

# Crossfire: 'The potential of data for development is overstated. The negative impacts on vulnerable populations outweigh the benefits at this point'

SOPHIE ROMANA and EMMANUEL LETOUZÉ

*'Big Data' can be useful in the battles to solve development problems such as climate change and Ebola pandemics, but they may also be leaked and used for purposes that the subjects of the data would not authorize. Sophie Romana and Emmanuel Letouzé debate whether the advent of Big Data presents enormous opportunities for good, or invites calls for a greater degree of caution.*

*Dear Emmanuel,*

The international development community's way of doing business is entering the era of data, yet few organizations have clear guidelines on how to handle data responsibly. The tension lies between a 'do no harm' approach, minimizing the data we collect from our programme beneficiaries, and the promise that 'Big Data' can help us solve the Big Problems we pledge to tackle, such as climate change, poverty, Ebola epidemics, and problems that we don't know about yet, and which may be solved in the future thanks to vast amounts of data we'll be collecting today. There is a legitimate case to be made around Big Data's capacity to help solve big problems, but I will argue that the state

of our industry today requires us to be cautious and collect only the data we need as we don't know how to handle data responsibly and we hold a duty to our beneficiaries to protect their data.

*Data need not be digitalized to be compromised.* First, we need to educate our teams, partners, and programme beneficiaries/participants on what data is. A few years ago, upon joining Oxfam, I discovered a huge suitcase under my desk, full of questionnaires from the baseline survey we had run in one of our programmes in Central America. At the time, this was the 'best way' to send the questionnaires back to headquarters. The names of savings group members, field agents, villages, and their financial data were all nicely bound in piles, not even in envelopes, for anyone who opened the suitcase to see, a suitcase full of data, and some of the most personal data. We were a long way from a simple password-protected file, single key IDs, IDs for villages instead of names, etc. This example may be extreme, but in reality, programme officers, field agents, partners, numerators, and evaluation consultants and even ourselves, carry a lot of personal data collected from

---

*Sophie Romana (sromana@oxfamamerica.org) is Director, Community Finance Department, Oxfam America; Emmanuel Letouzé (eletouze@datapopalliance.org) is the Director and co-Founder of Data-Pop Alliance, the first think-tank dedicated to Big Data and development.*

---

*Copyright © Practical Action Publishing, 2014, [www.practicalactionpublishing.org](http://www.practicalactionpublishing.org)  
<http://dx.doi.org/10.3362/1755-1986.2015.002>, ISSN: 1755-1978 (print) 1755-1986 (online)*

programme participants, whether on paper, tablets, or in unprotected Excel sheets, on motorbikes and buses all around the world. Paper data is probably even more sensitive than digital data because of its immediate accessibility. However, that is not to say that digitizing data makes it safe!

*The more data collected, the bigger the safety issue.* Protecting the confidentiality of data is directly linked to the amount of data collected. This is a problem banks and credit card companies are acutely aware of. Credit card companies typically break down the data on each customer in different databases to make it extremely difficult for a hacker to reconcile a customer's full record. Most recently the US mega retail chain Target experienced a massive breach which led to the cancellation of thousands of store credit cards because they had not taken all the security measures needed to guarantee that pin numbers would be separated from card numbers. We don't face that issue as development organizations. However, we work with thousands of people who trust us with intimate information on their finances (income, assets, savings, debts, real estate, cattle holdings, etc.), their families (number, age, sex of children, number of wives), and their personal beliefs (religion); we also collect their ethnic origins and their affiliations to local organizations (savings groups, economic interest associations, cooperatives, local citizens' organizations). The more information they give us, the more vulnerable they make themselves. How can we protect activists when we put our projects on maps? How can we protect savings groups when the GPS coordinates of

one village could be in the column next to 'accumulated savings' in a spreadsheet? As we collect more data, how can we ensure that we don't put the lives of the people we work with in danger?

*Yours,  
Sophie*

*Dear Sophie,*

Thanks for this; let me start by building on and generally seconding your main points before making additional observations.

You mention how 'we' have collectively entered a new data-rich era that presents both great promise and perils – overall, lots of uncertainties and uncharted territories, tensions and contradictions. As you note, ethical issues around privacy, informed consent, data, or statistical literacy predated the digital age, but my sense is that their magnitude and nature may well have changed. One reason is that the bulk of what we term 'Big Data' (as data, rather than as tools and methods and those implementing them) is personal – often personally identifiable – data that is passively emitted, with little understanding of the potential and actual use made of them by third parties. Because of pressures to make 'evidence-based' decisions, private and public sector organizations, and others, will become increasingly data-hungry, compounding risks of abuses along the data life cycle. There are many ways, as you suggest, for data to harm, even with the best initial intentions, and there are also legitimate concerns being expressed that differential access to data and capacities may enhance inequalities. Ultimately, the main challenge will

be to distinguish what can be done from what should be done with data, why, when, where, how, by, and for whom. That will require adopting or adapting legal, technical, and practical frameworks and systems to allow the perspectives of different constituencies to co-exist – those of private sector companies, governments, individuals, and communities – while upholding key ethical principles of beneficence, respect for persons, and justice, rooted in human rights – as stated in particular in the Menlo Report. I hope we can come back to these issues in more detail – notably their implications and requirements in terms of awareness raising and capacity building – in our exchange.

But I will also say that I often find misconceptions – or simplistic assumptions, rather – about the role of data in development, which I think are worth unpacking and challenging for the complexity and richness of the question at hand to fully emerge. In other words I think the wording of the question may need a bit of reframing or clarifying. I call these mechanistic or Bismarkian views of societies and the world – the notion that data are and should be primarily used (once turned into information, one is always reminded) by policy-makers or ‘decision-makers’ to make good (better) decisions for the benefit of the people. And we would be asked whether the expected marginal ‘benefit’ derived from better policies or programmes does or doesn’t outweigh the possible risks. But I will argue here, as I have before, that lack of data has historically played a limited role in shaping (explaining) bad policy and programmes. So, if we consider that

aspect and argument alone – better or more data in the hands of some elusive ‘decision-makers’ (who isn’t?) – as the only component on the ‘plus’ side of the picture, we may be missing a big part of it; namely the role that data can play to empower people to make daily decisions and challenge existing power structures and systems, even beyond simple arguments of ‘greater accountability’ enabled by data.

Over to you!

*Yours,  
Emmanuel*

*Dear Emmanuel,*

Many thanks for sharing your thoughts enriching our debate. You raise two very interesting points: 1) Are data-driven decisions really more impactful? And 2) How is data empowering people? Allow me to comment and build on both.

Your first point reminds me of the famous experiment about the stock market which consisted of comparing a portfolio designed by the smartest traders to another composed of stocks picked in a game of darts. The second performed better, highlighting the influence of randomness in the stock market, despite the fact that the trader had much better information than the dart thrower.

In the development world, a marginal impact can sometimes yield huge personal benefits. A randomized controlled trial that Oxfam and Freedom from Hunger ran in Mali between 2009 and 2012 with 500 Saving for Change groups showed that having access to a financial cushion in the form of savings made the lives of 100,000 women and their families less stressful and allowed them to start planning for their future.

I could not agree more with your second point. Recently, we have seen farmers in sub-Saharan Africa becoming price makers thanks to the access to market data they gained with mobile phones. This access totally changed their relationship with a middle man/buyer who until then would impose his price and conditions on them. Similarly, farmers in Ethiopia are placing over 30,000 calls a day to a 'green-line' to access weather and market data allowing them to make decisions about managing their agricultural production. In the future, I would love to see citizens using mobile technologies to track taxes and budget flows and make sure that, for example, oil revenues positively impact the budget of their producing villages.

However, those two examples are taking place in countries where there is little legal protection addressing the rights of citizens to access their own data and making sure that the companies (telcos and others) who collect, store, and share data do so while protecting the privacy of their clients. I believe the legal vacuum is even deeper when it comes to 'exhaust' data: the trail of data we all leave behind us and over which we have little power/control.

Finally, I would not like to single out developing countries: Sony Pictures has recently been the victim of what appears to be the single largest cyber-attack, which directly impacted the private lives of thousands of employees by throwing their emails, personal addresses, and names of family members into the public arena, revealing how vulnerable our digital lives have become. And we haven't even discussed the issue of data ownership yet!

Overall, I believe that we should start by focusing on securing *the data we need* and not try to become data alchemists at the peril of our clients and beneficiaries, and ourselves as well: I can't imagine how damaging a data leak would be to an international NGO whose primary line of duty is to 'do no harm'.

I look forward to your response.

Warmly,  
Sophie

*Hi Sophie,*

Thanks,

On your last point, I am sceptical about the possibility of curbing data collection by private corporations. It is possible that there will be public pressure – of citizens and their representatives – to try to impose stricter limits on the kinds of data, and their granularity etc., that they may collect and store, but overall I think private corporations will in the future 'hold' a greater volume of personal data. Interestingly, most people seem fine with sharing large amounts of data if they perceive they will get a personal benefit – whether psychological in the form of getting 'likes' on photos on Facebook, financial by getting coupons for some goods after they searched for them, or 'practical' benefits as in the case of Amazon suggestions. I also think that there are really important insights that can be gleaned from these data sets to better understand and positively affect human ecosystems.

But we also know very well that the definition and conditions of 'informed consent' in the age of data may not be very clear; we all consent to our data being used in ways we are actually unaware of; and, as we talked about, there may be unexpected and

unintended negative consequences as a result of these data being used, especially if they fall into the wrong hands. Increasingly, we'll experience hard questions about competing human rights – for instance in the case of Ebola or another pandemic. Should cell phone data have been made available to research teams more swiftly and widely to support response efforts (to the extent that this may have helped track and predict population movements and the spread of the disease)? Or should considerations of individual privacy and possibly safety (if the data has been used to identify and target an ethnic group not respecting mobility restrictions for instance) have prevailed?

I do not think these hard questions warrant a crackdown on Big Data, but they definitely and forcefully call for greater awareness and participation of citizens (who happen to be data emitters) in debates and deliberations about the use of their data. Ultimately, it comes down and back to issues of literacy and empowerment. It also raises the fundamental question of who owns the rights to these data. Several scholars like Sandy Pentland, our Academic Director, have argued that people should have greater control over their data – and I strongly agree with that. I think a 'real' data revolution is one where people, even the most marginalized, are granted greater control over their data, are willing (incentivized) and able (educated) to use these data in their daily lives, including to hold their representatives to greater account.

Too often, still, and this gets to my remark about the 'Bismarkian'

approach, 'we' equate decision-makers and policy-makers and think that all social problems require a policy solution. So to make the point clear, I see a lot of the value of the new ecosystem happening outside of traditional decision-making processes, either by improving people's decisions and lives directly, or by giving them an incentive and the means to challenge existing power structures. This will require a lot of investment in training, advocacy, and months or years of debates. But the number one human right to me is political participation and I think the emergence of the new ecosystem of data requires and allows this to happen. In other words I think the 'new oil' of Big Data may help grease the wheels of democracy, not just help politicians and corporations make better decisions.

Now, another aspect is that in my opinion many organizations (including governments) do need to engage in a profound cultural shift to let evidence drive more of their decision-making processes. It is not a silver bullet, and it is not in contradiction to my previous point. There are scores of cases – and you point to a few – where the use of sound evidence on causal effects has led to significant improvements in human outcomes. New data and new tools do shed light on human processes we often had little or no understanding of. We need to see more of these examples, but the uptake of a sound data culture that balances competing considerations and priorities is a tall order and should start with heavy investments in education in and about data.

*Yours,  
Emmanuel*