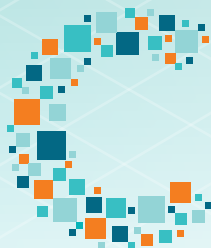


# DATA FOR GOOD

*Promoting Safety, Health, and Inclusion*



**TEC**

U.S. Chamber of Commerce  
Technology Engagement Center

**PERC**

**RESULTS AND SOLUTIONS**



# TABLE OF CONTENTS

Executive Summary .....	4
Introduction.....	6
Fraud Prevention and Security.....	10
An explosion in data flows, no real trend in ID theft, and declining fraud costs ....	10
Online Fraud Detection and Security .....	13
ThreatMetrix: An Example of a Data Driven Solution to Prevent Online Fraud .....	17
Data-Driven Law Enforcement and Security Tools .....	20
Thomson Reuters CLEAR: An Example of Data Driven Solution to Assist Law Enforcement ...	21
Example: Helping Save Communities from Wildfires.....	25
Financial Inclusion .....	26
Expanded Use of Data Increases Access to Credit.....	28
What goes into a credit score, and how does it predict behavior? .....	30
Proven Payment Data: Expanded access to credit, credit visibility .....	31
Who benefits from expanded credit data? .....	34
Other alternative data solutions.....	38
Economic Development .....	40
Health Care .....	49
The Transition from Paper to Electronic Health Records: Progress and Problems ..	49
Filling a Gap: SDOH Data as a Complement to EHR Information.....	53
Artificial Intelligence for Good.....	56
Google’s AI Tools.....	56
Example: AI For Good .....	58
Conclusion.....	59
About .....	61



# EXECUTIVE SUMMARY

---

The use of data is expanding at a rapid pace. Today a smartphone can process more data than what was used in the Apollo spacecraft that landed on the moon. The Data Revolution creates many concerns about issues like privacy and security but it also poses many new opportunities to improve society by enhancing public and personal security, expanding financial inclusion, promoting economic development, bolstering public health and fostering innovation through tools like Artificial Intelligence to solve the challenges of tomorrow. *Data for Good: Promoting Safety, Health, and Inclusion* highlights how the use and sharing of data is leading to societal benefits in these areas and that public policymakers must consider these benefits when deliberating about issues such as privacy, security, and data analytics policy. A well-thought national standard for privacy and security is vital to promoting cutting-edge life-saving and life-improving technology.

## **FRAUD PREVENTION AND SECURITY**

New data-driven tools are being developed to detect and prevent online fraud as well as to authenticate online transactions more generally. These new tools allow consumers and companies to keep one step ahead of would-be fraudsters. Such new tools may help explain why fraud costs have been decreasing. Data tools developed to aid law enforcement work to improve speed and efficiency that in turn, can translate to increased public safety.

## **FINANCIAL INCLUSION**

Data-driven solutions, such as credit scoring and automated underwriting have improved lending, reduced origination costs, increased financial inclusion, and have made lending fairer. New, innovative data-driven tools are being developed to fill remaining data gaps and further improve financial inclusion.

## **GREATER ECONOMIC DEVELOPMENT**

The twenty-first century has been the century of data-driven economic development. Digital Platforms enable small businesses to have a presence and connect with their customers. Payment processing companies allow small businesses to actually make a sale and make a profit. Data allows identification of needs when determining the deployment of disaster aid.

## **PUBLIC HEALTH**

Doctors are increasingly turning to data to improve the quality of care they deliver and improve their ability to practice medicine the way they want.

## **ARTIFICIAL INTELLIGENCE**

Artificial Intelligence among other things provides new ways to combat human trafficking, fight wildfires and connect people to needed mental health services. It is also dependent on the open channels of data to improve the lives of disabled citizens and help underserved communities obtain access to credit.

# INTRODUCTION

---

Most Americans, if asked how the information technology revolution and data impacts and benefits their lives, society, and the economy, would likely focus on a relatively narrow set of consumer-facing services. These would probably include social media, search engines, email, consumer devices, and online sales. If then asked how companies in these technology and data sectors were using data, most would likely say for marketing and advertising purposes. While this would be true, it also overlooks the many other ways data are used and the many ways data benefit them and the economy.

This report is the third part of our U.S. Data Ecosystem series. The focus of this part of the series is on the oft-overlooked uses of core data assets typically associated with other, non-core applications. Many of the particular cases and examples described in this report could be thought of as “data for good” examples. But these should not be confused with simply one-off charitable cases. While many data and tech companies do provide data and resources for charitable purposes—and a few such ancillary applications are featured in our report given their use of core data assets—the examples provided here are to demonstrate the very wide application of data across sectors to, among other things, promote social justice, enable inclusive access to financial services, guide community economic development, and protect people against bad actors and/or bad outcomes.

These socially beneficial applications can be seen either as “data for good” or “AI for good,” but they are also important lines of business for data companies. Instances of socially beneficial applications of data can be found in nearly any sector. While this report draws examples from healthcare, financial services, economic development, and law enforcement, for the report authors, the abundance of compelling examples was an embarrassment of riches.

Finally, it is hoped the illustrations contained in this report will prove valuable to lawmakers currently considering national privacy policy legislation. Examining the multiple ways data are used (and can be used) should be a crucial aspect of any discussion of potential national privacy legislation and how data should be regulated. For instance, while data, Artificial Intelligence (“AI”), and technology platforms are used for marketing and e-commerce they are also used in medical research, to assist law enforcement and to help locate and assist missing and exploited children. And the same data is often used for multiple purposes. So policy makers will need to understand how potential regulations and legislation would impact the entirety of our very interrelated, complex, and dynamic data ecosystem. And then how this would impact society.

The first two parts of the U.S. Data Ecosystem series are found in the paper “Data Flows, Technology, and the Need for National Privacy Legislation.”<sup>1</sup> In that paper we argued Congress needed to act and pass national privacy legislation. We argued that “[t]he public, consumer advocates, privacy groups, members of Congress, major US corporations, the US Chamber of Commerce, the GAO, and a growing list of organizations are calling for a national privacy and data protection law,” and that “[a]ll recognize that data privacy and data protection are national issues and consumers are seeking greater clarity, control, understanding, and where necessary, regulation, around data collected and shared about them.”

Underscoring the need to act is the fact that state legislatures are filling what they see as a legislative vacuum by passing state privacy laws. The proliferation of a patchwork of uncoordinated state laws will likely result in undue business uncertainty, especially within the technology sector, and that may unnecessarily impede data flows, erode American competitiveness and innovation, and ultimately harm overall economic performance.

---

1. Michael Turner, Patrick Walker, & Kazumi Moore, *Data, Technology Flows, and the Need for National Privacy Legislation*. Durham: Policy & Economic Research Council (PERC) & the U.S. Chamber of Commerce Technology Engagement Center (C\_TEC), July 2019, available at <http://www.perc.net/wp-content/uploads/2019/07/ctec.pdf>

Finally we argued that federal privacy legislation needs to be “deliberative, proactive, principles-based, and flexible in the face of a modern and dynamic 21st Century information economy.” Legislators need to balance the aims of consumer protection with the needs of consumer services and other services that utilize data and benefit consumers and the economy.

Part 1 of the U.S. Data Ecosystem series focused on the technology sector’s contributions to the American economy. It concluded that the US technology sector leads the world, in part due to a large, unified domestic market with a regulatory environment conducive to a vibrant and dynamic national technology sector. It found that the tech sector is vital to the current and future competitiveness of the overall American economy, with the positive impact the tech sector and data flows have on *other* non-tech sectors of the economy.

Part 2 of the U.S. Data Ecosystem series utilized national statistics from the government and private organizations over the past 14 years on fraud, ID theft and data breaches. It also utilized millions of anonymized records from a Credit Reporting Agency (CRA), that included data on credit scores, rates of having data on the dark web, credit monitoring activity alerts, and credit file locks for both a group of consumers known to have been part of a data breach and comparison samples.

This second part of the series found that while databases and data flows have exploded and grown exponentially over the past 14 years, known data breaches rose (but to a lesser extent), rates of ID theft and fraud remained little changed and fraud costs actually declined. We were unable to identify a strong, meaningful relationship between the overall rate of data breaches (or records breached) and the overall rate of ID theft and fraud. Using the CRA data, the study concluded that consumers who were part of a data breach were *not* found, as a whole, have more negative impacts in terms of credit score declines, a dark web presence, and other measures relative to the comparison groups. One of the explanations for this seeming disconnect was the fact that much of the incidents of ID theft and fraud are unrelated to data breaches, such as fraud resulting from lost or stolen credit cards, stolen wallets, stolen mail, or malware. Another explanation is that increased investments in fraud detection and prevention over this period, with accompanying new data tools, have made committing fraud more difficult. This would be consistent with the drop in fraud costs witnessed.



In this Part 3 of the U.S. Data Ecosystem series, focus shifts to a deeper dive, via case studies and examples, into the benefits that the non-technology sectors derive from technology as well as data analytics and sharing. Specifically, there are cases and examples from law enforcement, public safety, health care, lending, and community economic development. That is, this is a deeper dive with concrete examples into what was examined Part 1 of the series.

The first case explored in this report, in the section that follows, is on online fraud detection and provides an example of a new data-driven solution developed by LexisNexis Risk Solutions. This provides a concrete example of the better fraud defenses posited in Part 2 of the series.

# FRAUD PREVENTION AND SECURITY

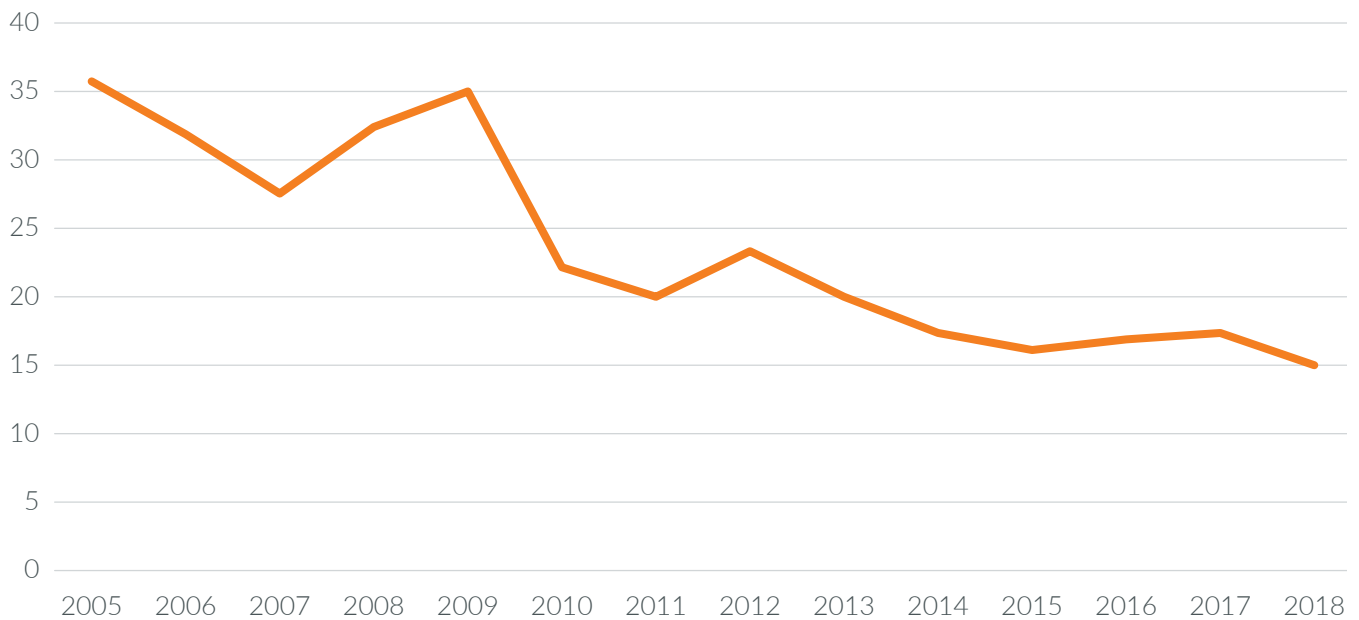
## SECTION SUMMARY

*New data-driven tools are being developed to detect and prevent online fraud as well as to authenticate online transactions more generally. These new tools allow consumers and companies to keep one step ahead of would-be fraudsters. Such new tools may help explain why fraud costs have been decreasing. Data tools developed to aid law enforcement work to improve speed and efficiency that in turn, can translate to increased public safety.*

## AN EXPLOSION IN DATA FLOWS, NO REAL TREND IN ID THEFT, AND DECLINING FRAUD COSTS

Data flows and the size of databases have grown exponentially over the past decades. Not surprisingly, the rate and size of data breaches have also grown, though at a much slower rate. This was shown in “Data Flows, Technology, and the Need for National Privacy Legislation.” That report also demonstrated that despite the growth in data flows, overall rates of ID theft and fraud have remained fairly flat since 2005. Even more intriguing is that since 2005, the overall trend has been a decline in the total costs of ID theft and fraud (see Figure 1).

**FIGURE 1:**  
**Cost of ID Theft and Fraud (billions of US\$)**



Source: Javelin Strategy & Research press releases

The mystery of the decline in the costs of ID theft and fraud (and little change in the rate of ID theft) is likely explained by a few factors.

First, while major data breaches may garner much national media attention, evidence shows that many if not most ID theft and fraud incidences result from non-digital sources, such as stolen wallets, stolen mail, phone scams, or incidents carried out by family members or acquaintances.<sup>2</sup> And when the source is digital, it may involve malware or other non-data breach sources.

It is also worthwhile to note that data breaches existed prior to the internet (think of missing or stolen files or electronic tapes). And, of course, ID theft and fraud is as old as society.

---

2. University of Texas at Austin, *Identity Threat Assessment and Prediction Report 2018*. Austin: University of Texas at Austin, Center for Identity, May 2018, accessed at: [https://identity.utexas.edu/assets/uploads/publications/ITAP\\_Report\\_2018.pdf](https://identity.utexas.edu/assets/uploads/publications/ITAP_Report_2018.pdf)

Second, there may be a limited number of “bad actors” or fraud opportunities, so that while data flows and breached records have increased, only a certain number can be utilized.

Finally, the same information technology revolution and growth in data flows and databases have engendered new ways to identify and prevent fraud and ID theft.

These new ways to prevent fraud can act to prevent online/electronic fraud as well as more traditional, “paper-based” fraud. Compared to today, fraud carried out by forging paper, documents, and IDs in the 1960s may have been easier to commit due to the lack of real-time databases that could be used to verify identify (consider the instances in the book and movie “Catch Me If You Can”).

It may be that because of these three factors that the report “Data Flows, Technology, and the Need for National Privacy Legislation,” found no real trend in ID theft and fraud over the past 15 years and no strong relationship between data breaches and ID theft/fraud.

However, it should be emphasized that there are specific examples of costly ID theft and some breached data is no doubt used to carry out ID theft and fraud. And it should also be pointed out that bad actors are consistently devising new scams and tools, just as new protections and defenses are developed.<sup>3</sup>

---

3. Vijay Balasubramanian, “Deepfakes and Voice as the Next Data Breach.” *Helpnet Security*. 21 October 2019. Accessed at: <https://www.helpnetsecurity.com/2019/10/21/deepfakes-voice/>; see also Allen Kim, “A scam targeting Americans over the phone has resulted in millions of dollars lost to hackers. Don’t be the next victim.” *CNN*. 27 October 2019. Accessed at: <https://www.cnn.com/2019/10/27/business/phishing-bank-scam-trnd/index.html>

## ONLINE FRAUD DETECTION AND SECURITY

Consumers are increasingly utilizing and expecting real-time access to financial services, online shopping, and other online and digital services via computers and mobile devices. Consumers can check their bank balance and transfer funds on their mobile phone, start a shopping order on their work computer, continue shopping on their phone on a train ride home, and then after they get home finish on their home computer. The consumer would also have access to their accounts when traveling on vacation. Consumers take this seamless ease of access to their accounts and online services for granted.

What consumers typically do not realize is how data is used and which databases are utilized to enable safe and secure access to their online services and accounts. Without such efforts and the use of fraud detection, online services would likely become much more vulnerable.

Most of us have likely had an experience in which a credit card or debit card has been temporarily deactivated or a transaction is denied due to suspicious activities. That is, the card issuer suspects fraud. The card issuer then tries to contact the account owner to review recent purchases or to confirm purchase attempts. The suspicion of fraud likely arose from one or more anomalous purchases or purchase attempts that did not match the pattern of purchases for the account holder. This can include purchases made overseas (if the card issuer was not notified that the account holder would be traveling). This analysis of patterns protects the card issuers and account holders. The exact details of what would trigger a fraud concern are typically not revealed (and would differ between issuers). This is so would-be fraudsters do not have a guidebook on how to go undetected.

This last point is an important one. This report will not discuss specific details of fraud detection because doing so could aid would-be fraudsters. Card issuers and other entities that engage in fraud detection will also be circumspect, general, or a bit vague in discussing fraud detection for this reason. This is in the best interest of all stakeholders, including consumers/account holders. This differs from credit origination in which there is a great deal of transparency. In mortgage underwriting, for instance, consumers can know exact credit score cutoffs for acceptance or different pricing tiers, needed debt-to-income ratios, and so on. In lending, that level of transparency *is* in the best interest of stakeholders, particularly consumers.

While such fraud detection efforts are effective and essential in the increasingly digital economy, they are not perfect. Some fraud will not be detected and sometimes fraud will be suspected when there is none. It can be very frustrating for consumers if their credit card or debit card is disabled when they truly are traveling overseas or attempting to make a big or atypical purchase. Because of this, as well as changes in technology and the economy, and evolving threats, companies are constantly seeking new ways to effectively identify and prevent fraud.

The addition of chips in cards makes in-store card purchases more secure; ever-greater use of multifactor authentication improves authentication; use of small charges is used to verify bank accounts; and so on. With multifactor authentication and other security techniques, simply having login credentials becomes less valuable for would-be fraudsters.

As Table 1 below indicates, spending on information technology security amounted to over \$100 billion US in 2018 and was expected to rise faster than overall economic growth or inflation in 2019.<sup>4</sup>

**TABLE 1:**  
**Worldwide IT Security Spending in 2018 (billions of dollars)**

Market Segment	2018
Application Security	2.7
Cloud Security	0.3
Data Security	3.1
Identity Access Management	9.8
Infrastructure Protection	14.1
Integrated Risk Management	4.3

4. Gartner, *Gartner forecasts worldwide information security spending to exceed \$124 billion in 2019*. 15 August 2018. Accessed at: <https://www.gartner.com/en/newsroom/press-releases/2018-08-15-gartner-forecasts-worldwide-information-security-spending-to-exceed-124-billion-in-2019>

TABLE 1 *continued*

Market Segment	2018
Network Security Equipment	12.4
Other Information Security Software	2.1
Security Services	58.9
Consumer Security Software	6.4
<b>Total</b>	<b>114.2</b>

Source: Gartner (August 2018), <https://www.gartner.com/en/newsroom/press-releases/2018-08-15-gartner-forecasts-worldwide-information-security-spending-to-exceed-124-billion-in-2019>

This information technology security spending is across a whole range of market segments, though about half was for professional security services. What may not be as easily captured is that applications, business processes, and the like are being built or redesigned to be more secure. Soft costs of management emphasizing greater adherence to security protocols, for instance, may not have a dollar figure attached to them but can make organizations less vulnerable to fraud and hacks.

Major financial and data companies, especially those with web-based services for consumers and businesses need no reminding of the need for data security. Such entities are under a constant assault of information technology attacks from bad actors, the vast majority of which are blocked. Some companies have set up elaborate command centers to monitor web traffic, counter attacks, and detect intrusions.

An example of how fast the scams and fraud approaches evolve to keep up with defenses (or to take advantage of them) was a type of scam publicized in October 2019.<sup>5</sup> In this particular scam, a bank account holder is called and told they are speaking with the fraud department of their bank and the bank wants to verify whether a charge made on their card was legitimate. In reality, the caller is a fraud-

---

5. Kim, Allen. "A scam targeting Americans over the phone has resulted in millions of dollars lost to hackers. Don't be the next victim." Oct. 27, 2019. <https://www.cnn.com/2019/10/27/business/phishing-bank-scam-trnd/index.html>

ster. When the account holder confirms that the charge was not legitimate, the caller then says that they are sending the account holder a text to verify their identity via two-factor authentication. They then ask the account holder to read the code sent to them. In reality, the text with the code is from the actual bank of the account holder because the fraudster is attempting to change the account holder's password. Once the fraudster has the code, they have access to the account.

The fast evolution of threats is why it is important to be ever vigilant with regard to ID theft and fraud and for fraud defenses to constantly improve and keep one step ahead of would-be fraudsters.

The following is an example of a data-driven fraud prevention tool by LexisNexis® Risk Solutions.



## THREATMETRIX®: AN EXAMPLE OF A DATA DRIVEN SOLUTION TO PREVENT ONLINE FRAUD

ThreatMetrix® is a LexisNexis® Risk Solutions Company that produces tools that utilize a combination of data to verify the integrity of transactions.<sup>6</sup> Specifically, these are examples of tools that use the fact that there is a relationship between people, their accounts, their devices, their locations and the businesses with which they interact. By comparing the elements of a digital transaction to prior transactions containing those same elements, the risk of continuing with a new transaction can be assessed. Importantly, this can be accomplished without storing or relying on raw personal information. For instance, ThreatMetrix produces a unique anonymous alphanumeric identifier for users of devices (there are 1.4 billion recognized users). This anonymous alphanumeric identifier is then linked to devices, locations, etc. Devices are recognized via a number of digital fingerprints that are sent when online. In this way, ThreatMetrix is not based on actual personal identities as the “data subject,” like a credit bureau, but a combination of devices, their locations, and their uses. The actual identities associated with those devices are kept anonymous in the data.

Currently, the ThreatMetrix® Digital Identity Network® works with thousands of digital businesses, looks across billions of devices and, each day, analyzes (millions of) transactions in real-time. Through machine learning and other advanced analytics, the ThreatMetrix tools use this data to build patterns and in turn improve fraud detection capabilities.

ThreatMetrix tools identify devices, keep track of which ones may have been compromised or are associated with fraudulent or suspicious behavior (combined with shared global threat information data). They also detect location cloaking or Internet Protocol spoofing, Virtual Private Networks, or unusual changes in behaviors or patterns. And they seek to identify automated bot attacks, malware, remote access Trojans, and other such device takeovers and compromises.

---

6. Materials for this example came from an interview with Steve Emmert, Senior Director, Government & Industry Affairs RELX Group and information from <https://www.threatmetrix.com/>

That is, these tools look over a vast array of data on a constellation of devices, activities, and transactions (such as use of online shopping sites) and search for patterns associated with fraud.

In this way, when a trusted device is used to make a purchase that is not anomalous, with no red flags detected, the transaction proceeds frictionlessly and hassle-free for the consumer. If suspicious activity is detected, then either additional authentication procedures could be utilized or the transaction could be blocked.

These tools work on the same general principle as pattern analysis with credit card and debit card transactions, however, they leverage a much wider scope of data, and hence can be much more effective at detecting fraud and have greater uses.

These tools can also be integrated with other fraud detection and prevention processes, such as multifactor authentication or even biometrics, or other in-house processes.

Since this pattern analysis tool goes beyond a single card or payment system it can be used as a way to authenticate and secure many online activities. So, while these tools can be used to detect fraud in an online purchase, they can also be used to prevent account takeovers and to verify the veracity of applications for *new* services.

These tools can also help verify the authenticity of remote access of employees to their companies' computer systems. More generally, they can help verify logins to all manner of accounts and help detect compromised accounts. And they can even be used to help authenticate customer/consumer reviews and online comments — giving consumers greater confidence in these valuable ratings. The applications for these tools go well beyond financial transactions and extend to online activities in general (insurance, healthcare, government, education, media, remote logins, etc.).

These data-driven anti-fraud and online authentication tools demonstrate how it is that over the past 14 years the volume of data online can increase exponentially, that data breaches can rise, but the cost of fraud can decline with a lack of a strong relationship between data breaches and fraud and ID theft. Namely, the same increase in online activity and data that *could* fuel fraud also empowers fraud detection and defense.

Companies such as Experian, LexisNexis Risk Solutions, and Thompson Reuters among others have developed real-time sophisticated solutions using many core data assets built for other purposes to detect potentially fraudulent activity, to verify personal identity for online transactions, and to do all this in a privacy-sensitive manner.

Private sector spending on cyber-security was projected by the Gartner to exceed \$124 billion in 2019, while public sector spending was estimated to be approximately \$17 billion.<sup>7</sup> While it is difficult to quantify the subset of this investment dedicated exclusively to protecting consumer data as opposed to corporate intellectual property and other internal assets, even if only 10% of total annual cyber-security spending were allocated to protecting sensitive personal information, that still represents over \$14 billion annually. Given this level of commitment, it is not hard to understand why fraud losses associated with identity theft have been trending downward for the past decade.

Any national privacy legislation that unduly restrains access to these valuable data elements would likely reduce the ability of data firms to so efficiently protect consumers from fraudsters. This must be considered by lawmakers when deliberating over the details of national privacy legislation.

---

7. Ratnam, Gopal. "Cybersecurity budget up 5% in 2020, according to White House." Roll Call. 19 March 2019. Downloaded at: <https://www.rollcall.com/news/whitehouse/cybersecurity-up-5-percent-in-2020-budget-white-house-says>

## DATA DRIVEN LAW ENFORCEMENT AND SECURITY TOOLS

Law enforcement has been utilizing new technologies and IT advances to improve operations, reduce crime, and better identify and apprehend suspects. Some of the flashier new technologies such as the use of unmanned aircraft, DNA testing and matching, and sensors that detect and triangulate gunshots, have been written about extensively. The use of data by law enforcement has also received attention.<sup>8</sup> The precipitous reduction in violent crime in NYC in the 1990s (by nearly 60% compared to 28% for the nation as a whole, including a whopping 73% reduction in homicides) was attributed in part to Compstat—a tool showing clusters of crimes on maps to more efficiently target law enforcement activities.<sup>9</sup> This tool had its origins in mapping pins for subway crimes, which was also attributed with a dramatic reduction in crimes.<sup>10</sup> More recently, many may have also come across the greater data-driven law enforcement efforts by exploring online GIS crime maps such as one created for Chicago.<sup>11</sup>

Like many activities, law enforcement, particularly the investigative aspects of it, rely on the processing of information, for instance determining the current or past addresses of a suspect. Information technology advances and modern, searchable databases have made such tasks faster and more efficient.

To aid law enforcement a number of data-driven tools have been developed, what follows is an example of one set of tools called Clear developed by Thomson Reuters.

- 
8. Some of the media attention has focused on the more sensationalistic aspects of data driven law enforcement, see Jeff Asher & Rob Arthur, “Inside the Algorithm that Tries to Predict Gun Violence in Chicago.” *The New York Times*. 13 June 2017. Accessed at: <https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html> or The Economist, “How data-driven policing threatens human freedom.” *The Economist*. 4 June 2018. Accessed at: <https://www.economist.com/open-future/2018/06/04/how-data-driven-policing-threatens-human-freedom>
  9. Corman, Hope and Naci Mocan. “Carrots, Sticks and Broken Windows.” National Bureau of Economic Review. NBER Working Paper No. 9061. July 2002. <https://www.nber.org/papers/w9061.pdf>
  10. Dussault, Raymond. “Jack Maple: Betting on Intelligence.” *Government Technology*. 31 March 1999. Downloaded at: <https://www.govtech.com/magazines/gt/Jack-Maple-Betting-on-Intelligence.html>
  11. City of Chicago, *Crimes Map*, Chicago Data Portal, accessed at: <https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6>

## THOMSON REUTERS CLEAR: AN EXAMPLE OF A DATA DRIVEN SOLUTION TO ASSIST LAW ENFORCEMENT

Thomson Reuters CLEAR integrates a large number of high quality and up-to-date data resources useful to law enforcement in a single platform with an easy-to-use interface, advanced search and data aggregation capabilities.<sup>12</sup> This enables law enforcement to focus on the core tasks of public safety versus needing to learn complex software or interfaces or pulling together data from disparate databases or time-consuming sources.

Combining multiple data resources in a single integrated platform with advanced searching capabilities is crucial for law enforcement. For instance, unlike business transactions that would have first and last name, date of birth, social security number, and other variables, record searches carried out by law enforcement might need to utilize only a disparate set of incomplete data to locate subjects of interest. For example, law enforcement may have a person's possible first name, the city they may potentially live in, and a couple of characters of their license plate. Or the search may involve a different set of data altogether. For this reason, an integrated platform that brings together many different data elements with advanced search capabilities allows law enforcement to easily perform a query on what information they do have. This may result in a single subject being identified or several. If the pool of potential subjects is large, then other ways to refine the search can be explored. If list of potential subjects is sufficiently narrow, then traditional investigative methods can be used to further hone the list and eliminate potential subjects.

A good example of the importance of these capabilities is a case presented on the product website for CLEAR.<sup>13</sup>

---

12. Information for this example came from an interview with Kate Friedrich, Vice President, Global Government Affairs, Thomson Reuters and information from <https://legal.thomsonreuters.com/en/products/clear-investigation-software>

13. Thomson Reuters, Clear Case Studies, *Detroit Amber Alert*. Accessed at: <https://legal.thomsonreuters.com/en/insights/case-studies/clear-case-study-detroit-amber-alert>

An 11-month-old child, diagnosed with sickle cell anemia and needing regular medical treatment, had just completed a doctor's appointment where the foster care mother and birth mother, who was not allowed unsupervised visits, attended. Following the appointment, the birth mother abducted the child.

The Detroit Police Department arrived on scene and immediately realized the added time-sensitive nature of the situation, due to the child's age and medical condition. The police canvased the area, interviewed known associates and issued an Amber Alert. They also pulled every report they had in relation to the birth mother. CLEAR provided numerous past addresses and the police began incorporating that information into their investigation, going house to house.

One officer recalled being involved with the birth mother as part of a foster care situation and that she had called him via a mobile phone. Upon entering that number into CLEAR, an unfamiliar name and address was displayed. This was the break needed and the police were able to locate the foster mother and abducted child.

As the emergency medical services and backup police arrived, they convinced the birth mother to let them examine the baby. They moved the child to the EMS vehicle and then immediately transported it to the area hospital, where the child was administered its medication and released back to the foster mother.

As is pointed out in this case, law enforcement may often have scattered pieces of a puzzle, and thus needs tools like CLEAR to assist them.

This is one of many cases in which tools like CLEAR are helping law enforcement improve operations, solve crimes, and apprehend suspects. Most of these instances, much less the fact that advanced data tools are aiding the investigations, are not publicized or otherwise garnering media attention. This helps protect the data tools from being gamed by would-be perpetrators and, in turn, contributes to the ongoing protection of the general public by law enforcement officers.

Another example of how core data assets—public records data, publicly available information and third-party data assets gathered from private sector sources—can be combined to greatly enhance law enforcement capabilities was demonstrated in the aftermath of a mass shooting at an event in the Inland Regional Center for San

Bernardino health department employees in San Bernardino, California. <sup>14</sup>

An attendee of the event and county health inspector left the event and returned with his wife, both dressed in combat gear. They then killed 14 people with assault rifles.

After police arrived, witnesses told them that the health inspector was upset and had left the event prior to the incident. A witness also told police that the suspected perpetrators had escaped the scene in a black SUV. The police department received a report that a vehicle matching the description of the black SUV was seen and had Utah plates (with the license plate number identified). It was a rental car and the name on the rental car contract was that of the health inspector.

Law enforcement then employed CLEAR to locate and verify the home address of the suspect. CLEAR produced an additional address that was different than the one found on the suspects rental car agreement and other verifiable sources. It turned out that the different address produced by CLEAR was more consistent with the location of a cell phone ping from the individual's county-owned cell phone than the other address, so law enforcement focused on the CLEAR produced address.

As law enforcement was arriving at the address that was only found in CLEAR, they noticed the black SUV was just leaving the location. When police intercepted the SUV, a firefight broke out. Both suspects were killed, and an officer was injured.

The suspects were heavily armed and also had a detonator device for an explosive left at the site of the mass shooting. If law enforcement would have taken more time to identify the suspects' addresses, or went to the other addresses they had on file from other databases and documents, they would have been many steps behind and the suspects would have had opportunities to carry out additional attacks—including potentially detonating bombs in a crime scene now populated by dozens of law enforcement individuals and many first responders.

---

14. Thomson Reuters, Clear Case Studies, *San Bernardino Mass Shooting*. Accessed at: <https://legal.thomsonreuters.com/en/insights/case-studies/san-bernardino>

Access to disparate public sector and private sector resources and combined with meaningful data analytics is key to CLEAR's ability to offer real-time actionable intelligence to the law enforcement community.

These case studies and others can be found on the CLEAR product website.<sup>15</sup> They demonstrate how data driven tools can aid law enforcement and increase public safety.

In context of national privacy legislation, the important role played by ongoing access to an array of data assets to protecting the public from bad actors must be given full consideration. Restricting access to this data via legal obligations that would have a dampening effect on enabling access, or that would give bad actors the right to opt-out to having their data shared, could contribute to the creation of blind spots that would hamper the ability of law enforcement to quickly and accurately locate bad actors and save lives. While the examples provided here are for illustrative purposes, a growing range of law enforcement applications make these data flows critical infrastructure for law enforcement and public safety.

---

15. Thomson Reuters, *Clear Case Studies*. Accessed at: <https://legal.thomsonreuters.com/en/products/clear-investigation-software/case-studies>





## EXAMPLE

# ***HELPING SAVE COMMUNITIES FROM WILDFIRES***

Wildfires have become a global crisis with severe economic, environmental, and public safety impacts. Destructive and deadly fires in California, Australia, Southern Europe, and the Amazon have generated headlines in recent years. In order to better plan for, predict, and respond to wildfires, innovative firefighters and technology companies are turning to data-driven solutions.

There is a tremendous amount of data relevant to making better decisions during a wildfire. Data on weather (past, current, and forecasted), topography, vegetation (aka fuel), aspects of a fire, location of houses, hospitals, schools, roads, historical and real-time traffic, power lines, and available firefighting resources, could all be taken into account in determining the optimal ways of responding to a particular wildfire. Also, newly available forms of data related to sensing and alerting for wildfires – everything from ground-based sensor networks, satellite imagery, and mountain top cameras – are being explored to automatically alert for the existence of the fire itself. Innovators are currently working to integrate this data effectively into technology solutions that firefighters and other emergency responders can utilize to save lives and property.

Splunk, the global data analytics software company, has been partnering with software startup Zonehaven, that integrates and analyzes “real-time data from weather stations, field sensors and other sources to detect fires earlier, pinpoint the ignition location and dispatch faster, more informed evacuation recommendations.” Speed matters in fighting fires, and in evacuations, minutes can be crucial. Zonehaven is working with several Bay Area, California fire jurisdictions on this important data application.

This cutting-edge use of data integration and analysis is highlighted in a video created by Splunk. In that video a firefighter notes that while the hard work of putting out fires will still need to be done by firefighters, data and technology can aid and speed the efforts of firefighters and carry out the crucial task of monitoring for fires.

Source: [https://www.splunk.com/en\\_us/resources/videos/Using-Data-to-Gain-an-Edge-Against-Wildfires.html](https://www.splunk.com/en_us/resources/videos/Using-Data-to-Gain-an-Edge-Against-Wildfires.html) and interview with Corey Marshall, Director, Splunk for Good.

# FINANCIAL INCLUSION

## SECTION SUMMARY

*Data-driven solutions, such as credit scoring and automated underwriting have improved lending, reduced origination costs, increased financial inclusion, and have made lending fairer. New, innovative data-driven tools are being developed to fill remaining data gaps and further improve financial inclusion.*

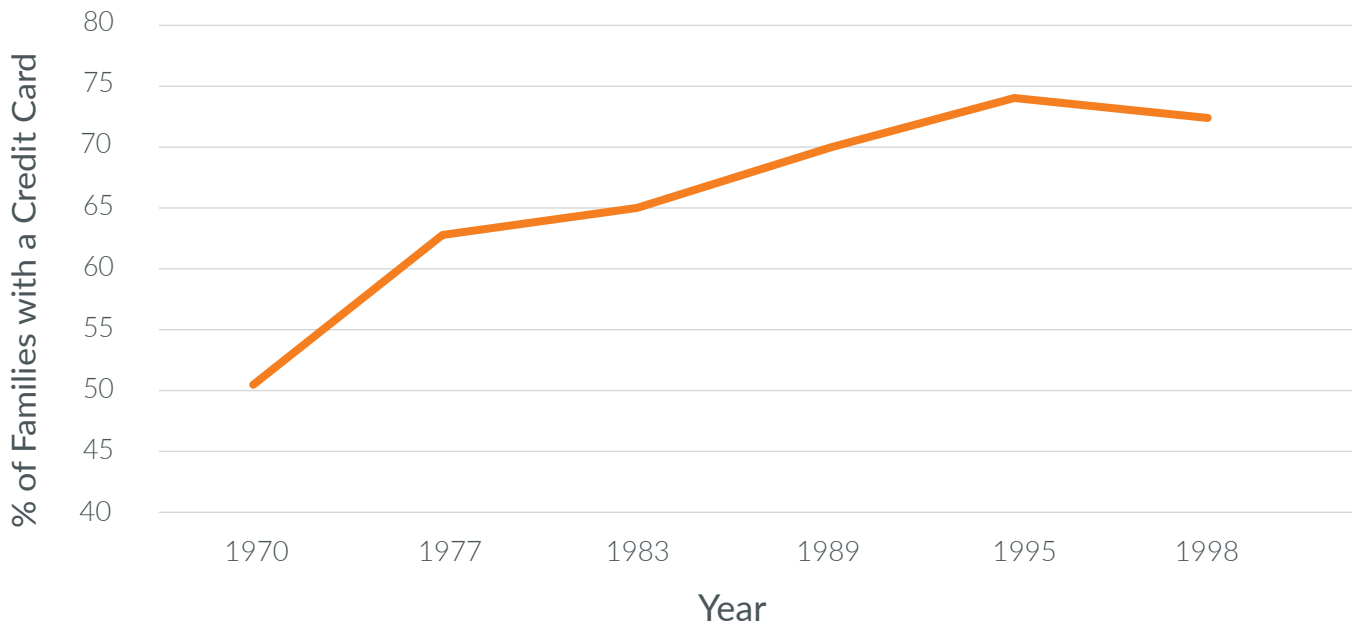
Banking is a sector that has evolved to be increasingly data-driven, culminating in its current form, FinTech (Financial Technology). Data addresses the imbalance in information between lenders and borrowers, the gap or lack of information a lender has about whether the borrower will repay their loan, often referred to as “information asymmetry.” Information asymmetry results in credit rationing,<sup>16</sup> which is a market failure when borrowers who are willing to pay higher interest rates are excluded from the credit market.

Data used to fill this gap has been standardized and packaged in credit reports, now heavily regulated due to their importance to consumers and the economy. This data is collected by credit bureaus (CRAs) and shared with lenders, and contributed to access to credit and financial inclusion expanding in the US. Figure 2 shows how credit card ownership grew in the US, and Figure 3 shows how the price of credit dropped for everyone as lenders were able to tailor interest rates to individual risk, a practice known as “risk-based pricing.”

---

16. Dwight Jaffee & Thomas Russell, “Imperfect Information, Uncertainty, and Credit Rationing.” *The Quarterly Journal of Economics* 90 (4), November 1976: 651-666. Retrieved from <https://www.jstor.org/stable/1885327>.

**FIGURE 2**  
**Prevalence of Credit Cards Among U.S. Families**



Source: Federal Reserve

**FIGURE 3**  
**Distribution of Account Balances by Interest Rate**

Interest Rate Tier	1990	2002
<5.5	6	15
5.5-10.99		31
11-16.49		25
16.5-17.99	20	3
18%+	73	26

Source: Barron & Staten

## EXPANDED USE OF DATA INCREASES ACCESS TO CREDIT

The data in credit reports is analyzed with statistical models to predict the likelihood a borrower will default on their loan. There are two aspects to this: a borrower's capacity to pay, and a borrower's willingness to pay. By observing many borrowers' characteristics and how they behave, lenders and credit bureaus make connections and use them to predict future behavior.

Barron & Staten compared the U.S. and Australian models of credit information sharing – Australia does not share positive data, only negative data.<sup>17</sup> They found that including positive data improved the performance of credit scoring models (and by extension lending) in every metric:

- Acceptance rate by target default rate: if the target default rate of a loan portfolio is 3%, more loans can be accepted with positive data included
- Default rate by target acceptance rate: if the target acceptance rate of a loan portfolio is 75%, less loans are defaulting loans with positive data included
- Type I errors: “false positive” or a loan thought to be good that turns out to be bad; the incidences of these decreases with positive data included
- Type II errors: “false negative” or a loan thought to be bad and therefore excluded, although it is actually good (loss of revenue); the incidences of these also decreases with positive data included.

With more information, credit scoring models are better able to identify risk. With fewer bad loans, less money is lost, and this money is deployed to make more money on more good loans. Credit is expanded. Pagano and Jappelli have also found that lower default rates as an effect of information sharing lead to increased lending, with both positive and negative data sharing being significant.<sup>18</sup>

---

17. John Barron & Michael Staten, “The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience,” 273-310, in Margaret M. Miller ed., *Credit Reporting Systems and the International Economy*. Cambridge: The MIT Press. 2003.

18. Tullio Jappelli & Marco Pagano, “Information Sharing, Lending, and Defaults: Cross-Country Evidence.” *Journal of Banking & Finance* 26, 2002: 2017-2045. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0378426601001856>

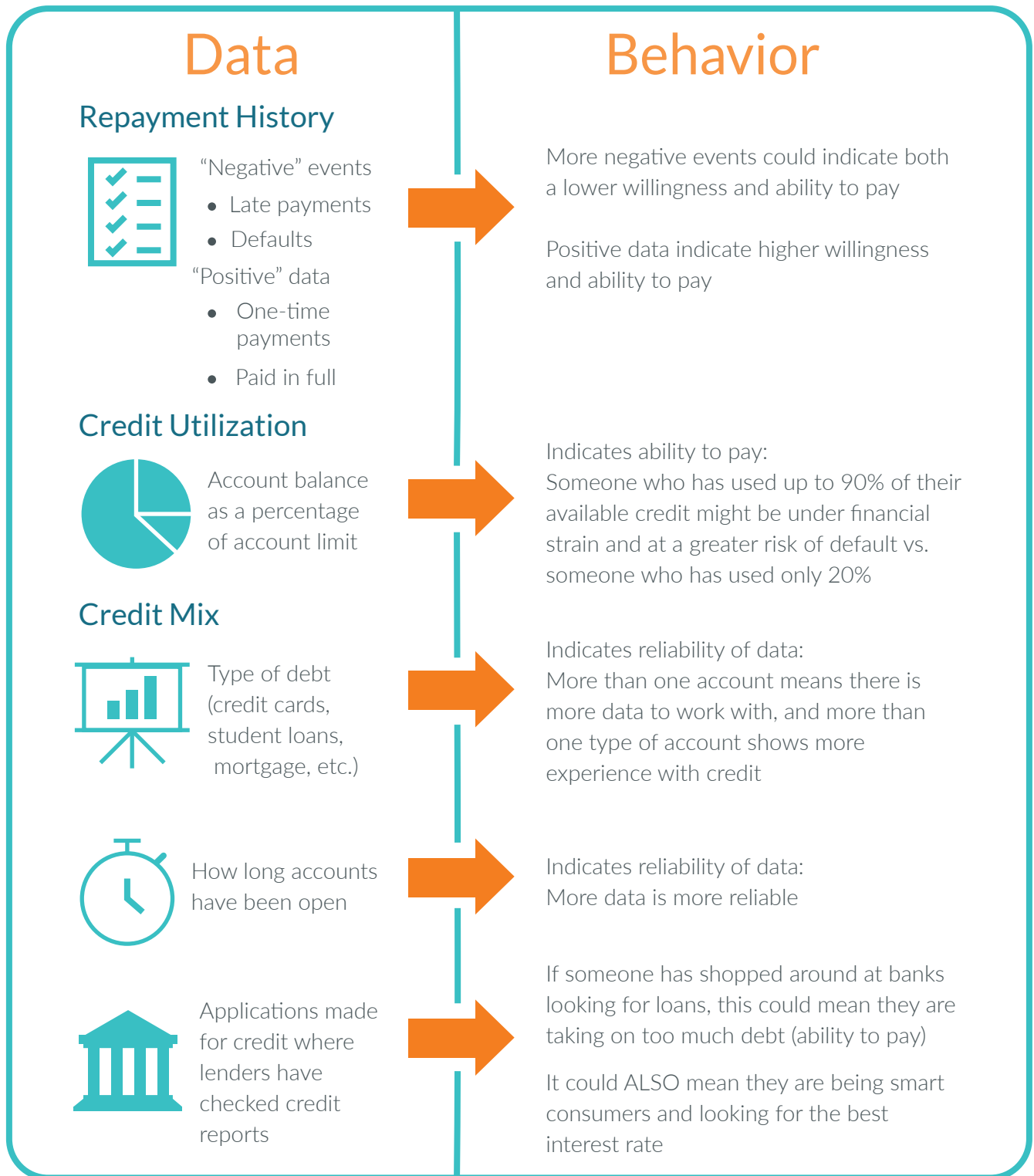
Small businesses also benefit from this effect, although commercial credit bureaus may share information that differs from that contained in consumer credit reports. Love and Mylenko have found that countries with strong credit reporting systems lead to higher shares of bank financing of businesses and lower perceived financing constraints.<sup>19</sup> Frame, Srinivasan, & Woosley find that even the smallest businesses benefit from credit scoring, which is associated with an 8.4% increase in the portfolio share of small business loans, or \$4 billion per institution.<sup>20</sup>

In many ways credit data is the *original* big data. Large, high quality CRA datasets have been in use for decades now in the US. These data have helped drive data-driven processes in lending with the resultant benefits described above. Risk analysis and credit scores have inspired similar developments in other sectors.

Given credit data's relatively long history in terms of large datasets, its regulatory evolution, and the Fair Credit Reporting Act, may also be useful guides to policy-makers.

- 
19. Inessa Love & Nataliya Mylenko, *Credit Reporting & Financing Constraints*. World Bank Policy Research Working Paper 3142, Washington D.C.: World Bank Group, 2003. Retrieved from <http://documents.worldbank.org/curated/en/773601468761701929/pdf/wps3142.pdf>
  20. W. Scott Frame, Aruna Srinivasan, & Lynn Woosley, "The Effect of Credit Scoring on Small Business Lending." *Journal of Money, Credit and Banking* 33(3), August 2001: 813-825. Retrieved from <https://www.jstor.org/stable/2673896>

# WHAT GOES INTO A CREDIT SCORE, AND HOW DOES IT PREDICT BEHAVIOR?



Relatively few Americans now remember the days of being unable to pay for things with money they didn't have on them at the moment (charging something to their credit card), and fewer still remember the days of dressing up in their Sunday best to meet their loan officer. Data-driven processes have enabled this lending evolution.

But what happens when a system of data sharing is built up, becomes critical financial infrastructure, and consumer data is not captured?

## **PROVEN PAYMENT DATA: EXPANDED ACCESS TO CREDIT, CREDIT VISIBILITY**


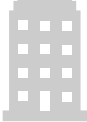





Once the information sharing system is established, the lack of data itself can become an issue. Data is expected to be in the system. This leads to what is called a “credit catch-22” where an individual needs a credit history to obtain credit, but that individual needs to get credit to obtain a credit history.<sup>21</sup> People who have no credit history or thin credit files, and are unable to get into this system, are called “credit invisible.”<sup>22</sup> Lenders are unable to determine whether a credit invisible person is high-risk or low-risk – so they typically treat them as high-risk. Higher-risk people pay higher interest rates or get rejected for credit, which is particularly problematic for lower-income people who are more likely to be credit invisible.

---

21. Michael Turner & Alyssa Lee, *Give Credit Where Credit is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data*. Washington, DC: The Brookings Institution, December 2006, available at [http://www.perc.net/wp-content/uploads/2013/09/alt\\_data.pdf](http://www.perc.net/wp-content/uploads/2013/09/alt_data.pdf)

22. Kenneth Brevoort, Philipp Grimm, & Michelle Kambara, *Data Point: Credit Invisibles*. Washington D.C.: Consumer Financial Protection Bureau, May 2015, available at [https://files.consumerfinance.gov/f/201505\\_cfpb\\_data-point-credit-invisibles.pdf](https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf)



What is reported?		What isn't?	
	<b>Mortgage</b>		<b>On-time rent payment</b>
	<b>Auto loans</b>		<b>On-time utility payment</b>
	<b>Credit cards</b>		<b>On-time telecom bill payment</b>
	<b>Student loans</b>		<i>*Negative data is reported!*</i>

Including data that is currently not reported in credit reports, such as utility payment data, telecommunications payment data, and rent payment data, will act to improve credit-scoring models and their coverage of consumers.<sup>23</sup> These categories are termed “proven payment data” because the effects of this data have been proven in both research and practice. Services like utility and telecommunications are given before payment is made (credit like), which makes them useful as indicators of payment behavior.<sup>24</sup> There may be other data out there that would be beneficial to Americans and should also be included, such as rental payments. Many companies are testing and experimenting on ways to include more Americans in the financial system with such data.<sup>25</sup>

23. Turner & Lee, *Give Credit Where Credit is Due*.

24. Michael A. Turner, *Giving Underserved Consumers Better Access to the Credit System: The Promise of Non-traditional Data*. New York: Information Policy Institute, July 2005, available at <http://www.perc.net/wp-content/uploads/2013/09/nontrad.pdf>

25. Michael Turner, Patrick Walker, & Kazumi Moore, *Alternative Data in the US: Progress, Promise, and Paralysis*. Durham: Policy & Economic Research Council (PERC), August 2019, available at <https://www.perc.net/wp-content/uploads/2019/08/AltDataUS.pdf>



Utilizing more (useful) information generally leads to an increase in acceptance rates and a decrease in default rates through improved risk modeling and improved lending. And making credit data more *inclusive* and reducing the rates of credit invisibility can also make lending fairer and more inclusive (see Figure 4).

Adding new types of data to the system allows people *outside* the formal credit system to be seen, scored, and have access to lower cost mainstream credit.<sup>26</sup> While not everyone owns a car, almost everyone pays for electricity, gas, and water.<sup>27</sup> Most Americans use a telecommunications service of some kind; though they may not have credit cards.<sup>28</sup> People who do not own homes rent them. Expanding data used in credit determination captures more Americans and enables better decisions. These new data sources could become more important as societal behavior changes – for example, millennials use more debit cards relative to credit cards, rent more, and have lower home and car ownership rates.<sup>29</sup>

---

26. Michael Turner, Patrick Walker, & Chet Wiermanski, *Changing the Lending Landscape: Credit Deserts, the Credit Invisibles, and Data Gaps in Silicon Valley*. Chapel Hill: Policy & Economic Research Council (PERC), November 2017, available at [http://www.perc.net/wp-content/uploads/2017/11/Credit\\_Desert.pdf](http://www.perc.net/wp-content/uploads/2017/11/Credit_Desert.pdf)

27. Turner & Lee, *Give Credit Where Credit is Due*.

28. *Ibid.*

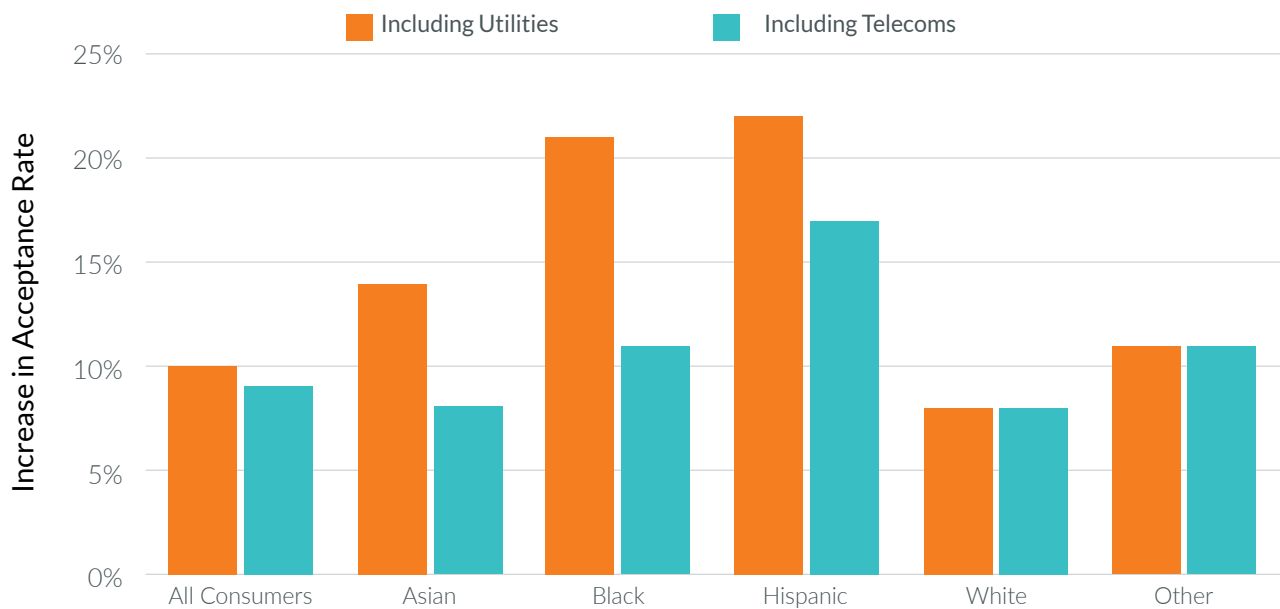
29. Michael Turner, Robin Varghese, & Patrick Walker, *Research Consensus Confirms Benefits of Alternative Data*. Chapel Hill: Policy & Economic Research Council (PERC), March 2015, available at <http://www.perc.net/wp-content/uploads/2015/03/ResearchConsensus.pdf>

# WHO BENEFITS FROM EXPANDED CREDIT DATA?

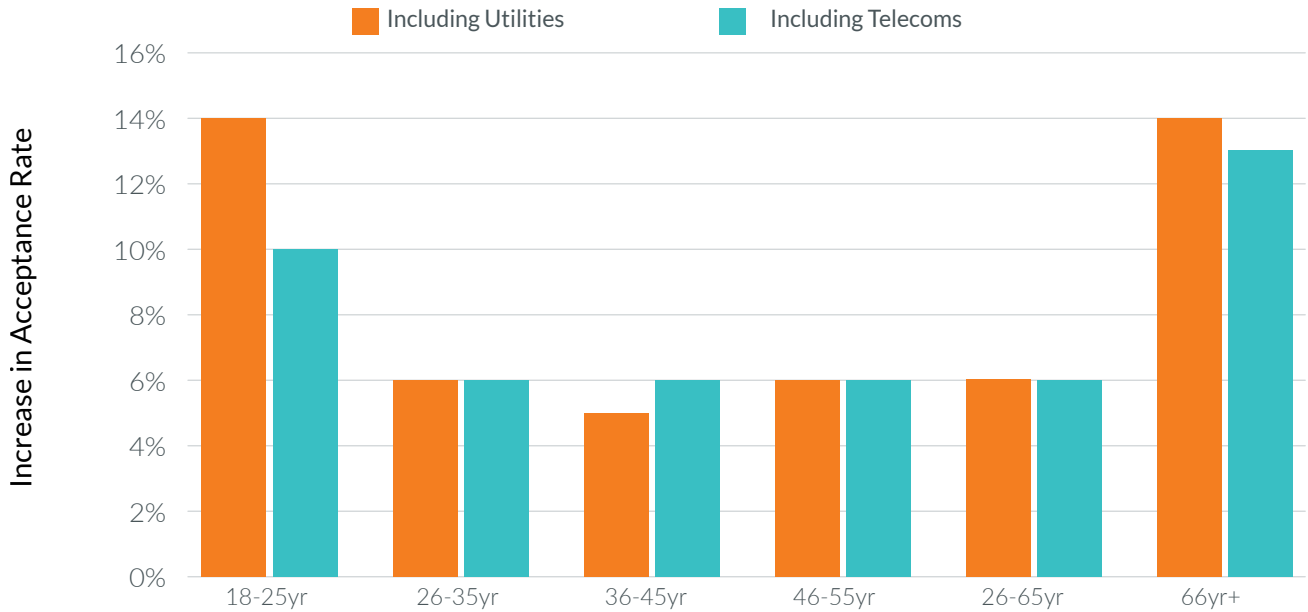


**FIGURE 4**  
**Effects of Including Utility and Telecom Payment Data**

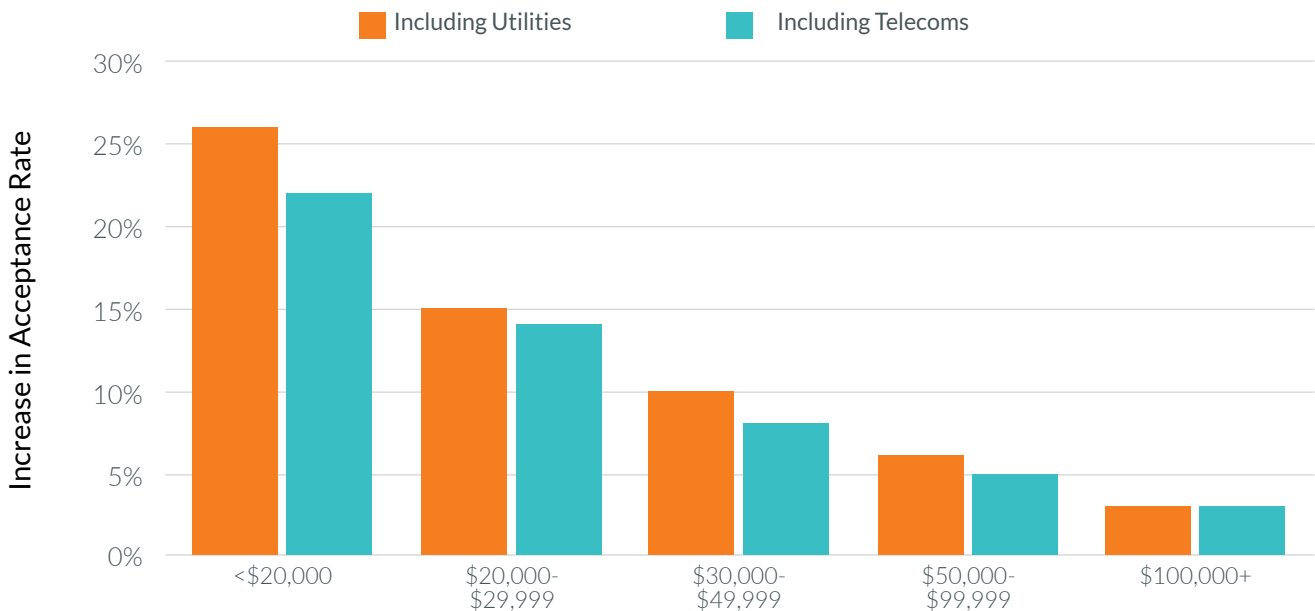
Consumers by Race with Utility and Telecom Trades (Assume 3 percent Serious Delinquency Rate)



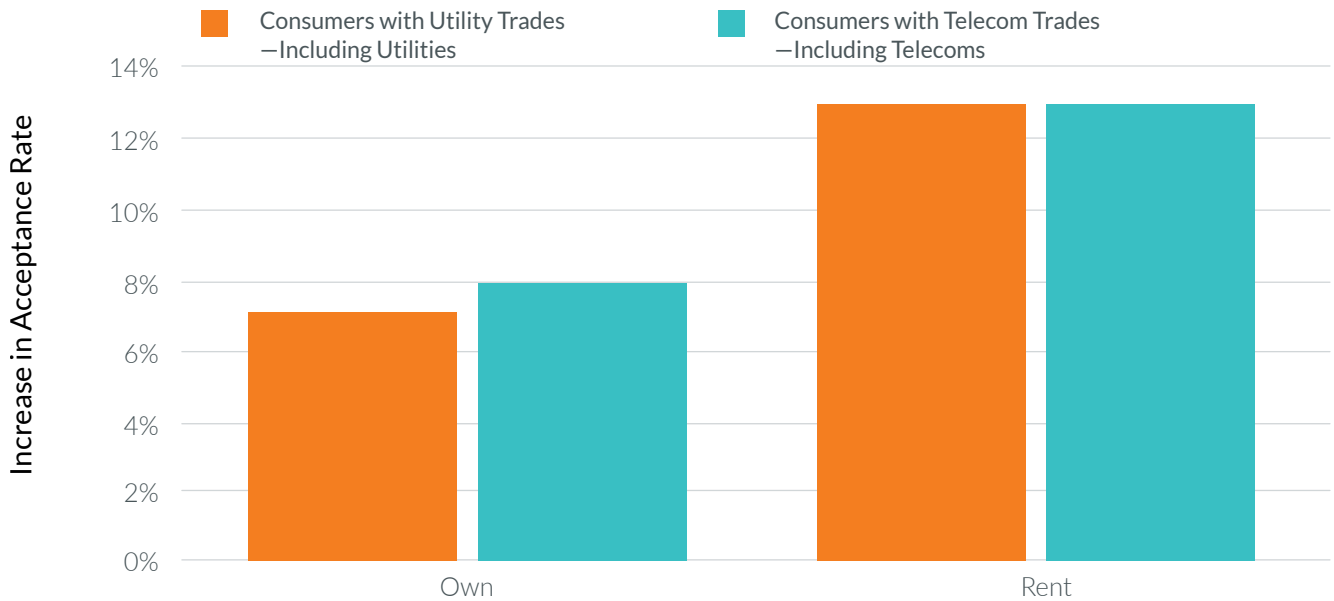
### Consumers by Age with Utility and Telecom Trades (Assume 3 percent Serious Delinquency Rate)



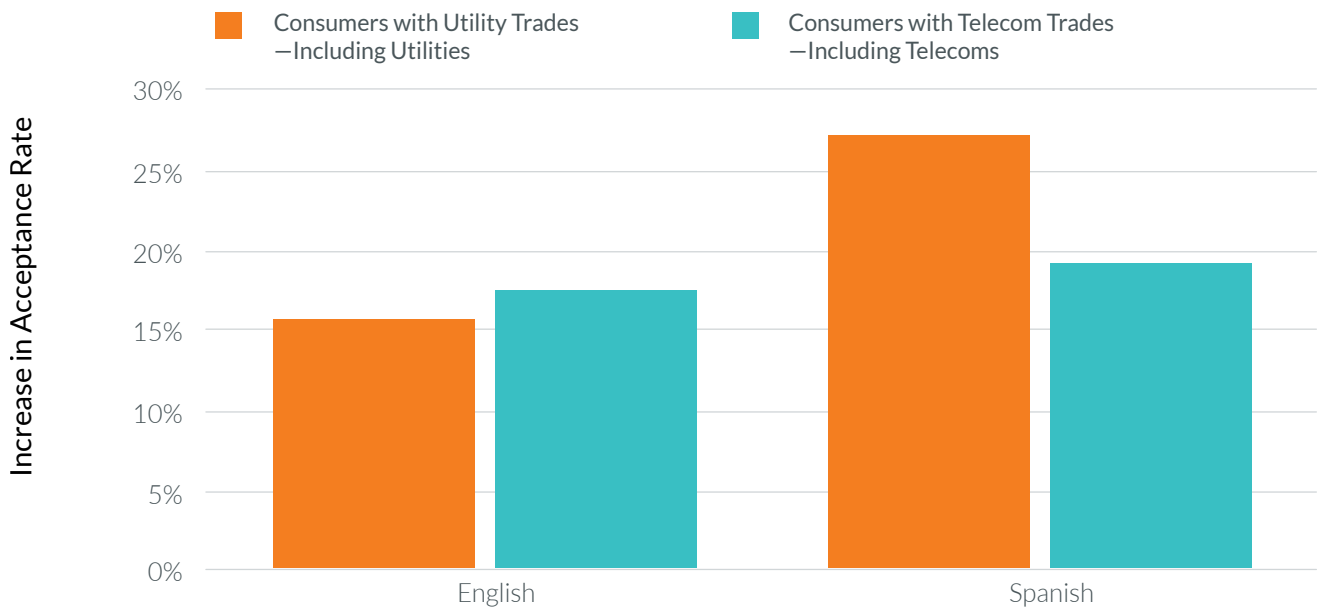
### Consumers by Income with Utility and Telecom Trades (Assume 3 percent Serious Delinquency Rate)



## Consumers by Homeownership Status with Utility and Telecom Trades (Assume 3 percent Serious Delinquency Rate)



## Consumers by Language Preference with Utility and Telecom Trades (Assume 3 percent Serious Delinquency Rate)



Source: Give Credit Where Credit Is Due

Most Americans pay their bills on time. And research has shown that including proven payment data would raise credit access, and this effect is particularly true within traditionally under-served populations.<sup>30</sup>

Barron & Staten also modelled the effect of restricting credit reporting data by lender, such as when financial and non-financial institutions do not share information about borrowers, and found that full-information comprehensive models outperform models with only sector-specific data.<sup>31</sup> This adds to the strong argument for the inclusion of non-financial payment data such as proven payment data, which is essentially credit in nature, to be included with traditional credit data and used in credit origination.

---

30. Turner & Lee, *Give Credit Where Credit is Due*.

31. Barron & Staten, "The Value of Comprehensive Credit Reports."

## OTHER ALTERNATIVE DATA SOLUTIONS

As mentioned above, many companies are exploring ways to expand financial inclusion. Some of these companies are utilizing proven payment data, some are using other types of alternative data, some are using expanded application data, and some are using so-called consumer permissioned data. Ideally, more information is better, and so layering one type of data on top of another is the best way to improve lending and increase financial inclusion. So, if CRA data exists, it should be used, with other types of data layered on if they are available, for instance. The FICO XD credit score, for instance, combines LexisNexis® Risk Solutions data elements, with Equifax CRA data, and National Consumer Telecom & Utilities Exchange data on utility and telecom payments.

Upstart is an example of a lender using new data and new approach in lending. It was issued the Consumer Financial Protection Bureau's first no-action letter regarding their use of education data to underwrite loans.<sup>32</sup> LexisNexis® Risk Solutions uses enhanced public records data, such as address stability and asset data to score credit invisibles.<sup>33</sup> Data such as cash flow and transactions from bank accounts are also being tested.<sup>34</sup> It also may be the case that while many such data elements may be useful in underwriting and credit scores, they may not fit neatly in the typical credit bureau framework.

Some of these solutions are alternatives to reporting data directly to credit bureaus. One of these is the consumer-permissioned model, where consumers authorize a third-party to access their bank account or bill statement data. Typically, consumer-permissioned data from direct deposit and other financial accounts is sent directly to a lender, or with Experian Boost for instance, is added to a credit report where the data is combined with traditional CRA data and can be used by credit scores like FICO and VantageScore. Unlike traditional credit reports, where data is automatically added,

---

32. Consumer Financial Protection Bureau, "CFPB Announces First No-Action Letter to Upstart Network," September 24, 2017, available at <https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/>

33. Matthew Heaslip, Enterprise Case Study: Solving Underbanked Credit Scoring. Ovum, September 2016, available at <https://risk.lexisnexis.com/insights-resources/article/enterprise-case-study-solving-underbanked-credit-scoring>

34. Turner, Walker, & Moore, *Alternative Data in the US*.

consumer-permissioned data requires opt-in and explicit consent from the user for the data access. Information technology advances and online consumer access to account information has made this model possible (if not inevitable, given the ease of sharing data). Fintech (Financial Technology) is an entire industry dedicated to disrupting the traditional financial services sector. These market developments are making the US financial services sector fairer and more inclusive.

## **LEXISNEXIS® RISKVIEW™ CASE STUDY**

LexisNexis RiskView uses and analyzes non-tradeline data and LexisNexis® Risk Solutions estimates that it can provide ~80% of previously unscorable consumers with a way to be assessed for risk in order to access needed financial products. RiskView can be used by different types of lenders to generate scores and attributes on consumers. There are a number of data elements that can enter RiskView to be used to assess a borrower's propensity to pay. These include educational status, professional licenses, derogatory public record history, address stability, and assets owned.

### **PARTNERSHIP**

Kinecta Federal Credit Union partnered with LexisNexis® Risk Solutions to use the RiskView score to underwrite loans for borrowers at Nix Neighborhood Lending (formerly Nix Check Cashing).

### **PRODUCT**

Payday Payoff Loans designed to help borrowers get out of the cycle of debt by consolidating high-cost debts at more affordable rates.

### **NEED**

Kinecta understood that many applicants would not have a traditional credit profile but still needed a way to assess borrower risk.

### **SOLUTION**

RiskView scores were found to correspond to default risk on Payday Payoff products, meaning RiskView could be used to fill in data gaps for credit invisibles applicants at Kinecta.

These innovations reduce credit invisibility and expand access to credit, providing solutions to the problems of financial inclusion. Given the importance of credit reports not only to credit, but to employment, insurance, and renting a residence, it is imperative that all Americans are credit *visible* and have access to the mainstream financial system.

# ECONOMIC DEVELOPMENT

## SECTION SUMMARY

*The twenty-first century has been the century of data-driven economic development. Platforms enable small businesses to have a presence and connect with their customers. Payment processing companies allow small businesses to actually make a sale and make a profit. Data allows identification of needs, such as Square's decision to accept cash or PERC's assessment of Louisiana Disaster Recovery Fund aid. Mastercard's Data for Social Good program combats the inequality in the collected data itself.*

Small businesses are the lifeblood of the US economy. Decker et al estimate that start-ups and high-growth firms (most of them young), in an average year, together account for approximately 70% of firm-level gross job creation,<sup>35</sup> although not all of those jobs will survive as some small businesses fail. The U.S. Small Business Administration estimated that businesses with less than 500 employees contribute approximately \$6 trillion to the economy.<sup>36</sup>

In Part I of our report *Data Flows, Technology, and the Need for National Privacy Legislation*, we saw how data itself powers the economy, as technology not only empowers other sectors to become more efficient, but becomes an increasingly important sector in itself. Although measuring the tech sector is often difficult, some estimates attribute to the technology sector 4% of total U.S. employment (5.9 million workers) and 6.5% of total GDP (\$1.2 trillion), and perhaps most importantly, an annual growth rate of 5.6%,

---

35. Decker, Ryan et al. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28(3), p.7

36. US SBA Small Business GDP 1998-2014. U.S. Small Business Administration, Office of Advocacy. December 2018.



nearly 4 times the rate of the rest of the economy.<sup>37</sup>

The information industry creates jobs both directly and indirectly, as their highly-skilled, highly-paid workers drive economic activity in their own communities. The multiplier effect is estimated to be 5.73, meaning for every 100 jobs in the tech sector, 573 jobs are created in other sectors.<sup>38</sup> As technology becomes more integrated in other industries, this will enable them to collect more data and become more efficient, a factor that will interact with the tech sector multiplier (as they supply the technology). Other industries' multipliers, of course, play key roles in economic development. The table of industries and their multiplier effects is reproduced below.

**FIGURE 5**  
**Employment Multiplier by Select Sectors**

Column	Direct Jobs	Supplier Jobs	Induced Jobs	Total Indirect Jobs
Utilities	100	515	442	957
Mfg (durable)	100	289	455	744
Mfg (non-durable)	100	185	330	515
Real estate	100	397	483	878
Information	100	252	321	573
Construction	100	88	138	226
Retail	100	47	75	122
FinSrv/Insur	100	150	215	365
Education	100	64	130	194
Health Care	100	70	136	206
Hotel/Food Srv	100	54	107	161

Source: Economic Policy Institute

37. Barefoot, Kevin, et al. *Defining and Measuring the Digital Economy*. Bureau of Economic Analysis, U.S. Department of Commerce, Washington, D.C., March 15, 2018. Accessed at: <https://www.bea.gov/sites/default/files/papers/defining-and-measuring-the-digital-economy.pdf>

38. Bivens, Josh. "Updated employment multipliers for the US." *Economic Policy Institute*. January 29, 2019. Accessed at: <https://www.epi.org/publication/updated-employment-multipliers-for-the-u-s-economy/>

Some of these highly publicized Silicon Valley startups have created platforms used by most Americans every day, and entire industries now exist around using the data these platforms generate. Businesses thrive by being present on these platforms – e-commerce companies have no physical storefront, and platforms have created tools to help them navigate the platforms. Google publishes an annual *Economic Impacts* report, and its 2018 edition revealed that Google provided \$335 billion of economic activity for over 1.3 million organizations nationwide, and drove 1 billion direct connections to businesses each month.<sup>39</sup>

Facebook helps small business owners connect with their current and potential customers. Survey results showed that 73% of small business owners use Facebook to market to new customers, and 28% of businesses that export said Whatsapp helped them achieve higher sales.<sup>40</sup> 24% agreed that Facebook’s products and services drove business creation in the US.<sup>41</sup> Results also showed the importance of feedback coming through communication platforms to increase customer satisfaction and customer retention.<sup>42</sup>

This report is chiefly concerned with how data makes outcomes efficient, and this section will explore data as an input into generating economic development.

Square is a payment processing company centered around small businesses, which Mary Meeker’s 2019 *Internet Trends* reported as a successful strategy.<sup>43</sup> Square provides many products seamlessly integrated into its system for small businesses to manage their money. According to Venture Beat, “Since 2011, Square’s main customers have been businesses with annualized gross payment volume of less

---

39. Google, *Economic Impact Report*, United States 2018. Accessed at: <https://economicimpact.google.com/>

40. Facebook & U.S. Chamber of Commerce Technology Engagement Center, *How Social Media Tools Benefit Small Businesses*. 25 November 2019. Accessed at: <https://www.uschamber.com/report/how-social-media-tools-benefit-small-businesses>

41. *Ibid.*

42. *Ibid.*

43. Mary Meeker, *Internet Trends 2019*, Bond, June 2019. Accessed at: <https://www.bondcap.com/report/itr19/#view/1>

than \$125,000.”<sup>44</sup> Square’s distinct approach to community economic development can be seen in its recent report, *Making Change*.<sup>45</sup>

Square’s payment system allows small business owners (referred to as “sellers”) to accept debit, credit, online wallets, and cash. While perhaps counterintuitive for a digital payments processing company to accept cash, this insight into the need to take cash arose from the data it collected from its small business customers.

Square’s data showed that in 2015, consumers used cash in 46% of transactions under \$20, but in 2019, this proportion decreased to 37%.<sup>46</sup> While the use of cash in transactions has been decreasing steadily, and 10% of Square sellers are cashless, the *Making Change* report concludes that the US is not a cashless society.<sup>47</sup> According to Square, “83% of small business owners say they will never stop accepting cash, and 2 in 3 small business owners say their customers would react negatively if they went cashless.”<sup>48</sup>

The report also highlighted a social issue with going cashless: financial exclusion.<sup>49</sup> While this issue is discussed separately in this report in the context of access to credit, financial exclusion is a multi-faceted phenomenon that interacts with economic development. Products and services of cashless services would be inaccessible to people who are unbanked or under-banked, a significant aspect of financial exclusion. These populations were estimated to be 6.5% and 18.7% of Americans respectively, by the 2017 Federal Deposit Insurance Corporation National Survey of Unbanked and Underbanked Households.<sup>50</sup>

---

44. Ken Yeung, “Square’s success story in two words: small businesses.” *Venture Beat*. 18 October 2015. Accessed at: <https://venturebeat.com/2015/10/18/squares-success-story-in-two-words-small-businesses/>

45. Square, *Making Change: Payments, Politics, & Perspectives*, 2019. Accessed at: <https://squareup.com/us/en/making-change>

46. *Ibid.*

47. *Ibid.*

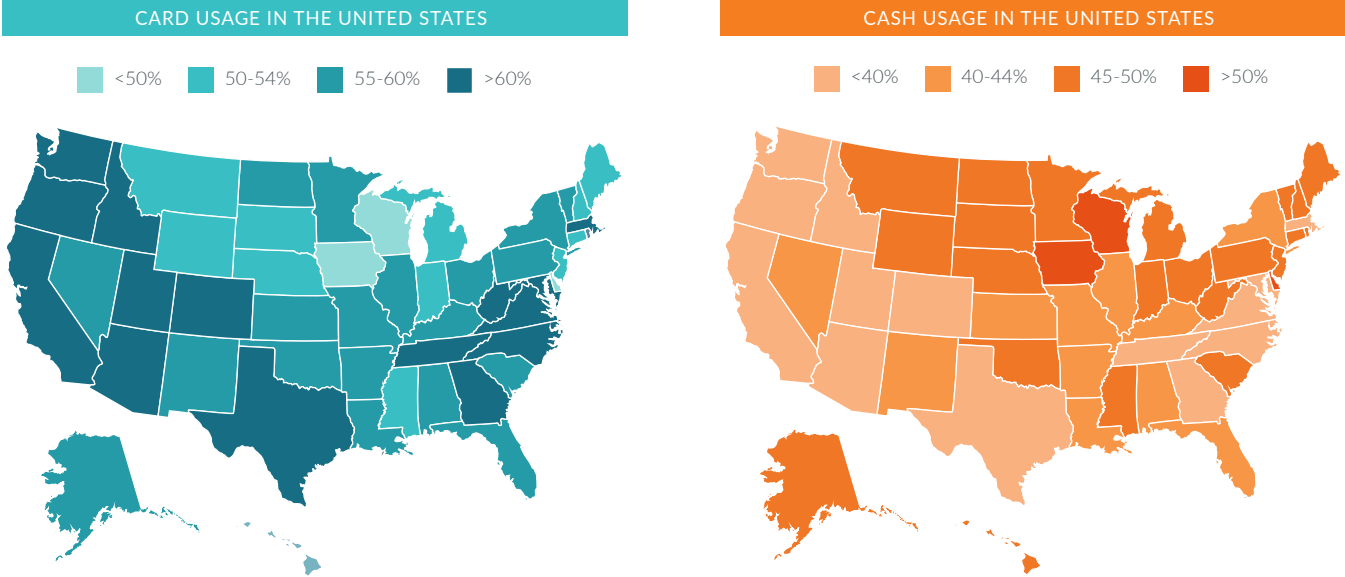
48. *Ibid.*

49. *Ibid.*

50. Federal Deposit Insurance Corporation, *2017 FDIC National Survey of Unbanked and Underbanked Households*. Washington D.C.: Federal Deposit Insurance Corporation, October 2018, accessed at: <https://www.fdic.gov/householdsurvey/2017/2017report.pdf>

These people lacking debit or credit cards would be completely locked out of a cashless society. This is particularly important when considering economic development outcomes. Some areas are considering legislation banning cashless businesses to ensure accessibility in their communities and equal participation in the economy.<sup>51</sup>

### FIGURE 6 Card vs. Cash Usage in the U.S.



Maps show the percentage of credit card transactions by state (left) and percentage of cash transactions by state (right).

Source: Square

The report featured some businesses who made the transition to cashless, some who did not, and some who did and reversed. One of the main reasons female entrepreneurs considered going cashless was time efficiency (speeding up register times, reducing time spent handling cash), but other reasons included employee safety.<sup>52</sup> As the survey results showed, many small business owners knew that feedback from consumers would be negative if they switched to being completely cashless. Other Square sellers received negative feedback and switched back to accepting cash, but some businesses implemented innovative cashless systems. For example, Mercedes Benz stadium, home of the Atlanta Falcons, installed kiosks

51. Square, *Making Change*.

52. *Ibid.*

where customers can load cash onto a prepaid debit card, which can also be used at businesses outside the stadium.<sup>53</sup> This lowered transaction times, wait times, and concession prices for fans.<sup>54</sup>

Small business owners who use Square can also use the payment data to make their companies more efficient. For example, if the details of every purchase is recorded, the most popular products or busiest hours of the week can be easily identified. Then more of those popular products can be stocked in inventory, or an extra staff member can be assigned on the busiest day. Analytics is one of Square's offerings.<sup>55</sup>

Another payment processing company using data to improve economic development outcomes is Mastercard. Its research is directed to three key areas: economic growth, economic mobility, and economic resilience. Mastercard also partners with the Centre for Inclusive Growth for their Data for Social Good program, which focuses on big data inequality.<sup>56</sup> This idea refers to the fact that the technology that collects much of the data used as inputs, such as smartphones, can still be inaccessible to the lowest-income members of our communities. If data does not get collected in the first place, it cannot be used in development.

One illustration of this phenomenon is the City of Boston's crowdsourcing of pothole data. Initially, they released an app for Bostonians to download called Street Bump that used accelerometer and GPS data to record potholes.<sup>57</sup> However, lower-income community members are likely to have fewer people with smartphones to download the app. Relying solely on Boston drivers to download the app and collect the data meant potholes in already under-served lower-income commu-

---

53. *Ibid.*

54. *Ibid.*

55. Siraj Dato, "Small businesses can get valuable insight from Square -and the company should boast about it more." *The Guardian*. 7 November 2013. Accessed at: <https://www.theguardian.com/news/datablog/2013/nov/07/square-analytics-help-small-businesses-jack-dorsey>

56. Shamina Singh, "A Call to Action on Data Philanthropy." *Mastercard*. 4 October 2016. Accessed at: <https://www.mastercardcenter.org/insights/call-action-data-philanthropy>

57. Kate Crawford, "The Hidden Biases in Big Data." *Harvard Business Review*. 1 April 2013. Accessed at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>

nities would be less likely to be detected. The city solved this problem by having city workers, who cover all parts of the city equally, also download the app.<sup>58</sup>

Another example is Hurricane Sandy and Tweeting about the destruction during the storm. *Harvard Business Review* reported on a study that found that “the greatest number of tweets about Sandy came from Manhattan,” which “creates the illusion that Manhattan was the hub of the disaster. Very few messages originated from more severely affected locations, such as Breezy Point, Coney Island and Rockaway. As extended power blackouts drained batteries and limited cellular access, even fewer tweets came from the worst hit areas.”<sup>59</sup> So, data, or more precisely, lack of data, can be deceptive.

Mastercard addresses the issue of big data inequality with many different initiatives, and provides support in three ways.

1. *Data grants* provide necessary data to organizations working to solve critical problems.
2. *Data analysis know-how* gives organizations with data the necessary expertise to make use of it.
3. *In-house data analysis* contributes actionable insights to Mastercard’s non-profit partners advancing social goals.

For example, the Center’s collaboration in the White House’s Data Driven Justice initiative involved both in-house data analysis and data analysis know-how, where Mastercard data scientists participated in the project and contributed an analysis of Mastercard’s data. It showed how crime impacted economic development in Baltimore, utilizing data on Mastercard merchant locations and local job opportunities.<sup>60</sup>

---

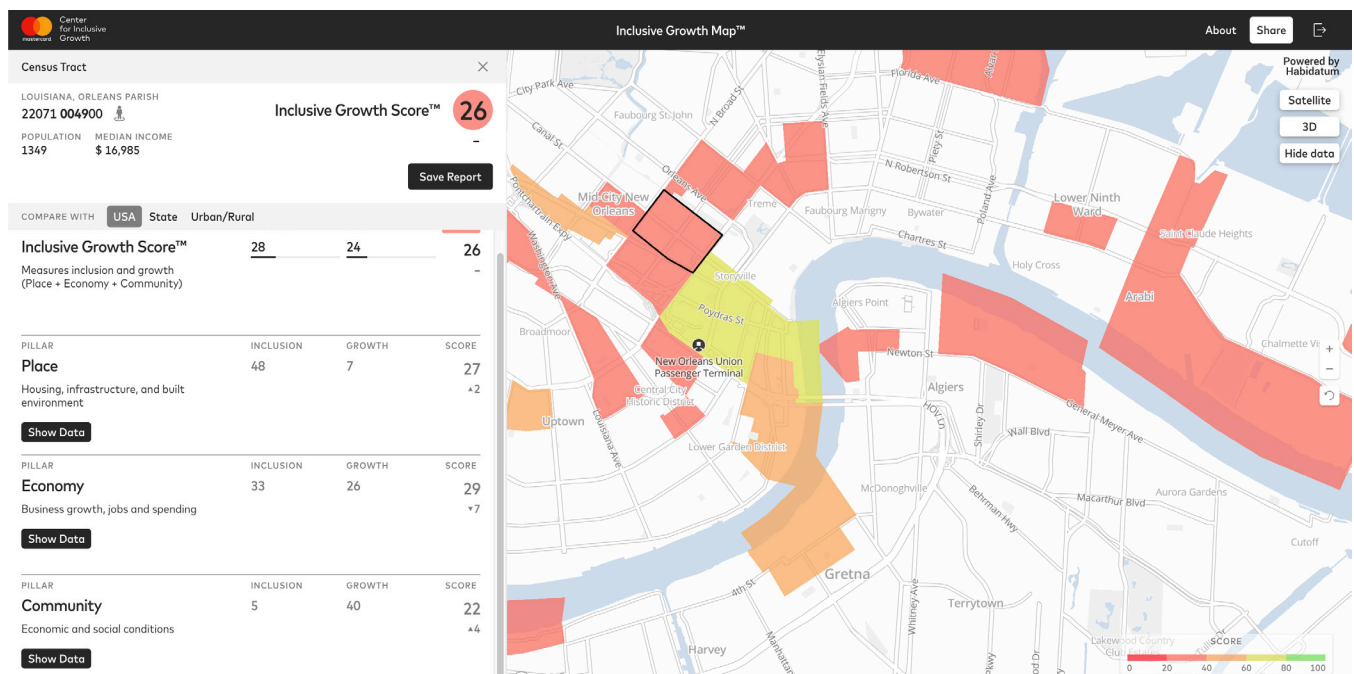
58. Lena Groeger, “When the Designer Shows up in the Design.” *ProPublica*. 4 April 2017. Accessed at: <https://www.propublica.org/article/when-the-designer-shows-up-in-the-design>

59. Crawford, “The Hidden Biases in Big Data.”

60. Singh, “A Call to Action on Data Philanthropy.”

Another Mastercard initiative involves opportunity zones,<sup>61</sup> which provides mayors and local citizens in areas ripe for revitalization with the strategic data and knowledge required to attract inclusive investment opportunities. Using historical data, trend data, and side-by-side economic performance comparisons, Mastercard helps identify products and services that reach the underserved and better fulfill their needs. Tracing cash flows can also provide a broad view of the financial ecosystem, seeing how incomes grow, how jobs are created, and which events affect the economic and social welfare of people and communities. The data can also be used to pinpoint outcome-based indicators to ensure revitalization is succeeding.

**FIGURE 7**  
**Mastercard’s Inclusive Prosperity Score**



Source: Mastercard

One opportunity zone is located in New Orleans,<sup>62</sup> which was ravaged by Hurricane Katrina in 2005. PERC conducted research on rebuilding efforts, which was

61. Mastercard Center for Inclusive Growth, “Inclusive Growth Map.” 2019. Accessed at: <https://inclusivegrowth-score.com/>

62. Mastercard Center for Inclusive Growth, “Inclusive Growth Map.”



released in a series of reports. PERC's 2010 report, *Louisiana Small Businesses Five Years Post-Katrina: Assessing LDRF Program Impacts and Measuring Existing Needs* used data and surveys to evaluate the efficacy and impact of Louisiana Disaster Recovery Fund aid given to community development organizations.<sup>63</sup>

The main findings were that aid was critical and it worked. Slightly over half, 52%, of those who applied for aid but did not receive it indicated that their business would have been much better off or would have survived.<sup>64</sup> Furthermore, every \$100 in aid resulted in an estimated \$103 of wages from net new jobs produced in a year.<sup>65</sup> These jobs themselves would have a multiplier effect in the Louisiana economy. Survey results also showed that community-based aid was critical, with close communication between the lender and borrower.<sup>66</sup> Finally, data also identified young and minority entrepreneurs to be most in need of aid.<sup>67</sup>

Unfortunately, 5 years later Louisiana still had not recovered completely from Hurricane Katrina, and rebuilding had also coincided with the 2008 Financial Crisis and the British Petrol oil spill in the Gulf of Mexico. Reduction in credit access was a major effect of the Financial Crisis.

Technology companies build platforms that simplify Americans' everyday lives and create jobs. These platforms also generate data that can be used as inputs to direct advertisements to the right consumers and connect customers with companies. They can also be used to drive economic development more generally. Payment processing systems help small business owners thrive and contribute to economic development in communities across the country. Data also allows these companies to make business decisions that work for their customers, evaluate those decisions, and commit resources to the fight social ills and inequality.

---

63. Michael Turner, Robin Varghese, & Patrick Walker, *Louisiana Small Businesses Five Years Post-Katrina: Assessing LDRF Program Impacts and Measuring Existing Needs*. Chapel Hill: Policy & Economic Research Council (PERC), March 2011, accessed at: [https://www.perc.net/wp-content/uploads/2013/09/LDRF\\_stat\\_report.pdf](https://www.perc.net/wp-content/uploads/2013/09/LDRF_stat_report.pdf)

64. *Ibid.*

65. *Ibid.*

66. *Ibid.*

67. *Ibid.*



# HEALTH CARE

## SECTION SUMMARY

*The past decade has seen the federal government and the healthcare sector invest massively in transitioning from paper to electronic health records (EHR). Despite an overall commitment of an estimated \$3 trillion dollars to date, roughly 10 percent of healthcare providers meaningfully use EHR while just under 50% are light users. Problems ranging from high transaction costs, physician burn out, and systems incommensurability plague the national EHR system. Evidence increasingly suggests that many of the problems have to do with government pressure to rapidly transition before the market had a chance to evolve and details of requirements.*

*While the process of transitioning from paper continues, a complementary set of tools is emerging to help physicians. Doctors are increasingly turning to data on Social Determinants of Health (SDOH) to improve the quality of care they deliver and improve their ability to practice medicine the way they want. Consistent with early findings on the cost and efficacy benefits of EHR, early evidence on the use of SDOH data by physicians is similarly promising. As with EHR data, there remain regulatory challenges, principally around data privacy, that must be addressed to facilitate the full potential of SDOH data.*

## THE TRANSITION FROM PAPER TO ELECTRONIC HEALTH RECORDS: PROGRESS AND PROBLEMS

Over the past decade, the US government has spent an estimated \$30 billion promoting the digitization of healthcare data, and the use of EHR by physicians and healthcare providers nationally.<sup>68</sup> This is just government support. The systemic costs of transitioning from paper are much higher, and one estimate suggests that

---

68. Dan Gorenstein. "The government may want a refund for its \$30 billion investment in electronic medical records." *MarketPlace*. 14 April 2016. Accessed at: <https://www.marketplace.org/2016/04/14/government-may-want-back-its-30-billion-investment-electronic-medical-records/>

over \$3 trillion has been allocated to the use of EHR since 2009.<sup>69</sup>

Given the promise of EHR, it is easy to understand why the federal government made a large-scale investment, and why practitioners initially fully embraced the transition from paper to electronic health records. The use of EHRs can dramatically reduce drug interaction rates, decrease mortality rates among the chronically ill, cut nurse staffing needs, and lower costs.<sup>70</sup> In fact, a recent meta-analysis of EHR efficacy studies found over 9 in 10 published reports to date had largely positive results.<sup>71</sup>

As with most massive systems changes—and make no mistake, the transition from paper to electronic health records is a sea change within the healthcare sector—progress often comes with setbacks. Two recent articles summarize the myriad problems plaguing the EHR system in the US, about which patients are generally ignorant but practicing physicians are painfully aware.<sup>72</sup>

---

69. Brian Schilling. “The Federal Government Has Put Billions into Promising Electronic Health Record Use: How is it Going?” The Commonwealth Fund. Downloaded at: <https://www.commonwealthfund.org/publications/newsletter-article/federal-government-has-put-billions-promoting-electronic-health> University of Pennsylvania Sociologist Ross Koppel estimated that hospital and physician practice spending on developing and implementing EHR technology, combined with federal subsidies, has now reached \$3 trillion.

70. *Op. Cit.*

71. M.B. Buntin, M.F. Burke, M.C. Hoaglin et al., “The Benefits of Health Information Technology: A Review of the Recent Literature Shows Predominantly Positive Results,” *Health Affairs*, March 2011 30(3):464-71. Downloaded at: <https://www.ncbi.nlm.nih.gov/pubmed/21383365>

72. The two referenced articles are: Shulte, Fred and Erika Fry. “Death By 1,000 Clicks: Where Electronic Health Records Went Wrong” *Fortune*. 18 March 2019. Downloaded at: <https://khn.org/news/death-by-a-thousand-clicks/> and, Gawande, Atul. “Why Doctors Hate Their Computers: Digitization promises to make medical care easier and more efficient. But are computers coming between doctors and their patients?” *The New Yorker*. 12 November 2018. Downloaded at: <https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers>

Among the most prominent issues identified are: (1) practicing physicians find the systems burdensome; (2) systems do not meet needs of practicing physicians but rather those of government health officials not involved in direct patient care; and perhaps most alarmingly, (3) EHRs contribute significantly to physician burnout.<sup>73</sup>



**I'M AN ENTHUSIASTIC SUPPORTER OF THIS OF TECHNOLOGY, BUT A LOT OF THE PRODUCTS NEED SERIOUS IMPROVEMENT.**

-Ross Koppel, University of Pennsylvania

University of Pennsylvania Professor of Sociology Ross Koppel, who is also senior fellow at U Penn's Leonard Davis Institute of Health Economics, noted that "in a well-intended but misguided effort to quantify quality of care, we've clogged up the EHR with all kinds of measures that irritate doctors, don't make them better at their jobs and don't produce better healthcare."<sup>74</sup>

One study from Harvard University found that poorly-designed systems of EHR require physicians to spend increasingly more time on tasks with no direct patient benefit. A recent study from the Mayo Clinic concluded the strongest predictor of physician burnout was the amount of time spent completing computer documentation. In other words, owing to the transition from paper to digital medical records, physicians have less time with patients, are spending more time on work of

---

73. Hsieh, Paul. "How Government Policies Created the Current Disaster of Electronic Health Records." *Forbes*. 31 March 2019. Downloaded at: <https://www.forbes.com/sites/paulhsieh/2019/03/31/how-government-policies-created-the-current-disaster-of-electronic-health-records/#310597d14666>

74. "Penn Sociologist Tackles Electronic Health Records, Cybersecurity, Passwords." *Penn Today*. 20 December 2016. Downloaded at: <https://penntoday.upenn.edu/news/penn-sociologist-tackles-electronic-health-records-cybersecurity-and-passwords>

seemingly no direct benefit to patients, and are at greater risk of burning out. Consequently, any tools that could aid physicians in their efforts to provide better quality patient care, and be more effective with the time they do have with patients, is likely to be embraced by a burdened physician population.

Again, these types of responses to change are perfectly natural. There is no doubt that when other industries and sectors such as lending shifted to computer systems, electronic records, and data-driven processes there was also some backlash and missteps taken. And in these sectors that have embraced data-driven processes and are now reaping the rewards, the idea of turning back the clock would be unthinkable.

Further, more and more patients now have easy access to their health records, test results, appointment schedules, and the like via online portals. In many more cases patient data can easily migrate from system to system so a patient visiting a new physician or a different practice can have their health records reviewed.

The real promise of EHRs goes *well* beyond easy access to health records. Consider having data on patient vital statistics, histories, medical interventions, and health outcomes. Relationships between factors, symptoms, and treatments on the one hand can be analyzed with outcomes to see who is more likely at risk for certain health problems and which treatments or combinations of treatments are most successful in treating the patient. Researchers will have a new resource to improve medicine. To the extent that records also capture costs or can be linked to costs, administrators may be in a better position to identify where expended resources are utilized more wisely and where patient outcomes are better. That data has the potential to both improve and rationalize American healthcare.

Combining these data with genetic information and data from wearable devices such as Fitbit or Apple Watch could also aid medical researchers significantly.

However, we are still years away from realizing the potential harnessing all of this information.

Other data already collected, typically, initially, for other purposes, can also aid health-care. Some of this data can be used to describe what is called social determinants of health.

## FILLING A GAP: SDOH DATA AS A COMPLEMENT TO EHR INFORMATION

What are social determinants of health and why are they relevant to physicians who directly interact with patients? SDOH is a term that is widely understood in health-care circles, and is the subject of great attention among multilateral organizations focused on global healthcare issues.

While there is no universally accepted definition of SDOH, the World Health Organization (WHO) and the Centers for Disease Control (CDC) each utilize a similar one. Namely, conditions in the places where people live, learn, work, and play affect a wide range of health risks and outcomes. These conditions are known as social determinants of health.<sup>75</sup> Put differently, SDOH information includes data about all the things that affect a person's health that come OUTSIDE all the things one would be asked by a physician during a check up with a primary care physician or visit with a specialist.

For example, a patient visits their primary care physician due to a problem—their knee hurts. The physician would normally ask questions such as “What are you doing? Any new exercise or unusual behavior that may have affected your knee?” The physician would not ordinarily be asking questions outside of the context of the patient's knee issue. However, maybe the patient lives in area with an above average crime rate so it is not safe to exercise outside. If so, this would be helpful information for the physician to know. Where you live, socio-economic factors, environmental factors such as social isolation—all of which can be gleaned from public records and proprietary sources may provide information to a physician or other care provider that would be helpful for direct patient interactions.

There exist an enormous number of health factors that a patient does not necessarily mention when they visit a physician. Are you socially isolated? Do you have rides to get here to see me and/or other physicians? If no, there may be options to help

---

75. The CDC maintains an excellent archive of scientific studies examining the relationship between specific SDOH and health outcomes, such as between the financial resources people have and their health, or the contexts in which people live, learn, work and play and their health. For access to this resource, please see: <https://www.cdc.gov/socialdeterminants/research/index.htm>

you. These more intimate doctor/patient discussions can be started non-judgmentally in order to administer help. If you have financial stress, your physician may be able to provide you with access to assistance programs. Access to tools that utilize SDOH data could enable physicians to better identify patient risk factors and gain a more holistic view of the patient.<sup>76</sup>

In terms of data used in this space, the same core data for other non-health applications—including credit risk decision-making, individual consumer reports for tenant and employment screening, and fraud prevention/anti-money laundering as well as public safety is used. Using SDOH data to improve the quality of healthcare is a newer application. It has been discussed and researched for decades—especially by multilaterals in the context of emerging markets—but has not made its way into mainstream until recently. Discussions are now heating up about how to acquire and apply the data.

LexisNexis® Risk Solutions and others have been analyzing this data and have published some intriguing findings in this area. For instance, the Kaiser Family Foundation found general relationships pointing to the potential usefulness of healthcare integrating SDOH inpatient care, such as the relationship between educational achievement, income, likelihood of smoking and life expectancy.<sup>77</sup> And social factors contribute or account for 1 in 3 deaths, a positive relationship between parental education and growing up in an environment that poses a risk to health, and a negative relationship between income and premature death.

Where a LexisNexis® Risk Solutions white paper suggests that SDOH could play a particularly useful role is in helping determine the optimal care management plan that, for instances, reduces hospital readmissions.<sup>78</sup>

---

76. LexisNexis Risk Solutions. “How data on social determinants of health is revolutionizing healthcare.” <https://risk.lexisnexis.com/-/media/files/healthcare/white-paper/how%20data%20on%20social%20determinants%20of%20health%20is%20revolutionizing%20healthcare%20pov%20pdf.pdf>

77. Kaiser Family Foundation. Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity. Nov 4, 2015.

78. LexisNexis Risk Solutions. “The Population Health Data You Need Right Now.” <https://risk.lexisnexis.com/-/media/files/healthcare/white-paper/population-health-data-you-need-right-now%20pdf.pdf>

LexisNexis® Risk Solutions currently provides Socioeconomic Health Scores using SDOH data.<sup>79</sup> These scores include a Readmission Risk Score, a Medication Adherence Score, and a Motivation Score (which predicts how motivated an individual will be to engage in their own care). These scores are built using SDOH data elements and data on actual patient outcomes. In 2018 LexisNexis® Risk Solutions reported that “Research conducted by LexisNexis indicated that the Readmission Risk Score is on par with—or better at predicting risk than—readmission scoring models that use only clinical data, which is often difficult to obtain.”<sup>80</sup>

These new SDOH tools will no doubt continue to develop, and as they do, the quality of patient care should increase while systemic costs decrease. Doctors will be better empowered to meaningfully engage with patients in a non-judgmental, sensitive context in order to customize care to the patient’s holistic needs.

As data used for promoting public safety through improved law enforcement applications, the best solutions don’t typically use one type of data versus another but instead integrate many types of relevant and useful data. This is the case with health and healthcare data tools as well.

As EHR and other patient data becomes more available (such as data from wearable devices), the integration of these datasets with SDOH data will likely produce powerful new tools to improve patient care, medical outcomes and medical research.

---

79. See the following LexisNexis Risk Solutions website: <https://risk.lexisnexis.com/products/socioeconomic-health-score>

80. LexisNexis Risk Solutions. “LexisNexis Assists Hospitals in Predicting Readmissions with Non-clinical Data.” Press Release. 3/01/2018. <https://risk.lexisnexis.com/about-us/press-room/press-release/20180228-readmissions>

# ARTIFICIAL INTELLIGENCE FOR GOOD

## SECTION SUMMARY

*Artificial Intelligence is being used to combat human trafficking and also connect people to needed mental health services. It is also dependent on the open channels of data to improve the lives of disabled citizens and help unserved communities obtain access to credit.*

## GOOGLE'S AI TOOLS

Google's AI for Social Good program has supported a number of artificial intelligence uses that concern public safety. For example, Google collaborated with the organization Thorn, a non-profit aimed at eradicating child trafficking and the spread of child sexual abuse material. With a \$2.5 million grant and a team of Google engineers, they developed the machine learning tool Spotlight. It looks at over 150,000 escort ads posted online every day for data about children bought or sold for sex, and that data is combined with law enforcement data to locate victims and dismantle trafficking networks.<sup>81</sup> Three-quarters of those who were victims say that at some point they were advertised online, but up to now sifting through the daily deluge of posting to find information about child victims would have been very challenging and resource intensive. In other words, law enforcement lacked the capacity to take advantage of this critical data asset in efforts to thwart child trafficking and aid victims.

---

81. Thorn, *Spotlight*. Accessed at: <https://www.thorn.org/spotlight/>



Spotlight delivered compelling results. Specifically, 83% of cases reported by users of Spotlight identified a victim (over 9000 victims in total), and over 10,000 traffickers have been identified.<sup>82</sup> In addition to finding bad guys, Spotlight also saves law enforcement resources, as investigation time is reduced by 67%.<sup>83</sup> The immediate benefits of Spotlight have made it invaluable. Today, law enforcement officials routinely refer to Spotlight as their go-to tool for child trafficking cases.<sup>84</sup>

Google has also supported several initiatives in natural language processing. For example, Crisis Text Line is a non-profit organization that connects people in crisis to trained counselors through text messages. It receives more texts than it can handle at once, and Google's contribution prioritized and assigned texts that came from people most at risk of doing harm to themselves for crisis counselors to respond to first, and overall reduced wait times. This has saved lives and improved the novel non-profit organization's core deliverable—real-time access to crisis counseling.

To accomplish this, Google's software utilized 65 million messages and those deemed to be at the highest risk for suicide were tagged as such in a survey taken post-conversation.<sup>85</sup> From this, the software was able to identify 86% of those at very high risk for suicide from their initial text conversations. Using Spotlight, Crisis Text Line is now able to service 94% of the highest risk texters in under 5 minutes. Crisis Text Line also makes use of other data-driven insights. With the sheer amount of data just generated by text messages, Crisis Text Line has for example, found that positive responses are associated with first-person statements (such as "I'm worried about how upset you seem").<sup>86</sup> Another organization Google has supported is The Trevor Project, which aims to prevent suicide in LBGQT youth by supporting them through phone, text, and chat.<sup>87</sup>

---

82. *Ibid.*

83. "Spotlight." Accessed at: <https://spotlight.thorn.org/about>

84. Thorn, *Spotlight*. Accessed at: <https://www.thorn.org/spotlight/>

85. Ankit Gupta. "Detecting Crisis: An AI Solution." Crisis Text Line. March 28, 2019. Accessed at: <https://www.crisistextline.org/blog/ava2>

86. Alice Gregory. "R U There?" *The New Yorker*. 2 February 2015. Accessed at: <https://www.newyorker.com/magazine/2015/02/09/r-u>

87. "The Trevor Project is a Google AI Impact Grantee." *The Trevor Project*. 7 May 2019. Accessed at: [https://www.thetrevorproject.org/trvr\\_press/the-trevor-project-is-a-google-ai-impact-grantee/](https://www.thetrevorproject.org/trvr_press/the-trevor-project-is-a-google-ai-impact-grantee/)

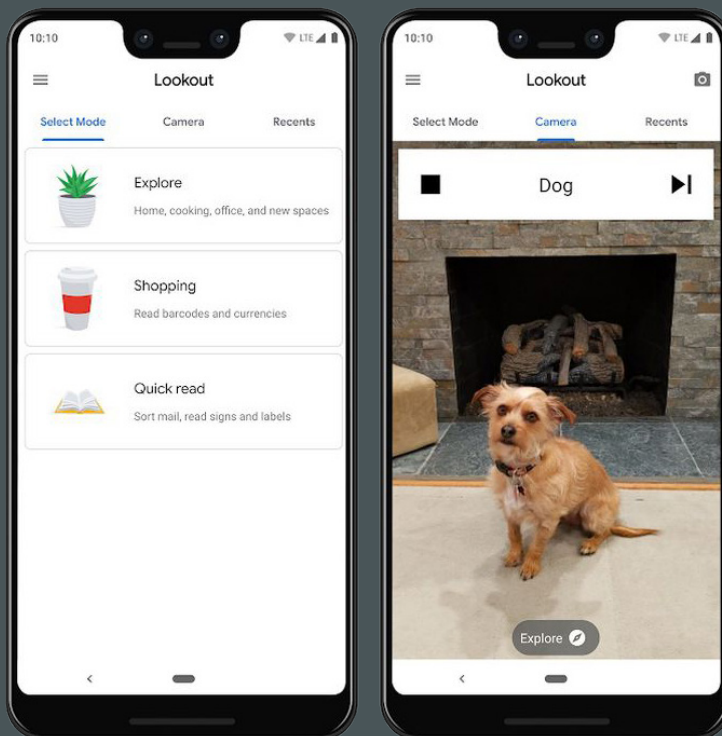
## EXAMPLE

# AI FOR GOOD

## EXPERIAN LIFT

Over 40 million Americans are credit invisible. Experian recently released a new series of products to help more Americans access affordable credit. On the borrower side, Experian Boost is a consumer-facing tool that allows consumers to opt into self-reporting positive utility, telecom, and rent payments to their Experian credit profile. Consumer-permissioned solutions are relatively new to the market, enabled by new technology allowing access and verification of this data with consumer authorization. Experian Lift is the lender side, which allows lenders to score consumers with thin credit files or even no credit files at all, who would have otherwise been excluded from the credit system. The score uses the latest technology in machine learning, combining traditional data, alternative data assets exclusive to Experian, and trended data to build a more comprehensive consumer profile. Compared to other scores used to underwrite traditionally under-served populations, Experian Lift enhances the model's predictive performance by 23%.

Source: Experian



## GOOGLE LOOKOUT

As part of Google's AI for Social Good program, Google has made efforts to make its services more accessible. Its Lookout app, which is an artificial intelligence tool to help the visually impaired. Using cutting-edge computer vision technology, the app recognizes what it sees through the phone camera, and reads out identified objects in the user's surroundings. The app also scans barcodes, reads text, and identifies currencies. Nearly 253 million people worldwide are blind or visually impaired.

Sources: Materials provided by Google, interviews conducted on December 5, 2019.

# CONCLUSION

---

The examples and cases presented in this third installment of our “U.S. Data Ecosystem” series underscore the observations and findings from our first installments. Namely, it highlights that technology and data flows are crucial to the performance of many sectors outside of tech. Future improvements in these sectors will rely heavily on data-driven solutions being developed and improved now or expected to be developed in the near future. The cases and examples from law enforcement and public safety, lending, community economic development, and health care covered in this report clearly show this. Improvements, better outcomes, cost reductions enabled by technology and data flows in sectors such as healthcare or those not covered in this report, such as education, already have yielded enormous positive public policy, economic, and quality of life benefits, and have profound implications for the future of these sectors and society as a whole.

Another observation that follows from the cases and examples in this report is that data flows and data-driven processes are already deeply intertwined and crucial in a multitude of sectors, and this appears to be a continuing trend. Data and data tools useful in a sector today may not have been thought of as crucial or even contemplated decades ago.

These cases make it clear why it is so important that policymakers take great care when crafting needed national privacy legislation. Legislation will have the potential to impact much more than a few industries or uses that might be highlighted in the media or well-known to the public. It will have the potential to impact nearly all sectors of the economy and in ways that aren't readily intuitive.

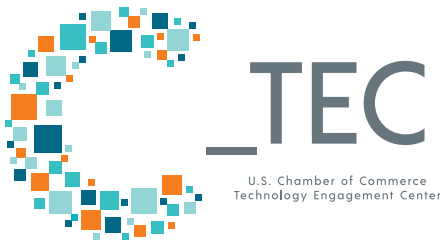
The presented examples of the new ways online fraud is battled by *harnessing* data provides a compelling case of how new and improved fraud defenses may be the reason fraud costs have been declining. This decline in fraud cost was shown in the second part of our U.S. Data Ecosystem series, and occurred despite ever-greater data flows and a higher frequency of data breaches.

This example also makes clear the potential for perverse, unintended consequences from ill-crafted privacy legislation. For example, legislation that *unduly* limits data flows in order to protect consumers from data breaches and the perceived resulting ID theft and fraud, may also hobble key ways consumers are protected from fraud. This underscores why it is necessary for legislation to be evidence-based.

Just as it would have been difficult 30 to 40 years ago to predict how various sectors and industries would utilize data today, it is very difficult to predict how data will be used across sectors in the future, much less the details of how technology will change. To address this reality, national privacy legislation must be flexible and principles-based to enable needed adjustments.

## ABOUT C\_TEC

The U.S. Chamber of Commerce is the world's largest business federation representing the interests of more than three million businesses of all sizes, sectors, and regions. Four years ago, the U.S. Chamber of Commerce launched the Chamber Technology Engagement Center (C\_TEC) to advance technology's role in strengthening business by leveraging tech innovations that drive economic growth in the United States. C\_TEC promotes policies that foster innovation and creativity and sponsors research to inform policymakers and the public.



## ABOUT PERC

PERC is a non-profit (501c3), non-partisan research and development organization headquartered in Durham, NC. Founded in 2002, PERC has undertaken projects in over 25 countries on 6 continents, and has contributed to national policy changes in over 10 countries. PERC's mission is to increase financial inclusion through the responsible use of information and information solutions. Our constituency includes the 45 million Credit Invisibles in the US and the billions worldwide.





**Dr. Michael A. Turner, Ph.D.**

FOUNDER & CEO

*Policy & Economic Research Council*

Michael Turner, Ph.D. is the founder, President and CEO of PERC. He is a prominent expert on credit access, credit reporting and scoring, information policy, and economic development. He has advised two Presidential administrations, testified before Congress and numerous state legislatures, and has partnered with the Bureau of Consumer Financial Protection, the Department of Housing and Urban Development, and the White House (Council of Economic Advisors/National Economic Council).

Dr. Turner has worked on projects in over 25 countries on six continents, and is a global thought leader on consumer and commercial credit information sharing. He has worked on projects for the OECD, the Inter-American Development Bank (IDB), the International Finance Corporation, the Consultative Group to Assist the Poor, and US AID among others. He also launched the Asia-Pacific Credit Coalition and has collaborated with the Asia Pacific Economic Cooperation (APEC) Business Advisory Council to establish the Asia Pacific Financial Forum (APFF) to promote financial inclusion among the 21 member economies of the APEC. He is a core group member of the APFF and serves as the credit information sharing “Sherpa.”

Prior, Dr. Turner served as Graduate Fellow at the Columbia Institute of Tele-Information at the Columbia Business School in New York City, Executive Director of the Information Services Executive Council, manager of government affairs for the North American Telecommunications Association and staff assistant in the U.S. Senate.

Dr. Turner received his Ph.D. from Columbia University in International Political Economy and his B.A. from Miami University in Economics. He was awarded a Yeck Fellowship from Harvard Business School, and was awarded the Ashoka Foundation lifetime fellowship in 2009.



## **Patrick D. Walker, M.A.**

**DIRECTOR OF RESEARCH**

*Policy & Economic Research Council*

Walker serves as Director of Research at PERC. Walker's concentration is econometrics and statistical methods. Walker has built commercial and research grade credit scoring models, designed complex multi-country longitudinal analysis on issues involving credit information sharing, micro-finance, payment systems, and consumer and commercial credit access. He has carried out ground-breaking work on the consumer impacts of alternative data. Working with a team of economists from the World Bank and the Brookings Institution, Walker pioneered an effort to gauge the financial vulnerability of a geography to natural disasters, and created a dashboard for recovery. Walker received his M.A. in economics from Duke University. He has taught both undergraduate microeconomics and econometrics while in Duke's economics Ph.D. program (ABD).



## **Kazumi C. Moore**

**ECONOMIC POLICY RESEARCH MANAGER**

*Policy & Economic Research Council*

Moore recently joined PERC Canada. She develops and maintains PERC's information resources on everything from theoretical economic models on credit rationing to government reports on data policy. She contributes quantitative and qualitative data and analysis to PERC reports. She studied International Business and International Relations at McGill University in Canada and Kyoto University in Japan. Moore has past experience in journalism and market research.



