Private Solutions for Public Problems

How startups are harnessing Big Data to tackle social challenges

March 2017

Engine

Table of Contents

Introduction

Our 21st century global economy is powered by data. In the same way machinery enabled the industrial revolution, recent improvements around our ability to capture, store, and analyze massive quantities of data have ushered in an incredible period of innovation. Take the Internet of Things (IoT), for example: with almost 8 billion mobile connected devices (more than the entire human population), we are constantly generating data in a way that drives economic and social value.[1] These types of innovative "Big Data" applications are enabling novel approaches to solving persistent social issues. In industries ranging from education to healthcare to agriculture and beyond, American startups are harnessing data in ways that provide tremendous value to citizens and policymakers alike, addressing challenges ranging from chronic illness to food production.

The term Big Data, while difficult to distill, essentially refers to "the information asset characterized by a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value."[2] This focus on the "high volume" of data at issue often results in heightened anxieties around the mismanagement or misuse of sensitive information. As a result, much of the policy discussion thus far has centered on how to minimize the potential harms of Big Data, rather than how to maximize its potential benefits.

In particular, policymakers have focused on privacy and security-related issues, emphasizing the protection of sensitive, personal information. However, despite the breadth of the Big Data sector, only a very limited subset falls into this category. Much of the data being used by startups are significantly less sensitive, ranging from weather information to agricultural sensor data to food availability. Additionally, the applications of these data vary tremendously and span across a number of industries, making broad, sweeping regulations highly problematic.

While policy conversations around privacy, security, bias, and discrimination are essential, focusing solely on potential negative outcomes ignores the enormous promise of Big Data technologies and the role that government can play in accelerating their growth. This report seeks to highlight the ways in which startups are capitalizing on emerging Big Data technologies to drive innovation and have a positive, public impact. It highlights the diverse uses, outcomes, and challenges of Big Data, while keeping in mind some of the policy questions that remain. It also addresses a number of regulatory roadblocks that are currently limiting data-driven growth across various sectors.

Overall, this report offers insights for policymakers seeking to foster innovation and social transformation while maintaining sufficient protections for the privacy of the American public. Big Data has the potential to completely transform the way we live and work, and both startups and polcymakers will play a key role in unlocking its benefits.

Civil Maps

Of the more than 35,000 traffic-related deaths in 2015, almost 95 percent were caused by human error.[3] For many Americans, the promise of autonomous vehicles rests with their potential to eliminate a large portion of these automobile accidents, and therefore save lives. Proponents of self-driving cars also highlight their ability to decrease congestion and increase vehicle efficiency. However, in order for autonomous vehicles to function effectively, they must be able to understand the world around them. Enter Civil Maps, a San Francisco-based startup that develops cognition for self-driving cars so they can develop and crowdsource dynamic, 3D maps.

Founded in 2013, Civil Maps is creating a new kind of map. While most people tend to think of maps as 2D, Civil Maps is changing that: mapping on-road features like lane markings, as well as static and semi-static off-road features such as traffic lights, signage, construction zones, pedestrians, and delivery vehicles. As the startup's Vice President of Strategy & Business Development Richard Hurlock explains, "When you turn sixteen and get your driver's license, you've had years to learn about and understand the world around you. But what happens when an AI system, which hasn't had the benefit of growing up and contextualizing the world around it, is the one driving the car?" Civil Maps provides these AI systems with context, taking the massive amounts of raw sensory data collected by the vehicle and applying deep learning techniques to turn it into something more meaningful that allows the vehicle to perceive, orient, and respond to the physical world.

"What happens when an AI system, which hasn't had the benefit of growing up and contextualizing the world around it, is the one driving the car?"

In order to create their maps, Civil Maps' system consumes sensor data from as many points as possible, but relies most heavily on LiDAR, a remote sensing method that uses pulsed laser light to measure distances and identify objects. In combination with artificial intelligence and real-time algorithms, the technology is even able to identify signage such as street markers and speed limit signs. To collect this data, Civil Maps is doing something unique: rather than sending out a fleet of cars to map roadways, the startup has plans to crowdsource the mapping data. Among other approaches, it aims to partner with a number of automakers who would

install the mapping software in semi-autonomous vehicles that are already for sale. Under this "crowdsourced" or "swarm" approach, if the owner opts in to exchange and contribute map data, the information from his or her car will be aggregated and sent back to Civil Maps, where it will be reviewed and incorporated into the larger map that the owner can also use. The company's proprietary map compression technology uniquely enables the transmission over 4G networks.

"When integrated with other layers of data held by cities and municipalities, such as building footprint locations, neighborhood boundaries, and street line markings, Civil Maps' technology presents an opportunity to increase the efficiency and effectiveness of local governments."

In addition to the software's applications for autonomous vehicles, Civil Maps also shares portions of its data and maps with cities and municipalities. As Hurlock explains, "Our data is helpful to cities for a number of reasons. Some of the governments we work with use our software to monitor construction projects. Others use it for asset management—things like knowing the exact location of fire hydrants or parking meters can be incredibly complicated and

expensive for cities to manage." When integrated with other layers of data held by cities and municipalities, such as building footprint locations, neighborhood boundaries, and street line markings, Civil Maps' technology presents an opportunity to increase the efficiency and effectiveness of local governments.

As with any autonomous driving software, Civil Maps' system could be targeted for attack. As a result, the startup works with its partners to ensure their security protections are aligned. In terms of policy priorities, Civil Maps focuses most heavily on promoting a legal framework that enables testing and adoption of autonomous vehicles. "The next decade of transportation will be defined by autonomous vehicles," Hurlock argues. "We are hopeful that policymakers recognize the immense potential of this emerging technology to save lives, reduce congestion, and lessen environmental impacts. And we urge them to pursue a legal and regulatory approach that allows for the growth and proliferation of self-driving cars and the technologies that power them." The success of Civil Maps and other promising startups powering this technological revolution depends on it.

BEhelper

There is consensus among a range of scientists, scholars, and public policy professionals that over the coming decades, policymakers will be forced to confront climate change and its impacts on our environment, economy, and society. Reducing greenhouse gas emissions will be a key piece of this, with a particular focus on the building sector, which accounts for almost 40 percent of carbon dioxide emissions in the U.S. and consumes more than 70 percent of the electricity produced domestically.[4] While certain carefully crafted government policies can have an impact on climate change, there is also an important role for innovators in the private sector to help address these challenges.

Bradley Taris is one of those innovators. In late 2015, he founded BEhelper, a software platform that allows architects and building owners to run a variety of tests (including energy analysis, code analysis, and environmental impact studies) with the same precision as a consultant, but much earlier in the design phase—allowing them to build energy efficient, cost-saving design features in from the beginning. Having previously worked as a consultant to builders, Taris noticed a problem: the energy analysis typically was not performed until very late in the design process, when changes were virtually

impossible. At that point, "the client suffers, the architect suffers, and society suffers," he says. He knew that in order to design and build energy efficient structures, considerations had to occur much earlier. Armed with data, designers could uncover patterns, make better-informed decisions, and find new opportunities for reductions in energy usage and costs.

"While certain carefully crafted government policies can have an impact on climate change, there is also an important role for innovators in the private sector to help address these challenges."

BEhelper relies on data from a number of sources. A large portion of the data it uses comes from the government. For example, BEhelper uses thousands of weather files from the National Renewable Energy Laboratories (NREL) covering a range of cities and towns across the U.S. at each hour of the year (8,760 hours total), and including data points ranging from temperature to precipitation, cloud cover, solar irradiance, humidity, and wind. The startup also relies on publicly available building standards data. Arguably though, the most important data BEhelper uses are those that come from its users. These

inputs cover specific building characteristics such as location, type, area, number of floors, window coverage on each side of the building, and more. As Taris notes, "without the users' inputs, BEhelper's outputs would be so generic that they would not be of use to anyone."

Once all of the data points are entered into the system, BEhelper's software creates a detailed building model and runs it through a computationally intensive simulation to see how it performs. BEhelper then delivers these results to the customer in a way that makes them highly actionable, allowing users to make clear, informed decisions about their building. Taris notes that "with buildings, there are many ways to save money while also saving energy. We help to identify these opportunities, communicating at the same time that 'going green can save green.'"

Like many other entrepreneurs working with large quantities of sensitive data, Taris cites privacy as a key concern. BEhelper, which is still in beta, already has a number of security protections in place and plans to build in additional defenses as they scale. Taris notes that cloud services have also provided BEhelper with a number of built-in capabilities and security protections that they would not otherwise have had access to as a small startup.

Finally, Taris sees huge promise in combining Big Data with emerging artificial intelligence technologies to uncover patterns and new opportunities, but cites data ownership issues as a major challenge. He notes that while one must have baseline skills in order to capitalize on AI technologies like machine learning and neural networks, many of the algorithms and tools necessary to take advantage of these technologies are already widely available. Data ownership, on the other hand, is highly commoditized and presents a major roadblock to progress.

"Many of the algorithms and tools necessary to take advantage of [AI] technologies are already widely available. Data ownership, on the other hand, is highly commoditized and presents a major roadblock to progress."

"Depending on the complexity and type of neural network, a quality model could easily require a dataset with over 100 million samples. This typically is not problematic for larger companies. Take Nest, for example: they have millions of houses using their thermostats, each with thousands of temperature and occupancy data points," notes Taris. "However, most startups do not have access to datasets of that size. Open source access to large, publicly available datasets and AI technologies are therefore increasingly important if we want to give startups and innovators an opportunity to create innovative products and services."

HAU.AI

In the U.S., we spend almost $10,000 on healthcare per person per year, by far the most of any country in the world, and yet we have only the 42nd highest life expectancy rate.[5] While study after study has shown that prevention can greatly reduce costs over the long term, we spend only 5 percent on preventive care, whereas treating chronic illnesses accounts for almost 75 percent of spending.[6] HAU.AI, a hyper-local health data portal, was started out of the belief that targeting the social determinants of health (SDOH), namely environment and lifestyle, have the potential to improve health outcomes and reduce costs to a greater extent than simply focusing on treating illnesses.

HAU.AI combines Big Data with artificial intelligence and machine learning to create geographically-focused profiles of community health. It could tell you, for example, how your zip code compares to others in terms of air-quality, water-quality, blood pressure, average commute time, food availability, and other metrics. Levels of health can vary greatly by location, and HAU.AI helps healthcare providers, lawmakers, and other community stakeholders to better focus their efforts and reduce costs, essentially increasing the ROI on health spending.

"HAU.AI gives communities and governments the tools to identify where 'health derailments' are occurring, insight into why, and the ability to prevent or minimize them in the future."

HAU.AI founder Scott Phillips uses a train analogy to describe the way his company thinks about healthcare: "Imagine you have a section of a railroad track that kept causing trains to derail. Rather than investing in additional train maintenance and repair to get the derailed trains running again quickly, wouldn't it make more sense to figure out what was causing the derailment in the first place and fix that? Over time, the American healthcare system has increasingly moved towards putting people back on track after a health disaster, instead of thinking about the conditions that cause illness in the first place and how we might stop disasters before they happen." HAU.AI gives communities and governments the tools to identify where "health derailments" are occurring, insight into why, and the ability to prevent or minimize them in the future.

Unlike many other startups in the health IT space, HAU.AI does not focus on personal health data. Any personal health information they use has been geo-aggregated and de-identified through statistical methods. Instead, they rely heavily on public data from federal, state, and local sources, including the Department of Commerce, Centers for Medicare and Medicaid Services, Department of Housing and Urban Development, and Department of Justice.

As a result of HAU.AI's reliance on government data, Phillips is hopeful that policymakers will increasingly embrace the movement towards open data and transparency. "99.9 percent of the government data that can add value for entrepreneurs is 'boring' data, spread across the government with no real security or privacy risk. Making that data more readily available is a win-win for the government and entrepreneurs," Phillips argues.

"The government should maintain the pipelines through which data can flow, create a level playing field where aspiring entrepreneurs have access to the same opportunities as well-established institutions, and remain open to new opportunities for collaboration."

He also notes that the real cost with data is the cost of collecting it, which in most cases is already being done: "There is very little cost in sharing it, and the act of sharing it better justifies the cost of collecting it." In addition to pursuing open data policies, Phillips believes that the government should maintain the pipelines through which data can flow, create a level playing field where aspiring entrepreneurs have access to the same opportunities as well-established institutions, and remain open to new opportunities for collaboration with innovative companies working to solve various societal challenges.

Overall, Phillips believes that the pursuit of policies that facilitate data-driven innovation is an issue of American competitiveness. He argues, "the next technological revolution will involve the disruption of existing legacy industries. Big Data will be key to that disruption, and if the policy environment in America is hostile to innovators in this space, then the innovation will move to more data-friendly countries and America will be end up with the short end of the stick—importing these new technologies instead of inventing them."

The Crucial Intersection of Data and Technology in Healthcare: The Promise of Precision Medicine

Historically, our healthcare system has viewed disease as a series of symptoms. In recent years though, researchers and doctors have transitioned towards a more inclusive view of medicine that takes into account a person's environment, lifestyle, and genetics when diagnosing and treating patients. This new approach, called precision medicine, is more accurately able to predict how particular groups of people will respond to prevention and treatment strategies. As traditional "one-size-fits-all" medical approaches are often inadequate for those with specific genetic makeups or lifestyles, precision medicine helps doctors prescribe the best drugs and correct dosages with the fewest side-effects.

Not surprisingly, the promise of precision medicine has garnered significant interest from both the public and private sector. In 2015, then-President Barack Obama announced his Precision Medicine Initiative, allocating $216 million in funding for the program to be carried out at the National Institutes of Health (NIH), National Cancer Institute, and the Food and Drug Administration (FDA).[7] In the private sector, pharmaceutical companies performing precision medicine research and startups seeking ways to innovate with this new capability have generated a high level of investment.

While the practice of precision medicine holds tremendous promise, it cannot run without its chief source of fuel: Big Data. In fact, its success is dependent upon complete access to all of the data available on an individual.[8] Through new technologies such as fitness trackers and pedometers, as well as improvements in electronic medical records and efforts to better incorporate genomic information, we are well on our way to producing this kind of data, but the ability to tap into it remains uncertain.

Another roadblock to broadly realizing precision medicine is the high number of data breaches experienced by health plans and health systems in recent years. One proposed solution is to secure medical records through the use of a blockchain—a peer-to-peer, cryptographically secure database system popularized through the Bitcoin virtual currency that can be used to safely document a wide range of information, including health records. In a recently announced partnership, IBM Watson Health and the Food and Drug Administration are working together to implement blockchain technology in oncology research to assist in the discovery of new drugs and improve outcomes for specific patient populations.[9]

EnsoData

In the U.S., the American Academy of Sleep Medicine estimates that more than 29.4 million people suffer from sleep apnea.[10] However, despite being one of the most common sleep disorders, researchers estimate that more than 80 percent of individuals affected by sleep apnea go undiagnosed.[11] Left untreated, sleep apnea can lead to severe complications like stroke, cardiovascular disease, diabetes, and depression, and is estimated to double a patient's medical expenses over their lifetime. While there have been efforts to increase awareness of sleep apnea, the chief obstacle to progress is the capacity of sleep clinics, many of which currently operate close to 100 percent utilization, to test and treat those affected by the disorder.

EnsoData, co-founded by Chris Fernandez and Sam Rusk in 2015, is working to address this challenge. The startup's first software application, EnsoSleep, automates sleep data analysis, creating massive cost and time savings for frontline clinicians and allowing them to spend more time treating patients and less time manually sorting through massive amounts of sleep data. "We've gotten to the point where the current sleep clinic workforce physically cannot manage the population that needs care," explains Fernandez. "EnsoSleep helps to mitigate this, harnessing data to make healthcare more accurate, efficient, and affordable."

During a sleep study, real-time vital monitoring devices measure and record various indicators, including brain activity, eye and chin movements, heart rate and rhythm, respiration, leg movement, and more. This means that each sleep study typically generates a gigabyte of data. Once that data is collected, a sleep technician will comb through it and flag any instances where there is an abnormality or disruption—known as "sleep study scoring." Recognizing that this tedious analysis dominates a huge portion of a clinician's valuable time, EnsoData developed a software that automates this portion of the process: rather than going through trillions of bytes of data, clinicians can simply import the data into EnsoSleep's secure web application, which then applies advanced algorithms based on Big Data to identify patterns and abnormalities in seconds. "Our technology saves clinicians hundreds of hours per month by automating one of the most arduous and repetitive data annotation tasks in all of healthcare," Fernandez notes. "EnsoSleep allows clinics to score data 100 times faster for less than half of the cost, with the additional benefit of improved accuracy."

Since the company is dealing with sensitive, personally identifiable information, privacy and security are a top priority. The company hosts its application on a HIPAA compliant platform, and uses numerous technical security measures including full at-rest and in-transit encryption, access restrictions, secure and expiring authentication mechanisms, intrusion detection systems, and more.

Looking ahead, Fernandez believes that advances in Big Data and artificial intelligence technologies will play a transformative role in the way we deliver healthcare. While anxieties around "labor displacement" and "mass unemployment" often dominate conversations about these emerging technologies, Fernandez insists that policymakers should not fear the automation that Big Data and artificial intelligence enable. "EnsoSleep doesn't replace the need for clinicians," he notes. "At the moment, we've seen that the ratio of doctors to patients in this space is around one to 50,000. Our technology simply allows these resource-constrained clinicians to reduce the time they spend analyzing data and invest it in opening up capacity and providing treatment and care to patients instead."

Finally, Fernandez argues that policymakers have a critical role to play in advancing and encouraging innovation in this space. "Significant public investment in research,

as well as policies that support open data and improve the ease and liquidity of data collection and sharing, will be needed to pave the way for emerging artificial intelligence technologies and unlock the true potential of health data," he notes. "The Precision Medicine Initiative and the report on 'Preparing for the Future of Artificial Intelligence' were two important initiatives championed by the previous Administration. I'm hopeful that policymakers can stay ahead of the curve and embrace data-enabled innovation." As for EnsoData, the startup will continue to do what it does best: develop, refine, and deliver innovative solutions to some of healthcare's biggest challenges.

"Significant public investment in research, as well as policies that support open data and improve the ease and liquidity of data collection and sharing, will be needed to pave the way for emerging artificial intelligence technologies and unlock the true potential of health data."

Polco

A successful democracy requires consistent dialogue between citizens and their representatives. However, if the country learned one lesson from last year's historic election, it was that our system is currently failing both groups—many citizens are frustrated with their interactions with government, and policymakers lack a tool to better understand and engage with their constituents. In recent years, a number of "civictech" and "govtech" companies have emerged to address these challenges. One of these startups is Polco, which has developed a solution to improve the experience for citizens and local governments at the same time.

Co-founded by military veteran technologists in 2015, Polco's web-based platform offers polling and analytics to both better inform elected officials and increase civic engagement. "Modes of communication are changing and very few people answer their landlines or want to take the time to attend a town hall meeting anymore," explains Nick Mastronardi, Polco's Founder and CEO. "Right now, a city might send a postcard survey to 10,000 constituents and only get back ten responses. That's just not efficient. Citizens are looking for new tools and technologies to connect with their elected leaders and electeds are doing the same for their

constituents."

In addition to being ineffective, efforts to understand constituent sentiment and opinion have historically been forced to make tradeoffs between the breadth of responses and the statistical rigor. Online flash polls can quickly cast a wide net, but are very noisy in that they may allow repeat or automated voting and so are in need of major calibration. Analog polls are rigorous, but are slow and expensive, and only cast a small net that often missing key demographics.

Polco seeks to combine the best of both worlds by creating online verified polling. Polco's website and apps allow government leaders to pose questions and polls to constituents and receive insightful, reliable analytics to help shape policy decisions. Polco's policy polls are designed to be shared across multiple communication channels: in email, on Polco, on other relevant websites, or through social media. "By connecting with a broader audience, Polco helps to ensure that input is balanced and not guided by the same four people who consistently show up at meetings or hearings," Mastronardi notes. "Many local governments run operating budgets on the order of tens or hundreds of millions. There's no reason they should not enjoy

access to the equivalent of the market research analyses that are driving growth and decision making in the private sector."

The company only requires users to provide their name, email, and zip code during registration. This information is cross referenced against public voter registration files to ensure citizens do not respond to a poll more than once. Over time, users can provide additional, optional information, such as age or gender, to build a richer profile and get better tailored notifications and analyses. Polco also integrates open government data, including those from the US census, to overlay additional data points, such as average income level or housing price for a specific geographic area. The company believes these additional layers help decision makers better understand and forecast the impacts of their decisions, and help citizens discover how proposals will impact them individually. Finally, the startup uses social media data to build a deeper understanding of political social dynamics.

While Polco is collecting rich data sets, the company's privacy policy explicitly states that they do not share individual-level data with governments or third parties. Instead, government officials only get access to the aggregated and de-identified results. Mastronardi explains that many governments and citizens alike favor this approach: "Most city governments are not interested in individual-level data because they don't need it, it makes things more complicated for them if they have it, and

they understand the value of a trusted intermediary. Without a guarantee of security and privacy, we wouldn't be able to foster as healthy levels of participation."

While Polco is still relatively small (the startup is working with only a few local governments in four different states) and the data it uses may not be considered "big" by some comparisons, the platform is growing and is already seeing immense opportunities in increased data availability and collection. For example, they hope to one day capitalize on social media data be able to better trace the memetics of political preference—who has influence in the policy discussion, how they receive and transmit information, and why their narrative resonates (or fails to resonate) with specific groups.

"We want a useful, easy, secure, and private experience for citizens and government officials alike. With more data, computers are able to do more granular and accurate calculations," Mastronardi explains. "This is true across industries, and in politics this means the potential to produce individually tailored policy impact forecasts, as well as much more accurate descriptions and predictions of citizens' political sentiment formation." Polco's aim is to leverage new data from online polling-driven civic engagement, political social behavior, and existing open data sets to better educate communities on the impacts and outcomes of their initiatives, promote better-informed civic decisions, and thereby generate more effective government policies.

FarmLogs

Growing up on his family's fifth-generation farm in Caro, MI, Jesse Vollmar saw firsthand how the pace of innovation in agriculture was not keeping up with that of technology. Most of the farmers in his community relied on outdated, clunky software, if they used farm management technologies at all. But he also recognized that the farming industry needed technological advances more than ever. Farmers are faced with the challenge of increasing food production by nearly 50% by the year 2050.[12] And with the supply of farmland decreasing, farmers must find ways to make existing farmland more productive.

In an attempt to help farmers address this challenge, Vollmar founded FarmLogs in 2012. The startup uses data collection and analytics to help farmers grow row crops more efficiently and profitably, and is used by over 100,000 farms across the United States. "We've seen progress thanks to innovations in equipment, seed management, and genetics, but the next agricultural revolution will be tied to data science and precision agriculture," says Vollmar. "There is enormous value to be had in data-driven farming. Now more than ever, farmers must rely less on instinct and more on data to make informed decisions on the farm."

FarmLogs helps row crop farmers dramatically increase their crop yields while using fewer inputs and less energy. The startup does this by using software, data science, and machine learning technologies to capture, collect, and analyze massive amounts of data and transform it all into simplified, actionable field-level intelligence so that farmers can make better, data-driven decisions.

"We've seen progress thanks to innovations in equipment, seed management, and genetics, but the next agricultural revolution will be tied to data science and precision agriculture. There is enormous value to be had in data-driven farming."

The company leverages data from a variety of sources. One major input is open government data, such as precipitation data, crop history, and soil maps. Vollmar notes that the U.S. Department of Agriculture and their National Agriculture Statistics Service do a good job of making data available in a format that is accessible and usable. FarmLogs also relies on datasets from the National Oceanic and Atmospheric Administration and the National Weather Service. In addition to

government data, the startup uses private third party data like satellite imagery; physical sensor data from the field; and user data from the farmers themselves, such as yield information and seed, chemical, and fertilizer use measurements. But FarmLogs doesn't simply provide growers with this complex and often confusing data to process on their own. They also deliver the insights and recommendations farmers need to make smarter, more precise decisions to grow row crops more efficiently.

Beyond improving the productivity of farmland, FarmLogs' solution also helps to reduce the environmental damage caused by agricultural activities. For example, fertilizer run-off is a major challenge in many states. FarmLogs gives farmers enhanced insights into their fields, which in turn allows them to better control inputs and adopt new practices and tools that help them to mitigate environmental impact. "It's great, because the incentives are aligned here," says Vollmar. "We help to make it cheaper to produce good food and allow farmers to more easily keep up with increasing demands on food production, while also reducing harmful impacts on the environment."

Finally, Farmlogs is commited to its users' privacy and security, and is unique in its independence. "We don't sell seeds, inputs, or equipment, and we've built this software

for the growers, not commodity analysts or anyone else. This allows us to make unbiased recommendations. And we never share private user data," says Vollmar.

"Big data has the potential to entirely transform agriculture. Policymakers should embrace this progress and maintain a light-touch regulatory framework that allows all farmers to easily, securely, and inexpensively reap the benefits of these technological advances."

It's clear that in order to keep up with population growth and meet the growing global demand for food, farmers will need to turn increasingly to technology and data analytics. Vollmar cautions policymakers against pursuing reactive regulations that would limit the ability of third-party agtech companies like FarmLogs to provide their valuable services to farmers. "Regulating farm data more strictly would be devastating to technological progress and ultimately our farm economy. Big Data has the potential to entirely transform agriculture. Policymakers should embrace this progress and maintain a light-touch regulatory framework that allows all farmers to easily, securely, and inexpensively reap the benefits of these technological advances."

Propel

The food stamp program was created more than 50 years ago in an effort to help eligible low-income households more easily access nutritious foods. Today, the program—officially known as the Supplemental Nutrition Assistance Program, or 'SNAP'—represents a critical component of the U.S. social safety net, serving over 45 million enrollees (or roughly one in seven Americans) who spend about $70 billion a year through the program.[13] Yet, with so much money and so many livelihoods at stake, it is incredibly difficult for enrollees to simply check their balance in real-time. A majority of users must call a phone line or save receipts from previous purchases, meaning that despite recent technological advances, more than $70 billion in government funds and potential grocery industry income are tracked by paper receipts.

In an effort to bring the food stamp program into the 21st century, Jimmy Chen founded Propel and created its flagship product, Fresh EBT: a smartphone application that helps food stamp recipients more effectively manage their benefits. "I had been working at a large company in Silicon Valley and started to wonder—what if we could apply some the tech sector's proven best practices to modernize the way we administer benefits in this country," Chen says.

The app initially started as a way to streamline enrollment for the program. Today, Fresh EBT has grown into a free smartphone app that allows users to easily view their balance, transaction history, and nearby retailers that accept SNAP benefits. As of the publication of this report, the app has over 500,000 downloads and serves more than 200,000 users per week in all 50 states, plus four territories. Every day, it collectively saves users about three months of human time as a result of using the app instead of checking their balance over the phone. Beyond simply saving users time, Fresh EBT is also a distribution platform for products that make food stamp recipients more physically and financially healthy, helping users discover healthy recipes and save money on groceries, among other things.

Fresh EBT draws data from a number of sources to make this possible. To gather baseline data on enrollees and balances, the startup helps users access their data from EBT processors (the third party entities responsible for processing SNAP transactions). Data is also pulled from several open sources, including USDA's list of authorized EBT retailers, as well as third party providers that provide data on local stores, maps of food pantries, and other relevant information. Since Fresh EBT processes sensitive information (such as an individual's enrollment status and

balance), Propel has strong protections in place to guard user data. For example, instead of saving user information to an external server, the app saves any personally identifiable data directly to the hard drive of the user's smartphone. Further, all communications are fully encrypted.

Looking forward, Chen sees enormous opportunity for Fresh EBT to provide the government with important insights into the food stamp program and its administration. Since SNAP is administered by states, it has previously been difficult to aggregate usage data to inform policymaking. However, Fresh EBT is emerging as a potential aggregator of this data across states. As the app's user base increases, the amount of transactional data it receives also increases. "What wasn't interesting previously is now very compelling," Chen argues. Over time, as Propel continues to collect de-identified usage information, the company will be able to share that data with policymakers, providing unique and valuable insights into ways to improve the administration of the program.

Chen also believes that it is critically important that consumers maintain the right to view their own data in whatever form they would like and share that data with whomever they would like. "As a vehicle for users to view their EBT balance and transactions, we're an intermediary similar to a web browser, in that we format

information and provide a valuable tool for the user but we don't store personal credentials," he explains. "However, we rely on access to a consumers' data in order to provide them value. Any efforts to undermine or limit the ability for a consumer to share their data with third-parties like FreshEBT would have a negative impact on our business and disempower our users." Chen highlights the Consumer Financial Protection Bureau's call for financial institutions to offer consumers continuous, unfettered access to their data as a positive step, and notes that overall there has been a shift towards more openness since this key protection was codified in Section 1033 of the Dodd-Frank Wall Street Reform and Consumer Protection Act.

Overall, Chen sees his company as playing an essential role in helping food stamp recipients better manage and take advantage of their benefits, while also improving the effectiveness of government social safety net programs. "Simply making data or information public does not necessarily mean it will have its full impact," Chen notes. "But distributing information to the people who need it most, at the right time, in a consumable format is a critical component of helping open data to achieve its promise." Chen is hopeful that Fresh EBT can help bridge this gap, bringing solutions to those Americans who are often overlooked by traditional technological innovation and allowing the food stamp program to better serve them.

Understory

Each year, U.S. economic output fluctuates by as much as $485 billion as a direct result of weather variability.[14] Yet, weather insights gathered from traditional radar and satellite technologies remain inadequate for predicting what is actually happening on the ground, since they only analyze conditions observed in the atmosphere. Recognizing this, in 2012 Alex Kubicek founded Understory—named for the area in a rainforest beneath the forest canopy. The company fills an important gap in weather forecasting, collecting hyper-local, on-the-ground data from a network of sensors that allows companies and communities to better anticipate and respond to weather impacts.

Understory's sensors have been deployed in five communities across the U.S.: Kansas City, St. Louis, Dallas, Houston, and Denver. The sensors mount to existing infrastructure and detect rain, wind, hail, temperature, pressure, and other weather data points. The company then combines that information with data from third-party and government sources to paint a more accurate picture of what is happening directly at the earth's surface, where the risk to life and property is greatest. As Kubicek puts it, "we've created a dataset that has a better resolution and fidelity than anything else out there. By providing this data and graphical views of the movement and intensity of weather events to customers and communities, we're helping them to have better insight and detect risks earlier in the process. With this improved data and accuracy comes better investments and fewer losses."

"By providing this data and graphical views of the movement and intensity of weather events to customers and communities, we're helping them to have better insight and detect risks earlier in the process. With this improved data and accuracy comes better investments and fewer losses."

Understory works primarily with businesses that are most impacted by fluctuations in weather. For example, the startup sells its services to insurance companies, helping them to better understand which of their customers may have been impacted by a weather event and allowing them to proactively reach out to policyholders to ensure they are safe and develop a better relationship. Understory also works with farmers to detect weather conditions in their fields and helps with seed production and operations management of the field. Eventually, the company is interested in providing weather forecasting for

consumers.

In addition to selling its products and services to businesses, Understory helps improve public safety by providing its data free of charge to communities and academics. Understory is also dedicated to creating opportunities for classrooms across the country to study and learn about how weather impacts lives. Their "Weather Is Cool" program provides free weather stations to schools, which allows students to harvest the data generated from the weather station and learn about their local climate.

Much like other companies that handle large volumes of data, Understory protects its information with varying levels of security, depending on its sensitivity. For example, their exclusive hail data has much stronger safeguards. Customer information like emails, claims data, and site locations are also treated with a higher level of care. They use various measures, including strong keys, hashes, and firewall protections. Additionally, data are always transmitted via an encrypted connection.

In terms of policy, Kubicek appreciates the importance of not only opening up access to government data, but also ensuring that the datasets are in a format that allows for easy and efficient reuse by companies. This means ensuring that the information is released in an open, machine-readable format that is structured appropriately. "We rely mostly on the data generated by our

sensors, but we also capitalize on government data from agencies like the National Oceanic and Atmospheric Administration (NOAA)," Kubicek says. "We have skilled engineers on staff, and even they sometimes have a difficult time accessing and making use of publicly-available government data." He notes that Amazon's partnership with NOAA has greatly improved the usability of that data, and would love to see similar efforts from other agencies.

"For years, many industries relied on data that didn't provide a full picture of how weather was having an impact on the ground. While that data may have seemed good enough at the time, new technologies and insights can provide vast improvements."

Overall, Kubicek attributes Understory's success thus far to the ways in which the company is transforming the worldwide weather infrastructure and how weather data are collected. "For years, many industries relied on data that didn't provide a full picture of how weather was having an impact on the ground. While that data may have seemed good enough at the time, new technologies and insights can provide vast improvements." In the case of Understory, the company is redefining what it means to forecast the weather accurately, helping to improve lives, avoid disaster, and save money.

Sense

The Internet of Things (IoT) has exploded in recent years, giving consumers and businesses new ways to track various metrics in real time: activity levels, caloric intake, parking availability, structural integrity, and more. IoT has also unlocked massive potential for connected, or "smart," homes; enabling new ways for homeowners to track data that give them a better picture of what's happening inside of their houses. Enter Sense Labs, a startup that has developed an energy monitor that provides consumers with new insights about energy use and efficiency in their homes.

Sense Labs was founded in 2013 by Mike Phillips, Christopher Micali, and Ryan Houlette, three Cambridge-based entrepreneurs who had previously collaborated on Vlingo, a startup that pioneered speech recognition on Samsung and Apple phones and was later acquired by Nuance. With a proven track record working with Big Data technologies, Phillips, Micali, and Houlette decided to make their mark on the energy efficiency space. Sense Labs was born from the idea that people should be able to track and derive insights from what's happening inside of their homes, and in turn make their homes more energy efficient. The company has created a small piece of hardware that is installed in a home's electric panel. Then, Sense Labs applies machine learning algorithms to detect devices that are in the home (regardless of whether those devices are "smart"), what time they are on, and how much energy they use. In the future, Sense Labs hopes this information could become more granular in a way that would allow customers to compare energy usage (for example, notifying a user that their refrigerator uses more energy than the average home).

"Sense Labs was born from the idea that people should be able to track and derive insights from what's happening inside of their homes, and in turn make their homes more energy efficient."

Because of the personal nature of this information, privacy is a chief focus at Sense Labs. Sense's privacy policy explicitly states that they only share data when a homeowner has given them permission to do so, or if the information has anonymized. Additionally, the company has done extensive work on data security throughout their system, including annual audits by external consultants. Phillips says that complete transparency with customers about how their data are being used and

protected is critical to instilling trust in a company and a technology. "It is so important to stay on the right side of consumer privacy and security. From this confidence and understanding, innovation can flourish."

From a policy standpoint, Phillips believes that the best way for the government to protect consumers and their data is to ensure that companies are clear about telling customers what data are being collected, with whom they are shared, and how they are used in language that is simple and easy to understand. He also recommends a mindful approach to regulation. Maintaining consumer privacy is vital to the proliferation of any technology, he says, but for those that rely on data, it is critical that is done in a way that does not lose the potential benefits of the information. "The desired tradeoff between privacy and benefits will vary by consumer," he explained, "so lawmakers should focus policies on transparency and choice, rather than the institution of blanket regulations."

Conclusions

As policymakers wrestle with the issues raised by emerging Big Data applications, it is essential that they keep in mind the tremendous potential of these new technologies and abandon the misconception that Big Data simply enables large companies to use private consumer data solely for financial gain. As the each of the startups profiled in this report illustrate, there are numerous small companies from a diverse range of industries capitalizing on large data sets to drive innovation and develop targeted solutions for some of society's greatest challenges. Regulations and policies that do not account for this will stifle growth and delay important progress.

Additionally, the sources of data that are critical for Big Data applications are not limited to consumer information or privately gathered user data. Much of the most valuable information that startups can use to innovate is held by the government. FarmLogs relies on agricultural data from the USDA; Understory capitalizes on NOAA's weather data; HAU.AI uses community data from CMS, HUD, and DOJ. Policymakers should strive to make even more of the data held by the government available to the public in an open, searchable, and machine-readable format.

Big Data applications and the ability to store, collect, and process massive quantities of data in real-time have enabled a new wave of startups that are spurring economic growth and transforming society. While the government may not be the driving engine behind this innovation, it has an essential role to play in enabling transformative technologies and accelerating their development. While it may be easy to get caught up focusing on risk, it is essential that policymakers also explore and consider the myriad societal benefits that Big Data technologies can enable. Then, with these opportunities in mind, they can explore how to appropriately address risks while facilitating data-driven innovation and the benefits it brings society.

Resources

1. Annabelle Gawer, "Big Data: Bringing Competition Policy to the Digital Era," Organization for Economic Cooperation and Development, December 16, 2016,

2. A. De Mauro, M. Greco, and M. Grimaldi, "A Formal Definition of Big Data Based on its Essential Features", Library Review, Vol. 65., No. 3, pp. 122-135, 2016, http://www.emeraldinsight.com/doi/pdfplus/10.1108/LR-06-2015-0061.

3. "2015 Motor Vehicle Crashes: Overview," National Highway Traffic Safety Administration, August 2016, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812318

4. "Buildings Energy Data Book," U.S. Department of Energy, 2011, http://buildingsdatabook.eren.doe.gov/ChapterIntro1.aspx

5. NHE Fact Sheet, Centers for Medicare and Medicaid Services, December 2, 2016, https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html..

6. U.S. Senate Committee on Health, Education, Labor, Hearing on Prevention and Public Health: The Key to Transforming our Sickcare System, December 10, 2008, testimony of Donald Wright, Principal Deputy Assistant Secretary for Health, U.S. Department of Health and Human Services, http://www.help.senate.gov/imo/media/doc/Wright2.pdf.

7. "What is the Precision Medicine Initiative?" U.S. National Library of Medicine, February 28, 2017, https://ghr.nlm.nih.gov/primer/precisionmedicine/initiative

8. Paddy Padmanabhan, "How A.I. and Blockchain are Driving Precision Medicine in 2017," CIO, January 17, 2017, http://www.cio.com/article/3157477/healthcare/how-ai-and-blockchain-are-driving-precision-medicine-in-2017.html.

9. Lucas Mearian, "IBM Watson, FDA to Explore Blockchain for Secure Patient Data Exchange," Computer World, January 11, 2017, http://www.computerworld.com/article/3156504/healthcare-it/ibm-watson-fda-to-explore-blockchain-for-secure-patient-data-exchange.html.

10. Nathaniel Watson, "Health Care Savings: The Economic Value of Diagnostic and Therapeutic Care for Obstructive Sleep Apnea," Journal of Clinical Sleep Medicine, Vol. 12, No. 8, 2016, http://www.aasmnet.org/Resources/pdf/Watson-sleep-apnea-value.pdf

11. Sleep Apnea Information for Clinicians, American Sleep Apnea Association, https://sleepapnea.org/learn/sleep-apnea-information-clinicians/

12. "Food Security Overview," The World Bank, March 21, 2016, http://www.worldbank.org/en/topic/foodsecurity/overview

13. "Supplemental Nutrition Assistance Program Participation and Costs," U.S. Department of Agriculture, February 3, 2017, https://www.fns.usda.gov/sites/default/files/pd/SNAPsummary.pdf

14. Jeffrey K. Lazo, "Economics of Weather," APEC Research Center for Typhoon and Society, Vol. 2, No. 4, December 2012, http://www.ucar.edu/multimedia/articles/2013/ACTNewsletter2(4)27-29_Lazo.pdf