WORLD **PRIVACY** FORUM

Comments of the World Privacy Forum to the NTIA, US Department of Commerce regarding the Privacy RFC 2014

Sent via email to privacyrfc2014@ ntia.doc.gov.

National Telecommunications and Information Administration,
U.S. Department of Commerce,
1401 Constitution Avenue NW.,
Room 4725,
Attn: Privacy RFC 2014,
Washington, DC 20230

The World Privacy Forum welcomes the opportunity to respond to the request of the National Telecommunications and Information Administration, Department of Commerce, for comments on "big data" developments and how they affect the Consumer Privacy Bill of Rights. The request for comments appeared in the Federal Register on June 6, 2014, 79 Federal Register 32714, http://www.gpo.gov/fdsys/pkg/FR-2014-06-06/pdf/2014-13195.pdf.

The WPF is a non-profit public interest research and consumer education group. We focus our activities on privacy issues, and much of our work focuses on technology issues as they impact privacy. We have published substantive privacy research and studies over the course of more than a decade, and have testified before Congress and other federal agencies about these topics. You can see our publications and more information at www.worldprivacyforum.org.

I. Introduction

The White House issued its Big Data Report in May, *Big Data: Seizing Opportunities, Preserving Values*.[1] We broadly agree with the premise of the White House report that data can be useful in many ways. We also think that Big Data is wholly unremarkable in many ways. For the last several centuries, the amount of data available for economic activity and public policy purposes has increased significantly every decade. The capabilities to analyze that data also increased at a similar pace. Whether the current rate of expansion of Big Data and analytics is X% or 10X% or even 100X% compared to recent years hardly matters. Data and new analytics

[1] http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

will continue to grow at a great rate in the future, just as they have in the past. We see nothing fundamentally new in so-called Big Data. We will repeatedly use the phrase *Big Data* in these comments, but no one should take our use of the term as acceptance of the concept or of the need to change the world to accommodate Big Data.

Big Data is not a reason to abandon any of our laws or fundamental principles in exchange for the "promise" of Big Data. Our commitment to freedom, liberty, free speech, due process, equal rights, justice, privacy, and other basic ideals must remain unwavering. We find it interesting that in no sphere other than privacy is anyone being asked to modify their rights, benefits, or privileges because of the "promise" of Big Data.

We observe that no one is talking about sacrificing trade secrets or commercial confidentiality for the purposes of Big Data. Big Corporate Data might allow increase efficiency in the marketplace, allow the government to improve its tax collections, and produce better and safer consumer products. We do not understand why the push for Big Data policy adjustment only seems to extend to data about consumers and not to data about corporations. That is just an observation. We do not propose to sacrifice anyone's privacy or confidentially interest on the altar of Big Data.

We reject the notion that we can have either Big Data or privacy but not both. We need to approach Big Data rationally, and as such, we cannot and must not exchange our bedrock principles for vague promises that Big Data will somehow magically transform our lives.

In weighing the benefits of Big Data – and we acknowledge as we must that there are benefits – we must not overweight the benefit side of the scale with unsubstantiated and indistinct promises of infinite gains without cost or compromise. We will always make progress as we have in the past, and it remains wholly unproved that Big Data will increase the pace of that progress in ways that justify changes to any basic principle. We recognize that there are many uses of data, big, medium, or otherwise, that do not affect privacy rights and interests and that may result in benefits to science, medicine, or other endeavors.

We were gratified to see that the White House report recognized that Big Data "raises considerable questions about how our framework for privacy protection applies in a big data ecosystem" and has the potential to "eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace." This is, of course, among our concerns as well.

II. Response to Questions

We have answered selected questions from the Request for Comments. Each question is included here for clarity.

Q. 2 Should any of the specific elements of the Consumer Privacy Bill of Rights be clarified or modified to accommodate the benefits of big data? Should any of those elements be clarified or modified to address the risks posed by big data?

(We have answered Question 2 substantively in our response to Question 8 and in Question 18. We include the responses to those questions here by reference.)

Q 7. The PCAST Report states that in some cases "it is practically impossible" with any high degree of assurance for data holders to identify and delete "all the data about an individual" particularly in light of the distributed and redundant nature of data storage. Do such challenges pose privacy risks? How significant are the privacy risks, and how might such challenges be addressed? Are there particular policy or technical solutions that would be useful to consider? Would concepts of "reasonableness" be useful in addressing data deletion?

The right question is not whether the risks outweigh the potential gains. Privacy rights are *rights*. We do not abandon basic rights because someone alleges that they can build a better mousetrap if we will only give up our rights. Because of this basic problem in approach, this question needs to be reframed.

The Big Data argument is that we have the capability of slicing, dicing, and combining data virtually without limits. Yet the same people claiming infinite capability to process data often contend that they are without the technical ability to provide privacy rights. The disconnect here is striking. Better tools and better analytics can be used to protect privacy and to provide essential privacy rights. Anyone seeking greater use of consumer data must be required to include better protections.

If personal data from multiple sources is to be shared broadly and recombined in ways that will benefit society, we need the same kind of controls that we impose on research and similar protections for the privacy rights of data subjects. One model here comes from the health privacy rules issued under the authority of the Health Insurance Portability and Accountability Act (HIPAA). HIPAA allows non-consensual use of health records for legitimate scientific research. The HIPAA research rule is not perfect, but it strikes a balance of interests between individuals and society.

If other types of personal data are to be used in pursuit of societal benefits, then we should have a similar independent review of the activities, privacy and security policies for the activities, limits on the use of personal data, and assurances that the results will not lead to discrimination or unfair treatment. The research model of using institutional review boards (IRBs) for reviewing research protocols is one policy approach, but it is flawed. IRBs have an inconsistent track record on big data projects, especially in the area of human subjects research. See, for example, the IRB-associated issues with the Facebook mood study, where an academic IRB mechanism did not suffice to protect human research subjects in a number of important ways.[2]

[2] *What OKCupid and Facebook's experiments tell us about big data*, Progressive Media, July 30, 2014. See also *Facebook isn't the only Website running experiments on human beings*, Time, July 28, 2014.

Nevertheless, the IRB model provides a cautious starting point for a new review mechanism for reviewing uses of Big Data. The IRB model would need to be re-imagined, something which academics have been quietly working on for some time. These developments would need to be adjusted to corporate research, which does not fall under the Common Rule, but this work could be done and we see value in it. Again, with caution, and understanding past failures. We also should explore ways of using technological capabilities to find ways to asking consumers for their permission to use their data for other purposes. Consent will not solve all problems, nor will it always be practical, but it will resolve some conflicts.

Note that we suggest controls on *research* that may produce *societal benefits*. HIPAA defines research to be a systematic investigation designed to develop, or contribute to generalizable knowledge.[3] This does not include marketing or market research. We are willing to strike fair balances to support legitimate forms of health research and other beneficial forms of research. Finding better ways to sell toothpaste is a different activity that justifies no privacy concessions.

Further, we observe that too many users of personal data, big or otherwise, expressly seek to exploit, scam, or otherwise cheat consumers. Try as they might, our existing consumer protection institutions cannot protect consumers today, and they will not do any better in a more freewheeling environment with even more personal data. Data brokers today sell lists of consumers who pursue sweepstakes,[4] who respond to loan offers with usurious interest rates,[5] who have used credit repair services,[6] seek or have used debt-settlement companies,[7] and many more, including people who are described as "responders" and "impulse buyers" and are put on a list of people who have had their credit card turned down at point of sale.[8] There is no reason in the name of Big Data to give freer rein to these unscrupulous activities. Make no mistake. Even

[3] 45 C.F.R. § 154.501.

[4] For example, NextMark, Lottery and Sweepstakes Participants Mailing List, Nextmark ID #256114. The data card description of this list reads in part: "This file consists of sweepstakes and lottery players…" <http://lists.nextmark.com/>.

[5] For exmple, NextMark, Pay Day Loan Solutions Mailing List, NextMark List ID #246650. The data card description of this list reads in part: "These consumers are hungry for cash before their next paycheck. For a quick fix, they have applied for a short term - high interest loan to cover there cash flow problem." [sic]. < http://lists.nextmark.com/>.

[6] For example, NextMark, Credit Me I'm Ready! Mailing List, Nextmark ID #331697. The data card description of this list reads in part: "All of these consumers have paid for credit repair in order to re-establish their credit against derogatory credit files." < http://lists.nextmark.com/>.

[7] For example, NextMark, Debt Consolidation Seekers Mailing List, Nextmarke ID #227704 < http://lists.nextmark.com/>.

[8] For example, NextMark, EVantage Credit Card Turn Downs Mailing List, Nextmarke ID #274269< http://lists.nextmark.com/>. The data card description reads in part:"These responsive online EVantage™ Network buyers have responded to online and off-line sales offers, and had their credit card turned down at point of sale. These responders have well documented their high response for direct mail and telemarketing offers. These responders have also expressed an interest in additional promotions and opted-in to receive a variety of third party offers. Most of these impulse buyers also registered with a valid cell phone number while completing the purchase. This self reported age 18+, third party verified database is perfect for subprime financial or credit repair offers."

if it is a small number, a non-ignorable subset of those who argue that we must not let privacy stifle "innovation" appear to want better and more efficient ways to use information in ways that ultimately harm consumers, including those who use information to discriminate and to commit crimes. Any rules need to reduce the bad actors and protect good uses.

We do not think that Big Data offers any excuse to diminish or undermine privacy rights. Nothing we say here changes our view on that point. It is incumbent on those who seek concessions on the grounds of societal benefit to separate socially beneficial activity from mundane, commercial, and fraudulent activities. Not all forms of commerce are socially beneficial to the extent that compromises on privacy can be justified. We repeat that too many companies use data to exploit, scam, or otherwise cheat consumers. As the World Privacy Forum consumer scoring report shows[9] (and we discuss this report in more detail later), it is also easy for commercial enterprises to use consumer data to discriminate against classes of consumers as well, even if that discrimination is unintended. Predictive analytics applied to big data is nuanced, and even the simplest factors can end up creating problematic issues.

Can *reasonableness* help in deciding issues about data deletion? Reasonableness is a useful if elusive concept. Data deletion is helpful in addressing privacy rights and limiting the consequences of personal data to individuals. Many privacy issues are completely resolved when personal information is permanently deleted. Saving personal information that is no longer needed because of the prospect that someday, someone might possibly dream up a use for it and that use possibly might be beneficial cannot be justified for most information. Third degree contingent benefits cannot overcome the value of data deletion. We recognize, however, that in some areas such as health care, there is justification for the long-term maintenance of health records. Data retention must be the exception, not the norm.

We recognize the difficulties of completely deleting data with today's methods of data storage and data connectedness. One response is simple. If personal information can be practicably deleted, it should be. If, for good reason, the information cannot be deleted today, then its *use* should be expressly restricted so that the data cannot be used in any way that may affect the data subject until the data has been finally and completely deleted. Controlling use is one way to balance privacy interests until data is finally deleted or made truly nonidentifiable. There may be a point in time when a technology or process allows much great de-identification than is possible today. When that time comes, policies can then shift more toward de-identification.

Q 8. The Big Data Report notes that the data services sector is regulated with respect to certain uses of data, such that consumers receive notice of some decisions based on brokered data, access to the data, and the opportunity to correct or delete inaccurate data. The Big Data Report also notes that other uses of data by data brokers "could have significant ramifications for targeted individuals." How significant are such risks? How could they be addressed in the context of the Consumer Privacy Bill of Rights? Should they be? Should potential privacy legislation impose similar obligations with respect to uses of data that are not currently regulated?

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

(We include our previous answer to Question 7 here by reference, as there is some overlap.)

Regarding data brokers, we strongly support privacy legislation that requires all commercial personal information record keepers to comply with the full set of Fair Information Practices, and this includes data brokers large and small. Certainly the risks to consumers from data brokers is highly significant.

We strongly support the requirement of all data brokers to offer an opt out to consumers. We also strongly support a national standard of opt out, so that consumers who choose to opt out have consistency of approach across potentially hundreds of opt outs. We also strongly support a national data broker opt out portal that will facilitate consumers opting out once, instead of hundreds of times. This could be structured like the currently existing Opt Out Prescreen mechanism.

The World Privacy Forum maintains a data broker opt-out list of more than 300 data brokers. Only roughly a third of these data brokers offer opt out. Some of them charge fees. Some of them require scans of driver's licenses to opt out. Yet others allow only a partial opt out. Others yet do not allow opt out at all. This is a challenging situation for consumers. But victims of violent or other crime, domestic violence, and members of law enforcement are particularly vulnerable to potential problems associated with difficult or non-existent opt out.

Here are our specific data broker recommendations:

National data broker list

The Federal Trade Commission or the Consumer Finance Protection Bureau should require the industry to maintain a current list of all data brokers, with full identification, description, and contact information.

National consumer data broker opt out requirement

There is an urgent need for a national consumer data broker opt out requirement. All data brokers dealing in consumer PII would be required to allow consumers the opportunity to opt out.

The opt out needs to be standardized, and could operate like Prescreen Opt Out.

Consumers would opt out at a central portal, consumer data brokers would be able to download the list of those who had opted out, then data brokers would be responsible for using this dated list to scrub their lists.

National opt out standards:

No use of opt out data for marketing purposes
Standardized language around opt out
Prominent placement on home page of a button or link that says opt out

Notice to consumers that an opt-out request has been received and acted upon
Due process rights for consumers denied an opt out
Consequences for data brokers that do not comply
Opt outs for all without cost or prerequisites and with simple procedures

Reform and oversight of affiliate marketing of consumers' personally identifiable data. Affiliate marketing of consumer information creates very significant challenges for consumers. The businesses selling the data should exercise appropriate and reasonable oversight.

List brokers who are selling PII of consumers must allow consumers to see the lists they are on and opt out. If a consumer is on a list, the consumer should be able to become aware of that. The list could contain inaccurate information, and could have consequences if sold to an insurer or employer.

The sale of lists that endanger lives or safety or wellness should be stopped. There are lists that should not exist. This includes, but is not limited to:

- Lists of people with incurable diseases
- Lists of law enforcement professionals' home addresses and contact information
- Lists of domestic violence shelters
- Lists of people who are in severe financial distress

No secret consumer scores, no unfair factors, and full publication of data elements (but not weights) used in consumer scores.

The expansion of the FCRA to include modern eligibility options.

Proxy credit scoring or modeled credit scoring should be brought under the FCRA.

Limits on the use of sensitive information in scoring and on the sale of health data in all contexts.

Strict disposal requirements and time limits for all personally identifiable data held.

Fair Information Practices should be applied to consumer data broker practices and lists.

Civil and in some cases criminal penalties when there is a breach of the law. Private rights of action for aggrieved consumers

Effective Oversight by the FTC, CFPB, or both.

Suggested Changes to the CPBOR

Regarding the Consumer Privacy Bill of Rights, we would like to see a major problem corrected.

The first element of the Privacy Bill of Rights states:

> Individual Control: Consumers have a right to exercise control over what personal data companies collect from them and how they use it.

This principle suggests that consumer have rights only with respect to ***personal*** data companies collect from them. In other words, the Consumer Privacy Bill of Rights already provides commercial entities with the ability to use much consumer data without regard to consumer rights or consumer control. There is no reason to change the Bill of Rights to accommodate Big Data because it already rejects privacy rights when data is not collected from consumers. What we need is a better balance for consumers and not for data users. Big Data does not need more concessions. The Individual Control principle of the PBOR is flawed in its articulation from our perspective due to the reasons stated above.

The Commerce Department should use the opportunity presented by the White House report to cure the defect in the *individual control* principle. One way to accomplish the purpose would simply put a period after *personal data*. An alternative would strike all words after *personal* data and insert *processed by third parties* after *personal data*. There is no justification for restricting the principle to data collected from consumers.

The context principle suffers from a similar flaw. It says:

> Respect for Context: Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.

We understand why the Department included the context principle, but we have our doubts that the context is as meaningful when so much data is acquired, bought, sold, and otherwise processed by third party record keepers that are totally unknown to consumers. Applying the context concept only to data provided by consumers renders it substantially meaningless because the loophole overwhelms the policy. If a data broker obtains data from a third party, there is no context under the principle and therefore no restriction of any type on the processing of that data. The Commerce Department must return to the familiar Fair Information Practices principles of purpose specification and use limitation that must apply to all commercial data activities. The context principle should be dropped and replaced with standard use limits and purpose specification from Fair Information Practices.

We offer an additional point here. No one today thinks that consumer notice works to inform consumers meaningfully or to restrict the use and disclosure of consumer data. The White House report recognizes this limit of notice. Notice will not work any better with respect to Big Data than it does today on websites and elsewhere. There is no justification for relying solely on notice to consumers as a principal means of protecting privacy. Still, notice serves other purposes and retains its utility as part of a comprehensive set of Fair Information Practices. We observe in passing that n*otice and choice* is so flawed that it is not worth discussing as a means of protecting consumer privacy.

Q 10. The PCAST Report notes that "data fusion occurs when data from different sources are brought into contact and new, often unexpected, phenomena emerge;" this process "frequently results in the identification of individual people," even when the underlying data sources were not linked to individuals' identities. How significant are the privacy risks associated with this? How should entities performing big data analysis implement individuals' requests to delete personal data when previously unassociated information becomes associated with an individual at a subsequent date? Do existing systems enable entities to log and act on deletion requests on an ongoing basis?

The privacy risks of data fusion are significant. Some of these privacy risks are described in the World Privacy Forum's report on consumer scoring. Much of the risk occurs from the choice of underlying analytical factors, the number of factors used, and the specific contours of the data model(s) used to derive meaning from the data.

All big data analytics use some kind of consumer information – characteristics, behaviors, transactions, and/or attributes[10] – that describe an individual or sometimes a group of individuals and have demonstrative ability to predict some kind of behavior. Huge quantities of data may be collected and organized for this purpose – the finding of factors to feed into data models. As noted in the question, some of this data may be aggregate. Each type of analytic derived from big data uses consumer information in different amounts and combinations found in that data, including public and non-public data sets of varying degrees of identifiability.

Some analytical models, especially in the testing phases, can use as many as 1,000 variables or more to create a single score. Some use even more factors. Data fusion readily occurs in an environment with high numbers of factors, or which use factors that include rarity of information (ex, aggregate information about people with very rare disease) and scarcity of information (ex, aggregate information about residents of a small town in a less populous state).

Even though it may be unintentional, by layering and combining many data sets with high numbers of factors, it can be quite possible to re-identify individuals. This type of data fusion is becoming easier and easier as consumer data becomes ever more of a commodity that is inexpensive and can be readily purchased.

Whether anyone thinks that these activities involve Big Data, we think that the data processed today by data brokers is *big enough* for us. Regulating the deletion of re-identified data from previously aggregate data sets would be an extraordinary hurdle. We need to give individuals basic privacy rights over data broker data and over all other commercially processed data first. If we legislated Fair Information Practices, then issues of re-identification would be much easier to address. If we controlled socially beneficial uses of data through independent review and oversight, we could also address these same issues more readily. With these protections in place,

[10] Characteristics are, for example, the questions asked on the credit application. Characteristics can also be performance categories of the credit bureau report. Attributes are, for example, the answers given to questions on the application, or entries inn the credit bureau report. Education is a characteristic, college degree or highest level of education achieved is the attribute.

then re-identification could be readily controlled through contracts and chain of custody agreements defining allowable uses and disclosures.

Q 11. As the PCAST Report explains, "it is increasingly easy to defeat [de-identification of personal data] by the very techniques that are being developed for many legitimate applications of big data." However, de-identification may remain useful as an added safeguard in some contexts, particularly when employed in combination with policy safeguards.
How significant are the privacy risks posed by re-identification of de-identified data? How can de-identification be used to mitigate privacy risks in light of the analytical capabilities of big data? Can particular policy safeguards bolster the effectiveness of de-identification? Does the relative efficacy of de-identification depend on whether it is applied to public or private data sets? Can differential privacy mitigate risks in some cases? What steps could the government or private sector take to expand the capabilities and practical application of these techniques?

Good predictions require good data quality. Current datasets may be easy to collect, but inaccuracies may be significant and variable across datasets and can be made worse if data linkages are of poor quality. In looking at aggregated data, inaccuracy may be less of an issue. But if large datasets are used to make decisions around individuals, particularly identifiable individuals, then errors stemming from either underlying factors or analytic model error rate can be problematic and deserve policy attention. Privacy principles that call for data destruction (or de-identification) and tying data uses to original purposes remain important. Large datasets cannot be exempt from data quality principles. Ultimately, high data quality is good for all parties involved.

Even though de-identification cannot be assured in all cases, we can use legal and contractual approaches that will both allow legitimate data sharing and provide remedies to data subjects if their data is improperly used.[11] It is likely that the only effective way to control re-identification is through a chain of custody where data passes from one hand to another under a contractual or other type of agreement that establishes the terms of use and sanction in a meaningful way any inappropriate re-identification.

Sanctions should be more severe in cases where it is harder to detect re-identification. Any uses of data for discriminatory, unfair, or privacy-invasive activities should also be strictly sanctioned.

Regarding public or private data sets, this is an important point to explore. Public data sets can be very helpful to analyze for societal benefit. But gauging appropriateness of uses is of particular importance here, especially in regards to protecting vulnerable populations which may

[11] See, e.g., *The Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 Fordham Intellectual Property, Media & Entertainment Law Journal 33 (2010), http://iplj.net/blog/wp-content/uploads/2013/09/Deidentificatin-Dilemma.pdf (The main purpose of this article is to offer a legislative-based contractual solution for the sharing of deidentified personal information while providing protections for privacy. The legislative framework allows a data discloser and a data recipient to enter into a voluntary contract that defines responsibilities and offers remedies to aggrieved individuals. This idea won't solve all problems with the sharing of deidentified data for research or other activities, but it would establish rules and enforceable standards under a statutory scheme. The basic idea is for a federal law, but a state could enact the proposed statute for data within its borders.).

be the subject of data sets. We have discussed appropriateness of uses in our answer to Question 12, and we incorporate that answer by reference here.

Q 12. The Big Data Report concludes that "big data technologies can cause societal harms beyond damages to privacy, such as discrimination against individuals and groups" and warns "big data could enable new forms of discrimination and predatory practices." The Report states that "it is the responsibility of government to ensure that transformative technologies are used fairly" and urges agencies to determine "how to protect citizens from new forms of discrimination that that may be enabled by big data technologies." Should the Consumer Privacy Bill of Rights address the risk of discriminatory effects resulting from automated decision processes using personal data, and if so, how? How could consumer privacy legislation (either alone or in combination with anti-discrimination laws) make a useful contribution to addressing this concern? Should big data analytics be accompanied by assessments of the potential discriminatory impacts on protected classes?

The CPBOR must address the risk of discriminatory effects resulting from big data analytics. We agree that there is a need to worry about the potential discriminatory impacts on protected classes. We think, however, that the problem goes beyond that. In the current environment – and it will only get worse without meaningful controls – data may be used to unfairly affect everyone.

WPF published a deeply researched report about predictive analytics, The Scoring of America. We include here a brief summary of the report, as it is responsive to the RFC question:

> 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.

We repeat here the recommendations from the consumer scoring report[12] because most respond directly to concerns about Big Data. While there are other uses for Big Data, consumer scoring – when done in secret and using hidden, unfair, and potentially discriminatory factors – may be the poster child for anti-consumer and anti-privacy Big Data activity. Our consumer scoring

[12] Supra note 8.

recommendations track Fair Information Practices and emphasize transparency, purpose limitation, and consumer rights.

The **key recommendations** to mitigate this problem include the following ideas:

> Consumer scoring is not inherently evil. When properly used, consumer scoring offers benefits to users of the scores and, in some cases, to consumers as well. Some uses are neutral with respect to consumers. Consumer scores can also be used in ways that are unfair or discriminatory. The goal of these recommendations is to protect the benefits of consumer scoring, guarantee consumer rights, and prevent consumer harms.
>
> • No secret consumer scores. No secret factors in consumer scores. Anyone who develops or uses a consumer score must make the score name, its purpose, its scale, and the interpretation of the meaning of the scale public. All factors used in a consumer score must also be public, along with the nature and source of all information used in the score.
>
> • The creator of a consumer score should state the purpose, composition, and uses of a consumer in a public way that makes the creator subject to Section 5 of the Federal Trade Commission Act. Section 5 prohibits unfair or deceptive trade practices, and the FTC can take legal action against those who engage in unfair or deceptive activities.
>
> • Any consumer who is the subject of a consumer score should have the right to see his or her score and to ask for a correction of the score and of the information used in the score.
>
> • There are so many consumer scores in existence that consumers should have access to their scores at no cost in the same way that the law mandates credit reports be available at no cost, as mandated by Congress. Otherwise, if a consumer had to pay only one dollar for each meaningful score, a family could easily spend hundreds or thousands of dollars to see the scores of all family members.
>
> • Those who create or use consumer scores must be able to show that the scores are not and cannot be used in a way that supports invidious discrimination prohibited by law.
>
> • Those who create or use scores may only use information collected by fair and lawful means. Information used in consumer scores must be appropriately accurate, complete, and timely for the purpose.
>
> • Anyone using a consumer score in a way that adversely affects an individual's employment, credit, insurance, or any significant marketplace opportunity must

affirmatively inform the individual about the score, how it is used, how to learn more about the score, and how to exercise any rights that the individual has.

• A consumer score creator has a legitimate interest in the confidentiality of some aspects of its methodology. However, that interest does not outweigh requirements to comply with legal standards or with the need to protect consumer privacy and due process interests. All relevant interests must be balanced in ways that are fair to users and subjects of consumer scoring.

• The FTC should continue to examine consumer scores and most especially should collect and make public more facts about consumer scoring. The FTC should establish (or require the scoring industry to establish) a mandatory public registry of consumer scores because secret consumer scoring is inherently an unfair and deceptive trade practice that harms consumers.

• The FTC should investigate the use of health information in consumer scoring and issue a report with appropriate legislative recommendations.

• The FTC should investigate the use of statistical scoring methods and expand public debate on the proprietary and legality of these methods as applied to consumers.

• The Consumer Financial Protection Bureau should examine use of consumer scoring for any eligibility (including identity verification and authentication) purpose or any financial purpose. CFPB should cast a particular eye on risk scoring that evades or appears to evade the restrictions of the FCRA and on the use and misuse of fraud scores. If existing lines allow unfair or discriminatory scoring without effective consumer rights, the CFPB should change the FCRA regulations or propose new legislation.

• The CFPB should investigate the selling of consumer scores to consumers and determine if the scores sold are in actual use, if the representations to consumers are accurate, and if the sales should be regulated so that consumers do not spend money buying worthless scores or scores that they have no opportunity to change in a timely or meaningful way.

• Because good predictions require good data, the CFPB and FTC should examine the quality of data factors used in scores developed for financial decisioning and other decisioning, including fraud and identity scores. In particular, the use of observational social media data as factors in decisioning or predictive products should be specifically examined.

• The use of consumer scores by any level of government, and especially by any agency using scores for a law enforcement purpose, should only occur after complete public disclosure, appropriate hearings, and robust public debate. A government does not have a commercial interest in scoring methodology, and it cannot use any consumer score that is not fully transparent or that does not include a full range of Fair Information Practices. Government should not use any commercial consumer score that is not fully transparent and that does not provide consumers with a full range of Fair Information Practices.

• Victims of identity theft may be at particular risk for harm because of inaccurate consumer scores. This is a deeply under-researched area. The FTC should study this aspect of consumer scoring and try to identify others who may be victimized by inaccurate consumer scoring.

III. Possible Approaches to Big Data Suggested by the Reports and the Big Data Workshops

Q 13. Can accountability mechanisms play a useful role in promoting socially beneficial uses of big data while safeguarding privacy? Should ethics boards, privacy advisory committees, consumer advisory boards, or Institutional Review Boards (IRBs) be consulted when practical limits frustrate transparency and individuals' control over their personal information? How could such entities be structured? How might they be useful in the commercial context? Can privacy impact assessments and third-party audits complement the work of such entities? What kinds of parameters would be valuable for different kinds of big data analysts to consider, and what kinds of incentives might be most effective in promoting their consideration?

We already addressed many of the elements of these questions in earlier responses. We have additional observations.

Existing oversight and accountability mechanisms are insufficient to police the marketplace for consumer information, including the big data marketplace. We are aware that the Department of Commerce has stated that the FTC is the agency that will enforce violations of, for example, voluntary codes of conduct resulting from NTIA multistakeholder processes. However, without substantial impetus to adopt codes relating specifically to big data use and without a formal pathway for enforcement, this is not a realistic option for all enforcement activities related to big data. The use of big data is going to permeate the society, therefore, a single enforcement agency at work is not sufficient for oversight and accountability, even if the agency performed at 1000 percent efficiency.

Audits and assessments can be valuable tools if they are timely, transparent, and independent. Audits and assessments that are none of the above are exercises in box checking, and they accomplish nothing. The biggest issue with Big Data audits and assessments is who is doing the analysis? If the analysis is part of an independently reviewed research activity seeking generalizable knowledge, that is one thing. We distinguish legitimate socially-beneficial research from an analysis is done in the backroom of a pharmaceutical manufacturer to find more effective ways of marketing high-prices, patent-protected drugs to individuals with high-end health insurance. We would allow the first under a reasonable set of rules, procedures, and restrictions. We see no reason to make any concession for the second.

New policy guidelines that focus on the rules governing the appropriateness of the data uses to begin with will be important. WPF is extremely interested in developing such policies.

An important use case comes from the EPA. The EPA has enormous data sets, and they have been used for societal benefit without harm to individuals. In a groundbreaking series of articles, the Associated Press used EPA data[13] to map the air quality risk scores for every neighborhood in the U.S. The AP mapped existing EPA toxicity risk scores to socio-economic and racial factors for each neighborhood from the 2000 Census to determine who was breathing the dirtiest air in America. The headlines across the country read, in some variation, that minorities suffer most from industrial pollution.[14] The results established important understandings about neighborhoods and toxicity, and the resulting snapshot of where and how factory pollution affected neighborhoods and people was deservedly much-discussed. These results are examples of an informative and beneficial use of what today would be called large datasets or "big data."

The EPA has a set of meaningful best practice guidelines for analyzing its data. These best practice guidelines are contained in the *EPA Risk Characterization Handbook*. This book discusses EPA's use of risk characterizations in some detail. The EPA's discussion is valuable here:

> "Risk characterizations should clearly highlight both the confidence and the uncertainty associated with the risk assessment. For example, numerical risk estimates should always be accompanied by descriptive information carefully selected to ensure an objective and balanced characterization of risk in risk assessment reports and regulatory documents."[15]

The EPA also created excellent documentation on how the analysis of its own data is to be used.[16] The documentation is for its own researchers, but its quality suggests broader applications are appropriate.

It stated, in part:

[13] See <http://www.epa.gov/risk/health-risk.htm>. The EPA data in this instance help screen for polluted areas in the U.S. that may need additional study and vetting for potential human health problems.

[14] David Pace, *More Blacks Live With Pollution*, Associated Press (Dec. 13, 2005), <http://onlinenews.ap.org/work/pollution/wrap.py?story=./linked_story/part1.html>. See also http://www.nbcnews.com/id/10452037/ns/us_news-environment/t/minorities-suffer-most-industrial-pollution/>. The EPA uses toxic chemical air releases reported by factories to calculate risk for each square kilometer of the United States. The scores allow comparing risks from long-term exposure to factory pollution from one area to another. The scores are based on: the amount of toxic pollution released by each factory; the path the pollution takes through the air; the level of danger to humans posed by each different chemical released; and the number and ages of males and females living in the exposure paths.

[15] U.S. Environmental Protection Agency, Science Policy Council, Risk Characterization Handbook (. December 2000), <http://www.epa.gov/spc/pdfs/rchandbk.pdf >.p. A5.

[16] U.S. Environmental Protection Agency, *Policy for Use of Probabilistic Analysis in Risk Assessment*,

(May 15, 19970, <http://www.epa.gov/spc/pdfs/probpol.pdf>.

> "The methods used for the analysis (including all models used, all data upon which the assessment is based, and all assumptions that have a significant impact upon the results) are to be documented and easily located in the report. This documentation is to include a discussion of the degree to which the data used are representative of the population under study. Also, this documentation is to include the names of the models and software used to generate the analysis. Sufficient information is to be provided to allow the results of the analysis to be independently reproduced."[17]

These recommendations are excellent, and should also apply to large data sets applicable to other areas impacting consumers. Usage guidelines like EPA's, plus guidelines which discuss identifiability of consumers, create important fairness structures (like benchmarks) for many of the uses and applications of large datasets. These benchmarks would go toward improving privacy protections for other big data activities.

Handling problems arising from analysis -- vulnerable populations

Of particular importance in big data policy are guidelines governing vulnerable populations. When problems are uncovered using big data analysis, careful application of the information is necessary. For example, policies that would mandate identifying and protecting victims of abuse, or other crimes, could have an unfortunate reverse effect. No one wants to create a readily accessible list of identifiable or semi-identifiable victims of abuse, while at the same time, the promise of a proper analysis to pinpoint aid distribution and assistance in a timely way to those who need it most would be welcome. The tension here is real, and we have to acknowledge it and resolve it in a balanced way.

We suspect that different applications of large datasets to different populations will warrant slightly different approaches. Again, we are most concerned about privacy-related challenges in the use of big data when the data sets can be re-identified back to specific vulnerable consumer groups, or when the data sets are sensitive and are, or can become, personally identifiable to individuals.

Research is needed to understand how vulnerable populations in particular are affected by analysis and predictions based on such data, and what systematic biases could be potentially introduced into algorithms through faulty data and assumptions. In some cases, even loosely aggregate data has proven problematic, as discussed in our response to the data fusion question, RFC Question 10.

In working to ensure beneficial uses of large datasets in vulnerable or sensitive areas while mitigating potential harm, we share several thoughts.

[17] Id, p. 2.

The Common Rule, The Belmont Report, The Nuremberg Code and big data human subjects research

Of importance in determining large dataset policy in ***identifiable datasets*** (or datasets that become identifiable) is the Common Rule[18] for protection of human subjects of research, and the Belmont Report[19] regarding Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Also deeply applicable is the Nuremberg Code.[20] The Nuremberg Code sets forth an absolute requirement for voluntary consent of all ***human subject research participants***. If big data human subject research is allowed to move forward by collecting research consent passively via a loosely worded privacy policy on a web site, then human subject research ethics moving into the digital era is in very serious trouble. The consequences down the line will likely prove to be ugly. We note that several high-profile examples of big data and human subject research ethics problems already exist.[21]

Informational risks in research must be measured against a firm standard, one that is not affected by every change in technology or commercial practice. For example, the HIPAA privacy standard establishes a firm set of Fair Information Practices. While there is considerable flexibility in the application of the HIPAA privacy rules in some contexts, the standards themselves are not subject to change because of external factors. Patients can expect the HIPAA standards to protect their health information in the same way.

The same should be true for human subjects research, which is increasingly accomplished along with large datasets. The need for a baseline of privacy protection must be a constant for research even though the degree of informational risk can vary from project to project. The need for rules governing collection, use, and disclosure is constant. The need for openness and accountability is a constant. The need to consider individual participation rights (access and correction) is a constant. Thus, whatever the risk involved in a given project, the need for sufficient privacy protections for personally identifiable information is a constant.

We understand that the Common Rule only applies to research done under certain circumstances. But the ***ethical standards*** driving the rule, as articulated in the Belmont Report as well as the Nuremberg Code, are deeply applicable today.

Looking at this issue of identifiable data with more specificity would include for example, ensuring that recourse for discovery of accuracy-related problems is built in to the process. We are interested in policies that develop overall good practices in this area. Accuracy and recourse for correction for individuals identified in health care datasets is a foundational area for further inquiry. Some big data activities have been a part of health and other research for a long time, and there is nothing new in some respects. The demands of researchers can overwhelm existing

[18] < 45 CFR part 46, Subpart A-D. <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html>.

[19] <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>.

[20] < http://www.hhs.gov/ohrp/archive/nurcode.html>.

[21] *What OKCupid and Facebook's experiments tell us about big data*, Progressive Media, July 30, 2014. See also *Facebook isn't the only Website running experiments on human beings*, Time, July 28, 2014.

institutions (like institutional review boards) that do not have the necessary privacy or security expertise.

We support the use of large datasets in medical research, but researcher obligations to protect data and to protect vulnerable populations from problems resulting from analysis need to be defined in law. Any disclosure for health research in large datasets should be limited by law, regulation, and contract as appropriate.[22] HIPAA requirements that protect health information when held by providers and insurers may not apply to corporate researchers, for example, nor the Common Rule.

These ethical research principles derived from the bedrock of the Common Rule, the Belmont Report, and the Nuremberg Code need to be applied to other vulnerable populations undergoing large dataset analysis. For example, financially vulnerable populations are another group deserving of more attention. Aggregate credit scores applied to neighborhoods (versus individuals) are an example of how aggregate but specific predictions based on large datasets may be lead to potentially unfair practices based on primarily geographical factors. If the results of the analysis are not managed correctly or exposed to consumers, errors in prediction may never surface, and other usage issues can arise.

Other examples of important vulnerable populations exist, we do not attempt to be comprehensive here. The overall impetus of the policy guidance should be to identify potential risks for specific populations and in sensitive data, and plan for recourse and checks and balances to mitigate harm or abuse and encourage the best possible uses and results.

Q 14. Would a system using "privacy preference profiles," as discussed in Section 4.5.1 of the PCAST Report, mitigate privacy risks regarding big data analysis?

The proposal has some merit, and we welcome additional development of the concept. However, implementation could be complex, expensive, and overwhelming. Further, too much consumer data is obtained without any involvement of consumers, and it is not clear that there is any way for consumers to express a preference with respect to that class of data. If asked, consumers are likely to express conflicting views about secondary usage of their data, and reconciling those differences would not be sufficient.

Further, some consumers can be expected to change their views over time, and matching current views on the use of disparate data collected at different times by different entities could be so expensive that it is impossible. Simply tracking consumer preferences may create a new type of consumer data that some will seek to use to exploit consumers in a new way.

Q 15. Related to the concept of "privacy preference profiles," some have urged that privacy preferences could be attached to and travel with personal data (in the form of metadata), thereby

[22] See, e.g., Robert Gellman, *The Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 Fordham Intellectual Property, Media & Entertainment Law Journal 33 (2010), http://bobgellman.com/rg-docs/RG-Fordham-ID-10.pdf.

enabling recipients of data to know how to handle the data. Could such an approach mitigate privacy risks regarding big data analysis?

We include here our answer to question 14 by reference, as there is overlap. Additionally, in bringing up the issue of metadata, we urge the Department to provide meaningful privacy protection for metadata. The World Privacy Forum has signed on the Public Knowledge's comments to this RFC, which include a substantial discussion of metadata protections and issues.

Q 16. Would the development of a framework for privacy risk management be an effective mechanism for addressing challenges with big data?

We have not yet seen an example of a *framework for privacy risk management* that offers a complete approach in the area of big data. After much consideration, we have come to the conclusion that any risk framework that is developed for big data needs to focus on privacy ***rights*** and how to enforce those rights over time and across domains. We are concerned that somewhere in the bowels of a *framework for privacy risk management* is the notion that if the risks are deemed small enough, then privacy rights can be ignored. We are all for managing risks, but privacy rights must remain primary. The risk-reward balance is tricky, and can easily bleed over to impacting, subtly and no-so-subtly, rights.

We note that in the health world considerable development work has been done on frameworks for privacy risk management, including what we now call big data. We have worked in that world, and when we informed the health care community in 2005 of the risks of medical identity theft, we sought to update existing frameworks for privacy risk management in that field. Yet the frameworks that existed at that time in the health care community were so focused on risks versus rights, that they literally could not assist consumers with recovering from the crime of medical identity theft and its deep impacts. It has gotten better with national attention from HHS, but it has taken a decade. Things would have progressed much more quickly for both patients and for the health care industry if ***rights for consumers to address the problems*** had been the focus versus a narrow focus on ***risk***. We note that HIPAA's update in September 2013 includes an important expansion of patient rights, that is, the right to request (actually, demand) non-disclosure of medical records to a health plan – this right is complex, and has caveats. Namely, the right to restrict can be enacted if the treatment is paid for in full at the time it is rendered.

This is an example of a ***rights-focused*** approach to privacy that is working wonders. Key vulnerable populations have been asking for this right for many years, as it is highly protective of people who are victims of violent crimes or domestic violence. It is also protective of people who themselves are or who have children who are undergoing tests for serious diseases but may not want those tests on a record anywhere. People have many reasons for not wanting health treatment information going to a health plan, and the HIPAA right of non-disclosure finally accomplishes that. Prior to September 2013, this right did not exist and many problems resulted that no risk management solution could fix.

We also point out that ethics frameworks based on rights have been used with great success in the area of big data research. In the area of human subject research, the Common Rule, the

Belmont Report, and the Nuremberg Code are important, bedrock signposts. We discuss this issue in our answer to Question 13.

Q 17. Can emerging privacy enhancing technologies mitigate privacy risks to individuals while preserving the benefits of robust aggregate data sets?

We do not object to aggregate data if privacy rights and interests have been addressed. We suggested above the use of contracts, agreements, and chains of custody as one way to control the sharing of personal data. That same approach can work to create and control aggregate data. If proper protections can be achieved, we do not care if the technology and the controls are *privacy enhancing* or something else.

Q 18. How can the approaches and issues addressed in Questions 14–17 be accommodated within the Consumer Privacy Bill of Rights?

We have one very significant objection to the Consumer Bill of Rights, and it is regarding a deep flaw that negates the value of the CPBOR overall. The Consumer Privacy Bill of Rights has little value if it limits its scope in any way to data obtained ***directly from consumers***. Nothing else can be accomplished unless this limitation is expressly removed.

The idea of ***data obtained directly from consumers*** is severely outdated. Data about consumers can come from many sources, including data broker lists, retail purchases, social scores, census tract data, purchasing patterns, health conditions, ethnicity, book purchasing patterns, exercise patterns, and many other factors. Data used may be individual to a consumer or modeled (e.g., all consumers in a census tract).

We note again that the FTC data broker report and the WPF Scoring of America report,[23] both published in 2014 and containing years of research, found that data brokers exchange, sell, compile, and purchase detailed consumer information from each other routinely, all the down to what medications consumers are taking, their exact income, their religion, their book purchases and many other purchases, among many other data points. Data brokers and analytics firms can model individual consumer data from indirect sources, thus thoroughly nullifying the need to collect the information "directly from consumers."

Consumer data is a commodity now, and this has created a full marketplace of highly detailed, identifiable, and personal consumer data that can be acquired indirectly. This pool of consumer data is ***already*** part of the data broker marketplace. By consumer data, again, we mean highly identifiable personal data. (See appendix A of these comments for consumer data that is routinely exchanged by data brokers.) At this point, consumers do not need to directly give their information to have their information become part of the consumer data commodity marketplace. The CPBOR will not have any meaningful effect unless this specific defect is resolved. This is a significant data flow issue, one that is made clear by recent data broker research, and will not change anytime soon.

[23] Supra note 8.

We need equal treatment for all personal data, regardless of source or method of collection. We cannot divide the universe of personal data (big or otherwise) by separating personal data provided by consumers from personal data obtained in other ways. Online data has been merged with offline data in significant ways, and whether data has been collected directly from consumers is an artifact of a different world, a world from more than a decade ago and one to which we will not likely be returning.

A brief word about consent is warranted here. Data obtained directly from consumers implies that consumers have willingly and with their express consent given that data. Let us not be naïve. We all know that it is rare that consumers expressly consent to having their data used by data brokers. Police officers who have their home addresses sold[24] – to their potential peril – have they consented? When? How? People with diseases whose names and contact information is being sold[25] – people who revealed they may have drug addictions, mental illness, other diseases – when did they consent to this? Were these people told plainly, in large text prior to filling out a form online: "The information your are filling out will be sold on a data broker list, with your real name and information attached? (And possible matched to other data.)" Obscure privacy policies that claim passive consent by simply the use of a web site or filling in a survey are hardly a fair or sound basis for stating that data obtained directly from consumers is OK. Purchasing a pair of shoes from a brick-and-mortar store with a credit card should not mean that consumers consent to having their transactional information and purchases shared with data brokers for sale, resale, reuse, and more. Yet the FTC data broker report specifically documents this very practice.

How and with what kind of notice and opt-out or opt-in rights consumer data is directly – or indirectly -- obtained matters very much. Right now, it is a marketplace consumer-data-free-for-all, with currently ugly consequences and a substantial lack of rights for consumers. The CPBOR, any part of it, would improve the situation. But not with the loophole of ***data obtained directly from the consumer*** still in place.

Beyond that major objection regarding the need for equal treatment of all personal data regardless of collection method, the Bill of Rights could be improved if it mentioned expressly privacy impact assessments and similar approaches to identifying and assessment privacy consequences. Additionally, the discussion in the current provision on Security is inadequate, too narrowly focused, and does not call for sufficient transparency and stakeholder involvement in any assessment. If the Bill of Rights is to be altered to mention Big Data overtly, then a much greater emphasis on assessment should be included as well because the consequences for privacy are greater.

And finally, we want to address the telecommunications and other jurisdictional changes proposed in CPBOR.

[24] For example, NextMark, Police Officers and Troopers at Home Mailing List, NextMark ID # #253325. < http://lists.nextmark.com>. The list includes home addresses, and marital status and presence of child are selects available for this list.

[25] For example, NextMark, AILMENTS - INDIVIDUALS WITH HEALTH CONDITIONS, AILMENTS & MEDICATIONS Mailing List, NextMark List ID #83943. < http://lists.nextmark.com>.

The WPF is concerned that any jurisdictional changes made regarding telecommunications privacy or telecom-related issues must not supplant or displace FCC jurisdiction in any way. Ideally, primary jurisdiction would remain with the FCC. To be clear: WPF does not support moving telecommunications jurisdiction to the FTC. If there are any changes contemplated, at the very most equal jurisdiction would be the most WPF would support. Simplifying and clarifying the legal landscape, as iterated in the CPBOR, cannot extend to undermining existing consumer privacy protections that are working and shifting to a weaker protection regime.

Regarding jurisdictional authority and big data, WPF sees the FCC as having a more refined consumer protection regime in the area of telecom, and sees the FCC as better able to handle the challenges big data poses in this area.

Appendix: Consumer Data Available for Purchase and Use in Analytics

(This material is excerpted from the WPF Scoring of America report, pages 32-28.)

Below is a list of the most common elements of consumer data available and in circulation today. Most consumers would be stunned to learn the number of data elements available in the commercial marketplace. Not every consumer score and not every data broker file for sale includes each item on this list. Different scores use combinations of different elements and plug those into differing score algorithms and models.

This list includes independent data sets with both structured and unstructured data. This list is sourced in part from 2013 Government Accountability Report on information resellers. Other information came from a WPF review and analysis of data broker data cards viewed through NextMark [26] over the course of a year (primarily 2013), and also from WPF review and analysis of reliable data broker web sites that list data sources. For example, the Acxiom *About the Data* portal[27] lists many categories of information collected and used for consumer marketing.

The data sets available for purchase today listed here – along with others we did not identify – can create multiple layers of predictive analysis of how consumer behavior, finance, demographics, geography, and the other factors listed here interact. That does not necessarily mean that the results are better.

The range of consumer data available for use in data analytics is broad and deep. The categories listed here is not exhaustive, but it offers an idea of the range of consumer information that goes into consumer scores.

[26] Nextmark List Finder <http://lists.nextmark.com>.

[27] Acxiom About the Data Portal <https://aboutthedata.com>.

Demographic Information:

- Age
- Age range
- Date of birth
- Education
- Exact date of birth
- Gender
- Marital status
- Home ownership
- Own or rent
- Estimated income
- Exact income
- Ethnicity
- Presence of children
- Number of children
- Age range of children
- Age of children
- Gender of children
- Language preference
- Religion
- Occupation - category of occupation
- Examples: Beauty (cosmetologists, barbers, manicurists) civil servants, clergy, clerical/office workers, doctors/physicians/surgeons, executives/administrators, farming/agriculture, health services, middle management, nurses, professional/technical, retail service, retired, sales, marketing, self-employed, skilled/trade/machine operator/laborer, teacher/educator.
- Occupation - title of occupation
- Military history
- Veteran in household
- Voter party
- Professional certificates (teacher, etc.)
- Education level reached or median education

Contact Information:

- Full name
- Email address
- City
- State
- ZIP
- ZIP + 4
- Home Address

- Land-line phone
- Social IDs / social media handles and aliases
- Mobile phone number
- Carrier
- Device type
- Email address

Vehicles:

- Vehicle make, model and year
- VIN
- Estimated vehicle value
- Vehicle lifestyle indicator
- Model and brand affinity
- Used vehicle preference indicator

Lifestyle, Interests and Activities data (including medical):

- Antiques
- Apparel (women, men & child)
- Art
- Average direct mail purchase amounts
- Museums
- Audio books
- Auto parts, auto accessories
- Beauty and cosmetics
- Bible purchaser
- Bird owner
- Books
- Book purchases - plus types. (Mystery, romance, religious, etc.)
- Book clubs
- Career
- Career improvement
- Cat owner
- Charitable giving indicators:
- Charitable donor by type of donation (religious, health, social justice)
- Charitable donor by ethnicity or religion (Jewish donors, Christian donors, Hispanic donors)
- Charitable donor by financial status (wealthy donors)
- Children or teen interests
- Fashion and clothing (Multiple: sports, high fashion, shoes, accessories, etc.)
- Collectibles

- Collector
- Christian families
- Computer games
- Computers
- Consumer electronics (Many categories, including electronic fitness devices)
- Dieting and weight loss
- Telecommunications and mobile
- Dog owner
- Investing
- DVD purchasers
- Electronics - home, computing, office, other
- Empty nester
- Expectant parents
- Frequent mail order buyer
- Frequency of purchase indicator
- Getting married
- Getting divorced
- Gun ownership
- Health and beauty
- Health and medical: for example, Allergies, Alzheimer's disease, angina, arthritis/rheumatism, asthma, back pain, cancer, clinical depression, diabetes, emphysema, erectile dysfunction, epilepsy, frequent heartburn, gum problems, hearing difficulty, high blood pressure, high cholesterol, irritable bowel syndrome, lactose intolerant, ulcer, menopause, migraines/frequent headaches, multiple sclerosis, osteoporosis, Parkinson's disease, prostate problems, psoriasis/eczema, sinusitis/sinuses.
- High-end appliances
- Home improvement
- Household consumer expenditures — many categories.
- Jewelry
- Magazine subscriptions
- Mail order buyer
- Mobile location data (some analytics companies)
- Movies - attendance / collector
- Musical instruments
- Music
- New mover
- New parent
- Online and continuing education
- Online purchasing - many categories
- Parenting
- Pets - other
- Plus size clothing purchase
- Political affiliation
- Recent home buyer
- Recent mortgage borrower
- Retail purchasing - many categories.

- Science-related
- Sexual orientation
- Social media sites likely to be used by an individual or household, heavy or light users
- Spa
- Sports interests: (large category, these are examples)
- Birdwatching
- Equestrian
- Exercise and fitness
- Gardening
- Golf
- Fishing
- Outdoor interests - hiking, camping, climbing
- Swimming, diving, snorkeling
- Spectator Sports
- Stamps/coins
- Yoga
- Television, Cable, Satellite/Dish
- Travel: Vacations, domestic and/or international
- Purchase of international hotel or air flights
- Frequent flyer
- Types of purchases indicator
- Veteran in household
- Vitamins
- Volunteering

Financial and Economic – Property and Assets data:

- Estimated income
- Estimated household income
- Home value
- Length of residence
- Payment data: 30, 60, 90-day mortgage lates
- Purchase date
- Purchase price
- Purchase amount
- Most recent interest rate type
- Most recent loan type code
- Sales transaction code
- Most recent lender code
- Purchase lender code
- Most recent lender name
- Purchase lender name
- Fuel source
- Loan to value

- Purchase interest rate type
- Most recent interest rate
- Purchase interest rate
- Pool or spa
- Home - year built
- Air conditioning
- Boat ownership
- Plane ownership
- Motorcycle ownership
- Commercial assets or business ownership

Financial and Credit data:

- Bankruptcy
- Beacon score
- Credit score - actual
- Certificates of deposit/ money market funds
- Estimated household income ranges
- Income producing assets indicator
- Estimated net worth ranges
- IRAs
- Life insurance
- Low-end credit scores
- Mutual funds/annuities
- Summarized credit score or modeled credit score by neighborhood
- Payday loan purchaser
- Number of credit lines
- Tax liens
- Card data:
- Card holder - single card holder
- Range of new credit
- Debit or credit card present in household
- Card holder - brand (Discover, Visa, Mastercard, etc.)
- Card holder - type (Gas, bank, premium, luxury, prepaid, etc.)
- Frequent credit card user
- New retail card holders
- Underbanked or "thin file"
- Stocks or bonds
- Average online purchase
- Average offline purchase

In addition, a business may use enterprise data (historic data from its own customer files) to create proprietary or custom scores for its own use.