The specter of big data is haunting the world, but has the data revolution already occurred?

LSE blogs.lse.ac.uk/impactofsocialsciences/2014/03/05/specter-of-big-data-revolution/

3/5/2014

Changes to the supply and demand of data are restructuring privileged hierarchies of knowledge, with amateur hackers and machine-readable technology becoming a central part of its analysis. Traditional experts may be hoping for a gradual evolution, but a parallel revolution led by practitioners in the private sector may already be underway. **Prasanna Lal Das** *argues that partnerships will need to incorporate these new practitioners because for them, the data revolution is already a fact of life.*

This isn't the first age of revolution, but this one feels like it might not last 100 years. Our world is transmogrifying in front of our eyes – sometimes more forcefully than others – and the traditionally dry world of data, dominated by dons and 'experts', hasn't been immune to changes either. It might even be the spark for at least some revolutionary fervour, especially since the report of the high level panel of eminent persons on the post-2015 development agenda called for a 'data revolution' to 'strengthen data and statistics for accountability and decision-making purposes'. The official data revolution has however unfolded slowly, sometimes making one wonder if it's going to be a revolution of the bureaucrats, by the bureaucrats, and for the bureaucrats. Or if it will be a revolution that truly changes how we measure our world, what we measure in it, and who does the measurements.

The report describes the revolution at a fairly high level and there has been significant work since then to define what this revolution might entail and how emerging trends listed (disaggregation, crowdsourcing, new technology, and improved connectivity – that have the potential to empower citizens) may shape the contours of the revolution. Quite inevitably, at least some of the discussion has centred on well-known (but still important) issues such as greater capacity building, a more central role for national statistical offices, increased standardization of data collection efforts, smarter partnerships, and ineluctably, more resources for data agencies. But in the shadow has lurked – as the report itself recognized – the specter of big data and new data techniques, the recognition that the world of data may have undergone a revolution already and that some cherished truths may no longer apply.

Data mining. Credit: Matt Brooks (The Noun Project, Public Domain)

I found my way into a couple of conferences recently – one organized by Webdatanet and the other by UNECE, where the revolution – impending or underway depending on whom you asked – hung heavy in the air. These were gatherings of experts with authoritative reputations in the data community, and with the tools to discern that the ground was shifting beneath their feet. So while calm reigned on the surface, both the events roiled with questions.

Have machines taken over already?

Much of the data (over 90%?) in the world now comes from 'machines' (sensors, cell phones, satellites, cameras, drones, scanners…) and the imbalance is likely to grow (even in social sciences). Clearly there still is a role for researchers in the field talking to people and faithfully recording observations but the data they collect is dwarfed by what machines do and can gather. Talking about machines can be scary, but machines can record data from war zones, from inhospitable regions, from near and from far, about the visible and the invisible, the spoken and the unspoken, and they can do so repetitively and predictably (alas no blogs about the incredible ethnographic adventures of machines!). And machines often do a better job of analyzing this data and making sense of it – even learning and adapting as the data shifts. The second economy may or may not be about to hit the data scientist but

– in my opinion at least – a data revolutionary that isn't asking 'can machines do it' when considering any aspect of data may have missed the revolution

Have the barbarians entered the gates (don't you need to be an expert to handle data)?
Remember the days when journalists railed against bloggers and social media in general because the new content creators did not adhere to journalistic and style standards so painfully cultivated over decades by experts in the field? Or when video professionals mocked YouTube? Well, the same rude mass has gathered around data – collecting DIY data and publishing it, analyzing and visualizing it, and sharing its work with data practitioners, policy wonks, academics, and civil society alike. What makes these amateur data hackers powerful is their familiarity with context, their ability to ask the right questions of the data, and their purposeful approach to data – these are people for whom data is the means not the end, and suddenly they have a large variety of data tools to choose from, tools that almost commoditize/democratize many traditional data skills. I'm not for a minute saying that these amateurs can replace data scientists and specialists, and the latter remain indispensable in many contexts, but the work of these hackers does beg the question whether the real revolution shan't come from 'below, and if the new hackers aren't the real data revolutionaries

Are traditional data coalitions losing ground?
The data revolution conversation is largely dominated by traditional forces (despite the contradiction in terms!) – mostly official agencies, recognized think tanks, and universities that have the mandate to collect, curate, and distribute essential data (though there has been concerted effort to include new players). One could argue – though I don't have the numbers right now – that the proportion of data such entities contribute to the entire data ecosystem has been declining recently (essentially because of the data explosion from new data sources). What has also changed is the data mix – most 'modern' data sources are non-traditional (the machines above, social media, paradata, business transaction data, and more) and many official agencies have not developed the skills to collect and manage such data (though some are making efforts). The question thus inevitably is whether the revolution has already left such agencies in its wake or do they still have a chance to make up lost ground? I would like to say that the official agencies will continue to be vital, but some traditional partnerships may include several new faces.

Is everything we know about data wrong?
Let me be clear – it's not, but practitioners of new data techniques routinely lob provocative questions at the data establishment – aren't theories and models dead; isn't sampling a primitive technique better suited for a data-poor world rather than today's data rich world; are our favorite numbers, the so-called leading indicators like GDP, inflation, employment, balance of trade and the like just plain useless in the modern world; isn't traditional data too slow, cumbersome, and expensive to be useful in a decision intensive environment; is data accuracy really as important as we think it is (and has data ever been terribly accurate anyway) or should we accept 'messy' data as the new norm? Responses to questions such as these sometimes tend to be knee-jerk or worse, dismissive but if there's truly going to be a revolution, perhaps there's a case for throwing out a few old rulebooks and keeping an open mind

What's going to be the value of a data revolution?
Why do we need a data revolution? At least some of the discussion around the data revolution has focused on 'supply side' issues such as data gaps and quality, better documentation, technology infrastructure for data, the usability and openness of data, and the challenge of big data. The discussion on demand has centered on 'evidence', 'accountability' and 'decision making' but is this a gap that 'official' data revolutionaries can fill better than let's say the private sector? What are questions that official bodies can answer better than the 'market' does? Can they help farmers by anticipating agricultural yield better? Can they forecast disease outbreaks better? Can they help businesses make better investment decisions? Can they really help make sure aid goes to the right people and produces desired results? Answer yes – and that would indeed constitute a revolution. The private sector has responded to the demand for answers to such questions already – take a look at the work of Climate Corporation. Or Metabiota. These companies, and others like them, are building businesses on data, like insurance companies for

example did before them. As far as they are concerned, the data revolution is a fact of life already.

The groups in charge of the data revolution hope for an orderly transition, a change of order that follows today's rules, a gradual evolution that accretes into a revolution. Outside the development world, a parallel revolution may already be underway however.

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy *if you have any concerns on posting a comment below.*

About the Author

Prasanna Lal Das *leads the open financial data program (*finances.worldbank.org*) at the World Bank. He's especially interested in the role of open financial data in promoting civic engagement and results, and the potential of open data to foster commercial growth and innovation. Prasanna and his team are also investigating the possibility of using big data for international development, especially to increase effectiveness. Prasanna holds a Masters degree in Modern Indian History. You can follow him on Twitter at* @prasannalaldas *or read his* Bank blog*. Views in the blog are the author's personal opinion and don't reflect that of any organization.*