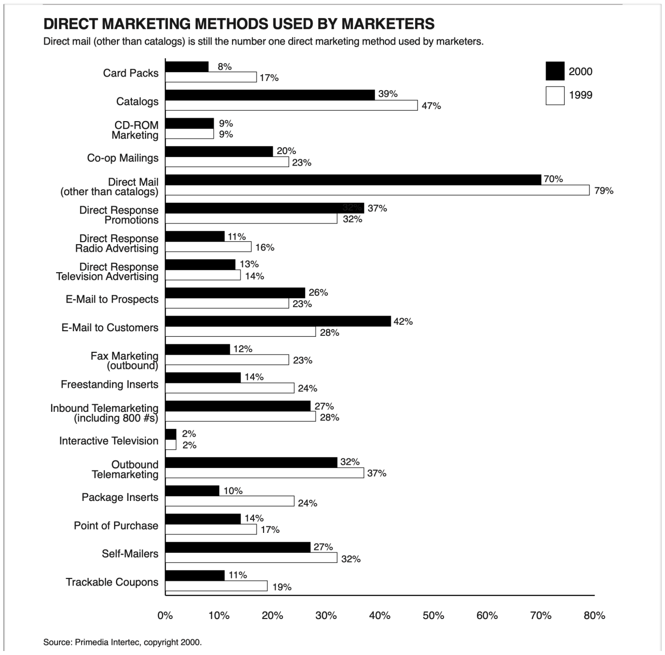
The following overview of data broker ecosystems now turns to a bird's eye view of data broker chronology to show in more detail the longitudinal growth aspect of the data broker ecosystem.

A. Data Broker Ecosystems: The past

Initial data broker ecosystems looked quite different before the Internet. These activities can be largely characterized as direct marketing, data cleaning, and support activities, such as printing and mailing envelopes. Lists of people and their details and preferences were sold via large paper books filled with data cards printed on paper.

These books used to be available in paper formats and were heavy, thick books. As time went on, these same data cards and marketing lists were digitalized, and became databases. Later still, the lists were offered via real-time or near real time APIs.

The data broker ecosystem can be plainly seen in the statistical recording of its activities at the time. Card packs, freestanding inserts, and other paper-based marketing were still in use, but even just from 1999 to 2000, their use was dropping.

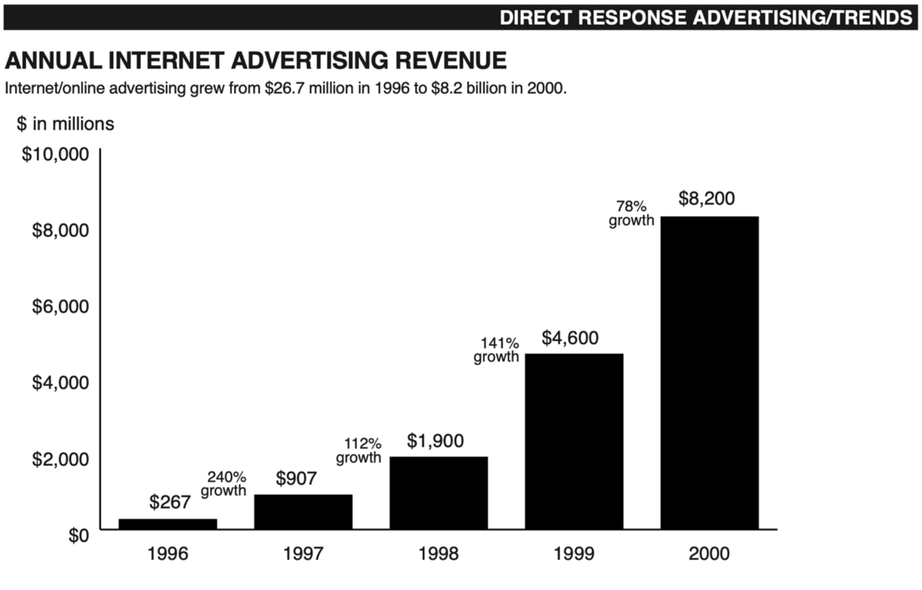

DIRECT MARKETING METHODS USED BY MARKETERS

Direct mail (other than catalogs) is still the number one direct marketing method used by marketers.

Source: Primedia Intertec, copyright 2000.

Source: The 2001 DMA Statistical Factbook, Archival copy via University of Washington. Available at: http://courses.washington.edu/dmarket/2001Factbook.pdf . Page 25.

In the same 2001 DMA volume, annual Internet advertising revenues increased from $26.7 million in 1996 to $8.2 billion in 2000. The growth curve is unambiguous, which tells the story of the late 1990's to early 2000's and the impact of the then relatively young Internet.



ANNUAL INTERNET ADVERTISING REVENUE

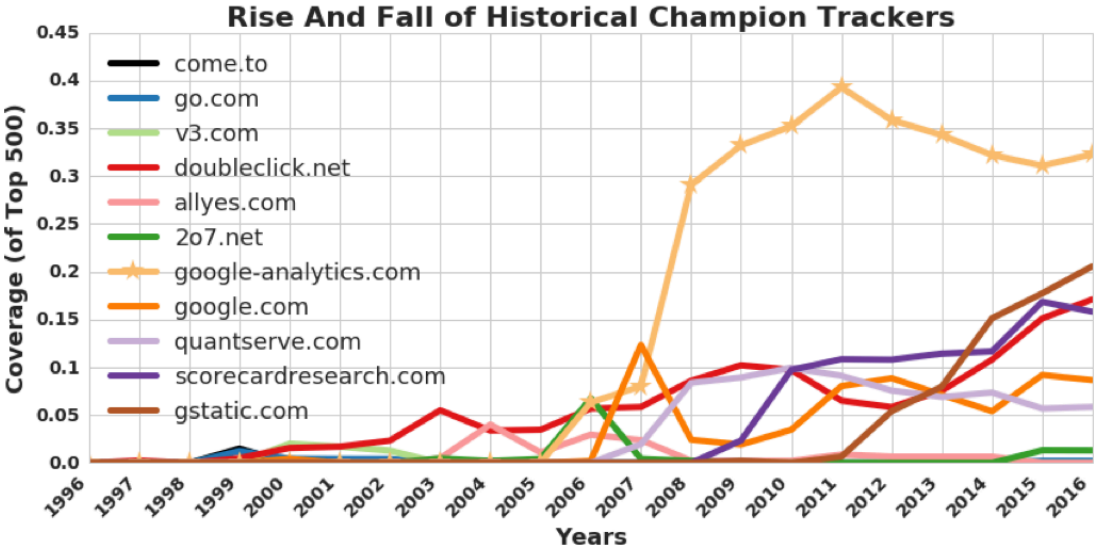
Internet/online advertising grew from $26.7 million in 1996 to $8.2 billion in 2000.

Source: Statistical Fact Book 23rd Edition, Direct Marketing Association, 2001. Archival copy available at University of Washington: http://courses.washington.edu/dmarket/2001Factbook.pdf . Page 29, Annual Internet advertising revenue from 1996 - 2000.

In 2016, a team from the University of Washington published a remarkable study of web tracking from 1996 to 2016.[15] This is an important study, because academic study of web tracking only began in 2005. The UW study utilized a complex technical process to document and fill in the gaps in knowledge for tracking prior to 2005.

Notably, this work documented that in 2016, 90% of the 500 top websites sent information about their visitors to at least one third party.

Here, one of the many data visualizations in the paper shows the rise and fall of historical champion trackers for the top 500 websites. The graph shows the outcomes are comprised of a variety of “tracking curves.” Most curves begin flat in 1996, as the web was just developing, then in 2007, Google-analytics.com trackers begin to demonstrate the start of an exponential curve up until 2011, when it cooled a bit. The rest of the trackers demonstrate more linear curves, which is not a surprising result.



Source: Internet Jones and the Raiders of the Lost Trackers: An archaeological study of web tracking from 1996 to 2016, University of Washington. Proceedings of the 25th USENIX Security Symposium, August 10-12, 2016. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner . Page 1008.

WPF is not remarking on the advertising ecosystem per se here. The point is to demonstrate the historic shifts of tracking online, specifically, on the web. These extraordinary shifts from analog marketing tracking to web tracking correspond to the growth and digitalization of data

[15] Ada Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner, *Internet Jones and the Raiders of the Lost Trackers: An archaeological study of web tracking from 1996 to 2016,* University of Washington. Proceedings of the 25th USENIX Security Symposium, August 10-12, 2016. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner .

brokering. Today, charting tracking would be more difficult, as the digitalization of the ecosystem goes far beyond the web and encompasses myriad devices, IoT, apps, and digital wallets. The broad strokes can be seen in these data: the data ecosystems of our time moved from analog to digital, and they did so fairly rapidly, with exponential curves that began in earnest in 2000, and became more obvious in less than 10 years.

In 2011, WPF testified before Congress regarding data broker harms to consumers. Skimming the testimony will quickly reveal that the general topics of discussion were far more basic at that time than today. The ecosystem was still emerging from the analog world. By 2013, WPF had spent 6 years researching modern data broker practices at that time. Our 2013 Congressional testimony was focused on what we had found. We found numerous data broker lists, which we documented in our testimony. We also discussed something new: consumer scores, and how data brokering was beginning to modernize and turn to machine learning and predictive algorithms to categorize people. This shift to the rise of machine learning in data brokering brings us to the present.

B. Data Broker Ecosystems: The Present

This section of the comments sketches some of today's characteristics of the data broker ecosystem. In a phrase, it is profoundly complex, and this has consequences for crafting legal and policy solutions.

Machine learning, not lists

As mentioned, in 2014, WPF published its *Scoring of America* report, which was 7 years in its research, and was the first report on data brokers to document and modernize the understanding of the data broker ecosystem. In this report, WPF did not focus on lists of consumers that data brokers were selling, because that practice was and still is receding. A new form of data brokering was becoming prominent, which was scoring people, and groups of people, and classifying them into categories and types of people, consumers, purchasers, etc. An era of AI and machine learning was coming, and the *Scoring of America* is a benchmark for the beginning of that era. The report forms a bridge between more analog data broker ecosystems and present-day data broker ecosystems.

The summary of the report states:

> This report highlights the unexpected problems that arise from new types of predictive consumer scoring, which this report terms consumer scoring. Largely unregulated either by the Fair Credit Reporting Act or the Equal Credit Opportunity Act, new consumer scores use thousands of pieces of information about consumers' pasts to predict how they will behave in the future. Issues of secrecy, fairness of underlying factors, use of consumer information such as race and ethnicity in predictive scores, accuracy, and the uptake in both use and ubiquity of these scores are key areas of focus.
>
> The report includes a roster of the types of consumer data used in predictive consumer scores today, as well as a roster of the consumer scores such as health risk scores, consumer prominence scores, identity and fraud scores, summarized credit statistics, among others. The report reviews the history of the credit score – which was secret for decades until legislation mandated consumer access -- and urges close examination of new consumer scores for fairness and transparency in their factors, methods, and accessibility to consumers.

Defining consumer scoring

The World Privacy Forum defines a consumer score as follows:

> A consumer score that describes an individual or sometimes a group of individuals (like a household), and predicts a consumer's behavior, habit, or predilection. Consumer scores use information about consumer characteristics, past behaviors, and other attributes in statistical models that produce a numeric score, a range of scores, or a yes/no. Consumer scores rate, rank, or segment consumers. Businesses and governments use scores to make decisions about individual consumers and groups of consumers. The consequences can range from innocuous to important. Businesses and others use consumer scores for everything from predicting fraud to predicting the health care costs of an individual to eligibility decisions to almost anything.

It is critical to understand this particular aspect of the evolution of data brokering: *data brokering is moving away from lists and databases of consumers.* It has moved toward scoring consumers in clusters, groups, households, and sometimes individually. Consumer scoring is already more widespread than most people realize. Thousands of consumer scores exist, perhaps more. How many Americans have them? Almost all do. Minors are less likely to be scored than adults, although they, too can have or influence some consumer scores. For example, household scores often reflect interests and activities of minors.

Among American adults, each individual with a credit or debit card or a bank account is likely to be the subject of one or more scores. Individuals who buy airline tickets have a score. Individuals who make non-cash purchases at large retail stores likely have a score.

Scores like the medication adherence score, the health risk score, the consumer profitability score, the job security score, collection and recovery scores, frailty scores, energy meter scores, modeled or "aggregate" credit scores, youth delinquency score, fraud scores, casino gaming propensity score, and brand name medicine propensity scores are but a few of the numbers that score, rank, describe, and predict the actions of consumers.

In short, almost every American over the age of 18 has at least one score, and most adult Americans have many scores. An individual could easily be the subject of dozens or even hundreds of secret consumer scores. We can safely predict that there will be many more consumer scores in the future. Fed by the masses of consumer data now available, consumer scoring is quickly becoming a simple shorthand to make sense of a sea of information.

How AI and consumer scoring complicates data control and opt-outs

WPF in the past promoted data broker opt-out mechanisms as a solution for consumers. The ideal of the early 2010 era was to find a way to create a "one stop opt out shop" for consumers. This ideal is no longer relevant today. AI has profoundly complicated the idea of consumers "controlling their data" or "opting out" of data broker ecosystems. The trends of the 2014 *Scoring* report have become much more pronounced and entrenched, and it has grown increasingly apparent that data broker opt out is not a realistic choice for a mitigation solution.

In January 2023 NIST published an influential management standard for AI risks, the *NIST AI Risk Management Framework*.[16] Looking at this framework, it becomes clear that opting out

[16] *NIST AI Risk Management Framework and Playbook*, NIST, January 2023. https://www.nist.gov/itl/ai-risk-management-framework .

can no longer be considered an effective remedy or mitigation for data broker challenges: the ecosystem is too complex, has too many layers, and the data is becoming more diffuse. AI and machine learning operate in systems. While much is written about algorithms, it is in reality the *system* of AI that counts, and the supporting or enveloping ecosystem.

If there is nothing else that comes from these comments, if there is just one thing these comments could impress, it would be that to solve the problems of data brokers today the problems must be seen as a system, and the solutions must be seen as part of a system. The understanding needs to also encompass the total ecosystem of data, technology, and AI-fueled analysis that wraps around these varying systems. The systems of today are stunningly advanced, facilitating analysis of even data that has been de-identified. How does opt-out work on de-identified data? It does not. Consumer data is embedded in systems that are within a larger and opaque ecosystem.

As can be seen in the NIST RMF, today's multi-layered and complex cloud data technology infrastructure combined with the AI processing and analysis that data brokers are utilizing becomes increasingly crowded and complex. The activities of protecting the security of people's data, ensuring it is accurate, and responding to opt-out requests has become increasingly challenging. It is not difficult to forecast that at some point soon, opting out will be understood to be an unworkable solution for consumers.

To give an example of why this is already a near-reality, consider how significant of a challenge it is to keep track of data after it has been replicated, split, and /or fed into algorithmic and machine learning systems. Individual's data or household or census block data might be incorporated into several different intersecting models and data sets, which are then crunched into a score. The score reflecting these groups and households then gets rolled into yet more algorithms and systems. The permutations are extensive, and it is not too much to state that they can be profoundly complex.

Tracking the data applied to produce or operate machine learning models requires an understanding of where and how the data has moved or processed since it was originally collected, and how it may have been used in downstream applications. This is not going to be likely something that data brokers are going to agree to engage in.

The relationship of the FCRA in the evolved scoring / machine learning context of modern data brokering

WPF's analysis is that as consumer scores proliferate, the majority of these new scores do not appear to fall under the narrow protections offered by the Fair Credit Reporting Act or the Equal Credit Opportunity Act for a variety of reasons. Scores built from factors outside a formal credit bureau file, scores designed to predict the behavior of groups of people instead of individuals, and new scores in emerging and unregulated areas may all fall outside of existing protections. For example, it is unlikely that energy consumption scores, churn scores, or identity scores would fall under the FCRA and other laws as currently written. Scores that identify the approximate credit capacity of neighborhoods instead of individuals also appear to be unregulated. As the CFPB knows, the FCRA only applies to individuals. The group / household / category workaround is an important part of the data broker ecosystem of today.

As a result consumers may have scant rights to find out what their non-FCRA consumer scores are, how the scores apply to them and with what impact, what information goes into a score, or how fair or valid or accurate the score is. Even if the input to a score is accurate, consumers do not know or have any way to know what information derived from their lifestyle, health status, and/or demographic patterns is used to infer patterns of behavior and make decisions that affect their lives.

The role of identity - particularly digital identity - in the modern data broker ecosystem

In a digitalized ecosystem, such as data broker ecosystems, digital identity is the key that unlocks just about everything, if not everything. This includes AI systems. This may surprise many, but the US does not have a formal federal digital identity ecosystem, nor federal -level regulatory governance for that system, nor regulatory leadership for that system. NIST has published its draft Identity and Access Management roadmap for digital ID in the US, but a standard alone cannot replace a formal governance structure with meaningful oversight and budgeting.[17] In today's digital world, the US actually is in an undesirable position of having to play catch-up. This has significant implications for how data brokers operate in the US.

This is particularly salient for the data broker ecosystem because identity practices in the US have been moving away from relying solely on personal contact information such as home address to link individual people to information about them. This has shifted as a result of several key factors.

First, technological capabilities have enabled the generation of new forms of data, new ways for people to interact with businesses, and new ways to make data connections. Second, regulatory restrictions on use and sharing of personally identifiable information such as names and postal addresses — along with gradual limits on the effectiveness of online cookies and other mobile identifiers — have compelled businesses to find alternative routes to establishing identity. A change that can be observed today in the data broker ecosystem is that data brokers and identity service providers are evolving to deduplicate and identify otherwise fragmented data about consumers.[18] It is notable that at least two CRAs have significant identity resolution functions.[19]

Much of this activity could be subject to regulation in other parts of the world, but it is not the case in the US. Countries in developing economies often have extremely advanced identity ecosystems, and these ecosystems are highly regulated, and come with authorities dedicated to enforcing those regulations. India, for example, has the world's most largest and most advanced identity ecosystem. It includes 1.4 billion enrollees, and it operates in near real-time to real-time. Biometric authentication facilitates strong authentication, and government services have been attached to the system. Each person has a unique identifying number.

Initially, the ID system, Aadhaar, introduced significant privacy problems, as researchers from WPF documented in research published in Nature - Springer, cited earlier in these comments. In 2018, the Indian Supreme Court overturned parts of India's ID law due to these problems, and required the government of India to put in place extensive technical and policy corrections

[17] *IAM Roadmap*, NIST. https://www.nist.gov/system/files/documents/2023/05/22/NIST IAM Roadmap_FINAL_For_Publication.pdf

[18] See for example Experian's discussion regarding automotive marketing, which is not a part of its FCRA-regulated activities. *Identity resolution: link data to get a better view of your customers*, Experian (Automotive - marketing) https://www.experian.com/automotive/identity-resolution . In its discussion it discusses bringing together fragmented data.

[19] As discussed in this section, Transunion (Neustar) and Experian appear to have developed meaningful identity resolution systems. It is unclear how or if these systems are used in FCRA-regulated activities.

to protect human autonomy and privacy. The government has done so, and the corrections are largely effective.

Now the Aadhaar identity database itself is federated, protected by a mandatory API, and enrollees can use distributed ID techniques that are built into the system to facilitate not having to share their actual ID number with businesses, however, they can still fully carry out transactions. It is an advanced digital backbone that is also privacy-preserving. A dedicated federal ministry manages the ID ecosystem, the ID system has specific federal legislation and regulations, and an ID Authority is at what would be called in the US a cabinet-level position.

This stands in contrast to the US, which has no such infrastructure in place. India's system is not perfect, but it is quite good in the post-2018 era of improvements. It is fair to say that the US has a major digital identity ecosystem problem brewing. Without leadership on how digital identity is managed and protected in the US, it will likely not be possible to solve data broker problems because of the way that data broker ecosystems are now becoming entangled with identity resolution ecosystems. The problems attached to an archaic and comparably unregulated digital ID system could become problematic fairly quickly. These same issues could also be critical in thinking about how to ensure that digital wallets do not become a playground for mischief by players that are not part of the financial services ecosystem or are unregulated.

The limits of privacy in the US ID ecosystem of today and how it relates to data brokering

Consider for example, companies operating in the identity ecosystem that have installed what they consider to be safeguards protecting user data privacy, such as obscuring email addresses and phone numbers through hashing or encryption techniques before the data are shared, [20] or conducting data sharing in so-called "clean room" environments.[21] This sounds good on first blush. However, in some cases, including cases involving regulated Credit Reporting Agencies, hashed email addresses can be used as an identity artifact. This means that an email that has been hashed can become just as good as a name, with work. To quote the literature, a hashed email address can be an "Authenticated starting point for cross-device identity resolution" that "can function like a digital passport that traces every behavior and action a customer takes when logged into an account that is authenticated with an email, making hashed emails a goldmine for customer data." [22]

Identity service providers are reliant on certain match partners to provide consumers with the ability to opt out from the use of their personal information,[23] but as discussed in these comments, opting out is not a serious option in data broker systems using AI, scores, and

[20] *Hashing Identifiers,* Liveramp. https://docs.liveramp.com/connect/en/hashing-identifiers.html.

[21] See for example LiveRamp's mention of clean rooms. *LiveRamp enables identity and advanced activation in Snowflake*, LiveRamp. 27 February 2023. https://investors.liveramp.com/news-and-events/press-release-details/2023/LiveRamp-Enables-Identity-and-Advanced-Activation-in-Snowflake/default.aspx .

[22] *Uncovering hashed email: you may be sitting on a goldmine of customer data and don't even know it*, Experian. 25 August 2021. https://www.experian.com/blogs/marketing-forward/uncovering-hashed-email/ .

[23] Privacy Policy, TransUnion.com / Neustar, note of "Match Partners" in categories of sources. https://www.transunion.com/privacy/neustar .

other machine learning techniques. Industry distinctions between “personal data” and unique or “pseudonymous” identifiers are blurry. And the increasingly multi-layered and interlinked design of the identity ecosystem (digital wallets - credit and debit cards - CRAs that have potential capacity to cross-walk data) could render such opt-out practices ineffectual as meaningful privacy protections.
ODNI

It is also worth reiterating that techniques employed in an attempt to de-identify or anonymize data are not always reliable. As noted above, even a National Intelligence Senior Advisory Group report on commercial available data use by national intelligence agencies states that commercially-available data “can also be combined, or used with other non-CAI data, to reverse engineer identities or de-anonymize various forms of information.”[24]

III. Conclusion

Looking at the past, WPF sees missed opportunities. Looking at the present, we see an extraordinary task before all of us if we want to solve problems for consumers regarding data brokers. Data brokering is built deeply into modern business processes today. The US government is using commercially available data from data brokers. The US lacks appropriate governance for its digital identity ecosystem, which is just now emerging. Of meaningful concern is the interactions between digital identity resolution, the data broker space, and the line between regulated and unregulated data.

WPF does see solutions.

- A key solution is to ensure that additional, modern forms of eligibility are added to the FCRA’s roster of what qualifies as eligibility.

- Also key is to ensure that the emerging “household” loophole is closed. If a score, like an aggregate credit score that is unregulated because it does not use regulated data elements, but it nevertheless acts as a form of a credit score, then these kinds of scores need to be brought under the FCRA.

- Modernized de-identification standards would help, as would rules that do prohibit the use of deidentified data for classifying consumers into certain types of groups which have eligibility or eligibility-adjacent implications. (Such as acceptance into an educational institution, such as a college.)

- A regulated digital ID ecosystem will be necessary, sooner rather than later.

- Also, public sector guardrails for the federal government would be welcome. There is no reason why rules for the public sector would not also apply to government agencies. WPF recognizes that National Security interests would likely need slightly different guidelines, but guidelines will still be necessary.

[24] *Report from the Office of Director of National Intelligence Senior Advisory Group Panel on Commercially Available Data*, approved for release 5 June 2023. https://www.odni.gov/files/ODNI/documents/assessments/ODNI-Declassified-Report-on-CAI-January2022.pdf . See pp. 1 and 5.

The World Privacy Forum urges CFPB to act on these issues; the World Privacy Forum appreciates the opportunity to respond to the RFI, and we look forward to offering assistance and working with you to make progress.

Respectfully submitted,

Pam Dixon, Executive Director, World Privacy Forum

Kate Kaye, Deputy Director, World Privacy Forum

Appendix

WPF developed a taxonomy in 2014 to understand how data brokers were using consumer scores. These scores are not regulated under the FCRA.

Score Taxonomy

In minds of consumers, there is just one score, the credit score. But the credit score is just one final outcropping of a layered and complex taxonomy of scoring. This taxonomy can assist consumers in seeing the full range and depth of scoring activities that exist, and may impact them.

I. Predictive Statistical Models

II. Formal Scoring Models

III. Consumer Scoring Models

IV. Consumer Scoring Model Type (application, behavioral, or combined)

V. Consumer Scoring Function: the broad function of the score card, as follows:

***Propensity* score** cards: will the consumer, for example, default, what is the propensity of a certain result. Credit scoring is a propensity scoring function. Health Scoring is a propensity function if it falls under the full taxonomy preceding this point.

***Response* score** cards: will the consumer respond to a direct marketing offer

***Usage* score** cards: will the consumer use the credit (or other) product if given the product

***Attrition* score** cards: will the consumer continue with the lender, especially if there is some special offer available for an introductory period only.
Customer profit scoring score cards: estimates the total profitability of the customer to the lender

Product profit score cards: seeks to estimate the profit the lender makes on this product from the customer

VI. Source of the Score Model and score (Generic, custom, or vendor supplied score) VII. The Specific Type of Score (fraud, credit, etc.) Here, the term credit refers to the broad type of score.

VIII. Application of Score (what purpose is the score used for)
Consumer-related: test: does the score impact a decision about an individual consumer or a group of consumers?
Research-related: (esp. Health research) test: is the score used to primarily to understand or explain a process or a disease and never used to make a decision about an individual consumer beyond a clinical medical decision? (If a financial or risk decision is taken, then the score becomes a consumer score, not just a clinical score.)

IX. Actual Scores (This includes all specific scores resulting from the taxonomy, Z score, Falcon score, FICO score, etc.) Note: this report is focused on Consumer- related scores, or scores that are used for consumer purposes. If at any point a pure research-related score is used in a consumer score model as a predictive factor and the resulting final score is used for consumer purposes, the final score would be considered a blended consumer score and would be included in the consumer category. See Taxonomy step VII.