DATA FLOWS, TECHNOLOGY, & THE NEED FOR

NATIONAL PRIVACY LEGISLATION

_TEC

U.S. Chamber of Commerce
Technology Engagement Center

PERC

POLITICAL & ECONOMIC RESEARCH COUNCIL

JULY 2019

TABLE OF
CONTENTS

EXECUTIVE SUMMARY

The 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. All 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. It is time for Congress to act and pass federally national privacy legislation.

The need to act sooner rather than later is underscored by the fact that state legislatures have already passed various forms of new privacy laws. The proliferation of uncoordinated state laws will result in undue business uncertainty—especially within the tech sector—and raise the specter of a patchwork of data privacy laws that unnecessarily impedes data flows, erodes American competitiveness, and harms overall economic performance. Privacy laws that unduly constrain data flows and target the tech sector are likely to cause predictable economic harms with consequences that could damage the tech sector, other sectors reliant on information technologies, and consumers.

WHY IS IT CRITICAL FOR CONGRESS TO PASS EVIDENCE-BASED, PROACTIVE AND NATIONAL LEGISLATION?

Tech Sector Major Contributor to US Economic Vitality

While relatively small by conventional metrics—employment, share of GDP—the tech sector is vital to the current and future competitiveness of the overall American economy. The tech sector has a vast domestic supply chain creating many jobs directly and indirectly. Moreover, the high-paying opportunities attract highly-skilled workers who contribute to the strong competitive position of many US industries.

Access to Data Critical to Competitiveness of American Economy

The US tech sector, and a growing share of industries leveraging data and IT, owe much of their competitiveness to the current structure of data regulation in the US—harms-based and done on the sector level (e.g. health care, financial). US firms are able to engage in sophisticated data analytics—the importance of which will only increase with broader applications of artificial intelligence (AI) and machine learning (ML). Put differently, the ability of US firms to innovate by using data analytics is a competitive advantage. Consequently, dramatic changes to the established network of laws and regulations governing the collection and use of personal information risks damaging the American economy and undermining a competitive advantage if not done right.

WHAT IS THE RELATIONSHIP BETWEEN DATA BREACHES AND DATA PRIVACY LEGISLATION?

On the surface, a major driver of privacy legislation in the states is large data breaches. Predictably, the ever-increasing collection, use, and transmission of sensitive and other personal data as part of the IT and information economy revolution has led to an increase in the frequency of data breaches and the volume of records breached. Over the past 15 years, once corporations were required to report breaches, mainstream media interest in and attention to data breaches spiked from a blip to a hailstorm. In part, the manner in which breaches have been covered—largely focused on major breachesand breaches with stories about victims—has resulted in a fear reaction concerning data sharing and data breaches. In turn, lawmakers in a growing number of states have passed data privacy and/or data security legislation that may be well-intended, but the uncoordinated state-by-state efforts risk causing serious economic harms for potentially little actual consumer protections.

One of the central themes of this report is that national data privacy legislation is needed, and that it must be sound, evidence-based, and proactive rather than reactive and hastily crafted. A primary contribution of this study is to explore the relationship between data breaches and consumer impacts. We hope this will help debunk widely-held misperceptions about consumer impacts from data breaches. According to the general narrative, following a breach, the risk of harm to individuals is high, immediate, and persists over time—perhaps forever. In fact, as this report demonstrates, there is no empirical evidence to support these beliefs. This study will argue that privacy legislation should take this fact into account, especially when crafting data security measures and data breach penalties into privacy laws.

In particular, privacy legislation that includes data security provisions must not feature draconian enforcement mechanisms such as automatic penalties and fines on breached entities without taking into account negligence and consumer harm. Such measures do nothing to further incentivize firms to improve data security measures, nor do they afford consumers any additional protections from data breaches. Instead, and far more likely, automatic and/or excessive fines and penalties will drive SMEs out of the data business, will reduce investments in innovation, and will dampen the overall competitiveness of the US tech sector and all other sectors reliant upon information services technologies.

Americans and groups across the political spectrum are calling for national privacy legislation. Lawmakers should respond by crafting well-designed and proportionate laws aimed at aiding and protecting consumers while at the same time preserving the nation's robust information and tech sector. Both can and must be achieved.

KEY FINDINGS

NATIONAL PRIVACY LEGISLATION

Federal privacy legislation is needed to prevent a confusing patchwork of state privacy laws. As the EU's development of regulations shows, fragmented data regulations in a market neither protects consumers uniformly nor helps companies compete. The need for informed, evidence-based, well thought-out policymaking is crucial in the data and technology space, and this is the impetus for our Data Ecosystem series.

WEAK RELATIONSHIP BETWEEN DATA BREACHES AND ID THEFT

Using data between 2005 and 2018, this study found the incidences of data breaches and the volume of records breached are unable to meaningfully predict incidences of identity theft. Data breaches have been increasing over time, while the rate of identity theft has stayed relatively constant, fluctuating between 4.35% and 6.63%.

WHILE DATA BREACHES ROSE, FRAUD LOSSES FELL

From a peak of over $35 billion in 2005, fraud losses steadily declined to just under $15 billion in 2018. During this time, reported breaches and the volume of stolen files climbed dramatically. This suggests that data breaches are not meaningfully driving fraud losses associated with ID theft/fraud, and that many other factors are key to determining fraud losses, including security and defences against fraud (which also use data).

NO NEGATIVE IMPACTS IN LARGE DATA BREACH POPULATION

This study examined 27 million subscribers to credit bureau credit monitoring, and their credit scores, credit file locks, and the presence of their personal identifying information (PII) on the Dark Web. The sample comprised nearly 5 million data breach victims, 8 million persons who subscribed to credit monitoring directly with the credit bureau, and another 14 million who subscribed via partner organizations of the credit bureau. The Breach sample witnessed an average credit score rise, not a fall, over approximately a 12 month period and had a slightly lower rate of PII detected on the Dark Web. In addition, the breach-affected population did not have higher rates of credit monitoring activity alerts or credit file locks. The report found no evidence of overall credit-related harms to those in the data breach sample.

WEAK LINK BETWEEN SPECIFIC DATA BREACHES AND ID THEFT

Extending the Government Accountability Office's (GAO's) 2007 analysis, this study examined an additional 24 notable reported data breaches in the US over the past 15 years and found the highest observed compensation claims rate linked to a data breach to be 2.5%. As way of comparison, this is nowhere close to the lowest rate of "natural" identity theft observed in the general population in 2010, which was 4.35%, and was far less than half the highest observed rate, 6.63% in 2017.

PREPONDERANCE OF EVIDENCE

The research design used in this study involved three levels of analysis, very large samples of data from credible sources drawn from the past 14 years in some cases, and contained a rich set of variables to empirically test widely held beliefs about the risks associated with data breaches. In every possible instance, research results from this study were compared to findings from other credible sources, including US government agencies such as the Federal Reserve Bank, the GAO, and the Federal Trade Commission (FTC); and in each case the C_TEC/PERC results were broadly consistent with the findings of other major studies. Taken together, the results from this research, as corroborated by external third-party research, present a compelling case that data breaches do not significantly contribute to ID theft and fraud losses, and that consumers affected by data breaches are not broadly harmed as is widely believed.

PART I

THE PRIVACY DEBATE

OVERVIEW

THE US DATA ECOSYSTEM SERIES

In the midst of a growing chorus of criticism aimed at the US information, communications, and technology sector (commonly referred to as "tech") is an emerging public policy debate around federal privacy and data protection legislation. States are also moving aggressively, risking broad economic damages. California and Vermont have already passed state data protection laws with laudable features, but that also contain so-called "company-killer" provisions deeply concerning to the business community.

In Part I of our Data Ecosystem series, we explore the need for a well-crafted national data protection law that achieves consumer protections while taking into account legitimate business needs. PERC cautions against reactive legislation for three interwoven reasons:

1. the importance of the tech sector;

2. the role of US federal data regulations as a competitive advantage for American firms globally; and,

3. the law of unintended consequences.

In Part II of this paper, this study contributes to evidence-based, informed policymaking, and analyzes the phenomenon of data breaches, perhaps the single most significant driver of recent privacy legislation in states. In this report, the relationship between data breaches and consumer impacts is examined, including the incidence of identity theft/fraud, losses associated with ID theft/fraud, data found on the dark web, and various other metrics indicative of potential consumer harms directly linked to data breaches.

In a forthcoming third part to our U.S. Data Ecosystem series, PERC and the U.S. Chamber Technology Engagement Center (C_TEC) will examine some lesser-known but compelling narratives around socially beneficial applications of information services. This will feature at least one examination of how small businesses use data. This will be our "Data for Good" series.

ONE

TECH IN THE SPOTLIGHT: VARIOUS CONCERNS ABOUT TECH DRIVING DATA PRIVACY LEGISLATION

Summary: While the tech industry is facing policy challenges on multiple fronts, the most serious ones center around data privacy and data security: the collection, storage, use, re-use and sale of data, especially data linked to personal identifying information. Principal among the data-centered policy debates is the prospect of a disparate patchwork of reactive and hastily-crafted state privacy laws, and emerging national debate on the need for a new comprehensive federal privacy law. While the logic for a federal approach is compelling, it is also fraught with risk if not done right.

Data privacy is again at the forefront in national and state policy discussions. In the wake of a series of high-profile data breaches—culminating most recently with the Equifax breach—and a string of notable privacy incidents connected to major social media platforms, a growing number of voices are calling for a new privacy law.[1] Pressure on lawmakers to act has been mounting for some time. Some pressure is coming from industry, some from consumer advocates, and some no doubt emanates from a more general backlash against technology and tech firms.[2]

Greater policymaker and regulatory scrutiny of the tech industry can be seen around the world. In the EU, large tech companies are under investigation for privacy violations and/or for antitrust concerns. Closer to home, state and federal privacy legislation is ascending to a short list of policy priorities.[3] In 2018, privacy laws were passed in California and Vermont in response to the growing public concern. Among other controversial provisions, these laws include potential statutory penalties for breaches ranging from $100 to $750 per subject in California.[4]

1. See Wu, Tim. "An American Alternative to Europe's Privacy Law." *New York Times*. May 30,2018. Accessed at: https://www.nytimes.com/2018/05/30/opinion/europe-america-privacy-gdpr.html; see also The Economist, "America Should Borrow from Europe's Data-Privacy Law." April 5, 2018. Accessed at: https://www.economist.com/leaders/2018/04/05/america-should-borrow-from-europes-data-privacy-law; see also Romm, Tony. "Democrats Vow Congress Will Assert Itself Against Tech – Starting With Silicon Valley's Privacy Practices." *Washington Post*, February 26, 2019. Accessed at: https://www.washingtonpost.com/technology/2019/02/26/democrats-vow-congress-will-assert-itself-against-tech-starting-with-silicon-valleys-privacy-practices/.

2. The backlash could result from many causes, many of which may be unrelated. For instance, displaced laborers who see automation as the enemy, white-collar workers who view technology as a ball-and-chain constantly tethering them to their work, parents who fear they are losing their children to videogames and smartphones, mental health advocates who see the growth of videogame addiction especially among America's youth, and economists who point to growing concentration across a range of industries, including the tech sector, and see potential for consumer harms.

3. Scott, Mark. "In 2019 Techlash Will Go from Strength to Strength." *Politico*. December 30, 2018. Accessed at: https://www.politico.eu/article/tech-predictions-2019-facebook-techclash-europe-united-states-data-misinformation-fake-news/

4. California Civil Code § 1798.155; The maximum fines are levied whenever a violation is found to be intentional or reckless and not cured within 30 days.

There is potential for such laws to be company killers—if the California Consumer Privacy Act (CCPA) were in place prior to the Equifax data breach, for instance, the breached firm could have been fined several billion dollars in California alone. With 143 million breached data subjects, at $750 per subject, Equifax would have been snuffed out had all 50 states been practicing the same law. A real fear is that if the CCPA does not have its private right of action provisions amended, it will also result in result in a wave of class action lawsuits.[5]

Extrapolating this to the broader economy suggests that entire sectors—indeed the entire American economy—could be severely harmed by a strict application of these new state-level privacy laws should a patchwork emerge. While we seriously doubt that state lawmakers intend to extinguish firms and up-end their own state economies owing to data breaches, even irresponsible ones, it creates this risk. And by involving the courts with a private right of action, such new state privacy laws could deter investments and start up funding in data-based innovation. Given the nature of the information economy, a new app or internet service startup could have millions of users but little revenue. The millions of users could suddenly appear as a massive legal liability by investors and act as a deterrent to funding. This would be the case if investors held a negative view or were simply uncertain how courts and juries would act. Given the critical role data analytics plays in innovation (which is only increasing with broader applications of AI and machine learning), unduly burdensome regulation could erode the relative competitiveness of a range of US firms, while possibly presenting an existential threat to data-intensive firms and entire sectors of the US economy.

Certainly, however, consumer protections must not be sacrificed before the altar of product innovation. Individual privacy and individuals having more control over and understanding of what data is collected on them, how it is shared, and how it is used is paramount. Decades ago, when large datasets were beginning to be put together, data protections, consumer protections, permissible purposes, and the like were put into place for specific activities, such as for use in credit origination (FCRA 1970) or data collected by the federal government (Privacy Act of 1974). Now with the exponential growth in data (its collection, creation and use) and data collected from all manner of software applications, electronic devices, cameras, and other objects (IoT), it is understandable that new, more comprehensive data protections are needed.

At the same time, imposing data restrictions that unduly erode the competitiveness of American enterprise and contain unnecessarily onerous penalties or regulations that do not demonstrably benefit consumers would be irrational. Rather, as 50 years past experience in the US shows, consumer protections and legitimate business use can be balanced within an effective, reasonable, and well-crafted governance framework. As the next section addresses, failing to get this right carries considerable economic risk given the extant and growing significance of tech and data analytics to the American economy.

5. Ballon, Ian, & Rebekah Guyon. "Anticipating the Flood of Cybersecurity Litigation Under the CCPA—What to Do About It." *Yahoo Finance*. January 25, 2019. Accessed at: https://finance.yahoo.com/news/anticipating-flood-cybersecurity-litigation-under-090055586.html

TWO

PUNCHING ABOVE ITS WEIGHT: TECH SECTOR CONTRIBUTIONS TO AMERICAN ECONOMY

Summary: While relatively small by conventional metrics—employment, share of GDP—the tech sector is vital to the current and future competitiveness of the overall American economy. The tech sector has a vast domestic supply chain creating many jobs directly and indirectly. Moreover, the high-paying opportunities attract highly-skilled workers who contribute to the strong competitive position of many US industries. The US tech sector, and a growing share of industries leveraging data and IT, owe much of their competitiveness to the current structure of data regulation in the US—which is done on the sector level (e.g. health care, financial) and is harms-based. Consequently, dramatic changes to the established network of laws and regulations governing the collection and use of personal information risks damaging the US economy and undermining the tech industry's competitive advantage if not done right.

It is inarguable that information technology is the industry of the current era. It attracts the brightest and best from across the globe, and has had dramatic effects on nearly every aspect of our daily lives. Tech has transformed the way we communicate with one another, the way we work and where we work, and has opened opportunities for vast populations to receive an education remotely, to access a full range of financial services even from remote locations, to live at home rather than an elder care facility, to start a small business and access a global market, and to monitor one's physical and mental well-being minute-by-minute and have emergency services or home healthcare workers at your door in the event of a crisis.

While a more detailed exposition on the widespread economic and social benefits of the tech sector is both useful and possible—PERC will showcase several of the more interesting beneficial applications in a series of white papers later this year—for now our focus is more limited, with a particular emphasis on just the tech sector.

The tech sector is vital by nearly any conventional metric. The tech sector:

- is healthy and growing, accounting for 6.5% of total US GDP ($1.2 trillion annual revenue);[6]
- directly employs 5.9 million workers, accounting for nearly 4% of total US employment;[7]

6. 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/system/files/papers/WP2018-4.pdf

7. *Op. Cit.*

- indirectly employs many more people—the tech sector multiplier is estimated to be 5.73. For each tech sector worker, 5.7 other jobs are created in other sectors, totaling 33.8 million jobs supported indirectly (another way to say this is that for every tech sector job lost, another 5.7 jobs will vanish);[8]

- energizes the economy—the digital economy is growing at an annual rate of 5.6%, nearly 4 times the rate of the rest of the economy.[9]

Measuring the importance of the tech sector, defined as "Information, Communications, and Technology" or ICT by the Bureau of Economic Analysis (BEA) on the broader economy is complicated.[10] Using simple metrics, the tech sector could be presented as non-trivial but hardly a juggernaut within the American economy. The BEA estimated that the tech sector accounted for just under 7% of the total value added to the American economy over the past several years.[11] Controlling for the contribution of computer manufacturing and the telecoms industry, the number is smaller still, as software accounted for just 3.6% of the total.[12] Further, the software industry accounts for less than 3% of employment in the US, a figure that has been fairly stable for much of the previous 20 years. As a point of reference, manufacturing, which is widely seen in decline in the era of a service economy, still employs more than 4 times the number of workers in the US tech sector.

So why is a relatively modest sector of the broader American economy perceived to be so important? There are several reasons. First, as mentioned above, the tech sector employs a large number of highly skilled persons. Consequently, it accounts for a disproportionate share of higher-paying jobs.[13] In general, the more opportunities for higher-paying jobs within an economy, the stronger is the overall economy relative to competing economies. Such economies tend to act as a talent magnet, attracting highly-skilled persons from around the globe who will contribute to sustaining an economy's competitiveness. Those that offer relatively few such high-paying opportunities are characterized by a net loss of highly-skilled workers, a phenomenon known as "brain drain."

---

8. 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/

9. Barefoot, Kevin. (15 March 2018).

10. There is no single definition of the tech sector, and we refrain from providing our own precise definition in this paper because various sources we cite use different definitions of the industry. We invite those curious to follow our citations to the definitions. There are deficiencies in every definition: in Forbes, Amazon and other companies in the "Internet and Catalog Retail" industry are not included in tech; in NAICS, data firms such as credit bureaus are classified as administrative and support services.

11. Klein, Matthew C. "The US tech sector is really small." *Financial Times*. January 8, 2016. Accessed at: https://ftalphaville.ft.com/2016/01/08/2149557/the-us-tech-sector-is-really-small/

12. *Op. Cit.*

13. Strauss, Karsten. "The Best-Paying Jobs and Industries in the US." Forbes. September 6, 2017. Accessed at: https://www.forbes.com/sites/karstenstrauss/2017/09/06/the-best-paying-jobs-and-industries-in-the-u-s/#71c0e1305038 Forbes' top-10 highest paying industries, complete with median annual wages plus bonuses, is as follows: Software & IT Services ($104,700); Hardware & Networking ($101,100); Manufacturing ($85,600); Healthcare ($84,600); Finance ($82,800); Consumer Goods ($80,000); Construction ($78,500); Corporate Services ($75,000); Legal ($72,600); Media & Communications ($71,900).

A second reason is the employment multiplier of the tech sector. Any given industry has both a backward linkage (a supply chain) and a forward linkage (how employees in an industry and their suppliers spend their wages). For example, the automobile industry has backward linkages to suppliers of tires, steel, auto parts, audio equipment, sensors, computing equipment for consoles, entertainment systems and so forth. Employees earn wages and spend those on rent, clothes, restaurants, hotels, and Netflix subscriptions among other things. Different industries have relatively more or less forward and backward linkages, depending upon the good or service they produce, and the wages paid to workers in that sector. A recent study of 179 major sectors in the US found that only three sectors—manufacturing (durable), utilities, and real estate—had a higher employment multiplier than the information sector (software).[14]

Table 1

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*

14. Bivens, Josh. "Updated employment multipliers for the US." (2019)

A third major explanation as to why the tech sector's significance to the broader US economy goes well beyond employment figures, mean/median annual salary, percentage of total value added, and even market capitalization is a bit harder to quantify, but fairly intuitive. Namely, software and data are used by nearly every industry in the American economy—including small businesses, farmers, and manufacturers—to improve efficiency and quality. For example, US automobile manufacturers use data analytics to better understand the wants and needs of their customers. This data is then used to impact the automobile design process.[15] As cars become increasingly computerized, the ability of US manufacturers to leverage data generated by a vehicle's GPS, sensors, and computers will also increase. This is but one example among many. It is this, the productivity-enhancing, innovation-driving aspects of data in industries outside of ***tech*** (such as farming, education, health care, manufacturing, entertainment, etc.) that have had the biggest economic impacts and will likely continue to do so.

The McKinsey Global Institute's *Digital Globalization: The New Era of Global Flows* calls these "digital wrappers," for "digital add-ons that enable and raise the value of other types of flows."[16] That report analyzes the economic impact of digital globalization, estimating that "global flows have raised world GDP by at least 10 percent; this value totaled $7.8 trillion in 2014 alone."[17] The US is one of the most active participants in this global exchange of data. Small and medium enterprises benefit disproportionately, as the cost of doing business globally decreases. In 1977, large American companies' share of exports was 84% but by 2013 it was 50%, and the share of the smallest US companies (less than 50 employees) is increasing rapidly.[18] Information makes the market more efficient and productive, and technology and data have given rise to an entirely new category of products, pure digital goods and services that are often free of charge.

However, one issue with information technology and data is that some consumer and economic benefits are not captured that well by traditional measures of gross domestic product or national productivity.[19] Economists and statisticians at the Bureau of Labor Statistics (BLS) and the BEA are constantly reviewing their metrics to make adjustments in their estimates of economic growth, employment, value added, and productivity among other indices of economic performance. For example, when the BEA decided that software and hardware should count as investment, it led to revisions showing sizeable increases in growth in the 1980s and 1990s.[20]

15. Schmarzo, Bill. "Big Data in Automotive and Machinery: Using Analytics to Deliver Better Products and a More Fulfilling Driver Experience." DellEMC. 10 September 2012. Accessed at: https://infocus.dellemc.com/william_schmarzo/big-data-in-automotive-and-machinery-using-analytics-to-deliver-better-products-and-a-more-fulfilling-driver-experience/

16. Manyika, James, et al. *Digital Globalization: The New Era of Global Flows*. McKinsey Global Institute, March 2016, accessed at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/digital%20globalization%20the%20new%20era%20of%20global%20flows/mgi-digital-globalization-full-report.ashx

17. *Op. Cit.*

18. *Op. Cit.*

19. DePillis, Lydia. "Technology helped America's economy more than we thought." CNN Money. August 3, 2018. Accessed at: https://money.cnn.com/2018/08/03/news/economy/gdp-economic-growth-technology/index.html

20. *Op. Cit.*

More recently, the BEA accounted for discrepancies in cell phones and cloud computing. The effect was to boost growth estimates, further highlighting the significance of the sector upon the broader economy.

The need for ongoing revisions in economic growth and performance metrics will continue insofar as the world is becoming ever more data-driven. IT developments are spawning new and rapidly-growing sectors and companies and disrupting the old. Consider online shopping and services, search engines, social media, biometrics, artificial intelligence, Big Data, machine learning and the Internet of Things (IoT). This data-driven revolution is not simply a new, single sector, it is also transforming large swaths of incumbent sectors and activities, from customer service, to retail, to farming, to manufacturing, to entertainment and so on.

As developments continue, this progress will have greater impacts and touch more and more activities and sectors. Economic benefits and strong consumer demand are helping to propel this greater use of data and data-driven processes. These processes are ongoing and will likely result in more upward adjustments in growth projections, as well as revised historic growth metrics reflecting an even greater impact still as new data are accounted for. Already, efforts are afoot to get ahead of the impacts of artificial intelligence and machine learning before they completely revamp the way many industries do business.[21]

Future productivity and economic growth prospects rest heavily on how this revolution plays out. All of this explains why there is increasing policymaker interest in data regulations and data protection rules but also underscores how essential it is to get data protection rules *right*.

The US tech sector is critically dependent upon access to data and data analytics for a broad range of uses and applications. Therefore, discussions of the tech sector and tech policy are bereft of meaning without including information services and data. This places the tech sector on center stage in debates over data privacy and data security policy.

21. *Op. Cit.*

In light of this, it is even more critical that lawmakers not dampen this impact by overly constraining the ability of the tech sector, and other data-empowered sectors, to use data and data technologies to innovate. One could credibly argue that America's dominant position in the industry of the era is partially, if not largely, attributable to the ability US firms have to access vast quantities of data for use in a broad range of applications. The competitiveness of many sectors, and their ability to innovate, is critically contingent upon this continued relationship with data and information technologies.

THREE

US VS. WORLD: THE NEED FOR REGULATION THAT ENABLES INNOVATION

Summary: American dominance in the tech sector vis-à-vis China and Europe is due in part to regulations that allow US companies to innovate with data. A growing number of state data protection laws are being put in place (or have been put in place) in an area that is truly national (if not international) in scope. This risks fragmenting the national US market, creating a more difficult and costlier environment for a range of national businesses. The relative competitiveness of US firms would be eroded in a wide array of industries, not just tech. A clear movement toward creating a flexible, proportionate data protection law that protects consumers (data subjects) without hampering innovation would preserve the competitiveness of American tech firms, other US industries utilizing technology for innovation, and the broader American economy.

The US is home to the largest share of top tech companies, as seen on the next page in Figure 2. In addition to the large presence of US tech companies, other noted characteristics of the global tech industry landscape is the relative strength of China and its relatively recent strong growth. Examples of large Chinese tech companies are Tencent, Baidu, and ZTE. In fact, by some measures, while the US has eleven of the top twenty, the remaining nine are Chinese companies.[22]

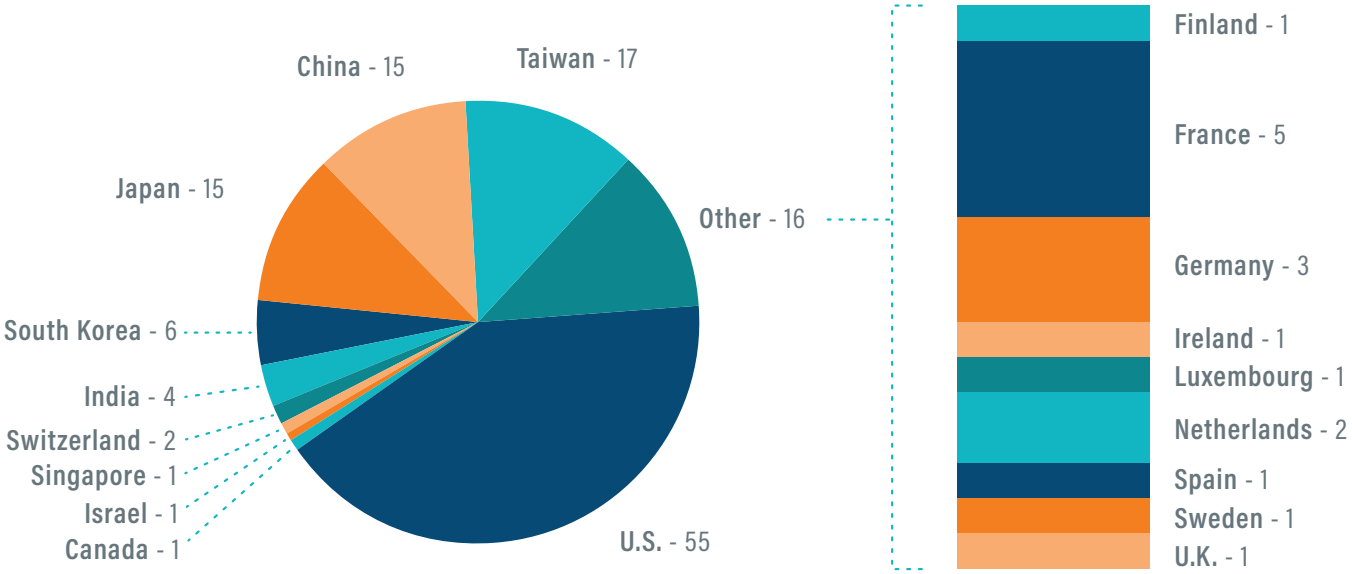
The flip side of these points is that the EU is underrepresented in the tech industry. The reasons for these differences between the US, the EU and China, in this regard, receives much speculation in meetings and conferences on data, data protections, and the tech industry.[23]

22. French, Sally. "China Has 9 of the World's 20 Biggest Tech Companies." *Marketwatch*. May 31, 2018. Accessed at: https://www.marketwatch.com/story/china-has-9-of-the-worlds-20-biggest-tech-companies-2018-05-31

23. While it would be very difficult to make a scientific, concrete argument for these differences, the following are common views. European fragmentation is one explanation. Another speculated cause is that the EU has more restrictive data protection laws compared to the US and China. Yet another cause may be the funding pipelines. The US has an extraordinary entrepreneur and private VC culture and China (while building this) has an interventionist government that protects and nurtures domestic industries.

Figure 1

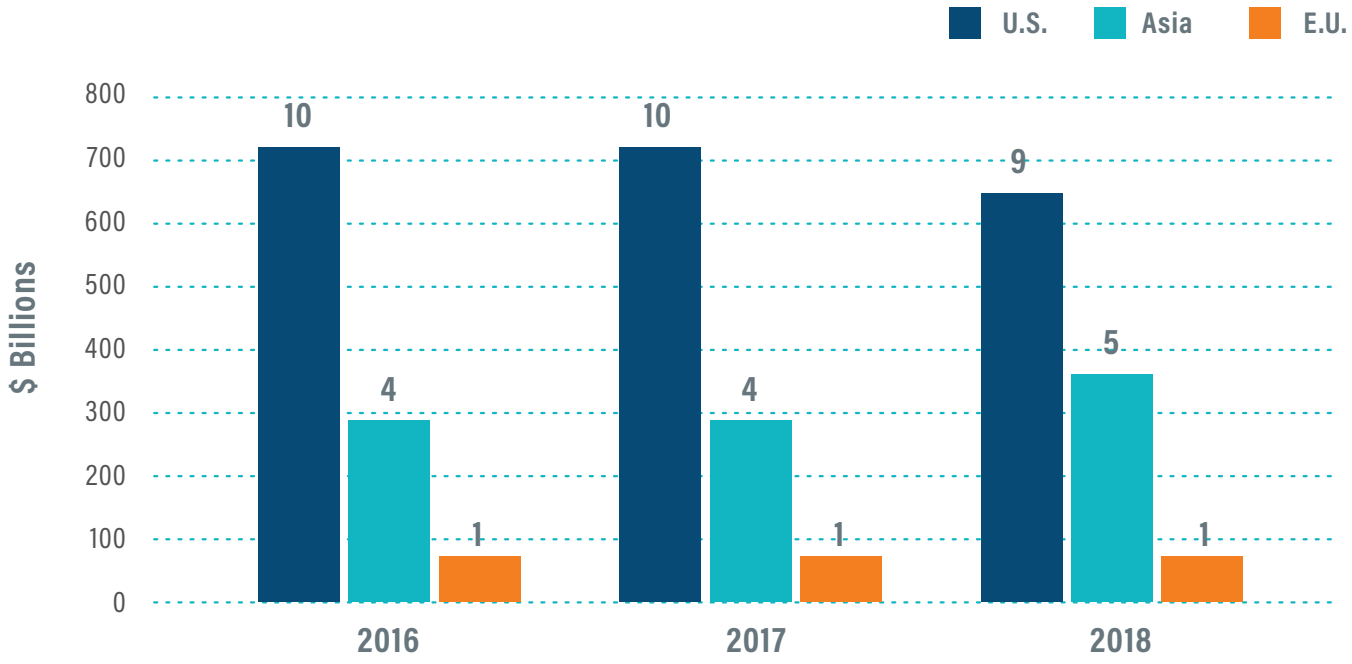
Number of Top Tech Companies by Country



Source: *Forbes*

Figure 2

Total Revenue of World's Top 15 Tech Companies by Region *($ billions)*



Source: *Forbes*

Generally, the US is sometimes thought of as having a "siloed," sectoral, or need-specific approach to data protection.[24] [25] The advantage of this approach is that rules and guidelines can be tailored to specific needs, and industries, and activities.[26] This may be most likely to produce appropriate rules for ***specific*** activities. However, there are instances in which data are collected and used outside of the silos, such as data collected on consumers regarding retail business transactions or from public records, and used for purposes "ungoverned" by laws such as the FCRA. There may also be ambiguity regarding whether an activity is within a silo or within multiple silos with differing rules. With the increasing importance and use of data in virtually all sectors, this has grown as an issue, as regulatory lines become increasingly blurred.

While the US and China are large national markets, the EU has been fragmented in terms of the data/tech industry over the past several decades due to a patchwork of national laws (and perhaps multiple languages and differences in cultural/social norms) covering not only data privacy, but e-commerce, cybersecurity, and audiovisual services. A stated aim of the EU's new General Data Protection Regulation (GDPR) has been to corral multiple national data rules created by the 1995 Data Protection Directive into a unified framework. European officials have touted that an added benefit of GDPR will be that European firms will be able to achieve the scale necessary to compete with US and Chinese tech firms. This omnibus or comprehensive approach covers data collection, protection, and exchange generally, across sectors and activities, and necessarily, does not have any ungoverned areas of personal data.[27] However, a potential downside to this approach is that it can be seen as "one-size-fits-all" and not properly gauged to specific needs.[28] Lack of specificity (given it is a ***general*** regulation) and the comprehensive nature of the GDPR means that Member States have begun to interpret and enforce the law in diverging ways, undermining the claim that GDPR creates a single unified European data privacy law. Moreover, the European Union has begun drafting an e-Privacy Regulation, a sector-specific law to require that electronic communications providers (so-called 'over-the-top' companies) and telecommunications companies comply with privacy requirements that may be inconsistent even with GDPR. Amidst this ambiguity, many organizations were not fully ready for the regulation when it went live May 25th, 2018, since all the specifics were not ironed out.[29]

---

24. In the US, for instance, the FCRA covers credit bureau data, HIPAA covers health data, COPPA for websites that collect data on children, the Privacy Act covers government data, and so on. Each of these may govern individual/data subject rights, how data may be used, what data may be collected, how it can be shared, and data security.

25. This section builds on a discussion from a previous PERC White Paper: Walker, Patrick. *Data Protection and Credit Information Sharing*. Policy & Economic Research Council (PERC), November 2017, available at: http://www.perc.net/wp-content/uploads/2017/11/Data-Protection.pdf

26. The "ad hoc" approach likely took root in the US since rules were built up over time addressing specific concerns. And given the weight attached to freedom of speech, it may be difficult to craft a single set of rules that would cover all data/information collection and flows (for example, collecting public record information or information from newspapers about data subjects).

27. This being said, some specific activities (such as some government data collection) are excluded from provisions.

28. In these simplistic descriptions, the EU is like a city with a single speed limit, no nuance in the rules. Whereas the US is like a city that has particular, well-thought out speed limits on some sections of its roads and no posted speed limits on remaining sections. And as with data protection rules and the explosion of data, as traffic grows in each city the flaws in both approaches will become costlier and ever clearer.

29. Henderson, Richard. "GDPR Compliance in the New Age of Data Consciousness." *Forbes*. June 11, 2018. Accessed at: https://www.forbes.com/sites/forbestechcouncil/2018/06/11/gdpr-compliance-in-the-new-age-of-data-consciousness/#1ba1ac5f623e

The reality of data protection governance in the US and the EU appears to be more flexible and reflective of pragmatism than might be assumed. In the EU, details of how the GDPR is to be applied to specific industries and activities is being worked through. What is considered necessary and appropriate data collection and retention for one activity will differ from another. In this way and in other flexible aspects of the regulation (setting aside some unhelpfully ambiguous parts of the regulation), the GDPR does in several ways bend to the realities of the market and individual needs. This is, of course, separate from whether the GDPR is *sufficiently* flexible, unduly burdensome, or always strikes the right balance between competing needs.

In the US, the Federal Trade Commission (FTC), a consumer protection agency with very broad powers covering virtually all national commerce, oversees many consumer data privacy and security rules (such as Children's Online Privacy Protection Act or "COPPA") and enforces the US-EU Privacy Shield framework, in which companies certify that they comply with the privacy principles required to meet the EU's adequacy standard.[30] The FTC is also a watchdog agency for "Unfair or Deceptive Acts or Practices," (UDAP), one aspect that helps to ensure that US regulators have comprehensive authority.[31] This authority enables the FTC to make sure companies are living up to their own data security, data protections, and data transfer statements made to consumers. The FTC has also moved further beyond rule-specific powers or enforcing statements made to consumers to more general enforcement of data security.[32] Under its "Unfair Practices" authority, the FTC has assumed regulatory authority over data security.[33]

Going beyond data security, the FTC during the Obama administration also proposed a more ambitious data privacy framework.[34] The proposed framework covers, among other issues, data retention, data collection, data security, data accuracy, consumer interactions, transparency, consent or choices, notices and data that, while not specifically personally identifiable, could become so if combined with other data elements.[35] This aspect is particularly relevant in the era of Big Data. A January 2019 GAO report highlights the FTC's role in internet privacy and notes that "Congress should consider developing comprehensive legislation on Internet privacy that would enhance consumer protections and provide flexibility to address a rapidly evolving Internet environment."[36]

30. See FTC, F*ederal Trade Commission 2014 Privacy and Data Security Update*. Federal Trade Commission, Washington D.C., 2015. Accessed at: https://www.ftc.gov/system/files/documents/reports/privacy-data-security-update-2014/privacydatasecurityupdate_2014.pdf

31. See FTC, *A Brief Overview of the Federal Trade Commission's Investigative and Law Enforcement Authority*. Federal Trade Commission, Washington D.C., July 2008. Accessed at: https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority

32. See the complaint and a supplemental memorandum of the FTC available on the FTC's website, https://www.ftc.gov/sites/default/files/documents/cases/2012/08/120809wyndhamcmpt.pdf and https://www.ftc.gov/system/files/documents/cases/150327wyndhamsuppbrief.pdf

33. In a suit against Wyndham Worldwide, the FTC has alleged that the company's data security was inadequate. The FTC argued, "the FTC has acted under its procedures to establish that unreasonable data security practices that harm consumers are indeed unfair within the meaning of Section 5. First, the LabMD Order directly states the Commission's considered determination that inadequate data security can be an unfair practice."

34. FTC, *Protecting Consumer Privacy in an Era of Rapid Change*. Federal Trade Commission, Washington D.C., March 2012. Accessed at: https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

35. The FTC is proposing that the framework cover all companies other than those that handle just a limited amount of non-sensitive data that is not shared with third parties.

36. GAO, *Internet Privacy: Additional Federal Authority Could Enhance Consumer Protection and Provide Flexibility*. Government Accountability Office, Washington D.C., January 2019. Accessed at: https://www.gao.gov/assets/700/696437.pdf

As discussed earlier, it is already clear that if the federal government does not act, states will. A benefit of enacting a general national data law would be the potential to unify and reconcile (at least to some degree) a patchwork of federal and state regulations in an area that is national in scope. Such a regulation (or regulations) must eliminate a confusing patchwork of state laws where appropriate, to prevent a fragmented patchwork of data protection laws, defer to well-developed laws governing specific data and activities (FCRA for credit data, HIPAA for medical data, etc.), and fill in gaps. Although it will be difficult given the emotions associated with data privacy and security, as well as the tech industry, lawmakers must also avoid overly burdensome regulations or a more-is-better approach. Such outcomes would likely offer little additional protection to consumers, and could have harmful economy-wide impacts.

As always, the details to such legislation are key. Ideally such legislation would be sufficiently flexible and well-designed so as to enable meaningful consumer data protections without unduly inhibiting useful data exchanges, data-dependent activities, and innovation. Since technology and services are changing so rapidly, it would likely be prudent to draft legislation that avoids specific technologies, focuses on principles, and provides regulators with adequate discretion in carrying it out. And given the fast pace of developments with IT and data processing, it is incumbent that the IT and data industries take a lead in this policy dialogue, or, at the very least, be a key part of the process.

FOUR

"TURTLING" THE TECH SECTOR AND THE US ECONOMY: THE LAW OF UNINTENDED CONSEQUENCES

The Law of Unintended Consequences describes an outcome caused by the action of a person or group of people (oftentimes policymakers) entirely unrelated to the desired objective. A classic example of this is the Eastland tragedy. In response to the sinking of Titanic, in which there were far too few lifeboat seats for passengers, support for a "lifeboats for all" law swelled resulting in passage of a law requiring lifeboat seats for at least 75% of all passengers.[37]

During Congressional debates, one expert warned that many Great Lakes vessels with their shallow drafts would "turn turtle" if required to bear the additional weight of lifeboats. Such warnings were ignored by Congress eager to act on the demands of the "lifeboats for all" movement. Tragically, owing to compliance with the new law, the ship capsized in the harbor still tied to the dock and drowned 844 passengers, 70% of whom were under the age of 25.[38]

37. Stranahan, Susan Q. "The Eastland Disaster Killed More Passengers Than the Titanic and the Lusitania. Why Has It Been Forgotten?" *Smithsonian.com*. October 27, 2014. Accessed at: https://www.smithsonianmag.com/history/eastland-disaster-killed-more-passengers-titanic-and-lusitania-why-has-it-been-forgotten-180953146/#WG6ukdATMhcLG2us.99

38. *Op. Cit.*

This is an apt albeit imperfect analogy for the policy environment in the US surrounding the tech industry. The Equifax breach and social media platform privacy incidents are akin to the sinking of the Titanic. National privacy legislation, unless crafted in a manner that balances the need for privacy protections with the legitimate needs of industry to access and use data, could saddle US firms with burdens sufficiently heavy to capsize the entire information-based economy.

The US dominates in the industry of our era—information technology—partially because American firms enjoy a competitive advantage in access to and the use of data. The current regulatory framework in the US, one that contains industry and use-specific regulations calibrating protections to the relative sensitivity of the data, permits US firms to use data for analysis and innovation. Countries with similar regulations are likewise competitive, including China, South Korea, and Israel, among others. Those that constrain access to and the use of data, such as the EU members and Canada, are largely laggards in this industry—an outcome with consequences for the competitiveness of many other industries in those countries.

While passenger vessels in the US ultimately became safe and profitable, it wasn't until after avoidable tragedies such as the Eastland. This is not to suggest that a national privacy law or state laws will capsize the US tech sector or economy, but they could do significant harm if poorly crafted. It is for this reason that it is critically important to get this right the first time.

We urge policymakers to recognize the need to take an informed approach and strive to maintain a balance between consumer privacy and data security protections, and the legitimate needs of a wide range of actors in American society. Our national economic vitality and security relies upon the continued success of American enterprises. The legislation should protect and empower consumers *and* enable a vibrant economy.

It is hoped that the US Data Ecosystem research succeeds in raising awareness of the pervasiveness of data throughout our modern economy, the real risks associated with unauthorized access to certain data, and the myriad benefits derived from access to and the use of PII and other data for innovative new applications in every sector of the US economy. We believe our studies will fill several glaring information gaps with well-reasoned conclusions derived from both logic and empirical analysis.

PART II

THE EVIDENCE ON DATA BREACHES AND CONSUMER IMPACTS

BRINGING FACTS INTO THE FRAY

Although data breaches are fundamentally a data security issue, current state data privacy laws have data breach clauses in them (California and Vermont), as do previously proposed legislation in other states such as Washington.[39] While technically data privacy broadly pertains to who may access data and the purposes for which it may be used, and data security focuses upon the prevention of unauthorized access and use of data, these distinctions are devoid of meaning in a hyper-emotional policy context.[40] That being said, the two are obviously linked since effective privacy outcomes are impacted by data security.

To be clear, data breaches have been linked to ID theft/fraud and other individual harms. Breaches of sensitive data in particular present a clear danger to affected persons, and for this reason all practical measures to prevent breaches should be implemented. Because of this risk, it is paramount that data custodians of all types of PII and other sensitive data remain ever-vigilant to the risks of unauthorized access. Investments in complying with strict data security standards are both a good business practice and the ethical course to take.

Once a breach occurs, sensitive health and financial PII can appear for sale on the Darknet or dark web,[41] enabling criminals to commit identity theft, tax fraud, and loan fraud. ID theft victims can incur significant and wide-ranging costs in an effort to restore their good credit standing, reclaim their identity, and protect themselves from future harms.[42] Fear of data breaches has never been higher in the US. One recent survey of 3,000 Americans found that fear of having personal data stolen in a breach (55%) eclipsed fear of having their wallet (26%), car (10%), cell phone (6%) and house keys (6%) stolen.[43] Confidence in data security is also at historic lows as a Pew Research Study Center found just 6% were "very confident" that government agencies could keep data secure, while only 9% were "very confident" that credit companies could secure the data.[44]

39. Washington Privacy Act, S.B. 5376, 66th Washington Legislature (2019).

40. There is one school of thought that includes the concept of autonomy as an element of data privacy. In this context, autonomy involves the desire to control information about us. Further, individuals wish to assure "fair processing," that is, that data collected about them cannot be used to cause them or their family harm. Data breaches intersect with fair processing because stolen data is only used to harm a data subject, if it is used at all. As such, under this construct, a data breach is a policy concern under the domain of data privacy via the notion of fair processing. To date, to the extent that lawmakers (mostly in the EU) concern themselves with fair processing, they focus on consent mechanisms such as "opt-in" or "opt-out." This will not matter in the case of a data breach. For a further discussion, see: Abrams, Martin. "Foundational Issues for New Privacy Law in the United States." The Information Accountability Foundation. 19 April 2019. Accessed at: http://informationaccountability.org/foundational-issues-for-new-privacy-law-in-the-united-states/

41. The **Darknet** is an encrypted network, built on top of the existing Internet, requiring specific tools or software to enable access. Darknet provides anonymity to users, which is why it is the preferred marketplace for stolen data. Examples of Darknet is The Onion Router or "TOR." TOR has a special protocol for accessing TOR only Web sites that have a ".onion" URL address. TOR-only sites are nicknamed "onion land." The **dark web** is comprised of Darknets like TOR. Paraphrased from: https://fossbytes.com/difference-deep-web-darknet-dark-web/

42. Ablon, Lillian, Paul Heaton, Diana Catherine Lavery, and Sasha Romanosky. *Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information*. RAND Corporation. Santa Monica, CA. 2016. Downloaded at: https://www.rand.org/content/dam/rand/pubs/research_reports/RR1100/RR1187/RAND_RR1187.pdf

43. "Americans Fear Their Data More Than Their Wallet, Radware Survey Finds." Market Insider at BusinessInsider.com. 7 August 2018. Downloaded at: https://markets.businessinsider.com/news/stocks/americans-fear-for-their-data-more-than-their-wallet-radware-survey-finds-1027437171

44. Madden, Mary and Lee Rainie. "Americans' Attitudes About Privacy, Security and Surveillance." Pew Research Center, Internet and Technology. 20 May 2015. Downloaded at: https://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

Trust in companies is eroding, as one recent survey found nearly 7 in 10 respondents believe companies are vulnerable to cyber-attacks and fewer than 3 in 10 felt that companies handle personal data responsibly.[45] Interestingly, the same survey found that more than 7 in 10 respondents thought businesses and not government were best equipped to protect them.[46]

With each new widely-publicized data breach, anxiety levels about personal security rise among the general public and pressure for government intervention mounts. One recent survey by PwC found that more than 8 in 10 respondents say the government should regulate companies' use of data, and that government regulation of new technologies is a crucial consumer protection.[47] Lawmakers in some states have already acted by passing measures ostensibly designed to protect consumers from harms associated with data breaches. Others still are poised to act, including federal lawmakers who are considering national privacy legislation with data breach provisions.

As part of the US Data Ecosystem project PERC recently launched in partnership with C_TEC, this study reflects completed original quantitative research focusing on the impacts of data breaches. The analysis on the relationship between data breaches and consumer impacts will provide lawmakers, regulators, and stakeholders with additional facts and data points to consider when crafting privacy legislation that includes components associated with data breaches. Particular emphasis was placed on quantifying data breach impacts on ID theft and ID fraud losses. Until now, there has been little systematic evidence and few studies on whether being a data breach victim meaningfully and measurably increases the likelihood of one's identity being stolen. This unproven assumption is a research gap that the Data Ecosystem paper will attempt to fill.

This report takes a macro to micro level approach, first analyzing the aggregate relationship between data breaches and general identity theft. Other putative relationships are also explored, including those between data breaches and fraud losses, and between investments in data security technology and fraud losses by sector and overall. Then, the lens is narrowed to individual-level impacts—the micro- level of analysis—examining credit report data from specific data breaches provided by a national credit reporting agency. Finally, the report zooms into specific data breaches, examining available data on a non-random sample of 24 large data breaches. This extends (and updates) the GAO's 2007 analysis, which found limited evidence of identity theft resulting from data breaches.

To prevent the US tech sector, other significant portions of the economy, and our national competitive

45. "Consumer Intelligence Series: Protect Me. An in-depth look at consumers want, what worries them, and how companies can earn their trust—and their business." October, 2017. Downloaded at: https://www.pwc.com/us/en/advisory-services/publications/consumer-intelligence-series/protect-me/cis-protect-me-findings.pdf

46. *Op. Cit.* Pg. 2.

47. *Op. Cit.* Pg. 2.

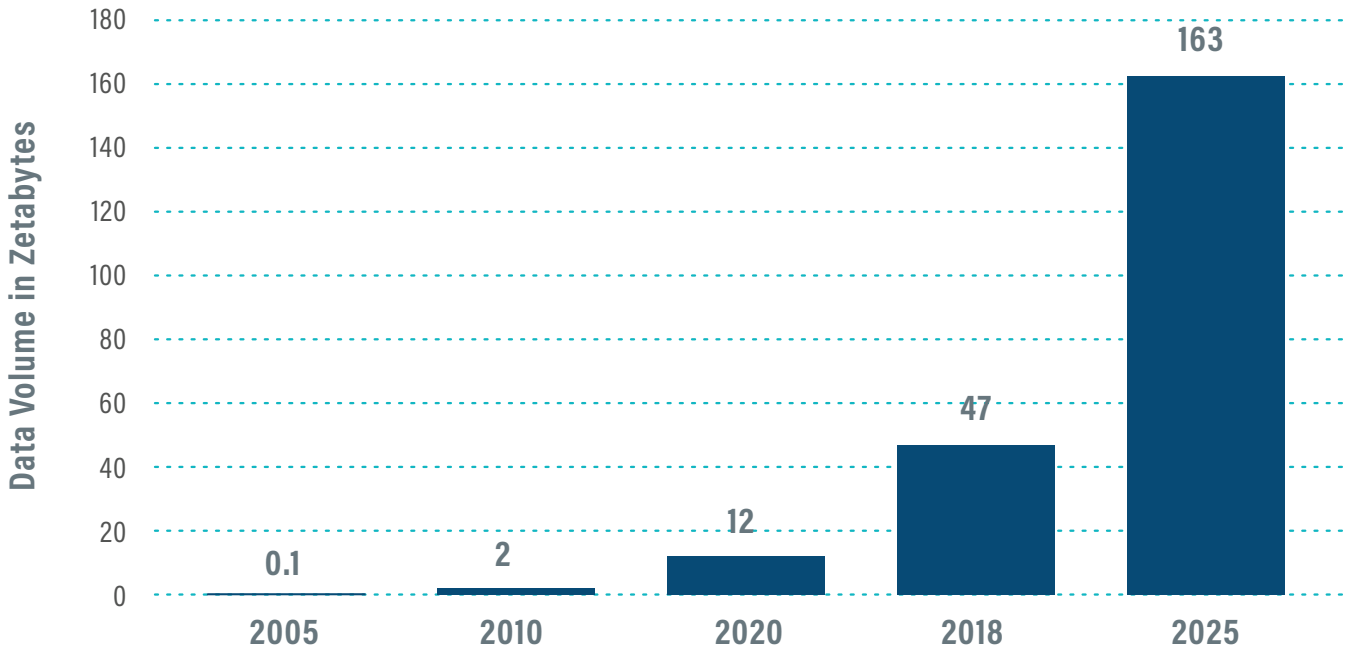
advantage from being capsized by ill-designed data protection legislation, it is important for lawmakers to act on evidence, and to go forward with a federal approach with clearly defined objectives rather than a piecemeal, state-by-state approach that is assured to maximize uncertainty and carry with it economic harm. As data breaches are a (perhaps the) primary driver of recent state privacy laws, gaining a better understanding of the impacts of data breaches and policy tools to reduce the probability of breaches and mitigate the consequences is essential.

DATA BREACHES

The collection of data is growing so quickly experts are continuously inventing new names for the volume of data created, collected, and stored. We are now up to zettabytes, which is equal to 1 trillion gigabytes. For comparison, the Library of Congress print collection is 10 terabytes, or approximately 10 000 gigabytes. One zettabyte equals 100 million Library of Congress print collections.

Figure 1

Volume of Data/Information Created Worldwide from 2005 to 2025 (zetabytes)



Source: *Statista, 2020 and 2025 are projected*

However, as more interactions become digitized, increasing amounts of sensitive data are stored online in interconnected systems. This increases the opportunities for criminals to target this data, as well as the opportunities for an employee to make a mistake and unintentionally expose or disclose the data. Given the exponential rise in the amount of data being collected, it is no surprise that the incidence of data breaches has been steadily increasing. While data breaches did not originate with modern IT (paper files can also be stolen or misplaced), the ease and speed with which data can now be gathered stored and transferred has made very large breaches possible and increasingly common.

A data breach is a very broad term, generally meaning an incident in which sensitive (personally identifiable) data has been lost, stolen, or viewed/used by unauthorized individuals. Breaches can involve electronic data or physical/paper records. So, a doctor's office emailing electronic patient records to the wrong email address or mailing physical files to the wrong address would both be data breaches. Data can be breached due to an accident (losing files, sending them to the wrong place, inadvertently making them public, accidently giving unauthorized parties access), and data can also be stolen. Data can be stolen from a traditional physical break-in, from a rogue employee, and electronically (from outside a targeted organization) by a computer hacker. Individual consumers can also lose data (or have it stolen) from their own computers. Individuals can also have their data stolen from their homes, cars, and mailboxes and taken from their garbage (among other ways). Information/data can be lost and stolen many ways.

A number of organizations record instances and details of data breaches, but given the fact that some breaches may go undetected (at least for some time period) and not all details of data breaches are released, the figures on data breaches should be viewed as being based on detected breaches for which information is publicly released. The Identity Theft Resource Center (ITRC), for instance, lists 1,244 data breaches in 2018 involving a little over 446.5 million records.[48] In the 2018 report many of the 1244 listed data breaches have "Unknown" listed for the number of records involved, so it is likely that 446.5 million records figure understates the true figure for the 1,244 incidents. Since 2005, the ITRC has recorded over 10,000 data breaches and over 1.5 billion records involved. Again, these likely understate the true figures.[49]

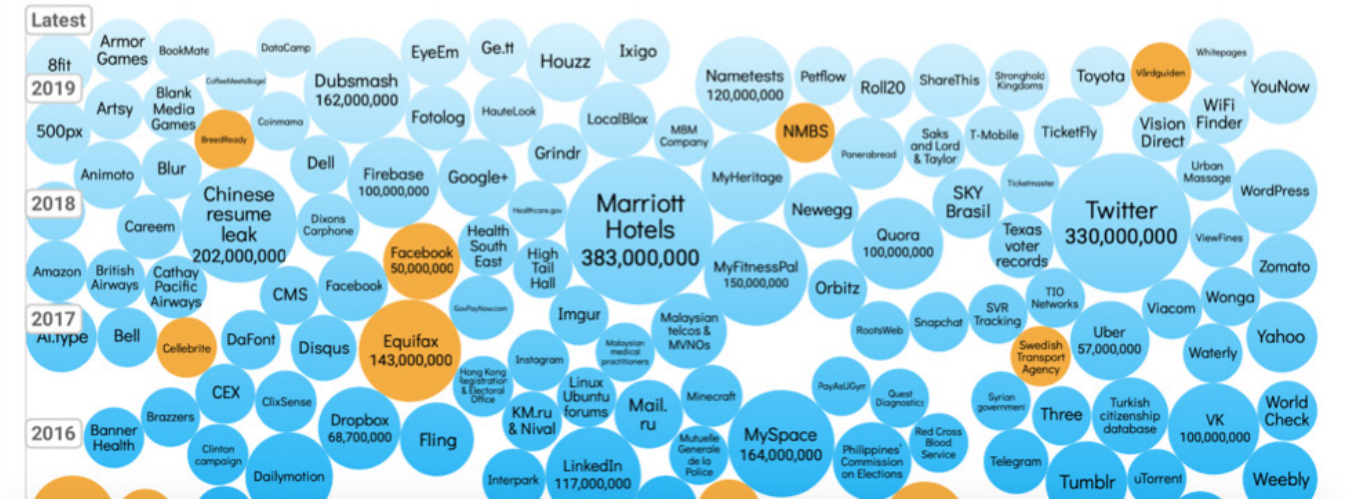
An interesting visualization of some selected data breaches has been compiled on the *Information is Beautiful* website. This infographic is shown in Figures 2a and 2b, which depict select data breaches of over 30,000 records. Figure 2a shows the most recent breaches, with the bubble size depicting the size of the data breach.

48. See Identity Theft Resource Center, *2018 End of Year Data Breach Report*. January 28, 2019. Accessed at: https://www.idtheftcenter.org/wp-content/uploads/2019/02/ITRC_2018-End-of-Year-Aftermath_FINAL_V2_combinedWEB.pdf

49. See Identity Theft Resource Center, Data Breaches. Accessed at https://www.idtheftcenter.org/data-breaches/

Figure 2a

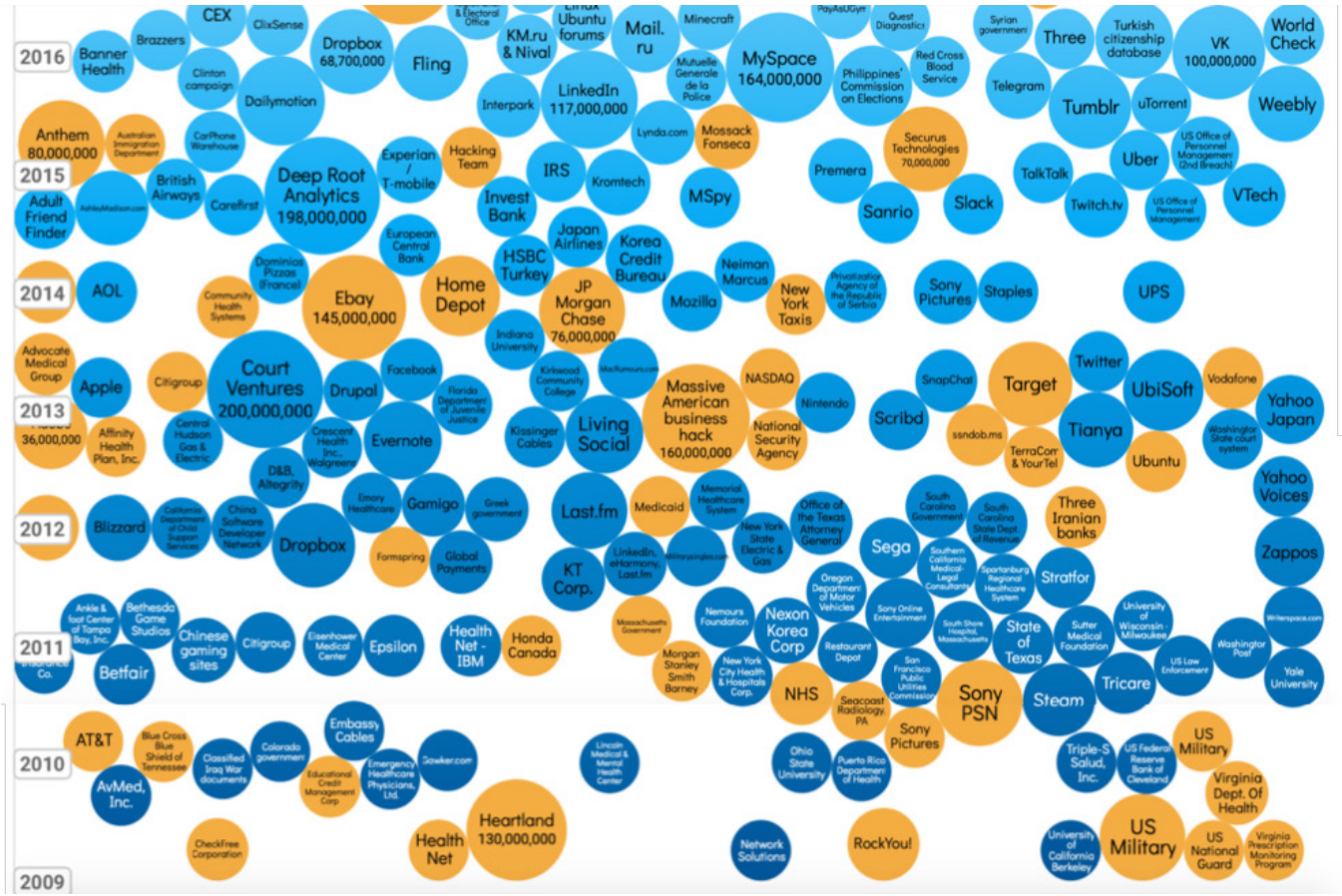
Select Latest Data Breaches



Source: *Information Is Beautiful*

Figure 2b

Select Data Breaches from 2009 to 2016



Source: *Information Is Beautiful*

The sheer scale of the number of records involved in breaches, not uncommonly in the hundreds of millions, is enormous (Twitter: 330 million, Quora: 100 million, MyFitnessPal: 150 million, Equifax: 143 million). In fact, over the past decade the 1.5 billion records breached (ITRC figures), is over four times the size of the US adult population. Moreover, these are just the reported data breaches.

InfoSecurity surveyed malware analysts working at US corporations, and 57% stated that they were working on unreported data leakages.[50] A Baseline Magazine survey found that 80% of firms hid breaches to protect their reputation, and that roughly 60% had "don't tell" policies in place concerning cyber-attacks. It is possible that new legislation could provide a further disincentive to report data breaches, or if there are fines and consequences for not reporting and/or increased reporting requirements, then new legislation could prompt greater reporting. Europe's GDPR had the impact of companies "over-reporting" a deluge of data protections concerns to the Information Commissioner's Office in order to comply with the new regulation.[51] These details of what should be reported and how should thus be thought through carefully.

Figures 3 and 4 use data from ITRC from 2005 to 2018 to depict trends in the number of data breaches and the number of records breached over the last decade. Given these numbers, it would be reasonable to conclude that many persons would have had their data breached multiple times, and it is likely most would have had their data breached at least once. A range of studies supports this conclusion. Interview highlights from the NPR show *All Things Considered* titled "Theft Of Social Security Numbers Is Broader Than You Might Think" notes a particular company "gets cyberattack data from dozens of organizations around the world, including federal agencies like the Secret Service and the Department of Homeland Security's Computer Emergency Readiness Team. Jay Jacobs, a leading data scientist, is a foremost expert who has been slicing and dicing this data for years. He estimates 60 percent to 80 percent of Social Security numbers have been stolen by hackers. NPR put the question to him multiple times and he stuck by this estimate."[52]

Over the period 2005 to 2018, there has been a clear upward trend in the number of reported data breaches. Note that such changes may also reflect, to some degree, changes in reporting. While, as can be seen in Figure 4, there is no obvious trend in the number of reported records breached. Contrary to the explosive growth in data (EMC claims it is doubling every two years[53]) and the Internet, there does not appear to be matching exponential growth in the trend of records breached.

50. Charles Leaver, "Data breaches under-reported: Figures may be worse than they appear." December 5, 2013. Accessed at http://ziften.com/data-breaches-under-reported-figures-may-be-worse-than-they-appear/

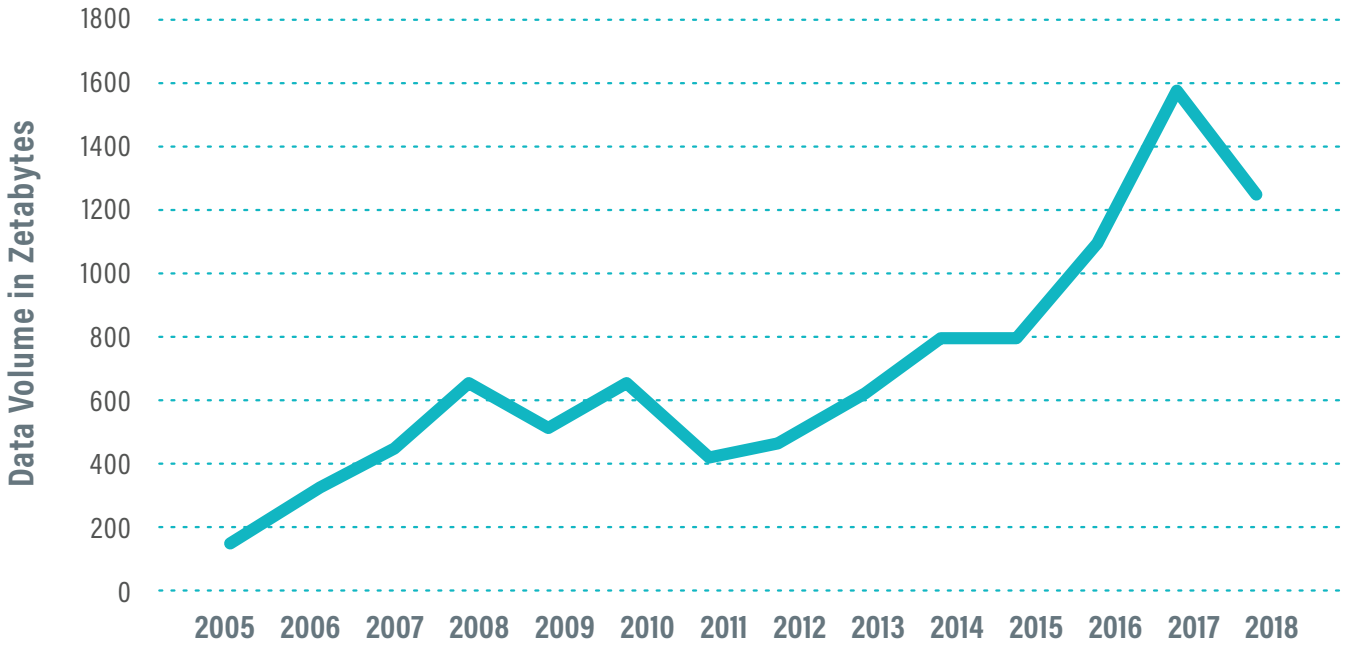
51. Hill, Michael. "Only 0.25% of Reported Data Breaches Have Led to Fines Since GDPR." *InfoSecurity*. May 10, 2019. Accessed at: https://www.infosecurity-magazine.com/news/only-025-breaches-fined-gdpr-1-1-1/

52. See Aarthi Shahani, *Theft Of Social Security Numbers Is Broader Than You Might Think*. June 15, 2015. NPR All Things Considered. Accessed at: http://www.npr.org/sections/alltechconsidered/2015/06/15/414618292/theft-of-social-security-numbers-is-broader-than-you-might-think

53. See Vernon Turner, *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things—Executive Summary*. April 2014. Accessed at http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

Figure 3

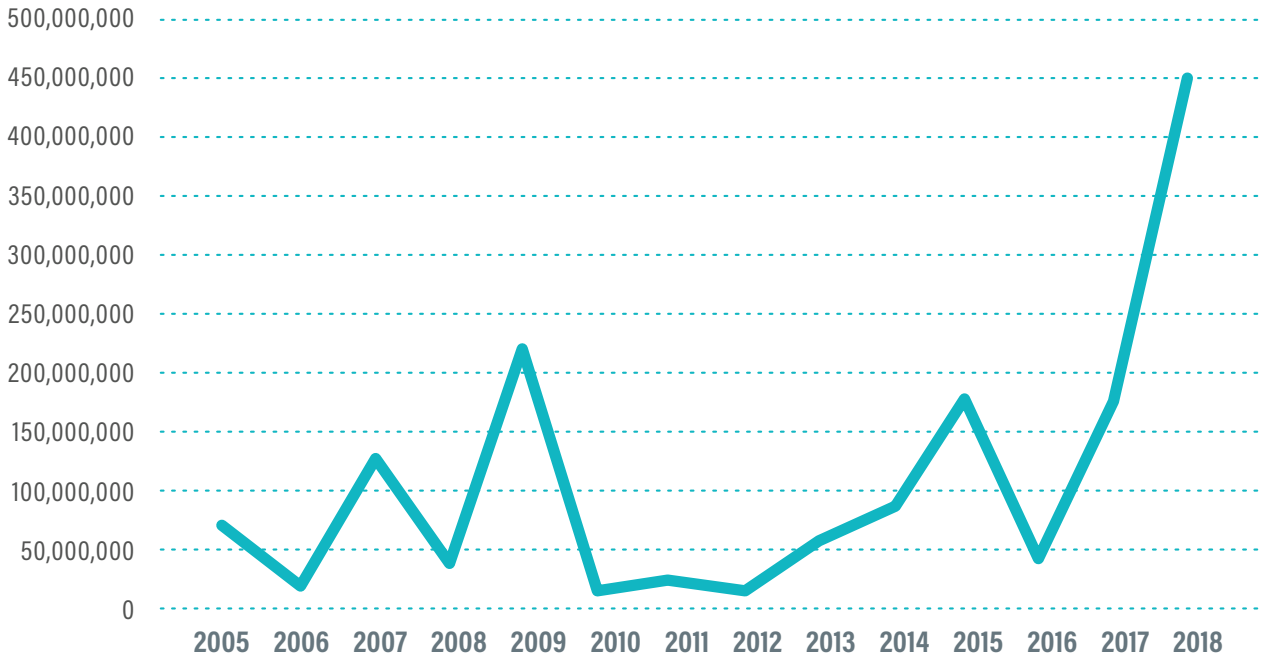
Number of Data Breaches



Source: *Identity Theft Resource Center*

Figure 4

Number of Records Breached



Source: *Identity Theft Resource Center*

In light of the evidence, whether or not they've ever received a data breach notification, it seems reasonable for a person to assume that their Social Security number and other PII has already been breached. This conclusion is further supported with logic and probability theory. Each year, a person's sensitive PII is shared electronically with an ever-larger number of third parties—banks, insurance companies, energy utility firms, telecommunications and cable companies, health care providers, dentists, employers, government agencies and more. It would seem that there is an ever-growing possibility for a person's PII to be breached.[54] That is, over time, people who are economically and socially engaged have ever-more possibilities for their PII to be breached once or (more likely) multiple times.

Despite widespread beliefs about data breaches, it is highly unlikely that a breach, is a breach, is a breach. For instance, the breach of certain types of sensitive pieces of financial data—like a checking account number, an ATM card number, a PIN code and/or a password—seem likely to pose more of a risk of ID theft/fraud than a *USA Today* or *Wall Street Journal* username and password. In other words, breaches involving greater volumes of sensitive information necessary to commit ID theft/fraud should pose relatively more risk to an affected individual than breaches with lesser amounts, *ceteris paribus*. By extension, larger breaches with more sensitive data should also present a relatively greater risk than smaller breaches involving the same sensitive data, all other things being equal. It is also possible that a relatively small breach-affected population with significant quantities of necessary sensitive data and personal identifying information (PII) could present a much greater risk of ID theft/fraud than a headline-grabbing massive breach of relatively benign data with PII.

54. Article citing 2014 study by Ponemon Institute. Ponemon survey found 43% of US firms surveyed experienced breach in 2013, up 10 percentage points over previous year. Same article also cited data breach of Korea Credit Bureau, the national credit bureau in South Korea, that affected over 75% of all adults in Korea—including all of their sensitive financial and much identifying information. Elizabeth Weise, "43% of Companies Had Data Breach in Last Year." *USAToday*. Sept. 24, 2014. Accessed at: http://www.usatoday.com/story/tech/2014/09/24/data-breach-companies-60/16106197/

ID THEFT AND FRAUD

An oft-cited harm stemming from data breaches is identity theft/identity fraud. This correlation is assumed on the premise that data is necessary (but not always sufficient) to steal someone's identity and commit ID theft/fraud. As with data breaches, there are a wide variety of types of ID theft and fraud. Most such theft and fraud can be broken down into a few main categories. For instance, **existing account fraud** is fraud conducted against an existing account, such as when a criminal obtains a victim's credit card number, credit card expiration date and the card verification value or "CVV number" and uses those to purchase items. In such a case, the criminal is less stealing the victim's identity than fraudulently accessing their accounts. According to statistics from Javelin for the year 2018, this is by far the most common form of ID theft and fraud, with 77% of victims reporting such incidents.[55] The second major category for 2018 (according to Javelin statistics) is existing non-card account fraud, with an overall rate of 2.17% in 2014 (compared to 5.66% for all ID theft and fraud).

Account takeover fraud, as the name implies, goes beyond simply using account information. The criminal actually alters account information, such as adding themselves as an authorized user or changing mailing or account addresses. In 2018, the rate of account takeover fraud was 1.43%. **New account fraud**, by contrast, goes beyond using and/or altering existing accounts and occurs whenever the criminal uses a stolen identity to open a new account. Javelin found the rate for this in 2018 to be 1.25% (or reported by 22% of ID theft and fraud victims).

Fortunately, the most common form of ID theft/fraud, existing account fraud, tends to be the least burdensome for consumers. A common circumstance is whenever a consumer is notified by a card company (or notifies the card company) of suspect charges. The suspect charges are then typically removed from the consumer's account and the fraud is investigated. In fact, of all ID theft and fraud examined in 2016, a Department of Justice report found that the vast majority of victims (88%) claimed to have suffered no out-of-pocket losses.[56] Thus, ID theft involving the opening of a new account and, more generally, suffering materially from any type of ID fraud and theft are rare events (with likelihoods well under 1%). Despite the typical low or no liabilities resulting from account fraud, large out-of-pocket costs can nonetheless arise if, for instance, the ID theft or fraud victim knows the perpetrator (friend or family member) and does not want the matter investigated and/or perpetrator charged.

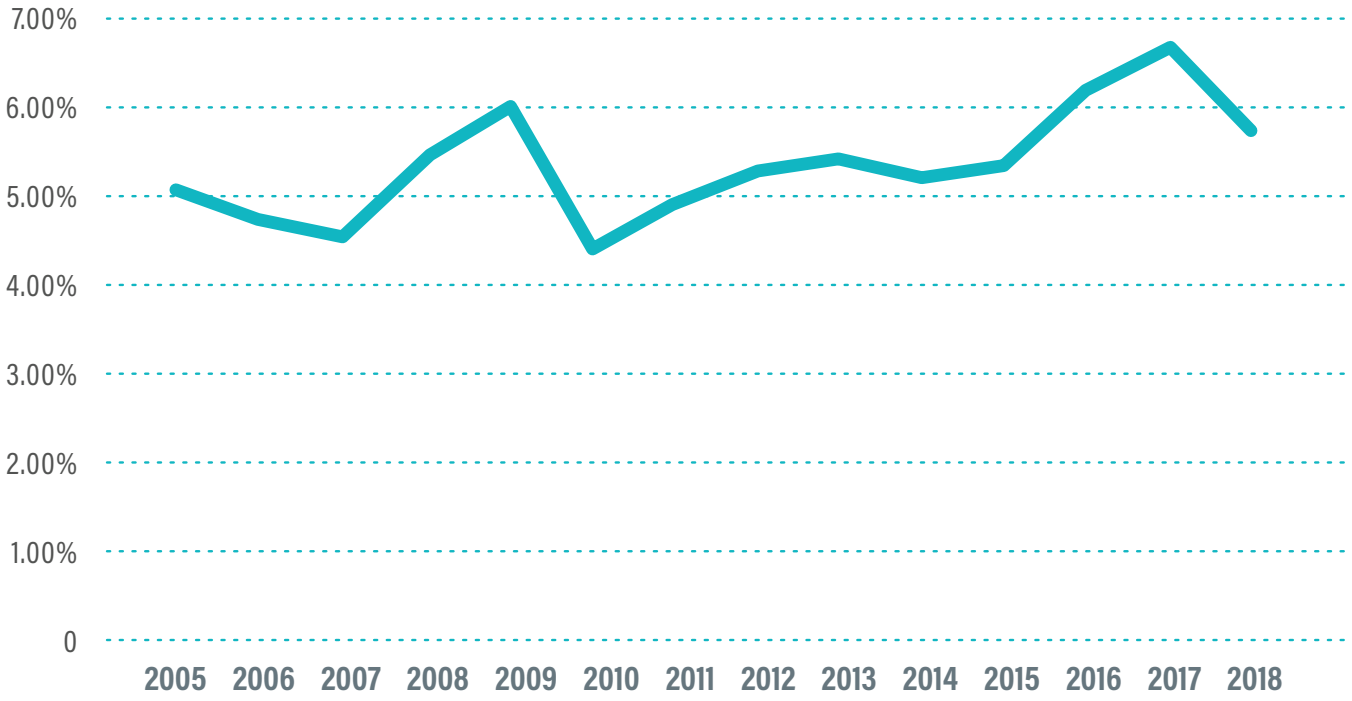
55. Where the rate of all identity fraud and theft is reported as 5.66% in 2018, the rate for existing account fraud is 4.40%. See Javelin Strategy & Research, press release for 2018 at: https://www.javelinstrategy.com/press-release/consumers-increasingly-shoulder-burden-sophisticated-fraud-schemes-according-2019

56. Erika Harrell, Ph.D. and Lynn Langton, Ph.D., *Victims of Identity Theft*, 2016. US Dept. of Justice.

The following figure shows the trend in ID theft and fraud for over a decade from Javelin data.

Figure 5

Rate of ID Theft and Fraud



Source: *Javelin Strategy & Research press releases*

As can be seen in Figure 5, there has been no obvious trend or large variation in the rate of ID theft and fraud over the past decade. And while the rate found in this survey was 5.66% in 2018, it should be noted that this cannot be taken as a uniform rate across different demographic groups. A December 2016 report from the US Department of Justice found variation in the rate of ID theft and fraud by socio-demographic groups.[57] For instance, they found that individuals from higher income households had twice the rate of ID theft and fraud than individuals from lower income households. And those ages 35-49 had an ID theft and fraud rate almost double those who were 18-24.

57. Erika Harrell, Ph.D. and Lynn Langton, Ph.D., *Victims of Identity Theft*, 2016. US Dept. of Justice. January 2019.

Table 1

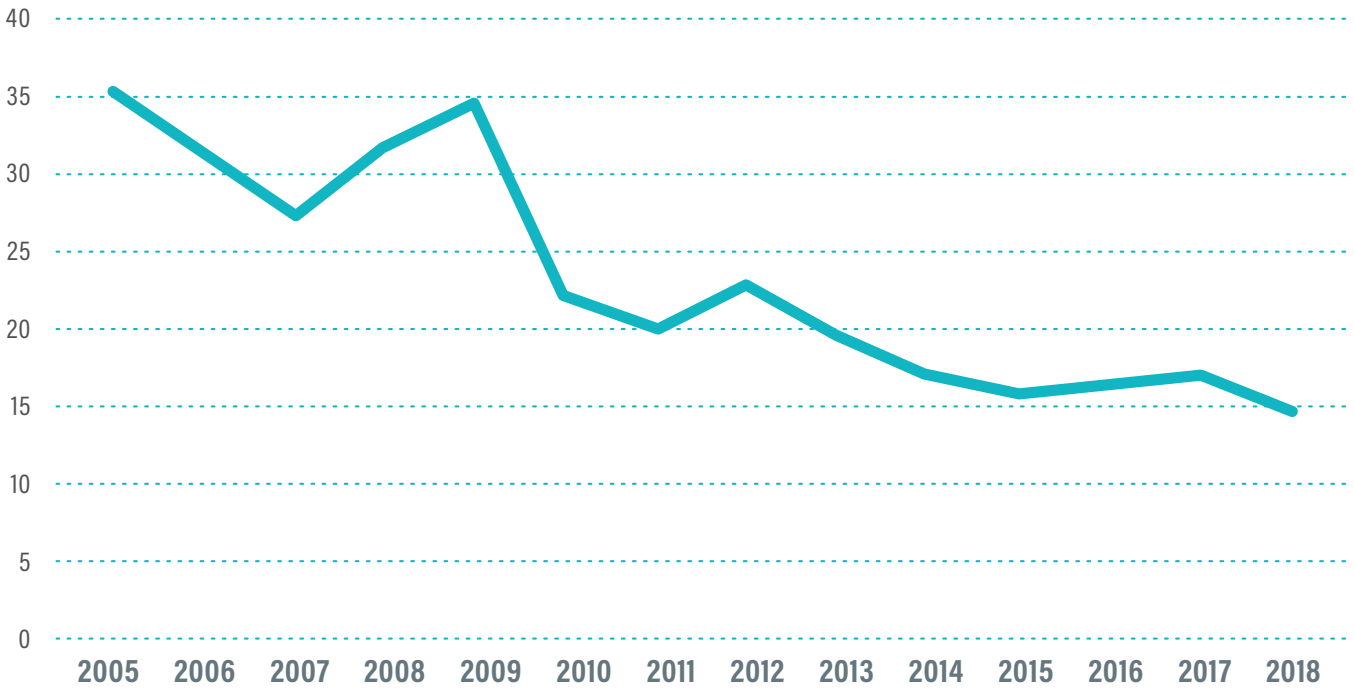
Rate of Identity Theft or Fraud in 2016

All			
Age		**HH Income**	
18-24	6.6%	<25K	6.2%
25-34	10.8%	25-49K	8.0%
35-49	12.4%	50-74K	10.2%
50-64	11.8%	75K+	14.1%
65+	8.5%		

Source: *Bureau of Justice Statistics*

Figure 6

Cost of ID Theft and Fraud (billions of US$)



Source: *Javelin Strategy & Research press releases*

Figure 6 shows that costs of ID theft and fraud (as surveyed by Javelin) have fallen over the past decade. While the $14.8 billion figure for 2018 is, nonetheless, a large figure, it is less than one-tenth of one percent of the nation's GDP.

The potential degree of harm from ID theft and fraud can vary greatly from person to person. For instance, as figures from the 2016 US Department of Justice report show, while most individuals would suffer little or no out of pocket losses, a very small fraction would suffer very large losses.[58] These figures are shown in table 2.

Table 2

Out-of-pocket loss for ID theft victims, 2016

Loss	Share
No Loss (<$1)	88%
$1-$99	5.8%
$100-$499	3.1%
$500-$999	1.3%
$1000+	1.8%

Source: *Bureau of Justice Statistics*

As such, it would appear to be very difficult to treat victims of ID theft and fraud as a monolithic group. While the typical victim experience no out-of-pocket loss, an extreme one-percent could sustain thousands of dollars of out-of-pocket loss.

58. Erika Harrell, Ph.D. and Lynn Langton, Ph.D., *Victims of Identity Theft*, 2016. US Dept. of Justice.

MACRO LEVEL ANALYSIS

As shown in the last two sections, while data breaches are common, by stark contrast, materially impactful ID theft and fraud are not.

Despite a preponderance of both evidence and logic, some seem to assert that breaches always pose great risks to affected persons. Consumer surveys have shown that recipients of a data breach letter reported that they were much more likely to also be victims of identity fraud. While such findings may *seem* commonsensical, it could simply result from ID theft victims being more likely to remember getting a data breach notification letter. A person who did not suffer an ID theft or fraud may not remember a letter they received several months let alone several years previously. As such, determining the relationship between the aggregate level of ID theft and fraud and the overall rate of data breaches (or breached records) with the use of consumer survey data alone would likely prove very problematic. Second, of course, it is not the case that those who did not receive a letter had no data breached, since a breach may not have been discovered or reported.

Since ID theft and fraud can result from non-data breach sources, such as friends and family, phishing, skimming, lost or stolen mail, etc., there may not be a very strong, determinative link between data breaches the rate of ID theft and fraud.

To this point, while Figure 3 shows an increase in data breaches over the past decade and Figure 4 shows wild swings in the number of records reported breached, Figure 5 shows very little change in the rate of ID theft and fraud over the past decade. If one expected a strong positive causal relationship, such as a significant rise in the quantity of reported breached records resulting in a comparably higher rate of ID theft and fraud, then one would expect greater variation in the rate of fraud or theft.

The regression results in Table 3 show estimates of the potential relationships between numbers of breached records, numbers of ID theft/fraud victims, and fraud costs. These results use data available from the Privacy Rights Clearinghouse (PRC) and Identity Theft Resource Center (ITRC) for data breaches and Javelin for ID theft rates and costs.[59] The period covers 2005 to 2018.

59. Data obtained from Identity Theft Resource Center (https://www.idtheftcenter.org/), Privacy Rights Clearinghouse (https://www.privacyrights.org/), and Javelin press releases on their top line findings (https://www.javelinstrategy.com)

Table 3

Regression Results

Dependent Variable	Independent Variable	Estimated Coefficient	Other Independent Variables
# of ID Theft/Fraud Victims	# of Breached Records (ITRC)	*0.00734*	
# of ID Theft/Fraud Victims	# of Breached Records (PRC)	*0.00101**	
# of ID Theft/Fraud Victims	# of Breached Records (PRC)	*0.00048*	*US Adult Population*
Fraud Costs	# of Breached Records (ITRC)	*-16.44973*	
Fraud Costs	# of Breached Records (PRC)	*-2.90026*	

N=14, * indicates Statistical Significance at the 95% confidence level

The simple linear regression estimating the relationship between the total numbers of ID theft/fraud victims underlying the rates in Figure 5 and the total number of breached records in Figure 4, is the first regression shown (using the ITRC data). The relationship here is not statistically significant. The year reported for the breaches with the ITRC data, however, is the year the breach was reported/publicized, not the actual year the breach occurred. The second regression uses the PRC data, which classifies the breach year as the year the breach occurred. In this second regression, the coefficient on number of breached records is statistically significant and equal to about .001. Roughly speaking, this would translate to about 1 ID theft/fraud victim per 1000 breached records, overall. Given the PRC reported around 1.4 billion records breached in 2018, this would be around 1.4 million, or 10% of total ID theft/fraud in 2018.

The third regression adds the US adult population as a control since ID theft may simply be changing due to changes in the population. In this regression, the coefficient on number of breached records is no longer statistically significant. This change in significance may, of course, be due to the limited sample size.

The last two regressions show (negative) relationships that are not statistically significant between fraud costs and the number of records breached. Of course, there is not really a negative relationship between number of breached records and fraud costs. This simply results from the fact that fraud costs have declined, likely due to better fraud detection and security measures (which, of course, are not included in the model).

But this is an important point. Fraud costs really cannot be modeled with just the number of data breaches or breached records. The defense side must also be taken into account. So, while more data and tools may be at the disposal of "bad actors," more data and tools are also at the disposal of the "good actors."

The exercise of producing the regressions was not aimed at precisely modeling a relationship between the variables discussed. The sample is sufficiently small and the data of is of unknown quality and completeness (not all data breaches are known and/or reported, and not all number of records breached are known or reported), making such estimates difficult. If the sample size were larger, other important variables could have been added. Importantly, controls for types of records breached and by types of companies could also be explored (this data is captured by the PRC and ITRC), lags could have been added, but this would have gone down a path of extreme data mining with just 14 years of observations.

Instead, this exercise simply formalized and confirmed what is clear to those looking at Figures 4, 5, and 6. First, there is no positive relationship between fraud costs and numbers of records breached, so there must be other important factors at play not included, such as other sources of tools/data used in fraud and defenses against fraud.

Second, the overall annual number of ID theft/fraud victims is not strongly related to the overall number of breached records. Other important factors likely involved are: the type of records breached, cause of breach (state actor vs. accident vs. criminal hack), the limited number of "bad actors" and fraud opportunities (perhaps), and, again, defenses used to prevent ID theft/fraud. And many ID theft/fraud occurrences may result from non-breach sources (such as stolen wallets).

This simply demonstrates that when there are years of very large data breaches involving hundreds of millions of records, the fact that the incidence of ID theft and fraud is low and only rises or falls very modestly means that individuals involved in data breaches (overall) are not at an especially high risk for ID theft or fraud.[60] Only a very small share of all breached records could possibly translate to annual incidents of ID theft and fraud (perhaps one in a thousand, as suggested by the regression with the statistically significant coefficient on number of records breached).

There is also a lack of evidence that source material for the ID thefts and fraud that do occur is overwhelmingly or even largely driven by data breaches. The University of Texas at Austin Center for Identity's Identity Threat Assessment and Prediction 2018 Report found that 51% of identities are still stolen from non-digital sources. Furthermore, the 48% of digital cases do not solely originate from data breaches – they also encompass other methods such as phishing and skimming.

60. It may also be the case that data on data breaches and records breached are actually higher in reality due to a lack of reporting. But this would likely only strengthen the case that only a small share of individuals involved in a data breach become victims of ID theft and fraud.

A 2009 Javelin report noted, "Despite the hefty blame... placed on the Internet and cyber-crime, online identity theft methods (phishing, hacking and malware) only accounted for 11% of fraud cases in 2008. The truth is, most known cases of fraud occur through traditional methods, when a criminal has direct, physical access to the victim's information. These instances include stolen and lost wallets, checkbooks, or credit cards, or even through the simple act of a criminal surreptitiously eavesdropping into your conversation as you make a purchase. "Friendly theft," reported by 13% of victims, occurs when friends, family or in-home employees take your private data and use it without your permission for their personal gain."[61] The survey of victims who knew how their information was accessed found that only 11% reported that the data came from a business data breach.[62]

Insurance companies that cover losses associated with ID theft and fraud have a clear interest in understanding the sources of the fraud and theft. A 2012 Travelers Insurance[63] study found that, based on its claims data, 73 percent of ID fraud and theft was due to burglary and theft of wallet/purse/personal identification/computer, 10 percent was due to forgery, 2 percent was due to change of address/postal fraud, and 15 percent was due to online or data breach sources.[64] Note that this 15 percent includes data breaches *and* personal online activity by the consumer.

The Ponemon Institute's 2015 report on medical ID theft found that, in 2014, to the knowledge of medical ID Theft victims in a survey, 47% occurred due to the victim sharing personal identification or medical credentials with someone they knew or family member taking their personal identification or medical credentials without their consent. Other reasons included loss of wallet, intercepted mail, phishing attack, and so on. Only 10% agreed that "My health care provider, insurer or other related organization had a data breach."

To summarize, the University of Texas at Austin Center for Identity's Identity Threat Assessment found 48% of ID theft originated from digital cases (phishing, skimming, breaches, etc.), the 2009 Javelin report found 11% of ID theft resulted from all data breaches, the Ponemon Institute's 2015 report on medical ID theft suggested 10% of ID theft originated from a breach, and the 2012 travelers study found 15% of the ID theft was due to a breach or online activity. These are broadly consistent with the previously discussed result from the only regression with a statistically significant coefficient on number of breached records—that around 10% of ID theft could be explained by data breaches (regardless of the fact that controlling for US population changes eliminated the statistical significance of that coefficient).

61. See Javelin Strategy and Research, 2009 Identity Fraud Survey Report: Consumer Version. Accessed on July 20, 2015 at http://www.search.org/files/pdf/IdentityFraudSurveyConsumerReport.pdf

62. *Op. cit.,* Figure 6

63. In the company release Travelers notes that it is "The first insurance carrier to offer identity fraud insurance." Travelers, Company Release, *73% of identity fraud cases resulted from stolen personal items.* November 26, 2012 13:00. Accessed on July 20, 2015 at http://investor.travelers.com/Mobile/file.aspx?IID=4055530&FID=15508447

64. Travelers, Company Release, *73% of identity fraud cases resulted from stolen personal items.* November 26, 2012 13:00. Accessed on July 20, 2015 at http://investor.travelers.com/Mobile/file.aspx?IID=4055530&FID=15508447

Most data security and identity theft experts agree that establishing specific linkages between data breaches (or a particular breach) and specific identity theft/fraud is a challenge.[65] There are a variety of reasons for this, including: lags between the breach and when the stolen data is used to commit ID theft/fraud; the same data being stolen in multiple breaches; breached data being used with data stolen by other means and then used to perpetrate ID theft/fraud; a victim having multiple vulnerabilities each of which could explain how the incident of ID theft/fraud occurred.

Unfortunately, there is an abundance of extreme claims made about the link between data breaches and ID theft—such as "31.7 percent of breach victims experienced identity theft."[66] Taken at face value, this statement implies that nearly one-third of all data breach victims will experience ID theft/fraud as a result of a data breach. Were this true, given the number of data breaches and files breached annually, and assuming data can build up over time, then each year nearly every adult person in the US would be likely to experience ID theft. That the average rate of ID theft/fraud in the US is roughly 5% strongly contradicts this notion.

Perhaps a better approach would be examining the rate of ID theft/fraud within a known breach-affected population. One of the upsides of the recent attention with data breaches is that many particulars about high profile data breaches are made public. This makes possible an analysis of the observed rate of ID theft/fraud among a known breach-affected population. This can provide one measure for extrapolating a likely probability of ID theft/fraud. It is also important to know what information was stolen, and whether the stolen information is necessary and sufficient for enabling different types of ID theft/fraud. Information on whether the hack was likely for financial gain or some other purpose would be helpful, as would data from large breaches that happened some years ago to account for a lag in the use of the stolen data.

A few points are important to consider.

First, only a small number of victims of ID theft/fraud appear to be victimized because of data breaches.

Second, there appears to be no positive relationship between overall fraud losses and the number of data breaches or the number of breached records.

Finally, the fact that the frequency of data breaches is increasing and the volume of breached data can vary from year to year, while the incidence of ID theft/fraud remains relatively constant and fraud losses are

65. For instance, the Director of the FTC's Bureau of Consumer Protection testified before the House Committee's Oversight and Reform's Subcommittee on Economic and Consumer Policy that "generally, the proximate causation of compromised data to any eventual consumer harm can be a difficult thing to show." Testimony of Director Andrew Smith, Bureau of Consumer Protection of the Federal Trade Commission. *Subcommittee on Economic and Consumer Policy Hearing on Improving Data Security at Consumer Reporting Agencies*. March 26, 2019. Accessed at: https://oversight.house.gov/legislation/hearings/subcommittee-on-economic-and-consumer-policy-hearing-on-improving-data-security

66. Keylor, Ben. "What Are Your Odds Of Getting Your Identity Stolen?" 2 January 2018. Downloaded at: https://www.identityforce.com/blog/identity-theft-odds-identity-theft-statistics

trending downward, suggests a more complex relationship than simply x number of breached records results in y number of ID theft victims. One possibility is that there are a limited number of bad actors with limited time, so if only X number of ID thefts are possible then it doesn't matter if 10X or 100X records were breached. It also matters what type of records were breached.

While much of the discourse on data breaches has focused on data as a problem, there are equally valid reasons to look at data as a solution to data breaches and associated risks. Specifically, it is fundamentally true that is that greater data sharing and use of information technology helps reduce ID theft. The investments in data security solutions over the past decade by industries targeted by hackers and that experienced considerable financial fraud losses are bearing fruit. These solutions are enabled by access to third-party data to improve identity verification, identify devices associated with past fraud, and generally reduce the probability of ID theft/fraud. This is important to consider since well-intentioned data protections that inadvertently constrain data access and data used in fraud and ID theft detection might have the perverse unintended consequence of diminishing the effectiveness of these tools.

Data Limitations

One severe limitation of the data stems from state data breach notification laws. Definitions of data breaches are specific to states, and can be expansive or restrictive depending on characteristics of the incident, such as whether the data was simply accessed or acquired by an unauthorized person, whether the data was encrypted, and the likelihood of harm (e.g. a stolen laptop containing sensitive information vs. a lost laptop).[67] There is no one uniform national definition of a data breach – a data breach in New York might not be defined as such if it happened in Florida. There does not appear to be a federal government agency tracking the number of data breaches or records breached nationally.

The organizations that compile the aggregate national data used in this report, Privacy Rights Clearinghouse (PRC) and Identity Theft Resource Center (ITRC), obtain their data from the media and state government agencies that track data breaches. ITRC reports data breaches the year they were discovered (not the year they happened).[68] However, this only partially accounts for the major discrepancy in PRC and ITRC's annual aggregate numbers. We attribute this to a lack of national standard for data breaches. PRC tracks "data breaches and the number of records breached reported through either government agencies or verifiable media sources."[69]

67. *State Data Breach Law Summary*. Baker Hosteler, July 2018. Accessed at: https://www.bakerlaw.com/files/Uploads/Documents/Data%20Breach%20documents/State_Data_Breach_Statute_Form.pdf

68. "Data Breaches." *Identity Theft Resource Center*. 2019. Accessed at: https://www.idtheftcenter.org/data-breaches/

69. "Chronology of Data Breaches: FAQ." *Privacy Rights Clearinghouse*. 2019. Accessed at: https://www.privacyrights.org/chronology-data-breaches-faq

"The ITRC defines a data breach as an incident in which an individual name plus a Social Security number, Driver's License number, medical record or financial record (credit/debit cards included) is potentially put at risk because of exposure... The ITRC will also capture breaches that do not, by the nature of the incident, trigger data breach notification laws. Generally, these breaches consist of the exposure of user names, emails and passwords without involving sensitive personal identifying information."[70]

Obviously, data breaches that go unreported or undiscovered are not reflected in this data. It is also unclear if data breaches that occur in other countries but affect US citizens' data are included.

Data on identity theft was also limited. The Bureau of Justice Statistics publishes an identity theft supplement to the National Crime Victimization Survey every so often, which points to an increase in the general rate of identity theft (6.7% in 2012, 7% in 2014, and 10.7% in 2016). This appears to contradict Javelin's findings, which show a relatively constant rate of identity theft. However, due to a number of methodological changes in the identity theft supplement to the National Crime Victimization Survey, and the infrequency of publishing, there is no trend we can reasonably generalize over time from the Bureau of Justice Statistics.

If the FTC or another agency is charged with enforcing a national data protection law, one of the first steps should be to begin collecting the relevant data at a national level. This is key to making informed decisions and measuring progress.

70. "Data Breaches," *Identity Theft Resource Center.*

MICRO LEVEL ANALYSES

In order to further establish the extent to which reported data breaches are strongly and positively correlated to identity theft and fraud, the previous analysis is rounded off with micro level data—that is to say, data on individuals. One rich source of data on identity theft impacts is credit reports. Credit reports contain abundant direct (extended fraud alert requiring the filing of a police report alleging identity theft) and indirect (other fraud alerts, credit file locks, credit file freezes, increased credit activity) evidence of identity theft as well as metrics on impacts of ID theft and fraud (credit score impacts and material harms). Further, credit reports contain seven years of credit history that could capture evidence of immediate identity theft, and identity theft perpetrated years after any given data breach. Importantly, credit bureaus have direct knowledge about some data breaches, including individual victims of a given breach. This happens whenever a breached entity provides breach victims with free credit monitoring services offered by one or more national credit reporting agencies.

This section utilizes data from a nationwide consumer reporting agency (NCRA, or "credit bureau") on roughly 27 million consumers. The analytic sample is comprised of close to five million consumers who were enrolled in the NCRA's credit-monitoring program as a result of different data breaches. To enable meaningful comparisons, PERC and the NCRA constructed two other files (control or comparison samples). The first control sample is drawn from consumers who came to the study NCRA for a direct-to-consumer credit monitoring service of their own volition. This group is not associated with a specific data breach, but they may be relatively more motivated than the general population, perhaps acting on suspicion and belief that they are at risk of identity theft, or wanting to improve their credit profile. The second control sample is comprised of persons who were offered credit monitoring from partner(s) of the NCRA. Together, these two files make up the remaining approximately 22 million consumers. As such, consumers enrolled in credit monitoring because of a data breach are being compared to consumers enrolled in credit monitoring more generally and who may or may not have been involved in a data breach.

The micro level data will be used to test a series of hypotheses about the relationship between data breaches and consumer impacts. In particular, this study tests the following hypotheses:

1. Breach victims are far more likely to be victims of identity theft/fraud;
2. Breach victims will experience credit score harms as a result of identity theft/fraud;
3. Breach victims are far more likely to have their sensitive data sold on the dark web;
4. Breach victims will be burdened by a constant struggle to protect themselves from ID thieves and must take extra precautions.

The following sections describe the results from testing of the above four hypotheses.

TESTING HYPOTHESIS 1
VICTIMS OF ID THEFT FROM DATA BREACHES

Regarding the first hypotheses—that breach victims are far more likely to subsequently experience ID theft/fraud than the general population (or non-breach victims), we have two primary means of testing this. First, we seek to measure the rate at which fraud alerts of different types, including extended fraud alerts requiring the submission of a police report alleging ID theft/fraud to a national consumer reporting agency, are present across the different samples. If the hypotheses were true, we would expect to see a much higher percentage of those in a breach affected population placing fraud alerts on their credit reports than in control samples of non-breach affected populations. Unfortunately, as this report went to press (July 11, 2019) we did not yet have this data owing to regulatory restrictions on co-mingling Fair Credit Reporting Act (FCRA) regulated data with non-FCRA regulated data. It is hoped that it will be possible to append the fraud alert data to existing samples and provide this analysis at a future date.

Fortunately, the national consumer reporting agency was able to provide data on credit monitoring activity alerts. These are communicated to persons enrolled in credit monitoring whenever something changes in their credit report—a new account is opened, an account is closed, an authorized user is added, a score changes, credit balance changes, etc. If those affected by a data breach do experience identity theft and fraud at a significantly higher rate than non-breach victims, we would expect to see measurable differences in credit monitoring activity alerts—namely, the breach-affected population should have relatively higher credit monitoring activity alert counts than the control samples. The study tests this hypothesis using credit monitoring activity alert counts for the samples.

As noted previously, we obtained the three samples of data from the NCRA described above. One is of a population that received credit monitoring ***because*** of a data breach. These are consumers that are known to be part of a data breach. This file has a sample size of 4,904,677 consumers. We call this the "Breach" file. The second file of a random sample of 8 million of the study NCRA's general direct-to-consumer credit monitoring customers. We call this the "DTC" file. The third file has 14,081,582 consumers who have a credit monitoring account with the study NCRA's partner. We call this the "EPS" file. While the Breach file contains consumers that are believed to be part of a data breach, the EPS and DTC files consist of more general consumers who may or may not have been part of a data breach (these are the control samples). The observations with the Breach file start at the point the consumers begin the credit monitoring following the data breach (going back at most three years) and continues to the present. Not all calculations were possible with all samples due to data limitations (not all services and hence data were available for all three samples).

It should also be noted that the DTC and EPS samples are not perfect controls for the Breach sample. As is clear from the credit scores in table 5, these are groups of consumers with different average credit risk profiles. In addition, there is also selection issues with consumers who seek out credit monitoring (such as in the DTC sample) and those provided credit monitoring (those in the Breach sample). So, some degree of caution should be taken when interpreting the results.

The following table displays the average number of credit monitoring alerts sent to members of the Breach and DTC samples and the average alert rate per day. As can be seen, the average alert rate for the breach sample is lower than that of the DTC sample. While most alerts are likely innocuous, some may alert consumers to potential ID theft.

Table 4

Credit Monitoring Activity by Sample

Sample	Average Number of Credit Monitoring Alerts	Avg. Number of Days	Alerts per Day
Breach	16.8	*311.7*	*0.05*
DTC	45.4	*408.5*	*0.11*

N: DTC=8,000,000; Breach=4,904,677

TESTING HYPOTHESIS 2

CREDIT SCORE IMPACTS

The second hypothesis—the credit scores of breach victims will be negatively impacted as a consequence of the higher rates of identity theft and fraud—can be directly tested by examining credit score distributions over time, average credit score over time, and score changes over time. In this case, data was accessed on breaches that were three years old or less. While this study is in the process of acquiring similar data on one or more large data breaches 5 to 7 years old, a breach that is 1 to 3 years old provides a sufficient lag to test the likelihood of identity theft and fraud, and the impacts of identity theft and fraud on a victim's credit score over time. Results from this analysis are contained in this report, while identical analysis will be conducted on older breaches and affected populations at a later date.

In a January 2019 Philadelphia Fed working paper titled "Financial Consequences of Identity Theft: Evidence from Consumer Credit Bureau Records," the authors explore the credit score impacts on consumers who had likely been victims of identity theft.[71] They identified these likely victims as consumers who placed "extended fraud alerts in their credit bureau files... because this type of fraud alert requires them to file a police report (with accompanying evidence of identity theft and penalties for misrepresenting this information)." Their findings were that negative credit score changes (of about 5 points) and other credit data impacts were detectable in the first one or two quarters after the placement of the fraud alert but after that, the long-term impacts were that their credit scores rose relative to a control group. And while the initial credit score shock was negative, it was modest and smaller relative to the longer-term positive impact. The authors believe this longer-term impact could result from victims of ID theft becoming more aware of their credit reports and credit scores following their ID theft. However, the authors also find that these consumers, despite having a better credit record, reduce their presence in credit markets. This, too, may be a result of being an ID theft victim. These results suggest that the actual initial impact of ID theft might be modest and limited, and has a smaller impact than the positive impact from the consumer's response.

The counter-intuitive data breach analogy to this that we hope to test in future work is whether consumers or some subset of consumers actually lower their overall risk of ID theft if they are alerted to being part of a data breach. That is, consumers may respond to being part of a breach by taking precautions such as changing passwords, monitoring their credit reports, and being more aware of phishing. The impacts of the responses may overwhelm any "initial" impacts directly related to the data breach. This could potentially result in a perverse finding that those who are alerted to being part of a data breach could ultimately be less impacted from ID theft/fraud. This would be an important finding in the context of optimal data breach notification law. However, this remains to be tested.

Table 5

Average Credit Scores by Sample

	Beginning Score	Ending Score	Change in Credit Score	Avg. Number of Months	Change in Credit Score per Month
Breach	771.8	772.6	+0.8	12	+0.07
DTC	634.8	641.2	+6.4	13	+0.49
EPS	695.0	696.1	+1.1	35	+0.03

N: Breach=17,351; DTC=6,072,027; EPS=5,375,349

71. Blascak, Nathan, et al. *Financial Consequences of Identity Theft: Evidence from Consumer Credit Bureau Records*. Federal Reserve Bank of Philadelphia, Research Department. Working Paper 19-02, January 2019. Accessed at: https://philadelphiafed.org/-/media/research-and-data/publications/working-papers/2019/wp19-02.pdf

The average credit score nationwide is about 700 and has been rising by about a point a year during the economic recovery. In table 5 we see that credit scores rose most among those with the lowest scores, in the DTC sample. This may be due to a *regression to the mean* where the below average scores tend to rise. It may also be due to the nature of the DTC sample, likely made up of individuals interested in their credit profiles and interested in improving them. The study NCRA's partner credit monitoring sample appears to be made up of more typical consumers, at least in terms of average credit score. The breach sample consists of consumers with above average credit scores. Yet even in this group, average credit scores also rose. Thus, we find no score declines or large credit score changes among the "breach" sample. This is not unexpected since it is likely that only a very small share of the breach sample (as suggested by the previous section) would suffer from ID theft and potentially have a modest score change as estimated in the Philadelphia Fed paper.

The score changes from the breach sample can be further broken down into the industry of the breached entity (provided by the NCRA). In the following table, we show score change results for consumers involved in breaches from the Financial sector, the Healthcare sector, the Reseller sector, and the Government sector.

Table 6

Impacts of Data Breaches by Industry of Breached Entity

Sample	Breach Industry	Beginning Score	Ending Score	Change in Credit Score	Avg. Number of Months	Change in Credit Score per Month
Breach	Financial	759.7	759.7	0	22	0.00
Breach	Reseller	776.4	777.5	+1.1	12	+0.09
Breach	Healthcare	758.3	757.5	-0.8	6	-0.13
Breach	Government	756.8	757.0	+0.2	4	+0.05

The only negative score change occurs among consumers caught up in a healthcare breach. However, given the relatively small sample size of this subsample, this small score decline is not statistically significant. As such, broken down by industry of breached entity, there is no evidence of average credit score declines among those involved in data breaches.

TESTING HYPOTHESIS 3

DATA ON THE DARK WEB

The third hypothesis tested in this report—that breach victims are relatively more likely to have their sensitive data on the dark web—readily lends itself to empirical testing. Of course, on the surface, this is a seemingly non-controversial statement. However, it warrants testing whether breach-affected persons do in fact have personal identifying information data on the dark web at higher rates. This report compares the leakage of personal identifying information onto the dark web for the three samples examined herein.

Table 7

Consumer Data Found on Dark Web

Sample	Consumer Data Found on Dark Web
Breach	66.4
DTC	70.6

N: Breach=4,632,510; DTC=1,234,023

The NCRA also captures data on whether a consumer's information has been detected on the Dark Web. Table 7 shows this for the subset of consumers with Dark Web monitoring. There is no practical difference seen in the rate of whether Dark Web data was detected from the consumer between the two groups. This may be due to the case that dark web is populated by data from more sources than just data breaches and that most people have likely had their information breached at some point in time.

Research on the Dark Web has produced many useful insights. For instance, prices for stolen or breached data on the black market reveal how valuable stolen data is to potential ID thieves (demand) as well as supply considerations. The National Counterintelligence and Security Center, part of the Office of the Director of National Intelligence, lists prices of such data on its website.[72] In the black market price list, name and password for an online bank account was listed at $1000. This could offer relatively easy access to someone's account, so long as the intrusion is not detected from the bank (the perpetrator may login via a proxy connection). Then "mag-stripe data from a 'secure' premium-level credit card" was listed at $80. Again, this contains useful information for theft.

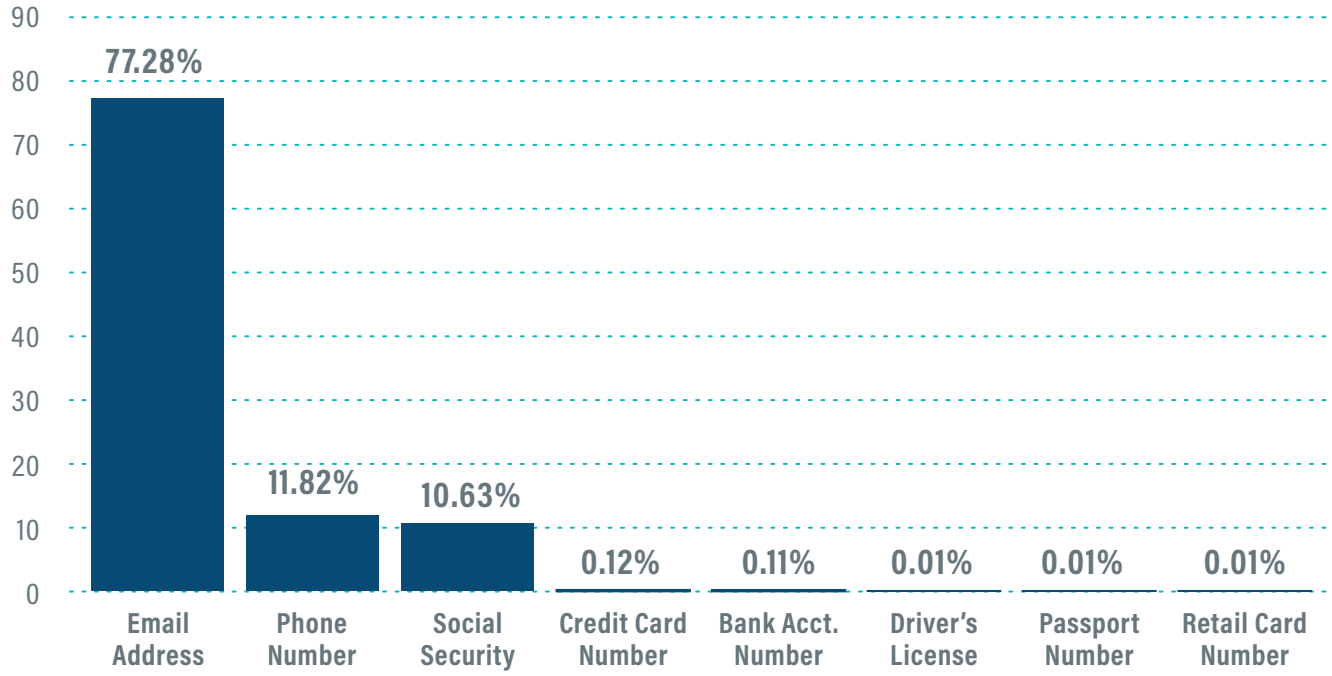
Mother's maiden name is listed at $6. This is much lower since mother's maiden name is useless by itself,

72. Office of the Director of National Intelligence, National Counterintelligence and Security Center, *How Much Do You Cost On the Black Market?* Accessed at http://www.ncsc.gov/issues/cyber/identity_theft.php

but it may be helpful as one ingredient in perpetrating fraud when combined with other PII. Even less than this is Social Security Number, at $3. Again, unlike complete credit card data or bank account data, Social Security Number in itself cannot be used to purchase things or transfer monies. What is surprising is that Social Security Number is listed lower than mother's maiden name, which with a little work can be uncovered by pulling up someone's birth records or announcement, and then using birth mother's name listed to look up her maiden name if she was married from a marriage record or announcement. This may be because mother's maiden name is a more useful identity verifier in helping a perpetrator commit fraud or because the Social Security Number market is flooded with data.[73]

Figure 7

Example of Data Found on Dark Web from the Study NCRA Sample



Sample size is 646,877 consumers, underlying data provided by the Study NCRA

Figure 7 shows that for a sample of consumers from a "Dark Web" monitoring service (separate from the three samples discussed previously), of those consumers with data detected on the dark web, 77% had their email address detected, close to 12% had their phone number detected, and close to 11% had their Social Security number detected. There were relatively few instances of other data detected, such as credit card and debit card numbers. Email address and phone numbers are not considered to be too sensitive, though if that data is combined with other elements, the risk potential rises. Phishing, of course, could start with an email address obtained on the dark web, as could phone scams.

73. The most valuable data packages for would be identify thieves are so-called "fullz," which can contain all the information needed for a quick effective fraud, such as subject's full name, email address, email password, drivers license number, bank name, needed bank or credit card account numbers, Social Security Number, physical address, phone number, and date of birth.

The third most common data was Social Security number. This, of course, is more sensitive than email. The ITRC reports that in 2018 about 48.7% of data breaches involved Social Security Numbers, down from 62% in 2010.[74] Nonetheless, between 2010 and 2018, slightly more than half of data breaches recorded by ITRC involved Social Security Numbers. Again, though, Social Security alone may not be useful without other data elements. But this nonetheless demonstrates that there is indeed sensitive consumer data on the dark web that may have originated from a data breach.

There are a few explanations for the discrepancy between the study NCRA's "Dark Web" analysis having emails being more common than SSNs, and the ITRC's types of data by breach. First, the Dark Web is no doubt populated with data from sources other than breaches (from phishing to schemes designed to capture email addresses), as mentioned previously. Second, the ITRC data on this is in terms of breaches, not number of records breached (the distribution of types of data by size of reported breach may not be uniform).

TESTING HYPOTHESIS 4
METRICS ON ID THEFT CONCERNS

To operationalize the final hypothesis tested in this report—that breach victims are subjected to a grueling and ongoing struggle with identity thieves and must take burdensome actions to protect themselves—this study explores two proxy data points. The first, counts of various forms of fraud alerts placed on credit files by data subjects, will be analyzed at a later date for reasons discussed above. However, the second, credit file lock rates may yield insights into the validity of this hypothesis. If the hypothesis were true, we would expect the data breach-affected sample to exhibit relatively higher rates of locking their credit file as a consequence of being ever vigilant against would be identity thieves. By contrast, if the rates between the breach-affected sample and the control samples were comparable, or if the credit file lock/unlock count is lower for the breach-affected sample than for the control samples, the notion that victims of data breaches are immediately and perpetually unduly burdened would be tested and strained.

Table 8

Credit Lock Rate

	Rate of Credit Locks
Breach	4.58%
DTC	12.73%

N: Breach=4,632,510; DTC=1,234,023

74. See Identity Theft Resource Center, *2018 End of Year Data Breach Report*.

We see that the DTC group is more active with regard to placing credit locks (currently in place). The rate for the breach sample is much lower. This does not suggest a large ID theft concern in this group.

In summary, we have found no evidence, of credit score declines among breached consumers. Nor have we seen other red flags. However, it is important to remember that the absence of evidence is not the same thing as the evidence the absence of a relationship. From earlier findings in this report we saw that the rate of ID theft due to data breaches is likely very small and the rate of material impacts from ID theft is also small. For instance, the January 2019 Philadelphia Fed working paper found very small average credit score impacts following ID theft. And this was on a population who would likely have filed a police report. So, only a small group of those in a data breach would experience ID theft/fraud and then only a small share of this group would likely have a material impact. As such, it is not surprising that no impact is seen in the overall group of consumers in a data breach. The average impacts are likely so small that they are swamped by other factors (including sample composition differences).

SPECIAL REPORT

OVERLOOKED VICTIMS OF HACKERS

The typical storyline associated with a corporate data breach focuses on the data subjects as the victims and the hacker(s) as the culprits. Those breached entities that behave irresponsibly are also cast as culprits.

Although this is sometimes picked up in the news media, in general, those firms who experience a data breach but also employ world class data security measures and are models of good behavior in the aftermath of a breach generally receive little mention as they do not easily fit into a conventional breach narrative.[75] It is, nonetheless important to understand that companies, especially those with world class security, are victims in a breach as well.

It is impossible to prevent every breach. In 2018 there were nearly 4 reported breaches per day in the US, and 441 million records were accessed without authorization.[76] Some estimate that a majority of breaches are unreported. Any organization with data of interest to unlawful actors is constantly at risk. Those seeking to steal data have a decided advantage over those custodians committed to protecting it. A *Bank Info Security* article on the eBay data breach quoted Forrester Research security analyst Tyler Shields saying, "There is an asymmetry of warfare going on where an attacker need only find one hole and defenders have to secure every point of entry."[77] As a result, the fact that the number of records stolen and the incidence of breaches isn't even higher is the true story, and is a testament to improvements in data security measures implemented by all responsible data custodians.

In this study's extension of earlier GAO analysis on data breaches and ID theft/fraud, 12 data breaches resulted in a disclosed settlement amount, with an average of $51.4 million (full discussion below). The average total cost of a data breach, available in 19 cases (taking the lower end of the range when a range was provided), is $207 million. This echoes the GAO's finding in 2007 that many companies incur significant losses due to data breaches.[78] Data breach notification costs make up part of these losses, which could be felt particularly in states with expansive definitions of data breaches (such as unauthorized access with no indication of theft).

Suggesting that breached firms—especially those with rigorous data security measures and policies in place at the time they were breached, and that behaved in an exemplary manner in the aftermath of a breach—should also be considered as breach victims does not belittle the real suffering that some individuals experience as a direct consequence of data breaches. Recognizing that responsible firms are also breached through no fault of their own may act to reward the best actors and promote improved security. Consumers ultimately pay the price when hackers and other data thieves raise the cost to industry of doing business.

75. See Kolbasuk McGee, Marianne. "A New In-Depth Analysis of Anthem Breach." *Bank Info Security*. January 10, 2017. Accessed at: https://www.bankinfosecurity.com/new-in-depth-analysis-anthem-breach-a-9627; see also Bradley, Tony. "Security Experts Weigh In On Massive Data Breach Of 150 Million MyFitnessPal Accounts." *Forbes*. March 30, 2018. Accessed at: https://www.forbes.com/sites/tonybradley/2018/03/30/security-experts-weigh-in-on-massive-data-breach-of-150-million-myfitnesspal-accounts/#27bcda323bba

76. See Identity Theft Resource Center, *2018 End of Year Data Breach Report*. January 28, 2019. Accessed at: https://www.idtheftcenter.org/wp-content/uploads/2019/02/ITRC_2018-End-of-Year-Aftermath_FINAL_V2_combinedWEB.pdf

77. Roman, Jeffrey. "eBay Breach: 145 Million Users Notified." *Bank Info Security*. May 21, 2014. Accessed at: https://www.bankinfosecurity.com/ebay-a-6858

78. GAO, *Data Breaches are Frequent, but Evidence of Resulting Identity Theft is Limited; However, the Full Extent is Unknown*, 32.

CASE STUDIES

AN EXAMINATION OF 24 DATA BREACHES

This report extends the GAO research done in 2007 that found "data breaches are frequent, but evidence of resulting identity theft is limited."[79] The GAO report reviewed 24 large data breaches between 2005 and 2006, and assessed consumer impacts using both primary and secondary research.[80] Although this report also examines 24 large data breaches (see Appendix for the full list), it uses a much longer time frame (2005 to 2018) and is limited to secondary research. The longer timeframe enables selection of a greater variety of data breaches, while the exclusion of primary research limits the details available for impacts analysis to those that are publicly available. Despite the methodological differences, this study's findings are highly consistent with those from the 2007 GAO report: namely, the best available evidence does not support the hypothesis of a strong positive relationship between the frequency of data breaches and the incidence of identity theft/fraud.

Of the 24 data breaches examined in the GAO report, 3 were connected to incidences of fraudulent transactions, and 1 to unauthorized creation of new accounts.[81] Together, these are the two most common forms of identity theft/fraud. A further18 breaches examined had no clear link to identity theft, and 2 lacked sufficient information to come to a conclusion.[82] It should be noted that even among those cases in which some degree of identity theft/fraud was identified following a reported data breach, it is difficult to confirm that the entirety of this activity was attributable to the breach as opposed to other means of perpetrating this crime (e.g. some of the fraudulent activity on Neiman Marcus may have been the result of swiping or privileged access as opposed to the breach ***per se***).

Of the 24 data breaches examined, we were only able to find 4 breaches that resulted in consumer claims, and none of these claim rates come close to the lowest observed "natural" rate of identity theft in the general population of 4.35% (2010) during the 2005-2018 observation period.[83] These four breaches were ChoicePoint (2006, 800 claims out of 163,000, also included in GAO sample) Office of Personnel Management (2015, 61 claims out of 22 million), Neiman Marcus (2013, 9200 claims out of 370,000) and Heartland Payment Systems (2013, 11 claims out of 130 million). One data breach resulted in an unspecified number of fraudulent charges (DSW, included in both reports), and another had claims of unreimbursed identity theft costs submitted but none found valid (TJ Maxx). In 8 data breaches, there was no evidence of any identity theft despite the passage to considerable time, and another 10 lacked sufficient data to make any definitive statements about any fraudulent activities associated with the theft of data.

79. *GAO, Data Breaches are Frequent, but Evidence of Resulting Identity Theft is Limited; However, the Full Extent is Unknown*

80. *Op. Cit.*

81. *Op. Cit.*

82. *Op. Cit.*

83. Javelin Strategy & Research.

One difference between GAO and this study's analysis is that our report is more concerned with quantifiable consumer harms rather than identity theft *per se*, because fraudulent transactions on an existing credit card is the most common form of identity theft and is often 100% reimbursed by the financial institution (88% of identity theft victims incur out-of-pocket losses of $0)[84]. However, if being a data breach victim increased risk of identity theft and identity fraud generally, then the expectation would be rises in the incidence of all types of identity theft and associated fraud losses corresponding to a rise in data breaches (accounting for a lag). Nothing in the data from either the 2007 GAO study, or from the extension of their analysis summarized above, indicates a higher risk of identity theft or material consumer harms resulting from data breaches.

The lack of evidence does not disprove a possible positive correlation between data breaches and identity theft and fraud, but those who argue that there exists a strong casual relationship between these two phenomena are burdened by this and should provide some supporting evidence. Otherwise, the posited relationship relies on an unsubstantiated hypothesis. Data is a necessary input for identity theft, but there is no evidence that it is *sufficient*.

Table 9 on the following page summarizes this analysis. It includes data from three high-profile breaches—ChoicePoint (2005), Heartland Payment Systems (2008), and Neiman Marcus (2013)—all containing the types of sensitive personal identifying information to enable a would-be identity thief to perpetrate the more common forms of ID theft/fraud. All three breaches also have large breach-affected populations, and occurred long enough ago to account for a lag in the use of stolen data. Finally, facts about the hackers' intentions are fairly well-known.

The ChoicePoint data breach is a particularly compelling case. First, it was the original high profile data breach. It was an overnight media sensation owing to the characteristics of the breach—socially engineered, enabled by negligence, involving a sketchy ring of Nigerian fraudsters illegally accessing vast quantities of sensitive personal identifying information for financial gain. This breach checks all the boxes—large sample, necessary and sufficient data for perpetrating ID theft/fraud, seeming intent to use data for financial gain, inadequate security measures and possible wrong-doing for intrigue. Taking the FTC's identity theft victim count of 800 at face value, this would yield an ID theft/fraud rate of one-half of one percent (ID theft rate of 0.005).[85] No specifics of how the FTC was able to link instances of ID theft/fraud to the Choice Point data breach were offered. Further, the FTC has not indicated a single claim filed against $5 million in consumer redress funds ChoicePoint were required to set aside as part of their settlement with the government in the aftermath of this breach.

84. Harrell & Langton, *Victims of Identity Theft*, 2016.

85. "ChoicePoint Settles Data Security Breach Charges; to Pay $10 Million in Civil Penalties, $5 Million for Consumer Redress." *Federal Trade Commission* press release. January 26, 2006. Accessed at: https://www.ftc.gov/news-events/press-releases/2006/01/choicepoint-settles-data-security-breach-charges-pay-10-million

Table 9

Comparative Analysis of Three High-Profile US Data Breaches

	ChoicePoint (2005)	Heartland (2008)	Neiman Marcus (2013)
Affected Population	163,000	130 million	370,000
Type of files breached	Consumer reports with personal information including SSN, credit and employment histories.	Credit card numbers, expiration dates, cardholder names	Names, credit card numbers, magnetic stripe information
Response	Provided free credit monitoring and ID theft insurance to affected persons. Agreed to implement new security measures, and undertake data security audit by independent auditor every two years. Hired independent credentialing and chief privacy officer. Stopped selling sensitive consumer reports in some markets.	Free credit monitoring and reimbursement for identity theft claims. Validated PCI DSS compliance with Visa.	Free credit and identity monitoring. Agreed to implement security measures to prevent similar breaches in future.
Impact on Breached Entity	$10m civil penalties and $5m consumer redress. $15-$20 million in lost revenues. $2 million additional costs.	$102.8 million in settlements (to consumers, Visa/Mastercard/American Express), $139.4 million in total costs.	$1.5 million settlement with 43 states
Impact on Affected Persons	FTC estimated between 750 and 800 persons (ID theft rate of 0.005) suffered ID theft/fraud. No specifics offered. FTC has not indicated a single claim filed against $5 million in consumer redress funds.	290 claims made, Heartland estimates just 11 are valid (ID theft rate of 0.00000008)	At least 9200 credit cards used fraudulently (ID theft rate of 0.025)

Sources: **ChoicePoint:** *PCWorld; FTC; Reuters; CBS News* **Heartland:** *NYT; ComputerWorld, In re: Heartland Payment Systems, Inc.;* **Neiman Marcus:** *Chicago Tribune; NYT; Iowa Department of Justice*

In the case of Heartland Payment Systems, a sophisticated crime ring that was also behind the 2006 TJ Maxx breach (another data breach analyzed by this study in the sample of 24) specifically went after the point-of-sale system and financial data for existing account fraud. 130 million credit cards were stolen, but only 290 claims were made for unreimbursed identity theft costs in the $1 million settlement, of which Heartland estimates 11 are valid.[86] This is an extremely low material impact rate of 0.00008%. It should be noted, however, that some customers may not have pursued a claim.

86. In re: Heartland Payment Systems, Inc. 851 F.Supp.2d 1040 (Southern District of Texas, 2012).

The Neiman Marcus case also involved a sophisticated crime ring that clearly targeted the point-of-sale system and financial data for existing account fraud. 9200 credit cards were used fraudulently,[87] the highest identity theft count of our data breach analysis, also resulting in the highest rate of observed identity theft linked to a data breach (2.5%). This is still much lower than the lowest observed rate of "natural" identity theft in the general population (4.35% in 2010). Furthermore, there is no data on how many consumers were not reimbursed by their financial institutions and suffered out-of-pocket losses from this data breach. Given that this breach involved credit and charge card information, in all likelihood the percentage of the population incurring financial losses from this breach was less than 2.4%.

The ID theft/fraud rate witnessed following the Neiman Marcus breach may also be somewhat high for today, as many ID theft/fraud detection and prevention solutions were not yet available at the time of the breach and for some years thereafter. In 2013, Neiman Marcus was still using magnetic stripes, and businesses were only required to accommodate the more secure EMV chip standard starting October 2015.[88] Even just 6 years ago it was much easier for ID thieves and fraudsters to use stolen information for existing account fraud and new account fraud—two more common forms of identity theft. Today, unauthorized access to the same information is simply insufficient to enable these types of crimes, largely owing to increased spending on data security and access to other databases used to verify a person's true identity.

The case study level of analysis conducted for this report yields two significant findings. They are:

- First, regardless of the size of the reported data breach, and regardless of the different types of data illegally accessed and exfiltrated, the harms experienced by members of the breach-affected populations sufficient to prompt a claim and that may be linked to a specific breach occur at a relatively low rate compared to the observed natural (annual average) rate of identity theft/fraud, and are overwhelmingly of the variety that has minimal financial impact on an individual breach victim (e.g. credit card fraud).

- Second, coverage of reported data breaches follows a general narrative: large breach size (many potential victims); bad actors (hackers, hostile state); unsympathetic breached entity (negligent firm); and additional intrigue (e.g. corrupt executives, poor decision-making). Sometimes, but not often (in just 2 of the 24 cases examined for this report), the script deviates and the breached entity is portrayed as a victim too, but usually only if they were a paragon of good behavior (world class data security measures and policies in place, quickly responded to identified breach, cooperated with government authorities).[89]

87. "Neiman Marcus Reaches $1.5 Million Data Breach Settlement." *Chicago Tribune*. January 9, 2019. Accessed at: https://www.chicagotribune.com/business/ct-biz-neiman-marcus-data-breach-20190109-story.html

88. Harris, Elizabeth, Nicole Perlroth, & Nathaniel Popper. "Neiman Marcus Data Breach Worse Than First Said." *New York Times*. January 23, 2014. Accessed at: https://www.nytimes.com/2014/01/24/business/neiman-marcus-breach-affected-1-1-million-cards.html

89. See Kolbasuk McGee, Marianne. "A New In-Depth Analysis of Anthem Breach." *Bank Info Security*. January 10, 2017. Accessed at: https://www.bankinfosecurity.com/new-in-depth-analysis-anthem-breach-a-9627; see also Bradley, Tony. "Security Experts Weigh In On Massive Data Breach Of 150 Million MyFitnessPal Accounts." *Forbes*. March 30, 2018. Accessed at: https://www.forbes.com/sites/tonybradley/2018/03/30/security-experts-weigh-in-on-massive-data-breach-of-150-million-myfitnesspal-accounts/#27bcda323bba

Taken together, these findings may shed a powerful insight into the persistence of widely-held misperceptions about the nature and magnitude of personal risks associated with reported data breaches. Further, breaches are far more often reported when they are "news," generally speaking, than after the passage of time which would allow for the additional discovery of facts and an assessment of impacts. When members of the general public are presented with coverage of a data breach, it is generally portrayed as something big and scary—something dangerous requiring protection—caused by corporate negligence. Under the circumstances, growing reluctance to share data, diminishing trust in corporate data custodians, and increasing calls for government and industry to do more to protect consumers is understandable.

Data Limitations

The case study was a non-random sample of 24 large data breaches that occurred between 2005 and 2018. It is not a representative sample of data breaches, which limits the generalizations we can make. We were also limited to publicly available information, which was severely lacking when it came to claims of identity theft and consumer impacts. Different types of identity theft exist, such as existing account fraud, new account fraud, and account takeover fraud, but these nuances were rarely covered in many reports. For example, while the rate of identity theft in the Neiman Marcus case was 2.5%, 9200 fraudulently used credit cards out of 370,000 affected credit cards, there was no data about how many of these transactions were not reimbursed by financial institutions. This restricts our analysis of the measurable consumer impacts of data breaches.

Furthermore, many studies make the case for how difficult it is to connect identity theft to data breaches. For example, emails are typically considered "public" information, but often make up a portion of records breached in a data breach. A phishing scam directed to these email addresses that tricks consumers into divulging bank account information would be difficult to link to that data breach.

Reported Data Breaches and the "Jaws Effect"

Much of the public attention on data breaches has led to a "Jaws" effect on the general public, which in turn has affected their views (and the views of many policymakers) on tech, data sharing, and data uses. At this point any reader is likely asking themselves "What is the 'Jaws effect' and how does it related to data breaches?"

While the movie *Jaws* is now 44 years old, this film induced a widespread fear in the US and globally about swimming in the ocean. Shortly after the film's release, the mainstream national media began reporting shark attacks and shark fatalities—something it had not done before.

So affected has humanity become by this fear reaction that nearly half of all persons who have seen the movie *Jaws*, 7 or more years later, report enduring issues with swimming in the ocean. In reality, the odds of being attacked by a shark are roughly 1 in 11.5 million and the odds of dying from a shark attack are 1 in 264 million (see Table 10 below).[90] For some context, a person in the US is more likely die from being hit by lightning (1 in 10 million), falling out of bed (1 in 2 million), or from drowning in a bathtub (1 in 685,000).[91]

Table 10

Comparing Incidents and Probable Outcomes, US 2017

	Data Breaches	%	Shark Attacks	%	Thunder Storms	%	Commercial Flights	%
Micro-units	33 zetabytes		2.4 billion swim/surf		100,000		1.15 trillion seat miles	
Americans	254 million	77.6	75 million	23	327 million	100	815 million passengers	249
Volume of Accidents	14.4 million ID theft victims (Javelin)	4.4	101 attacks	0.000043			5,720 passengers	
Injuries	172,800 Assume: 10% of ID theft from breaches (see Macro Analysis) then 12% of those with costs (DOJ)	0.053	75 bites	0.000001	300	0.0000009	18	0.00000002
Fatalities	NA	NA	1	0.00000001	93	0.0000003	0	0.0

Sources: **Data Breaches:** *Statista; Javelin; ChoicePoint;* **Shark Attacks:** *Statistia.com, Smithsonian.com;* **Thunderstorms:** *National Weather Service;* **Commercial Flights:** *US National Transportation Safety Board, The Telegraph*

Perhaps one of the worst outcomes of the "Jaws effect" has been the massive uptick in the recreational hunting of sharks as well as their commercial fishing. In part, owing to the widespread perception of sharks as dangerous predators, an estimated 100 million sharks are killed by recreational fishers each year.[92]

90. This estimation is quite conservative, as it merely divides the estimated total number of annual ocean swimmers in the US by the reported number of unprovoked shark attacks and associated fatalities. In reality, an individual may swim in a US ocean many times in the course of a year, and may spend many hours in the water surfing, snorkeling, and swimming. If one could measure the number of individual swimming events and person hours, the probability of being attacked by a shark would likely soar into the Powerball range, while fatalities would be rarer than being struck by lightning and winning the lottery on the same day. See: Naylor, Gavin and Tyler Bowling. "Yearly Worldwide Shark Attack Summary," The Florida Museum. Accessed at: https://www.floridamuseum.ufl.edu/shark-attacks/yearly-worldwide-summary/ See also: Schriever, Norm. "Coming Down from Shark Week: Facts Behind the Fear." 6 December 2017. *The Huffington Post*. Accessed at: https://www.huffpost.com/entry/coming-down-from-shark-we_b_3740495

91. *Op. Cit.*

92. *Op. Cit.* (Gavin and Bowling). See also: Rice, Doyle. "Sharks vs. Humans: At 100 million deaths against 6 each year, it's not a fair fight," USA Today. 11 July 2018. Accessed at: https://www.usatoday.com/story/news/2018/07/11/sharks-humans-no-fair-fight/775409002/

Sharks help maintain the health of ocean ecosystems, including seagrass beds and coral reefs. Healthy oceans undoubtedly depend on sharks.[93] This annihilation of sharks is having many unintended consequences, and may be irreversibly harming the world's marine ecosystem.

Extending our analogy, the same holds for technology firms and information services providers (or simply just data flows) in a modern information economy. Rash and ill-conceived attempts to protect consumers from data breaches—despite a preponderance of evidence to the contrary—will result in a disruption to the nation's economic ecosystem. Constraints on responsible data flows, differential treatment of data brokers from network systems, automatic fines on companies experiencing data breaches without evidence of either negligence or consumer harm, policies dictating data security practices, all will have unintended consequences that will result in consumer and economic harms, likely both, without achieving the stated policy objective of increased privacy protections. This is already happening as states enact state-specific privacy laws that are eroding the fabric of our national economic marketplace.

It appears that much public attention focus on select data breaches with any intensity (given over 1,600 data breaches in 2018 alone), usually ones conforming to the conventional narrative involving bad actors (hackers, hostile state), victims (loads of them), a breached entity (usually grossly negligent), and intrigue (delayed notification, insider trading). Repeated use of the conventional narrative has made many people reluctant to share personal identifying information with even highly reputable firms for clear personal benefits.

Data breaches compete with many other news items and therefore it appears that only the biggest, scariest, most corrupt/negligent cases make it into the national news spotlight. Unfortunately, the select news coverage likely feeds into broadly-held misperceptions about risks to individuals associated with reported data breaches.

It is important for lawmakers to cut through the misperceptions when designing data protection legislation. For instance, laws that would punish breached firms regardless of demonstrated consumer harm or negligence would be unreasonable, and would not promote better security practices as much as legislation which did distinguish between responsible and negligent firms.

National privacy law should eliminate a confusing patchwork of state laws, and should be focused on demonstrated and expected consumer harm. All breaches are not equal. A data incident involving hundreds of millions of persons, and little sensitive information is highly unlikely to result in any consumer harm, for instance.

93. Griffen, E., K.L. Miller, and M. Hirschfeld. "Predators and Prey: Why Healthy Oceans Need Sharks." Oceana. July 2008. Accessed at: https://oceana.org/sites/default/files/reports/Predators_as_Prey_FINAL_FINAL1.pdf

CONCLUSIONS AND

RECOMMENDATIONS

Congress must enact sensible federal data privacy legislation eliminates a confusing patchwork of state laws in the near future. Legislation that increases the authority and capacity of the federal government to protect consumers is both desired by a broad coalition of industry and advocates, and is needed to prevent unintended economic damages to the US tech sector and all other sectors that rely on data analytics and information technologies—which is to say every sector of the economy to varying degrees.

Federal legislation should be deliberative, proactive, principles-based, and flexible in the face of a modern and dynamic 21st century information economy. This can be achieved, but only by considering and weighing facts and evidence, and not by reactive fear-driven measures, however well-intended. A case in point is the role played by data breaches in crafting data privacy laws. Inarguably, fear of data breaches is the single most compelling driver of state privacy laws, and has motivated interest in national privacy law among a growing number of US Representatives and Senators.

This study argues that widely-held beliefs about risks associated with reported data breaches are a product of the way they are covered (breaches are covered when they occur and not years later, and only the very large breaches involving bad actors, sensitive data, negligent and/or corrupt companies). This can result in misperceptions.

In other words, today in the US and globally, we are witnessing the "Jaws effect" in relation to data sharing and the tech industry. The "Jaws effect" is a phenomenon whereby public fear of an activity is intensely shaped by socially-reinforced misperceptions. After the 1975 release of the film Jaws, an obsessive fear of sharks among broad swaths of the general public was exposed. Fear sells, and entertainment and other media began perpetuating to this fear, reinforcing misperceptions about the risk associated with swimming in the ocean—namely, that swimmers are highly likely to be attacked or even eaten by deadly sharks. Whereas in fact, a person is far more likely to be killed driving their car or by a vending machine, as the odds of being attacked or killed by a shark are infinitesimal.

The same holds with respect to data breaches. After the State of California passed the first ever data breach notification law requiring organizations to report data breaches to affected persons, the phenomenon of data breaches emerged in national news media.

As with shark attacks evidence shows that the probability of a harmful outcome to a breach-affected individual is miniscule. Consequently, the continued use of the conventional narrative framework around data breaches serves to reinforce widely-held beliefs about risks associated with data breaches, which in turn could contribute to the passage of emotion-driven and reactive data privacy laws instead of evidence-based and proactive ones. Such laws would present a clear danger to the continued competitiveness of the entire American economy without any demonstrable benefits to consumers in terms of data privacy or data security.

Our own research—examining data from three different levels of analysis (macro, micro, and case study) further demonstrates that there exists little direct risk of harms to the vast majority in any given breach-affected population. On the **macro level**, this study ran regressions on years of data breach data, data on ID theft/fraud, and data on fraud losses. The results were compelling. Namely, there is little evidence supporting anything other than a small relationship between data breaches and the incidence of ID theft/fraud or fraud losses. This study's **micro level** analyses on 27 million persons, including over 5 million persons affected by reported data breaches, found no evidence to support the conventional narrative and widely-held beliefs around the widespread risks to consumers from data breaches. Finally, the examination of 24 notable data breach **case studies** drawn from the past 14 years corroborated findings from an earlier GAO study and concluded that it is extremely difficult to link specific data breaches to incidences of ID theft/fraud. Using the broadest possible definition of ID theft/fraud, identity theft was only associated with 4 of the 24 data breaches examined, and even in these cases the highest rate of ID theft/fraud was roughly half of the lowest observed rate or natural rate of ID theft/fraud experienced by the general population over the past decade.

Given this, it is paramount that federal lawmakers move to pass sweeping national data privacy and security laws that balance the aims of consumer protection with the needs of industry so as to preserve our competitiveness in the 21st century.

ADDITIONAL POLICY RECOMMENDATIONS

In order for policymaking to be evidence-based, the evidence must first exist. Data breaches are not issues that are isolated by state. We strongly recommend a national definition of data breaches to allow a federal government agency to begin tracking how much data on Americans is truly breached. We also recommend that the Bureau of Justice Statistics identity theft supplement become a national annual survey on identity theft. This data is needed in order to come to accurate conclusions about the impacts of data breaches on Americans, and inform federal data privacy and data security legislation. We should also note that evidence (and suspicion) points to foreign and hostile nation-state carrying out hacking efforts, as such, understanding this landscape better may also benefit national security.

Finally, we underscore that it is imperative that data security enforcement provisions not be unduly punitive. They should be flexible, proportionate, and harms-based. Such measures should incentivize sufficient investment in appropriate data security protections, and enable the enforcement agency to send a signal if warranted. To accomplish this, penalties and fines must not be automatic whenever a breach occurs. Instead, they should be calibrated to the context and take into account negligence and consumer harms. Unnecessarily draconian measures that put breached companies out of business may also discourage investment in information technologies dampening the competitiveness of the entire US economy without affording any additional consumer protections.

ABOUT
THE AUTHORS

MICHAEL A. TURNER, PH.D.
PRESIDENT, SENIOR SCHOLAR

Dr. Turner currently serves as President of the Political and Economic Research Council (PERC) and was the founder of the Information Policy Institute. After serving as a Graduate Fellow at the Columbia Business School, he was named Executive Director of the Information Services Executive Council (ISEC). He is currently a member of the International Council on Credit Reporting (ICCR) of the International Monetary Fund (IMF) and International Finance Council (IFC) of the World Bank Group. Dr. Turner is the "Sherpa" on credit information sharing for the Asia-Pacific Financial Forum (APFF) of the Asia-Pacific Economic Cooperation (APEC) Business Advisory Council (ABAC), and is a core group member for the APFF.

Dr. Turner regularly testifies before Congress and numerous state legislatures, and presents studies to a host of government agencies—domestic and foreign—on consumer credit, credit risk assessment, data privacy and security, and economic development issues. He has served as expert witness in federal cases involving information policy and consumer credit issues, including one decided by the U.S. Supreme Court. Dr. Turner has served on the Brookings Institution Urban Markets Initiative Advisory Committee, was a policy advisor to the Obama Campaign on urban policy issues, and was appointed by Secretary Tom Ridge to a two-year term on the U.S. Department of Homeland Security's Data Privacy and Integrity Advisory Committee. He received his Ph.D. from Columbia University in Political Economy, and his B.A. in Economics from Miami University. He is a lifetime Fellow with the Ashoka Foundation, was a Yeck Fellow with Harvard Business School and an affiliate scholar at the Alfred P. Sloan Foundation. In his spare time, he serves as an executive partner for two FinTech startups that he hopes may help defray the cost of his daughter's education.

PATRICK D. WALKER, M.A.

DIRECTOR OF RESEARCH

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, microfinance, 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

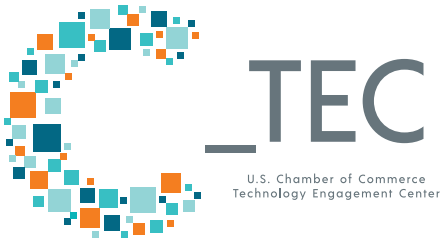
ECONOMIC POLICY RESEARCH MANAGER

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.

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.

TAKE SOME

NOTES