Regulatory Oversight of Insurers' Use of Big Data

Birny Birnbaum

Center for Economic Justice

April 3, 2016

NAIC Big Data Working Group of the

Market Regulation (D) Committee and

Presentation Outline

1. Big Data Defined
2. Why Insurers' Use of Big Data is An Important Issue for Insurance Regulators
3. Insurers Use of Big Data:
    a. Marketing, Sales, Pricing and Payment Plan Eligibility
    b. Claims
    c. Insurer Use of Big Data: Cybersecurity/Privacy
4. Insurer Use of Big Data: Restricting Competition
5. Insurer Use of Big Data: Policyholder Interaction, Loss Mitigation
6. Regulatory Use of Big Data: Monitoring Market Outcomes, Promoting Competition
7. Recommended Next Steps for the Working Group

Thank you for the opportunity to speak to you today on insurers' use of Big Data. I hope you will take away the following key insights from my presentation:

1. There has been a revolution in insurance pricing, marketing and claims settlement resulting from insurers' use of Big Data -- massive databases of new insurance and non-insurance, personal consumer information with associated data mining and predictive analytics and scoring.

2. Insurers' use of big data has huge implications for fairness and affordability of insurance and for regulators' ability to protect consumers from unfair practices

3. Market forces alone – "free-market competition" – cannot and will not protect consumers from unfair insurer practices. So-called "innovation" without some consumer protection and public policy guardrails will lead to unfair outcomes.

4. Oversight and limited regulatory intervention can promote more competitive markets and faster adoption of innovative technologies that benefits consumers and fulfill public policy goals.

5. The existing market regulation framework was designed for a different era and is unsuited and ineffective for addressing Big Data issues. Regulators need to collect and analyze Big Data from insurers to meaningfully evaluate market outcomes of insurer practices. The data needed for meaningful market analysis, evaluation of availability and affordability of insurance and the production of information to empower consumers and improve the competitive operation of insurance markets are available from the normal business records of insurers.

6. Regulators must be proactive to stop unfair and abusive practices or practices inconsistent with public policy. The NAIC and state insurance regulators are very late to the issue and need to act quickly to address existing and potential problems and to help realize potential benefits of insurers' and regulators' use of Big Data.

7. Insurers' Use of Big Data is a big issue, covering a variety of areas. We suggest that the working group identify the key topic areas and examine specific topic areas in more detail. We suggest the following topic areas:

    - Sales/Marketing/Pricing/Payment Plan Eligibility
    - Claims
    - Policyholder Interaction/Loss Mitigation Opportunities
    - Cybersecurity/Privacy
    - Promoting Competitive Markets/Stopping Anti-Competitive Practices
    - Regulatory Use of Big Data

8. Finally, as a preliminary step, we urge the Working Group to quickly develop a template for use by the states to collect information from all insurers on types of data used by insurers, the sources of those data and the uses of the data (sales, marketing, pricing, underwriting, claims, payment plan eligibility, other). This basic information will provide regulators with an overview of the Big Data landscape and help guide and inform the working group and states.

1. Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing and claims settlement. Big Data means:

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

In a series of reports on big data, the Federal Trade Commission and the White House have examined the potential benefits and harms to consumers from Big Data. For example, the White House report stated:

> Algorithms, Alternative Scoring and the Specter of Discrimination The business models and big data strategies now being built around the collection and use of consumer data, particularly among the "third-party" data services companies, raise important questions about how to ensure transparency and accountability in these practices. Powerful algorithms can unlock value in the vast troves of information available to businesses, and can help empower consumers, but also raise the potential of encoding discrimination in automated decisions. Fueled by greater access to data and powerful analytics, there are now a host of products that "score" individuals beyond the scope of traditional credit scores, which are regulated by law. These products attempt to statistically characterize everything from a consumer's ability to pay to whether, on the basis of their social media posts, they are a "social influencer" or "socially influenced."
>
> While these scores may be generated for marketing purposes, they can also in practice be used similarly to regulated credit scores in ways that influence an individuals' opportunities to find housing, forecast their job security, or estimate their health, outside of the protections of the Fair Credit Reporting Act or Equal Credit Opportunity Act.
>
> Details on what types of data are included in these scores and the algorithms used for assigning attributes to an individual are held closely by companies and largely invisible to consumers. That means there is often no meaningful avenue for either identifying harms or holding any entity in the decision-making chain accountable. Because of this lack of transparency and accountability, individuals have little recourse to understand or contest the information that has been gathered about them or what that data, after analysis, suggests.
>
> Nor is there an industry-wide portal for consumers to communicate with data services companies, as the online advertising industry voluntarily provides and the Fair Credit Reporting Act requires for regulated entities. This can be particularly harmful to victims of identity theft who have ongoing errors or omissions impacting their scores and, as a result, their ability to engage in commerce.

> For all of these reasons, the civil rights community is concerned that such algorithmic decisions raise the specter of "redlining" in the digital economy—the potential to discriminate against the most vulnerable classes of our society under the guise of neutral algorithms. . . . .But the ability to segment the population and to stratify consumer experiences so seamlessly as to be almost undetectable demands greater review, especially when it comes to the practice of differential pricing and other potentially discriminatory practices. It will also be important to examine how algorithmically-driven decisions might exacerbate existing socio-economic disparities beyond the pricing of goods and services, including in education and workforce settings.

2. Why is Insurers' Use of Big Data an Issue of Concern?

Insurers are rapidly increasing their use of large databases of non-insurance, personal consumer data for marketing, sales, pricing, claims, anti-fraud efforts and payment plan eligibility.

The regulatory framework established decades ago for insurance is that regulators had authority over and oversight of the information that goes into pricing and claims. Regulators enforce the requirements that rates not be unfairly discriminatory and that claims be settled fairly by stopping the use of information that would lead to violations for these requirements. For example, most states prohibit the use of race, religion, national origin. By reviewing rate manuals and underwriting guidelines, a regulator historically could have seen if any prohibited factor was used. If a company used a new risk classification in its rating plan, a regulator could ask for proof that the risk classification was related to risk of loss.

Regulators no longer have oversight of or even access to most of the new data used by insurers for all aspects of the insurers' business. And in most cases, insurers don't disclose the new data used to regulators, let alone to consumers. Market forces cannot discipline insurers and protect consumers without transparency.

For example, in the past 15 years ago, many states adopted insurance credit scoring legislation which brought insurance credit scoring under the oversight of insurance regulators – in addition to oversight of credit bureaus by federal agencies under the Fair Credit Reporting Act – and provided consumers with, among other things, disclosures and protection against certain unfair practices. Today, many types of non-insurance personal consumer information are used by insurers used with no disclosure and no accountability. In addition to obtaining consumer information from sources other than the consumer, insurers are collecting massively more information about consumers from consumers, their vehicles, their homes, drones and other means – with little, if any oversight or accountability.

Attached to my comments are a few slides from a presentation by consumer groups about "weblining" – the use of personal consumer data not covered by the protections of the FCRA to create a variety of scores used to determine what type of advertising or marketing is used for particular consumers, what type of products or services or terms of service are offered or shown to consumers shopping via web sites and the lack of transparency about what data are used and how those data are used to characterize and categorize consumers. As you can imagine, many of these databases and scoring practices reflect and perpetuate historical discrimination on the basis of race and income – just like insurance credit scoring.

Use of Big Data Scoring Models May Reflect and Perpetuate Historical Bias

New York Times, August 10, 2015: Algorithms and Bias: Q. and A. With Cynthia Dwork

> Algorithms have become one of the most powerful arbiters in our lives. They make decisions about the news we read, the jobs we get, the people we meet, the schools we attend and the ads we see. Yet there is growing evidence that algorithms and other types of software can discriminate. The people who write them incorporate their biases, and algorithms often learn from human behavior, so they reflect the biases we hold.
>
> Q: Some people have argued that algorithms eliminate discrimination because they make decisions based on data, free of human bias. Others say algorithms reflect and perpetuate human biases. What do you think?
>
> A: Algorithms do not automatically eliminate bias. . . .Historical biases in the . . .data will be learned by the algorithm, and past discrimination will lead to future discrimination.
>
> Fairness means that similar people are treated similarly. ***A true understanding of who should be considered similar for a particular classification task requires knowledge of sensitive attributes, and removing those attributes from consideration can introduce unfairness and harm utility.***
>
> Q: Should computer science education include lessons on how to be aware of these issues and the various approaches to addressing them?
>
> A: Absolutely! First, students should learn that design choices in algorithms embody value judgments and therefore bias the way systems operate. They should also learn that these things are subtle: For example, designing an algorithm for targeted advertising that is gender neutral is more complicated than simply ensuring that gender is ignored. They need to understand that classification rules obtained by machine learning are not immune from bias, especially when historical data incorporates bias.

A key takeaway for insurance regulators from this interview is this: The fact that insurers do not explicitly consider race or income in Big Data scoring models does not eliminate bias by race or income. Rather, it is only through explicit consideration of race and income in the scoring models that such bias can be identified and addressed.

3. Insurer Uses of Big Data: Pricing

There are many, many examples of Big Data scoring models for pricing, including price optimization, ISO's fireline score, the CarFax TransUnion Vehicle Score and Telematics. A recent entry is the TransUnion Criminal History Score, which is a glaring illustration of Big Data reflecting and perpetuating historical discrimination. TransUnion's brochure states:

> *TransUnion recently evaluated the predictive power of court record violation data (including criminal and traffic violations)*
>
> *While a court record violation is created during the initial citation, the state MVR is updated later and may be delayed depending on a consumer's response to the citation. For example, if someone pleads guilty and pays a ticket immediately, the state MVR will be updated in approximately two months. If the ticket is disputed in court, in contrast, the state MVR may not be updated for 6–19 months or longer.*
>
> *Also, as court records are created when the initial citation is issued, they provide insight into violations beyond those that ultimately end up on the MVR—such as violation dismissals, violation downgrades, and pre-adjudicated or open tickets.*

Many states specifically prohibit insurers from penalizing consumers for certain violations if they consumer takes a driver safety course – with the goal of promoting driver safety. The TransUnion score undermines this public policy.

More importantly, criminal history record scoring reflects and perpetuates historical discrimination in criminal justice. We need only look to practices of the criminal justice system in Ferguson, Missouri to see that African-Americans were targeted for a variety of minor violations.

4. Big Data and Rate Comparison/Shopping Web Sites

As the weblining slides show, price comparison/aggregator websites have the potential to steer consumers to higher-cost products based on personal consumer information unrelated to insurance risk and without disclosure to the consumer. For example, a consumer using a price comparison website for auto insurance might be channeled to non-standard insurers based on information pre-fetched by the web site. Again, there is little regulatory oversight or accountability to the consumers. Market forces cannot protect consumers in such a situation – regulatory intervention is needed to ensure fair competition.

5. Insurer Use of Big Data: Claims

At the end of my comments is a presentation to the NAIC I made last year on Big Data and Claims. The presentation includes a number of Big Data scoring models used by insurers – scoring models with little or no regulatory oversight by regulators or accountability to consumers. One example from that presentation is the LexisNexis Claims Tools

More Data Earlier: The Value of Incorporating Data and Analytics for Claims Handling at http://www.lexisnexis.com/risk/insights/value-incorporating-data-analytics-claims-handling.aspx

For third-party bodily injury settlements, the study found that more data earlier resulted in:
• 15–25 percent lower severity payments*
• 25–49 percent lower attorney involvement
• 5–15 percent shorter cycle times

Similar results were obtained for third-party property damage claims:
• 10–15 percent lower severity payments
• 8–15 percent shorter cycle times

LexisNexis (LN) seeks to provide a Single Point of Entry for delivering all of information directly back into a carrier's system whether from a marketing standpoint, underwriting process or especially the claims part.

LN has over 10,000 data sources that feed into its infrastructure each month and has contributed information from the industry.

"Claims Data Fill" – deliver data and analytics directly into claims system in the claims process regarding parties, vehicles and carrier information. Used to verify information provided to insurers and provide indicators beyond the data to identify whether a social security number is an indicator of fraud or whether an address provided is a good address. Has an analytic component at first notice of loss and throughout the claim, constantly monitoring the claim looking for fraudulent activities. Real time data verification and enhancement with fraud scoring and attributes

Example, insured calls in, rear-ended, all I got was license plate:

Claims Data Fill takes that license plate, reach out to DMV to get vehicle registration to get VIN number, we have policy database and get the carrier and policy information, take the registered owner, go out to public records, pull back their address, date of birth, telephone number, social security, wrap that into a package and put it back into our system, 88% of the time done in less than 5 seconds.

Take minimum information provided at first notice of loss, provide a fraud score at the initial notice of loss. Daily monitoring of claim every time new information comes in, able to run various scores: fraud scores, severity score

The LexisNexis tools are just one of many examples of claims scoring products used to channel consumers with claims into different claim settlement tracks. And, again, there is little or no regulatory oversight of these claims scoring models and no disclosure or accountability to consumers. And, again, market forces cannot protect consumers from unfair practices.

6. Insurer Use of Big Data: Restricting Competition

There are a number of ways in which insurers' use of Big Data can promote more competitive markets and policyholder interaction for loss mitigation. But there are also ways in which Big Data can thwart competitive markets. Consider the case of vehicle telematics data.

The largest insurers with the most vehicles enrolled in telematics have the telematics data. The ability for small to medium sized insurers to start telematics programs is limited by the lack of a telematics database, as is the ability of consumers to take their telematics data from one insurer to another. While there are industry-wide databases for claims and all insurers and consumers have access to consumer credit information, no similar database exists for telematics data.

Layer onto the inability of some insurers to access telematics data and the inability of consumers to take telematics data from one insurer to another the agreements between a small number of large insurers with some auto manufactures for exclusive telematics marketing agreements, concerns about anti-competitive practices arise.

Regulatory intervention – to promote a telematics database, to ensure consumer access to their telematics data for shopping and claims settlement, to ensure symmetrical use of telematics data by insurers for claims settlement and to ensure consumers' telematics data is not sold by insurers to third parties and is used only for stated insurance purposes – promotes competitive markets and fosters more rapid implementation of innovative technologies while ensuring fair treatment of consumers.

Insurers will argue that their use of Big Data provides only benefits to consumers, but many of their claims have no logical or empirical support

"We simply want to match price to risk."

Insurers typically argue that any limit on risk classification harms competition and, consequently, harms consumers. They argue that, for example, credit scoring has enabled insurers to offer insurance to more consumers and point to historical declines in the number of consumers in assigned risk plans. I'm a firm believer in the power of markets to benefit consumers, but regulation is essential to ensure that such competition is fair. The fact is that insurers' use of credit scoring has resulted in far higher rates for some consumers and lower rates for other consumers. Uninsured motorist rates have not declined as a result of insurers' use of credit scoring.

Insurers' claim that they are singularly focused on deriving the most accurate rates for the risk they insuring is belied by a recent story on MN public radio about price optimization. The industry spokesperson simply confirms what consumers have argued all along – that insurer pricing considers far more than expected claims and includes factors unrelated to cost-based pricing.

MN Public Radio, 11/16/2015
Insurance Federation of Minnesota spokesman Mark Kulda acknowledges some insurers in Minnesota use price optimization.
"It's also a practice used by many other industries," he said. "This is not really anything new in the economy. All companies really try to see how much can they charge for their product and what is the most that consumers will bear to pay. And it's no different in the insurance industry."

Insurance is not like other consumer products – state law requires that insurance rates and premium charges be not excessive and not unfairly discriminatory, that rates be cost-based. Few other products or services have similar pricing requirements. And such requirements are reasonable and necessary because insurance is not like other products – it is a required purchase and it is a contract for future benefits if certain events occur.

The confidence fairy

I must also address the confidence fairy argument – that unlimited risk classification allows insurers to have more confidence in writing insurance and therefore makes insurance more available. Putting aside the lack of evidence to support the claim, as discussed above, insurers routinely ignore the fact that the introduction of ever new rating factors – particularly those based on socio-economic characteristics and opaque to the consumer – lead to ever widening spread of rates with ever greater prices for those consumers viewed unfavorably by insurers. There is simply no empirical or public policy basis for unlimited risk classification and broad public policy support for limits on risk classification that promote affordability and empowerment of consumers for behavioral changes to reduce accidents and claims.

Insurers have claimed for years that their ability to use ever more refined and precise rating plans allows them to write more business because they can better match premium to risk. So they argue that any limitations on rating practices will cause them to write less business. This argument may have had some validity 40 years ago when insurers were using only a handful of rating factors. But we have long, long passed the point where more refinement in pricing leads to insurers writing more business.

7. Insurer Use of Big Data: Policyholder Interaction, Loss Mitigation

Big Data presents a radically-new opportunity for insurers to interact with policyholders for all aspects of the insurance transaction, but none more important than for loss mitigation. Through Big Data and the Internet of Things, insurers have the opportunity to communicate with policyholders to promote loss mitigation. Some of these opportunities are being realized by insurers. Examples are found in all lines of insurance. One example is the use of social media to alert consumers of a pending hail storm to move cars to shelter.

But not all such loss mitigation opportunity is realized. Telematics – auto and home – offers the opportunity for real time feedback to policyholders to avoid risky practices or situations. Unfortunately, the paradigm for insurers' telematics programs is to collect data, put it through a black box scoring model to develop a score and apply that score to the consumer – the same approach used for credit scoring.

Market forces may eventually lead insurers to develop Big Data tools for improved policyholder interaction and improved shopping experiences for consumers, but modest regulatory intervention could encourage faster development and adoption of Big Data tools for policyholder interaction and loss mitigation.

8. Regulatory Use of Big Data: Monitoring Market Outcomes, Promoting Competition

The insurance market regulation infrastructure represents a 20th century technology ill-suited to insurers' 21st century practices. A short history is useful.

Old Old School Big Data: Advisory Organizations and Loss Costs

Advisory Organization, like the Insurance Services Office, collected from many insurers exposure and claims information about millions of policies and developed projections of expected claims for pre-defined groupings of exposures (consumers) – advisory loss costs.

Information was organized and gathered according to pre-defined characteristics and used to evaluate those pre-defined risk classes.

Activities were / are subject to regulatory oversight, including licensing of an advisory organization, precautions against anti-competitive practices, public filings showing the loss-cost development methodology, regulatory review and approval of the loss costs.

Old School Big Data – Credit Scoring

Insurance credit scoring was a watershed event. First time a large non-insurance database was data-mined to predict outcomes for insurance companies.

Using a database from of consumer credit information of some 200 million consumers collected and maintained by the credit reporting agencies, Fair Isaac identified some 500 data elements within consumer credit reports, added some individual insurance outcome information to the data – associating insurance outcomes with the consumer credit information – and then data mining and applying predictive analytic models to identify the 10 to 20 data elements most predictive of the insurance outcome and created "credit-based insurance scores."

<u>Credit Scoring a Watershed Event – No Pre-Defined Risk Classes:</u>

Instead of starting with a pre-defined class, collecting data related to that class of insurance consumers and evaluating whether the characteristic defining that group of consumers had

different expected claims than consumers without that defining characteristic, credit scoring used data mining to root through vast amounts of data on individual consumers to identify characteristics that differentiated consumers in terms of insurance outcomes – expected claims, persistency, likelihood to purchase other insurance policies

Credit Scoring a Watershed Event: Vastly Increased Risk Class Segmentation

Dramatic increase in the segmentation of consumers into risk classes. Instead of, for example, four driving record risk classes of no accidents, one accident, two accidents or more than two accidents in the past three years, credit scoring provided scores that ranged over, say, a 500-point spread allowing insurers to create, for example ten 50-point groupings or 50 ten-point groupings. This led insurers to review traditional insurance classifications to develop more granularity. Instead of drive age classes of under 18, 18-21, 21-25, 25 to 65 and over 65, some insurers developed rate classes for individual age years.

Credit Scoring Oversight

Insurers' use of consumer credit information was quickly brought to the attention of consumers and regulators.

- To regulators because insurers used the credit information as rating factors included in rate filings submitted to regulators.

- To consumers because the insurers' use of consumer credit information was subject to the Fair Credit Reporting Act, which sets out specific requirements for both the credit reporting agencies who collect and maintain the consumer credit information and the users of the credit information. For example, the FCRA requires users of consumer credit information to provide a notice to consumers if the use of the credit information results in an adverse action.

Because of concerns about the fairness of insurance credit scoring, a few states banned its use, but most states passed laws regarding insurers' use of consumer credit information.

New School Big Data

Use of Big Data has exploded – financial service firms, including insurers, tap into a variety of non-insurance databases to steer and segment consumers with little or no transparency or oversight.

- Steering consumers without the consumers knowledge

- Skirting protections for consumers required by the FCRA

- In insurance context, use of a variety of data bases and related scoring models with little disclosure to regulators, let alone consumers of the use of these data and models with no accountability regarding the accuracy/completeness of the data, the objectivity / fairness of the models or the market outcomes of the use of these data and models, including unfair discrimination against low-income and minority consumers.

It is not feasible for regulators to monitor all the databases and scoring models used by insurers. If is unrealistic to expect regulators to provide the type of historical review of advisory loss costs to new pricing tools, what is the way forward? Two key initiatives are needed:

First, insurance regulators need regulatory Big Data – transaction level data from insurers on sales and claims – to perform more robust market analysis and better monitor the market outcomes for consumers of various insurer practices. These data will also allow regulators publish market information on insurer performance – information that currently does not exist, such as claims settlement performance by insurers. By publishing market performance data about insurers, regulators will promote more competitive markets by empowering consumers in the market transaction with insurers.

Second, the regulatory framework needs to change from insurers using any type of data for any purpose until explicitly prohibited to prior review of data sources and uses to allow public discussion of whether these data promote or defeat public policy goals of fair access and fair treatment of consumers, loss mitigation and competitive operation of insurance markets. Again, this is not about stifling innovating, but simply guiding that innovation towards outcomes sought by society.

9. Recommended Next Steps for the Working Group

Insurers' Use of Big Data is a big issue, covering a variety of areas. We suggest that the working group identify the key topic areas and examine specific topic areas in more detail. We suggest the following topic areas:

- Sales/Marketing/Pricing/Payment Plan Eligibility
- Claims
- Policyholder Interaction/Loss Mitigation Opportunities
- Cybersecurity/Privacy
- Promoting Competitive Markets/Stopping Anti-Competitive Practices
- Regulatory Use of Big Data

Presumably, the issues of Cybersecurity and Privacy are being addressed by the Cybersecurity Task Force, but the Big Data Working Group can and should inform the work of the Task Force. Each of the topics is substantial. We suggest that the working group focus on one of these topics at a time, starting with Regulatory Use of Big Data.

Finally, as a preliminary step, we urge the Working Group to quickly develop a template for use by the states to collect information from all insurers on types of data used by insurers, the sources of those data and the uses of the data (sales, marketing, pricing, underwriting, claims, payment plan eligibility, other). This basic information will provide regulators with an overview of the Big Data landscape and help guide and inform the working group and states.

Weblining and Other Racial Justice Concerns In the Era of Big Data

NCLC Webinar
June 3, 2014

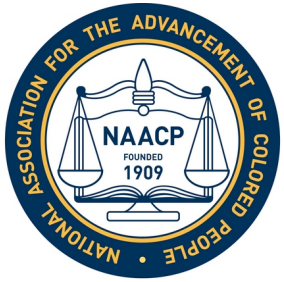
BY David Robinson
david@robinsonyu.com

AND

Harlan Yu
harlan@robinsonyu.com

robinson +yu

Civil Rights Principles for the Era of Big Data



HTTP://WWW.CIVILRIGHTS.ORG/BIGDATA

Civil Rights Principles for the Era of Big Data

1. Stop High-Tech Profiling.

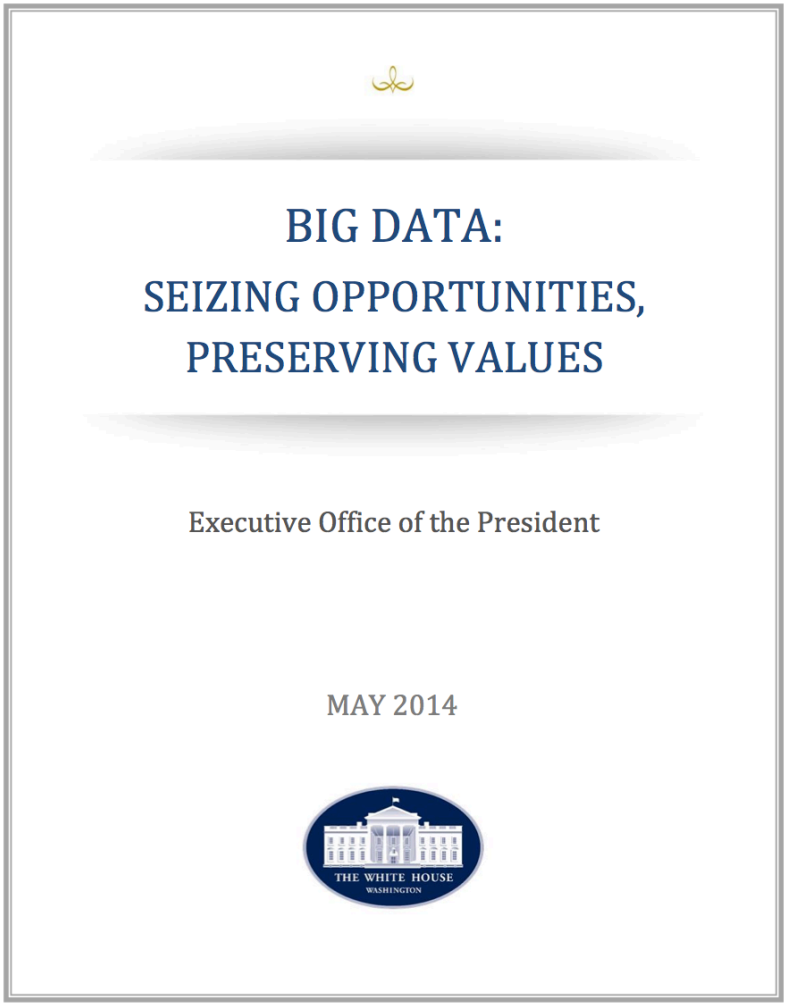
2. Ensure Fairness in Automated Decisions.

3. Preserve Constitutional Principles.

4. Enhance Individual Control of Personal Information.

5. Protect People from Inaccurate Data.

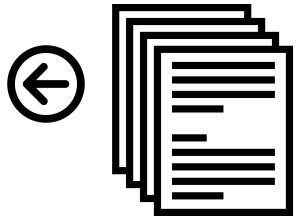
White House Big Data Report



HTTP://WWW.WHITEHOUSE.GOV/BIGDATA

UNDERWRITING

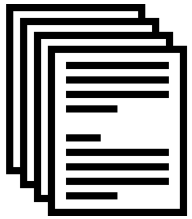
INDIVIDUAL CREDIT SCORES



CREDIT FILES

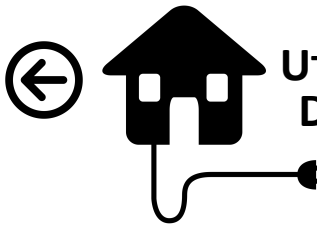
UNDERWRITING

INDIVIDUAL CREDIT SCORES



CREDIT FILES

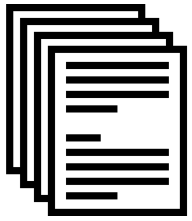
MAINSTREAM ALTERNATIVE CREDIT SCORES



UTILITY DATA

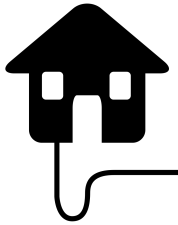
UNDERWRITING

INDIVIDUAL CREDIT SCORES



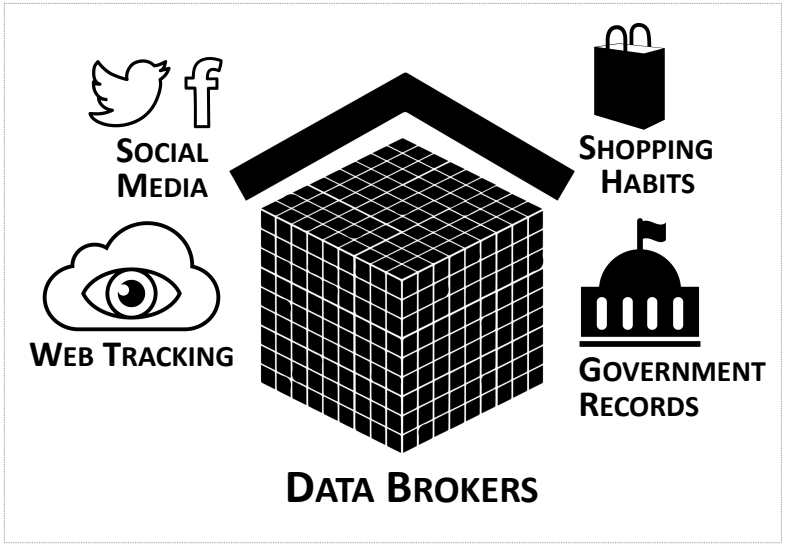
CREDIT FILES

MAINSTREAM ALTERNATIVE CREDIT SCORES

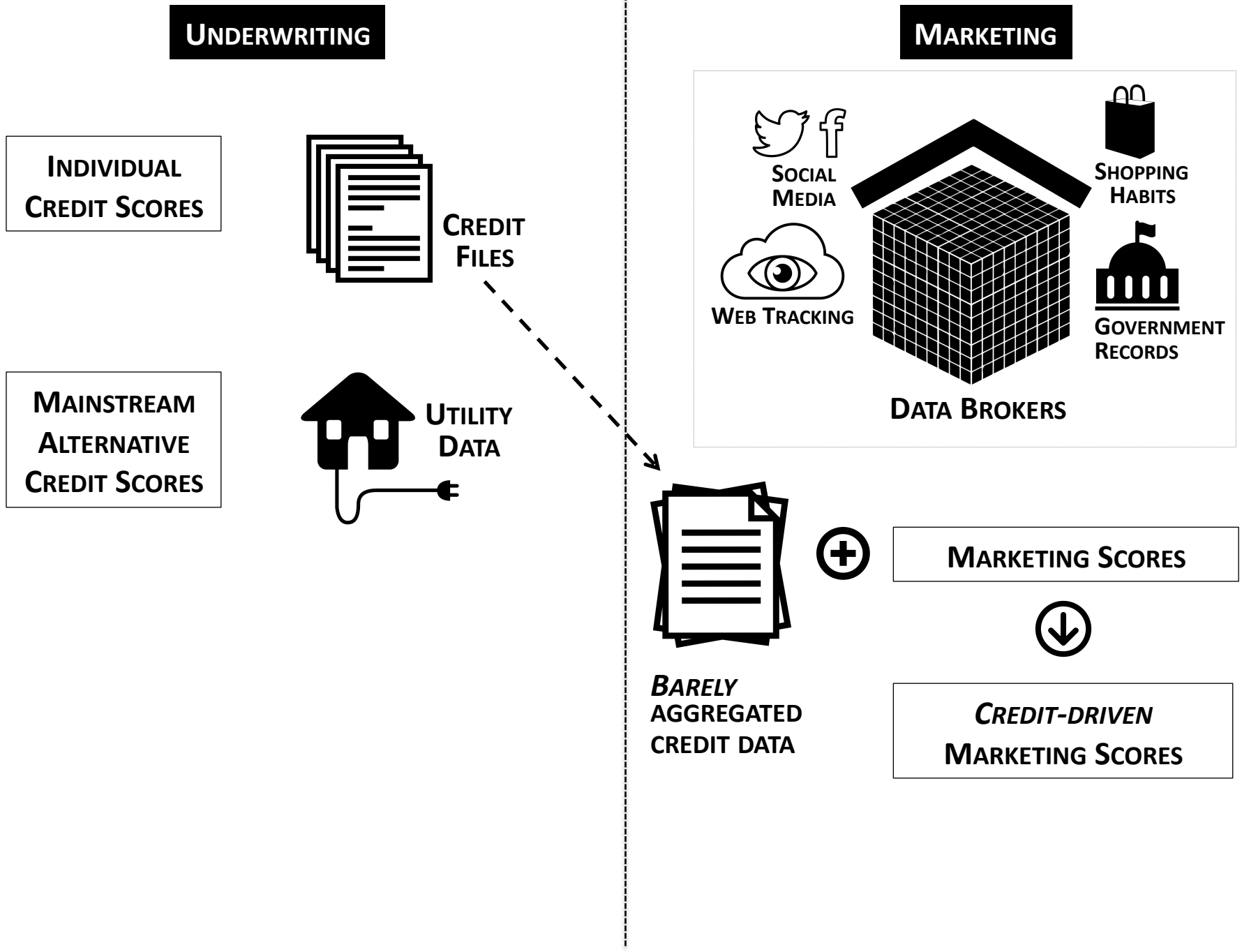


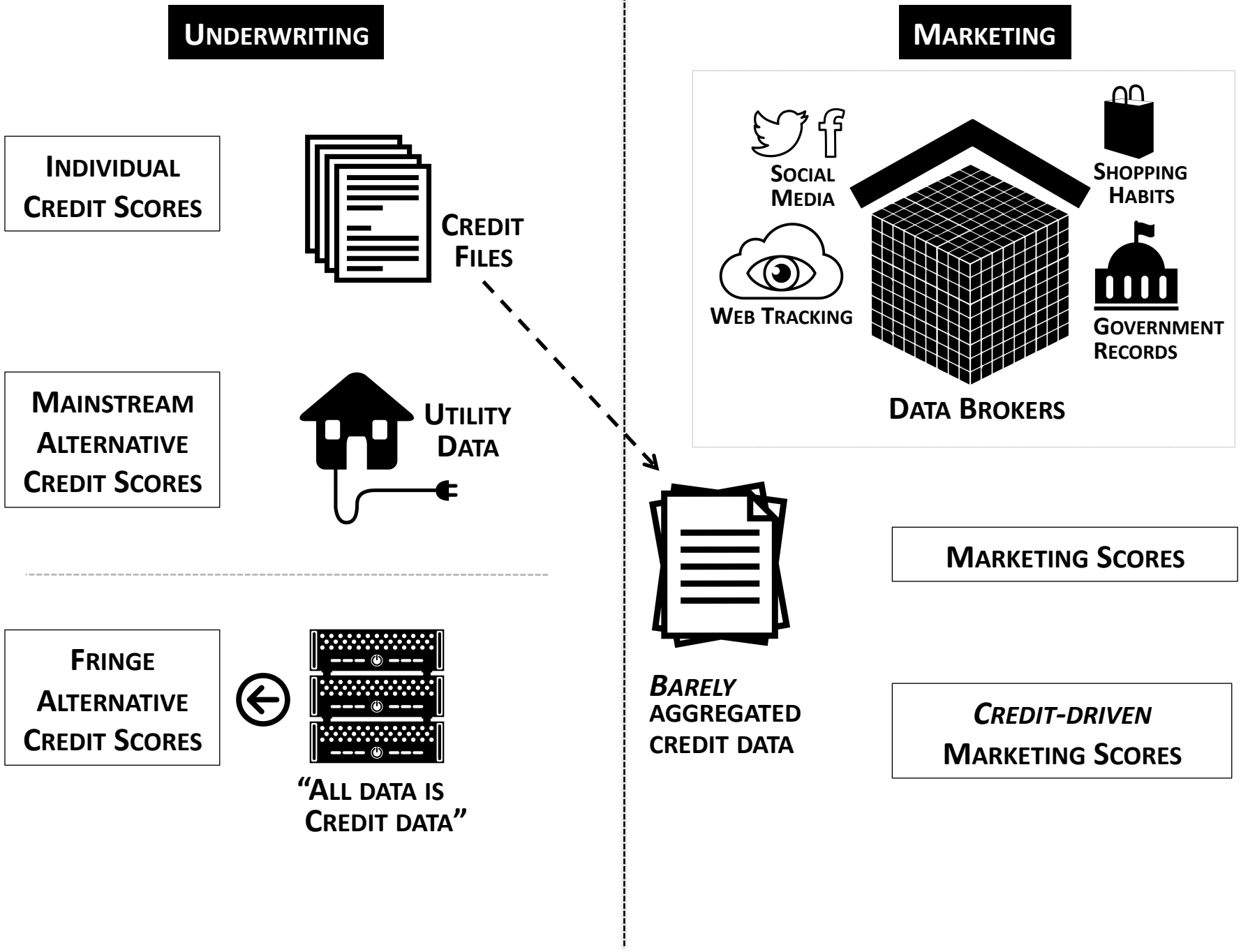
UTILITY DATA

MARKETING



MARKETING SCORES


UNDERWRITING
MARKETING
INDIVIDUAL CREDIT SCORES
CREDIT FILES
MAINSTREAM ALTERNATIVE CREDIT SCORES
UTILITY DATA
SOCIAL MEDIA
SHOPPING HABITS
WEB TRACKING
GOVERNMENT RECORDS
DATA BROKERS
BARELY AGGREGATED CREDIT DATA
MARKETING SCORES
CREDIT-DRIVEN MARKETING SCORES


UNDERWRITING
MARKETING
INDIVIDUAL CREDIT SCORES
CREDIT FILES
MAINSTREAM ALTERNATIVE CREDIT SCORES
UTILITY DATA
FRINGE ALTERNATIVE CREDIT SCORES
"ALL DATA IS CREDIT DATA"
SOCIAL MEDIA
SHOPPING HABITS
WEB TRACKING
GOVERNMENT RECORDS
DATA BROKERS
BARELY AGGREGATED CREDIT DATA
MARKETING SCORES
CREDIT-DRIVEN MARKETING SCORES

QUESTIONS?

BY David Robinson
david@robinsonyu.com

AND

Harlan Yu
harlan@robinsonyu.com

robinson +yu

II. Other Online Profiling/Data Collection Practices

NCLC Big Data and Racial Justice Webinar
on 3 June 2014, Ed Mierzwinski, U.S. PIRG

As the New York Times explained, PIRG/CDD uncovered use of invisible to consumer E-Scores, predictive financial algorithms and other scoring products, with impact on all—but esp. low income consumers.

Secret E-Scores Chart Consumers' Buying Power

Tim Gruber for The New York Times

Gordy Meyer, chief executive of eBureau, in its data processing center. EBureau calculates consumers' buying-power scores, which take into account details like occupation, salary, home value and spending patterns.

By NATASHA SINGER
Published: August 18, 2012

ST. CLOUD, Minn.

FACEBOOK
TWITTER
GOOGLE+
E-MAIL
SHARE

Related

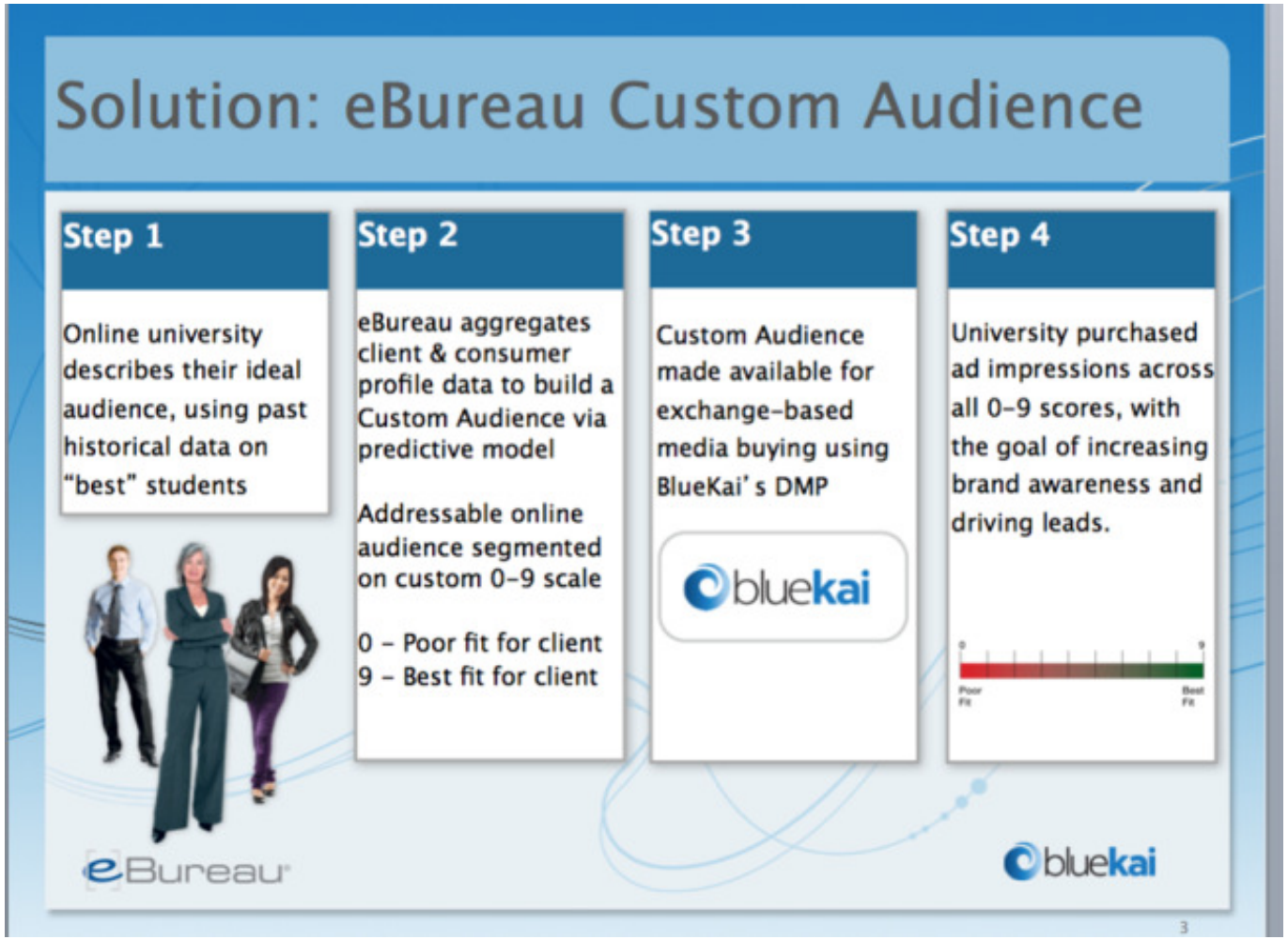
You for Sale:

AMERICANS are obsessed with their scores. Credit scores.

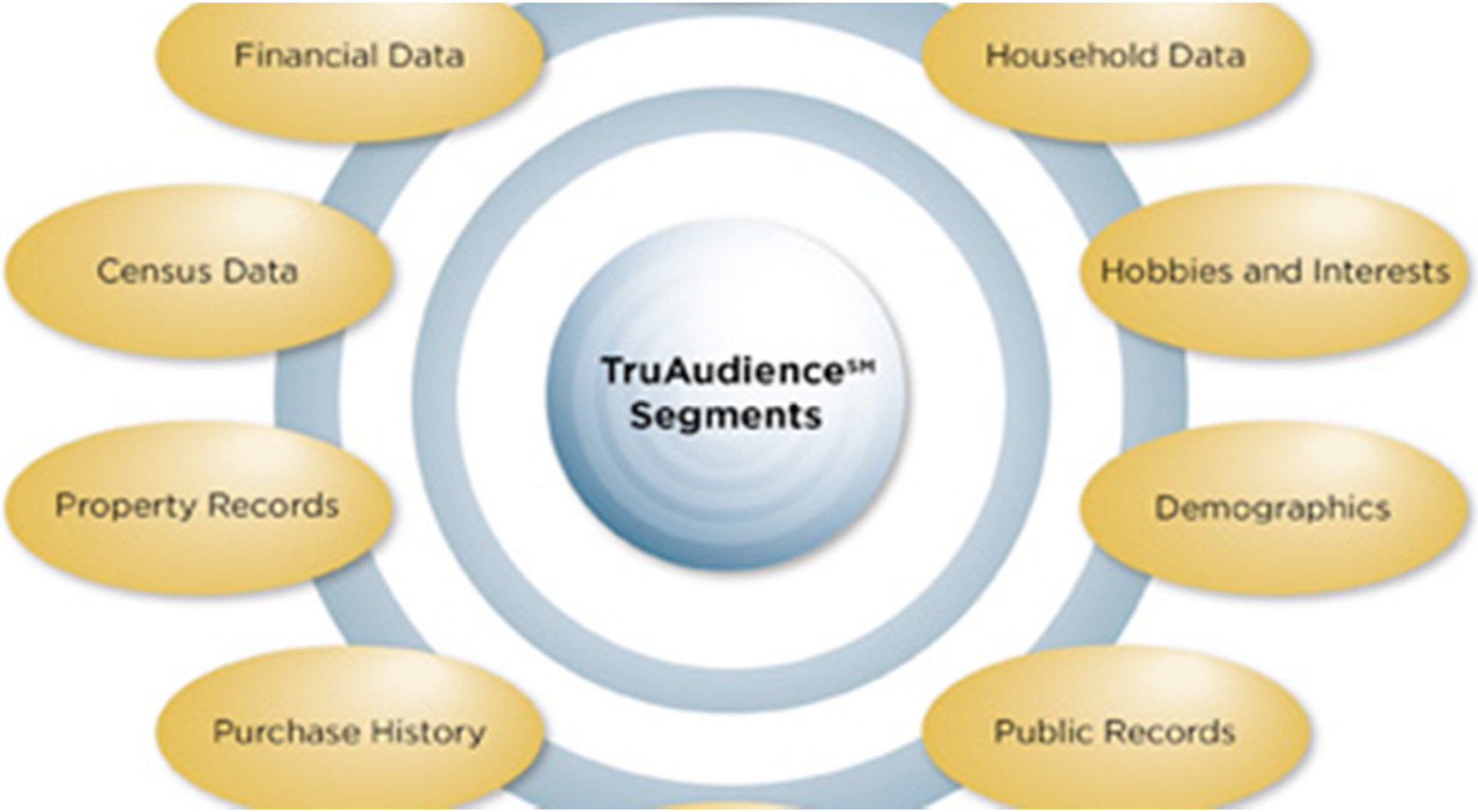
"These digital scores, known broadly as consumer valuation or buying-power scores, measure our potential value as customers. What's your e-score? You'll probably never know. …

… The result is a private, digital ranking of American society unlike anything that has come before."

Example: eBureau's score product for an online for-profit university. Invisible, non-transparent, unaccountable. Such scores can be used to determine pricing, including discriminatory pricing.



Digital E-Scores can be used to determine pricing or service; or even whether you are offered a product at all.



These powerful new capabilities have tremendous impact on consumers, esp. low-income/at risk or vulnerable populations. Data can be used to manipulate choices or offers or pricing structures

TruSignal™

DATA SHEET

Ideal Audience: Underbanked Consumers

Overview

TruSignal's Underbanked Consumer Ideal Audience includes people who have similar characteristics of consumers who maintain nontraditional banking relationships. This segment is valuable for advertisers looking to reach consumers using money transfer services, applying for short term loans or prepaid debit card products through online channels.

Top Predictive Factors:

- Length at residence
- Estimated home value
- Not owner occupied residence
- Musical tastes
- Financial strength

Built Using Real World Performance Data

TruSignal Ideal Audiences go beyond rudimentary demographics or pre-canned segments. By analyzing underbanked consumers with TruSignal's massive store of third party data, we discovered a proprietary formula of 127 predictors. Together, these data attributes detect which consumers are likely to use money transfer services, short term loans and prepaid debit card products. We used this audience formula to profile more than 100,000,000 US adult consumers, giving you all of the power and precision of many datasets distilled into a single, targetable audience.

Dynamic Pricing Based On E-Scores: Fair or Unfair?

- **FTC Data Broker Report: May 2014:** "the scoring processes used in some marketing products are not transparent … consumers are unable to take actions that might mitigate the negative effects of lower scores, such as being limited to ads for subprime credit or receiving different levels of service from companies."
- **Professor Joseph Turow, "The Daily You", 2012:** "Turow describes how our personal "reputations" related to our identity are being constructed by others--all out of the control of the individual. Some of us are regarded, he explains, as "waste"--because our incomes or life conditions may not make some marketer the profit they desire. We are secretly being labeled by others with various digital "scarlet letters" symbolizing our worth to the commercial marketplace (and the political one as well)."
- **Turow, "Niche Envy," 2010:** Merchants consider the online environment particularly ripe for such "*dynamic pricing*"—that is, for *price discrimination* driven by behavioral targeting.

THE USE OF CONSUMER DATA TO SETTLE AUTOMOBILE INSURANCE CLAIMS

Public Hearing of the
NAIC Market Regulation (D) Committee

Comments of
Birny Birnbaum
Center for Economic Justice

August 17, 2015

The Center for Economic Justice

CEJ is a non-profit consumer advocacy organization dedicated to representing the interests of low-income and minority consumers as a class on economic justice issues. Most of our work is before administrative agencies on insurance, financial services and utility issues.

On the Web: www.cej-online.org

Why CEJ Works on Insurance Issues

Essential Financial Security Tool for Individual and Community Economic Development: CEJ Works to Ensure Access and Fair Prices for These Essential Products and Services, particularly for Low- and Moderate-Income Consumers.

Primary Institution to Promote Loss Prevention and Mitigation: CEJ Works to Ensure Insurance Institutions Maximize Their Role in Efforts to Reduce Loss of Life and Property from Catastrophic Events.

Big Data Defined

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

Consumer Information in Big Data

- Social Media
- Shopping Habits/Purchase History
- Hobbies and Interests
- Demographics/Household Data/Census Data
- Government Records/Property Records
- Web Tracking
- Mainstream Credit Files: Loans, Credit Cards
- Alternative Credit Data: Telecom, Utility, Rent Payment

Insurance Big Data Example: LexisNexis Claims Tools

More Data Earlier: The Value of Incorporating Data and Analytics for Claims Handling at http://www.lexisnexis.com/risk/insights/value-incorporating-data-analytics-claims-handling.aspx

For third-party bodily injury settlements, the study found that more data earlier resulted in:
• 15–25 percent lower severity payments*
• 25–49 percent lower attorney involvement
• 5–15 percent shorter cycle times
Similar results were obtained for third-party property damage claims:

LexisNexis Claims Tools

LexisNexis (LN) seeks to provide a Single Point of Entry for delivering all of information directly back into a carrier's system whether from a marketing standpoint, underwriting process or especially the claims part.

LN has over 10,000 data sources that feed into its infrastructure each month and has contributed information from the industry.

"Claims Data Fill" – deliver data and analytics directly into claims system in the claims process regarding parties, vehicles and carrier information. Used to verify information provided to insurers and provide indicators beyond the data to identify whether a social security number is an indicator of fraud or whether an address provided is a good address.

LexisNexis Claims Tools

Has an analytic component at first notice of loss and throughout the claim, constantly monitoring the claim looking for fraudulent activities. Real time data verification and enhancement with fraud scoring and attributes

Example, insured was rear-ended, all I got was license plate:

Claims Data Fill takes that license plate, reach out to DMV to get vehicle registration to get VIN number, we have policy database and get the carrier and policy information, take the registered owner, go out to public records, pull back their address, date of birth, telephone number, social security, wrap that into a package and put it back into our system, 88% of the time done in less than 5 seconds.

LexisNexis Claims Tools

Take minimum information provided at first notice of loss, provide a fraud score at the initial notice of loss. Daily monitoring of claim every time new information comes in, able to run various scores: fraud scores, severity score

New contributory claims database, much deeper than prior claims databases – this is claims file submitted as new information added – allows us to track vehicles across carriers, medical providers across carriers – sharing of information much deeper than has been done before. Text mining, watch list mixed with LexisNexis data.

Take-Away: Many databases and scoring models with little or no transparency to consumers and regulators and outside the scope of consumer protection laws like the FCRA.

LexisNexis: "Severity Focus"

Identify claims with the potential to become severe: SeverityFocus utilizes advanced predictive modeling to identify claims with the potential to become severe as they develop claims that otherwise would go undetected until much later.

SeverityFocus does not constitute a "consumer report" as that term is defined in the federal Fair Credit Reporting Act, 15 USC 1681 et seq. (FCRA). Accordingly, Severity Focus may not be used in whole or in part as a factor in determining eligibility for credit, insurance, employment or another permissible purpose under the FCRA.

LexisNexis: "Severity Focus"

Due to the nature of the origin of public record information, the public records and commercially available data sources used in reports may contain errors. Source data is sometimes reported or entered inaccurately, processed poorly or incorrectly, and is generally not free from defect. This product or service aggregates and reports data, as provided by the public records and commercially available data sources, and is not the source of the data, nor is it a comprehensive compilation of the data. Before relying on any data, it should be independently verified.

LexisNexis: "Claims Discovery"

Understand claimants' prior auto or property claim histories: LexisNexis® Claims Discovery(SM) is a separate, contributory, nonFCRA database available to insurers that contains the same type of information submitted to C.L.U.E.®.

Product info includes same statements about non-FCRA data and data quality.

LexisNexis Fraud Focus

Detect possible fraud at the earliest possible moment: FraudFocus® is an effective fraud detection system that helps insurance companies proactively combat several types of fraud.

Product info includes same statements about non-FCRA data and data quality.

StatSoft's Predictive Claims Flow™,

A predictive analytics and reporting solution for property and casualty insurance companies, can help you reduce loss ratios and improve bottom-line profitability, often within a few months of implementation. StatSoft's Predictive Claims Flow™ solution incorporates predictive modeling at every stage of an insurance claim. This closed loop system has a unique scoring system that rates each claim at its inception on its propensity for fraud and then continually rescores the claim as it goes through each step of a claim's lifecycle.

StatSoft's Predictive Claims Flow™,

Using already-established fraud flags, plus intelligent variables that predictive models create, fraud scores are re-calculated every time a new piece of data is submitted, whether it's a verbal update from a claimant or a medical bill submitted by a vendor. Based on fraud probabilities, StatSoft's solution then determines the right level of servicing for the claim, including whether or not the claim should be assigned to a more senior adjuster or if there is a need for in-person-contact in order to reduce the overall duration of the claim. At the same time, predicted reserve estimates are updated and opportunities for subrogation are identified.

Infosys Social Network Analysis

The SNA tool combines a hybrid approach of analytical methods. The hybrid approach includes organizational business rules, statistical methods, pattern analysis, and network linkage analysis to really uncover large of amounts of data to show relationships via links. When one looks for fraud in a link analysis, one looks for clusters and how these clusters link to other clusters. Public records such as judgments, foreclosures, criminal records, address change frequency and bankruptcies are all data sources that can be integrated into a model. Using the hybrid approach, the insurer can rate these claims. If the rating is high, it indicates the claim is fraudulent.

Infosys: Social Customer Relationship Management

Social CRM is neither a platform nor a technology, but rather, a process. It is important that insurance companies link social media to their CRM.

Social CRM . . . gathers data from various social media platforms. It uses a "listening" tool to extract data from social chatter,. . . .The reference data along with information stored in the CRM is fed into a case management system. The case management system then analyzes the information based on the organization's business rules and sends a response. The response, from the claim management system as to whether the claim is fraudulent or not, is then confirmed by investigations independently, since the output of the social analytics is just an indicator and should not be taken as the final reason to reject a claim

Price Optimization Technology for Claims

Regulators have become aware of vendors promoting, and insurers using, price optimization tools for setting auto insurance premium rates. PO is a big data application that adjusts cost-based rate indications based on non-cost factors such as evaluations of consumer price elasticity of demand – consumer likelihood of shopping for another carrier in the face of a rate increase.

> Earnix explained that price optimization allows insurers to "[a]nalyze the price elasticity of each customer profile and uncover the efficient pricing frontier for each product in your portfolio." Source: Earnix.com "Price Optimization: Insurance Price/Rate Optimization"

> "Earnix best-in-class analytics and patent-awarded optimization technology empowers insurers to implement pricing strategies that go beyond traditional risk cost pricing, incorporating demand elasticity models to maximize profit and growth objectives."
>
> Source: Earnix.com "Insurance Pricing and Customer Value Optimization"

If insurers are utilizing tools for pricing based on consumer response to pricing offers, it seems logical that the same tools may be utilized for claims settlement – price optimization to evaluate claim settlement offers to consumers.

Telematics Data in Claims Settlement

Telematics devices can capture large amounts of information about a consumer's operation of a vehicle:

- Time and Dates Driven
- Where Driven
- Acceleration
- Braking
- Turning/Cornering

Telematics Data in Claims Settlement

Consequently, telematics data collected by insurers has potential for use in claim settlements when the nature of the accident is unclear or contested. However, insurers have the data and the ability to analyze and interpret very detailed information, ***raising the possibility that insurers might use data when beneficial to the insurer, but not disclose the data to claimant when helpful to the claimant.***

Insurer Big Data for Claims: Consumer Protections

- Accuracy and Completeness of Data
- Oversight of Data Bases
- Disclosures to Consumer About Data Used, How Used and Privacy Protections
- Compliance with Fair Credit Reporting Act
- Consumer Ability to Challenge False Information
- Evaluation of Bias in Models
- Regulators' Knowledge Of and Capability to Provide meaningful Oversight
- Prevent discrimination Against Low-Income and Minority Consumers and other protected classes
- Asymmetric Use of Data
- Greater Cybersecurity Danger for Consumers and Insurers

Use of Big Data Scoring Models Does Not Eliminate Bias

New York Times, August 10, 2015: Algorithms and Bias: Q. and A. With Cynthia Dwork

Algorithms have become one of the most powerful arbiters in our lives. They make decisions about the news we read, the jobs we get, the people we meet, the schools we attend and the ads we see. Yet there is growing evidence that algorithms and other types of software can discriminate. The people who write them incorporate their biases, and algorithms often learn from human behavior, so they reflect the biases we hold.

New York Times, August 10, 2015: Algorithms and Bias

Q: Some people have argued that algorithms eliminate discrimination because they make decisions based on data, free of human bias. Others say algorithms reflect and perpetuate human biases. What do you think?

A: Algorithms do not automatically eliminate bias. . . .Historical biases in the . . .data will be learned by the algorithm, and past discrimination will lead to future discrimination.

Fairness means that similar people are treated similarly. ***A true understanding of who should be considered similar for a particular classification task requires knowledge of sensitive attributes, and removing those attributes from consideration can introduce unfairness and harm utility.***

New York Times, August 10, 2015: Algorithms and Bias:

Q: Should computer science education include lessons on how to be aware of these issues and the various approaches to addressing them?

A: Absolutely! First, students should learn that design choices in algorithms embody value judgments and therefore bias the way systems operate. They should also learn that these things are subtle: For example, designing an algorithm for targeted advertising that is gender neutral is more complicated than simply ensuring that gender is ignored. They need to understand that classification rules obtained by machine learning are not immune from bias, especially when historical data incorporates bias.

Big Data and Insurance Claims: What Should Regulators Do?

Recommendation: Each state should require personal lines insurers to report all types of data used for sales, marketing, underwriting, pricing, conditioning payment plan use and claims settlement; the sources of the data; and the uses of the data. The NAIC should develop a template to promote uniformity across the states. The NAIC should also serve as the data collection and compilation agent for those states needing assistance to carry out this request for information.

This information will provide regulators with an overview of the types of uses of non-insurance data and enable regulators to identify big data applications that warrant further regulatory review.

Big Data and Insurance Claims: What Should Regulators Do?

Recommendation: Each state should monitor the outcomes in personal lines markets by collecting and analyzing transaction data on sales and claims. The NAIC should serve as the data collection and compilation agent for those states needing assistance to carry out this request for information.

Given that monitoring and evaluating all the data and assumptions that go into big data claims models, it is reasonable and necessary for regulators to collect detailed data on claims outcomes to evaluate whether there is disparate treatment of certain types of consumers in the claim settlement process – regardless of whether such treatment is intentional or unintentional.