Getting Data Right

Opportunities to better deliver services using “big data” have grown exponentially the last 10 years, but merely hoping it will not be misused to the detriment of vulnerable customers is naive at best.

AUTHOR
Katharine Kemp

“How is that possible? They can’t have it,” Parameshwari shakes her head and waves a bangled hand. “What is ours will belong to us. We give someone information because they need it. How can they give that to everyone? Why should everyone have my information?”

Parameshwari sits next to a sewing machine in a sparsely furnished room in Chennai, India, where she earns less than $10 a day, according to a 2017 research collaboration, Privacy on the Line, between Dalberg, Dvara Research and CGAP. She speaks for many consumers around the globe, equally baffled by claims made in the name of “big data” (watch the short film version of Privacy on the Line for more consumer voices in India). How is data about individuals collected and shared? Do they have any real say in the matter? Are their governments protecting them?

The current passion for data – its promises and perils – is everywhere evident in the press and policy making. There is little new about the practice of companies and governments collecting personal information about their customers and citizens. What is new is the enormous scale on which this information is collected and stored (largely in ways which are invisible to the individual concerned); the speed and accuracy of the information which may emerge from tracking individuals in real-time rather than questioning them later; and the extra insights which may be gained by combining information from numerous sources, both public and private. These are the features which make “big data” big.

Data issues concern everyone, and not only in their financial lives. This essay focuses on how the collection and use of data for the purpose of providing financial services affects consumer’s financial lives and other parts of their lives, with a particular focus on the most vulnerable customers – the base of the pyramid in frontier and emerging markets.

Enthusiasts vs. Skeptics

Big Data Enthusiasts: “The number one most important thing for any business ... is data.”

Big data enthusiasts warn that many opportunities to improve lives are wasted when we fail to collect, store and analyse data, or keep information “siloed”, refusing to disclose it to other entities. Big data can reveal a picture of the world previously invisible to human faculties. With the addition of machine learning, computers can decode patterns in this ocean of information, providing answers sometimes even before we know the questions.

These advances may deliver important benefits for financial inclusion. As a first step, data can help identify areas where consumers are not being served by traditional financial services and potentially determine why. Are women not opening or accessing accounts? Are they required to transact through their husbands? Are people from a particular ethnic background disproportionately excluded from insurance?

New ways of using data can help to close identified gaps. While traditional financial institutions may see little value in opening their branches in poorly-served rural areas or lending to small businesses with scant credit history, new players have shown a willingness to serve these consumers and bring rivalry to the market, using alternative sources of information.

A lender might, for instance, use customer location data collected by mobile network operators to offer credit to those with no formal credit history. If, for instance, location data reveals a woman travelled to the same marketplace every day from six in the morning until four in the afternoon for the last three years, that can support her claim that she has operated a food stall in that market for this period.

Lenders have used a broad range of mobile phone data for the purposes of alternative credit scoring, including contacts, geolocation data, apps installed, SMS messages and call logs. In Kenya, for example, Branch has used such data to provide uncollateralised credit to a range of underserved customers. Tala, also operating in Kenya, has found that people who make regular calls to their family are 4 percent more likely to repay their loans. This information is based on Tala's analysis of customers' call logs and the content of their text messages, for example, the use of the word "mama".

Credit providers also use bill payment data, social media data (such as the size of the customer's network), psychometric testing and e-commerce transactions to predict the likelihood that a borrower will make their repayments on time. The CEO of Zest Finance has gone so far as to say that "All data is credit data".

Seeking more data may allow providers to better meet the particular needs of a group of consumers who struggle to use conventional products. Understanding and tracking the seasonal fluctuations in the income of small-hold farmers, for instance, can allow a lender to tailor repayment terms to those cycles.

Both traditional banks and new players have also analysed large-scale data sets to discern patterns that identify and predict fraud. Acting on this information to reject fraudulent transactions reduces costs to providers and customers alike. In a number of countries, alternative data sources permit financial institutions and fintechs to identify customers previously excluded for lack of formal identity documents.

Big Data Skeptics: "Unless we look to change course in this sector, the risks and dangers to privacy loom large"

Privacy and data protection advocates have expressed serious concerns about these developments. This should come as little surprise. The underlying approaches of "big data" and data protection are, in a sense, fundamentally at odds with each other. At the heart of data protection is the limitation principle. Data protection says: Collect and use only that information which is necessary to serve the immediate purpose. Big data says: Give us everything you have and we'll see if we can find a purpose for it.

Is it safe to permit the unrestrained collection and use of data?

Harms

Harms from the Use of Personal Information

These concerns are not only hypothetical. The collection and use of consumer data by financial services providers has already caused harm in both well-known and relatively obscure instances. Opportunistic providers have exploited customers' personal information for their own ends with devastating effect. In Kenya, digital lenders published the details of defaulters on Facebook, using public humiliation as a debt collection tactic. In China, lenders have used information about students' financial hardship to offer loans which are easy to access but include high interest rates and severe penalties for default. There was a spate of suicides among the students when they later defaulted on those loans.

Data changes power in relationships. If a customer surrenders large amounts of personal information to a company, they become vulnerable to numerous forms of intimidation and exploitation which they cannot anticipate or control. An inherently unequal power relationship has become more unequal.

These opportunities for abuse are greater in jurisdictions where there is no general data protection law or responsible lending law, or little effective enforcement of these laws, as is the case in many developing countries. But even with such laws in place the use of personal information often does consumers no favors.

Ryan Calo, Assistant Professor at the University of Washington School of Law, has described the practice of vulnerability-based marketing, which uses personal data to target consumers based on their particular weaknesses. But as Calo points out, in the online environment, companies can also engineer moments of vulnerability by designing the timing, context and interface of an online transaction in a way that

creates frailty in that particular individual, influencing the consumer to act against their own best interests.

Providers who engage in such exploitative conduct exist, even on the frontiers of financial inclusion. I have heard the representatives of providers on stage at a conference or summit recite the mantra, "We would rather ask forgiveness than permission."

I have heard the representatives of providers on stage at a conference or summit recite the mantra, "We would rather ask forgiveness than permission."

But many providers and organizations are genuinely committed to the pursuit of financial inclusion and customer-centric business models, including the fair treatment of customer information.

Unfortunately, even responsible, well-intentioned players can expose their customers to risks. Here it is necessary to consider the risks of unintended harms and risks that arise directly from the collection of personal information.

The consequences of new types of collection and analysis of information enabled by rapid advances in technology are still being discovered. Many champion the use of algorithmic processing and particularly machine learning to gain insights from big data, and a number of the advances outlined earlier were achieved through such processing. However, researchers have also revealed that these algorithms may discriminate, exclude and produce otherwise inaccurate conclusions to the detriment of consumers.

Sometimes these tendencies are built into the program itself as a result of human bias. Algorithms are used to identify "high value" and "low value" consumers, presenting greatest risk for those who are already vulnerable and disadvantaged. Algorithms may also produce results which are plainly wrong when the data being processed is unreliable. This is a major issue in some developing countries where studies have shown that large percentages of the data held are inaccurate, incomplete or out of date.

In other cases, machine learning produces its own discriminatory tendencies. The fact of this flawed analysis cannot always be understood, particularly given increasing reliance on "black box" algorithms which produce results based on their own form of reasoning, not evident to their creators.

These data practices may bring the comfort of scientific terminology, quantitative analysis, and sharp-edged graphs and tables, but this does not make them immune from embedded bias, error and unjust outcomes.

Harms from Collection Alone

Acknowledging these risks and harms, some argue that we need only be concerned with the misuse of personal information. Collection alone is innocuous; misuse can be identified and addressed. This approach would permit businesses and governments to harvest personal data at will, unconfined by regulation, then determine at a later date how they might use the information and whether the proposed use is lawful and appropriate. This approach is flawed.

The simple act of collecting and storing an individual's personal information significantly increases the risk of harm to that person. As soon as we collect and store personal information, we increase the "attack surface" – that is, we increase the opportunities for that information to be hacked, stolen or used without authorization. The more data is stored and the longer it is stored, the greater that risk becomes.

These breaches can cause severe harm. Identity theft can cause a lifetime of expense and exclusion for the individual concerned. And harm is not limited to the individual. These events can have drastic consequences for consumer confidence, trust in the company holding their information and trust in financial services more broadly, working in direct opposition to the goals of financial inclusion.

Major breaches of Equifax, Facebook and the US Office of Personnel Management illustrate the reputational harm and losses to corporations and government from data misuse. Bruce Schneier, a security technology expert at Berkman Center for Internet and Society, Harvard University, has long pointed out that, for the firm holding the information, data can be a "toxic asset".

Even projects launched with the best intentions are subject to these risks. Taylor gives the example of the Harvard Signal Program on Human Security and Technology which aimed, in part, to identify forensic evidence of alleged massacres in Sudan with the advantage of unprecedented detail from satellite imagery analysis. However, researchers on the program discovered that hostile actors appeared to be hacking into the Harvard systems and using the project's data and communications to target their enemy.

Information that companies store about a consumer may also be accessed by governments, which do not always have due regard for the rule of law. In the East and West alike, the media has revealed numerous occasions where governments have required companies to surrender information about individuals in secret and without due process.

We should not forget that in some countries it is illegal to express dissent or criticize the government, to practice a certain religion or engage in homosexual activity. Information that seems harmless viewed in isolation – a person's transaction history, social media posts or location data – can reveal highly sensitive information, especially when combined with data from other sources.

The mere collection and storage of information can be profoundly unsafe for the individual concerned.

Consent

Is consent the answer to responsible use of data?

How then should we identify the boundaries of fair collection and use? Should we leave the decision to each individual, making data practices dependent upon their consent? One study asked consumers in Uganda whether they would be willing to trade some privacy for access to a loan or a better interest rate. It reported that many consumers were willing to make such a bargain.

Privacy terms have often been viewed as a bargain or trade of this kind. That is, companies propose certain privacy terms to consumers, usually via a privacy notice, and each consumer makes the decision whether to exchange some of their informational privacy for the benefits offered by the company. This approach is said to respect the individual's freedom and autonomy in making decisions about their informational privacy on the basis of their own privacy preferences.

However, we should be very cautious about drawing conclusions from the fact that consumers living in poverty say that they would give up some of their privacy for money. This is likely to say more about their straitened circumstances and their lack of alternatives than the legitimacy of the supposed bargain.

The nature of privacy policies themselves also frequently prevents consumers from making informed choices, seemingly designed to hide rather than reveal the most relevant or concerning data practices. Policies almost universally begin with reassurances about the company's concern for privacy, their diligence in protecting the customer's information, as well as obvious, innocuous uses of customer

information. More problematic terms come later, phrased in vague, open-ended language which guard the company against the accusation of unauthorised use without enlightening the consumer.

And as CFI Fellow Patrick Traynor of the University of Florida documented, the privacy policies of many traditional and digital financial service providers use language more suited to university graduates than ordinary consumers.

"Signing terms and conditions is not a matter of choice – it's something that you have to do because you have no choice," said interviewee Sanjay as quoted in Privacy on the Line.

"Signing terms and conditions is not a matter of choice – it's something that you have to do because you have no choice."

These terms breed incomprehension. Even with the benefit of high literacy rates and education levels, in Australia, around one in five consumers believes the existence of a privacy policy means that the company will not share their personal information with another company. Further, these policies are generally presented on a take-it-or-leave-it basis, and use "bundled" consents: that is, they do not provide separate options concerning uses of personal data beyond the immediate purpose of the transaction but require consumers to consent to all specified uses or none.

The informed consent – or "notice and choice" – model of privacy regulation is subject to more fundamental criticisms, which go beyond the deficiencies of privacy policies. This model was developed in the United States in the 1970s at a time when data practices were entirely different from those of today: collection of information was generally visible and actively involved the individual; the cost and difficulty of processing, storing and transferring data naturally reduced its exposure to misuse; machine learning, online monitoring and mobile phone location data did not exist.

Even if the clarity and usefulness of privacy policies are greatly improved, the nature of new data practices and their consequences will make it extraordinarily difficult for consumers to understand the privacy terms companies offer, let alone their consequences: consumers will lack the information necessary to make a rational choice. The mere process of consumer "consent" may be meaningless.

Prioritizing

Prioritizing Data Protection in Emerging Markets and Developing Countries

A recent study by Dalberg, Dvara Research and CGAP revealed that, even among some of the world's poorest, privacy is highly valued and protected to the extent that it will not be exchanged for financial incentives.

"Certain kinds of data are not tradeable. Even if you give me a 100% discount, I won't share my browsing history," notes Sushma, a customer in Delhi, in the report.

The Omidyar Network conducted a survey of customers of alternative credit scoring services in Kenya and Colombia which revealed that 82 percent of respondents in fact regard their mobile phone calls and texts as private information, and even more private than medical and financial data.

But is the need for basic financial services more pressing than these sensibilities? Where should our priorities lie when there are people in developing countries who have never had the ability to save using a bank account or transfer money digitally to far-away family or insure themselves against misfortune?

The contention that the right to privacy should be subordinated to the economic needs of the poor was considered by the Supreme Court of India last year in the landmark decision of Justice K S Puttaswamy v Union of India, where the Court held for the first time that there is a fundamental right to privacy in India. It would be hard to improve on the response of Justice Chandrachud, who delivered the Plurality Opinion:

"The refrain that the poor need no civil and political rights and are concerned only with economic well-being has been utilised through history to wreak the most egregious violations of human rights. ...The pursuit of happiness is founded upon autonomy and dignity. Both are essential attributes of privacy which makes no distinction between the birth marks of individuals."
There is no simple answer to the question of how data can be used safely and fairly to meet financial inclusion objectives.

While there is compelling evidence of potentially very serious risks to vulnerable individuals and groups, we lack certainty about the actual incidence and impact of the relevant harms. It is therefore tempting to suggest that high standards of data protection should be a second-order consideration, a "nice-to-have" which can be addressed once businesses have been persuaded to serve the underserved, free from the burden of extra regulation. Unfortunately, this is a situation where justice delayed would be justice foregone.

Once personal data is collected, disclosed and distributed to numerous parties, there is no retrieving it. The exposure of that data cannot be undone. With current technology, it can be stored and aggregated in perpetuity, throughout the lifetime of the individual concerned. When harms occur, they will undermine customer trust in the very services which might otherwise have improved their lives.

These factors weigh in favour of an approach that deliberately errs on the side of caution and restraint. Such an approach requires that we forego some uses and disclosures of personal data until those uses and disclosures can be safely made, as explained below.

Building Trust and Fairness

Global Developments and Lessons for Data Protection

There have recently been a number of positive global developments in data protection which should interest providers concerned to adopt innovative and responsible data practices. In this part, I outline the unfolding events as well as some proposed best practices we can glean from these developments.

In May 2018, the **General Data Protection Regulation (GDPR)** came into effect in the European Union (EU). The GDPR is intended to intended to create certainty for business and enhance consumer trust, placing additional obligations on organisations processing data in the EU or about individuals in the EU, including improved standards of consent, a right to erasure and very substantial penalties for infringement.

The GDPR has had ripple effects beyond its actual legal application. For instance, some organisations operating across numerous jurisdictions have seen economies of scale in adopting the same systems of data governance in all jurisdictions and adopted the GDPR requirements as the highest applicable standard.

The implementation of the GDPR has also driven an increase in the development and availability of **privacy enhancing technologies** (PETs) which allow organisations and their customers to use IT to manage privacy in more securely and conveniently. The GDPR has also led to a fresh focus on **privacy by design**, which makes privacy part of the foundational design requirements for systems, rather than a “bolt on” down the track.

Encouragingly, a number of organisations have recognised the importance of privacy to their customers and **compete on privacy quality**, including by the use of PETs and privacy by design. Apple, for example, has emphasised just-in-time privacy notifications which give users a choice about providing their data at the moment when an app is attempting to collect that data. Other organizations are using PETs such as transparent, user-friendly, easy-to-navigate online privacy policies and consent interfaces to earn their customers' trust.

Regulatory developments in the EU have also led to **growing consensus on data protection** regulation outside the EU. This has been driven in part by the fact that a number of countries wish to ensure their standard of data protection regulation is sufficient obtain an "adequacy assessment" under the GDPR, allowing organizations in those countries to process data about individuals in the EU. Professor Graham Greenleaf, a global privacy expert at the University of New South Wales, believes that a new global standard is actually emerging as a result of the widespread adoption of the standards in line with the Council of Europe data protection Convention 108, a regulation which he explains includes many, but not all, of the GDPR requirements – a "GDPR-lite".

Drawing on these developments there are certain data protection principles which providers should be implementing now in the interests of consumers, financial inclusion and their own reputations, whether or not they are currently subject to data protection regulation:

- **Privacy by Design (PbD)**. Critically, PbD recognises that privacy is not a zero-sum game: it does not require a trade-off. It is possible to provide innovative and affordable services which respect customers' privacy, particularly when organisations make privacy fundamental to the design of systems, from their very inception. This is not a costless exercise and some will argue that cost will deter providers from operating in developing countries. At the same time, some organisations will be establishing modern databases and IT systems for the first time, providing a prime opportunity to build privacy into these systems from the beginning, without the baggage of legacy systems.

- **Minimization**. According to the minimization principle, organizations should limit their collection and handling of customers' data to that which is actually necessary for the immediate purpose and, equally, delete or destroy that data when the purpose is complete. This minimizes the "attack surface" of the data, reducing the likelihood of harm to consumers but also the organization's exposure to liability and reputational harm in the event of data breaches.

• **Transparency.** Organizations must provide transparency about the fate of consumer data in the hands of the organization. Privacy policies need a major overhaul. These policies should not be used to excuse, but to inform. They should therefore highlight, and lead with, precise descriptions of the collections and uses that are likely to be least expected and most concerning for consumers.

• **Consent.** Customer consent should not be established on the basis of pre-ticked boxes, nor should it be implied from customers' continued use of a service where there is a privacy policy incorporated in some section of an app or website. Customer consent should be explicit, fully informed (in a language and mode of delivery the customer understands), unbundled, revocable, and require action on the part of the customer. An organization should not refuse to supply services based on the customer's refusal to consent to data collection or uses which are not necessary to supply the customer with those services.

• **Access, correction and erasure**. Customers should have a right to access the information which an organization holds about them, both so they understand the extent of the information collected about them and so they can require corrections where that information is inaccurate, incomplete or out of date. Increasingly, new data protection regulations also provide customers with the right to require erasure of their data when the purpose is complete.

• **Data breach notification**. Organizations should provide customers with notice of serious breaches that affect their data and their potential consequences. This requirement is increasingly a part of data protection regulation around the world. While organizations may instinctively seek to hide or diminish information about data breaches, a number of firms have learned through bitter experience that this approach only exacerbates reputational harm when the breach is inevitably discovered.

• **Liability**. Organisations should take responsibility for the protection of consumer data to the extent that they accept liability for harms caused by the organisation's failure to adequately protect that information from misuse in the hands of the organisation and in the hands of third parties to whom the organisation transfers that information. This would include liability for the re-identification of personal information which was purportedly disclosed on a de-identified basis.

These principles are not only fair to customers and likely to engender greater customer trust and confidence in the provider and financial services. They can also help us to be more rigorous in our assessment of the benefits and limitations of big data and to impose a sensible discipline on our use of data, reducing the risks of liability and reputational harm for data collectors. They may even remind us that there are still other ways of understanding our world. As Greenleaf has argued, "We must avoid the assumption that only a datafied world is understandable and valuable, and that more data is preferable to better data."

CENTER *for*
FINANCIAL
INCLUSION

ACCION

The Center for Financial Inclusion at Accion (CFI) is an action-oriented think tank that engages and challenges the industry to better serve, protect, and empower clients. We develop insights, advocate on behalf of clients, and collaborate with stakeholders to achieve a comprehensive vision for financial inclusion. We are dedicated to enabling 3 billion people who are left out of — or poorly served by — the financial sector to improve their lives.

www.centerforfinancialinclusion.org
www.cfi-blog.org
@CFI_Accion