Fast Data, Slow Policy: Making the Most of Disruptive Innovation

Aniket Bhushan

International affairs policy and practice are particularly ripe for disruptive innovation fueled by the rise of big data and open data. While there are several loci of disruption, this paper demonstrate how the speed and time dimension in policy-relevant research in international affairs and international development are being disrupted. Three illustrative case studies—real-time macroeconomic analysis, humanitarian response, and poverty measurement—are discussed. Finally, the concluding section explores how successful policy entrepreneurs can make the most of disruptive innovation in the age of big data.

Most trends are accompanied by a familiar cycle of hype. After the initial trigger come inflated expectations, then a trough of disillusionment followed by an upward slope of enlightenment, before finally resting on a mainstream plateau.[1] The heightened pace of open data and big data, and their potential impact on international affairs, is following a similar pattern. Whether one chooses to defend the hype or challenge "data fundamentalism,"[2] enough fodder exists to fuel both sides of the debate.

Proponents argue the rise of big data and open data fundamentally changes the way we think about the world. The sheer volume, velocity, variety and veracity of big data[3] means we can worry less about quality issues associated with narrower information sources. We can reframe our methodological orientation to focus on iterative learning and correlations as opposed to obsessing over causality. Doing this allows us to embrace the possibility of working with a plethora of untapped (and growing) data feeds to address challenges not even fully articulated yet. Doing this also means leaving in abeyance a host of new dilemmas in areas such as privacy, autonomy, and asymmetric coverage.[4]

Detractors on the other hand are quick to point out that data is not objective (indeed the term "raw data" is an oxymoron). Data cannot "speak for itself," as the proponents of big data would have us believe.[5] There are biases at all stages, from collection to analysis to presentation. Big data may be unbeatable when it comes to forecasting, but it is "dumb" when it "comes to doing science," as it is not underpinned by sophisticated research designs that aim to identify causal relationships.[6] The bigger the data, the more we are prone to lull ourselves into a false

---

Aniket Bhushan is senior researcher at the North-South Institute (NSI), a Canadian-based policy research institution. Mr. Bhushan heads NSI's work on the Canadian International Development Platform (CIDP), a data and analytical platform on Canada's engagement with the developing world. Mr. Bhushan's current research focuses on the impact of big data and open data on public policy. His academic background includes degrees in political science and commerce, and he completed his M.A. in political science at Carleton University (Canada).

sense of confidence in predictive analytics. Indeed, big data accentuates the "signal to noise" problem.[7]

There are a multitude of other issues associated with the use of data. Big data has its roots in the commercial sector.[8] The main intention behind generating sharper insights into customer behavior and profiles is to achieve better targeting and segmentation; in other words, smarter discrimination to ultimately drive profitability. When examined from a public policy perspective, this could be highly problematic. The kinds of targeting and discrimination taken for granted in many commercial sectors, like advertising, would be expressly forbidden in more regulated industries, like the insurance industry, and may be contrary to the aims of public policy and public service delivery.[9]

Whichever perspective one identifies with, the inescapable fact is that big data and open data are already disrupting several industries. Policy relevant research and analysis in international affairs is no exception. In fact, international affairs—and international development as a subset—are particularly ripe for data driven disruption. The speed of big data and open data has inherent disruptive potential that is already impacting a number of areas, from the highly mainstream business of macroeconomic indicators to new approaches to poverty measurement in difficult country contexts.

The aim of this paper is to describe what the emerging paradigm of international affairs in the age of big data looks like and the role of the policy entrepreneur at the center of this process. Successful policy entrepreneurs, both individuals and institutions, will be able to make the most of highly disruptive trends in data that will fuel future analyses of international affairs. Adapting to this emerging paradigm requires investing in new tools. While there are several loci of disruption, the emphasis of this paper is on the time and speed dimension.

International Affairs and Development are Ripe for Disruptive Innovation

Disruptive innovation is a business school concept often overused with little reference to the original idea. What does one mean by disruptive innovation in international affairs? The disruptive innovation paradigm argues that small, speculative innovations at the base of the pyramid can often leapfrog and disrupt established domains because top-tier players pursue incremental innovation with their most important, but also most change-resistant clients.[10] There are several examples from big business—Amazon's business model disrupted brick and mortar retail, Skype disrupted long distance telephony, Netflix is disrupting cable broadcasting.

The paradigm is also applicable to international affairs. The main clients of policy analysis and research are bureaucrats or political decision makers whether in government or international institutions. In this context, the potential disrupters are analysts who can make the most of the rise of big and open data.

Both as a field of practice and analysis, international affairs and development present an opportunity for disruption. Recent research into what senior policymakers in international affairs want and expect from researchers is revealing in this regard. The survey, unique in its sampling, targeted senior U.S. national defense policymakers from the George H.W. Bush, Bill Clinton, and George W. Bush administrations.[11] The findings paint an unflattering picture of the increasing gap between the scientific aspirations of international affairs scholarship and the needs of policymakers. The gaps are highest at the top. The study found that the more policymakers know about a subject (especially through direct experience), the less likely they are to believe the "experts."

While the gap between direct experience and academic knowledge may be unsurprising, the survey also points to just how wide some of the generational gaps can be. For instance, it is alarming that, for a survey conducted in 2013, senior policymakers still do not count the Internet as a useful source of policy relevant information. Instead they prefer to gravitate towards individual high profile scholars. A key take-away is that the ease of access to information does not necessarily translate into greater efficacy. Rather, the diffuse nature of the medium—usually seen as a positive in terms of democratizing information—is itself its weakness. A plethora of sources, many of questionable reliability and with no authoritative source among them, can become a barrier for time constrained policymakers who have little interest in "cutting-edge tools and rarified theory."[12] Senior policymakers instead prefer researchers and analysts act as informal advisers and creators of knowledge.

Senior policymakers neither trust nor have time for big data and open data. At the highest levels, most senior policymakers act precisely like incumbent industry leaders or the large risk-averse, change-resistant clients in Christensen's disruptive innovation model. They are highly embedded in the classical paradigm typified by clean, composite, largely linear, and mostly backward-looking indicators, underpinned by methodologically rigorous, often expensive heuristic frameworks that remain firmly within the closed-loop single source of truth paradigm. These clients are serviced by well-heeled advisors steeped in high quality (i.e. expensively vetted) information and data flows. Think of examples ranging from official unemployment figures to inflation to, worse, composite indexes of poverty, and inherently fuzzy concepts like "governance."[13] This paradigm is already being decisively disrupted by the rapid mainstreaming of big and open data, as outlined in the next section.

Much of the data we rely on in international affairs and international development research and analysis is fraught with serious problems and is so slow that it is almost a historical caricature by the time it is published, barely descriptive about the present, let alone insightful about the future. Some of the most important data that we take for granted as "real-time" is not only published with significant lags, but is subject to significant revision. For example, gross domestic product (GDP) is, at best, a quarterly series published with a two-month lag, and revised over the next four years.[14] This, along with the vastly asymmetric influence of a handful of influential senior advisors privy to the closed inner circle of senior decision makers, makes international affairs fertile ground for disruptive innovation.

The "MS Excel error heard around the world" provides a glimpse into what disruption in international affairs research in the age of big data looks like.[15] The most telling aspect about the now infamous Reinhart-Rogoff spreadsheet error, at least from the perspective of research in the age of big data, was not the furor it created by questioning whether high public debt (more than 90 percent of GDP) really has an unusually large effect on growth prospects. The issue was that the researchers were conducting the analysis manually in Excel instead of using reproducible code. Their specific data was not initially "open," but was made so when a graduate student requested it to replicate results—and obviously could not. The story was literally heard around the world, thanks not only to the speed of the spread of information but also due to the high profile of the "experts" involved in a highly polarized debate. A series of small, incremental and—when viewed in isolation—unexpected trends at the base of the pyramid were able to disrupt the asymmetric influence wielded by high profile individual incumbents, even in an unlikely domain such as international affairs research and analysis.

How Big Data and Open Data are Disrupting International Affairs: The Key Element is Time and Speed

In this section we discuss three highly summarized case studies on how big and open data are disrupting research, analysis, and practice in three very different data domains within international affairs: macroeconomic analysis, humanitarian crisis response, and the measurement of poverty and socioeconomic indicators. The common thread in each case is the focus on the impact of big and open data on time and speed.

Real-time economic analysis

Current or near real-time economic analysis is a highly data dependant enterprise. It is highly conservative, in that it is dominated by central banks, ministries of finance, and large private financial institutions. The more timely, accurate, and relevant the data, the better the current assessment and the more valuable it is from a policy perspective. Big data is already disrupting how we collect, compute, and project basic real-time macroeconomic indicators, ranging from GDP and inflation to financial, housing, and labor market indicators.

Recently, many central banks, including the Bank of England, European Central Bank, The Bank of Japan, and the Bank of Canada, have looked into the possibility of leveraging big data to enhance the timeliness of current economic analysis.[16] An interesting innovation in Canada is the use of big data to fill the gaps in the timeliness of official GDP statistics by developing a new short term GDP indicator that provides daily updates of real GDP growth forecasts. Existing monthly data is combined with big data to predict GDP growth before official national accounts data are released for a given quarter, thus bridging the gap period.[17] The example also demonstrates how big data traverses the "official" and "unofficial" domains.

In the case of Japan, the Abe government needed immediate information on time-sensitive policy changes, such as a major increase in the sales tax. What analysts found was that under the existing system there was no way to assess the situation until the household survey or sales data was released and analyzed months later—an eternity in terms of real-time economic analysis. In response, the government proposed the development of a new composite index that would use big data, including online searches and point-of-sale records that would shed immediate light on the impact of policies, albeit not without significant methodological challenges.[18]

Similarly, the Billion Prices Project (BPP) at the Massachusetts Institute of Technology (MIT) demonstrates how big data can be leveraged to provide a real-time gauge of inflation. BPP uses web-scrapers (a relatively simple approach, but one that is highly extensible and adaptable to several uses) to scour websites of online retailers for real-time prices on an enormous range of products. After the collapse of Lehman Brothers in 2008, BPP data showed how businesses started cutting prices immediately. In contrast, official inflation figures did not show deflationary pressures until November.[19] Given the importance of inflation and timely assessment of inflationary expectations from the perspective of monetary policy response, this information represents a significant improvement in response time.[20]

To assume that these innovations are limited to advanced economies would be a mistake. The UN Global Pulse initiative has partnered with BPP and Price Stats to apply the same web-scraping approach in six Latin American countries, specifically to monitor the price of bread and calculate a new eBread Index.[21] Nascent results from the project show the approach can be extended to developing country contexts, and that, in general, the eBread Index is highly

correlated with the official consumer price index for the food basket in these countries. However unlike official inflation data that is available monthly, the eBread Index is available daily. This again is a major improvement in country contexts where inflation and inflationary expectations can change rapidly.

Big data has also been successfully leveraged for a range of other macro indicators. For instance, online search data from Google has been successfully used to predict initial claims for unemployment benefits, consumer sentiment indexes in the United States and United Kingdom, and even car sales down to specific brands.[22] These trends show that companies like Google, Facebook, and Twitter[23] are as important to the future data flow that will fuel policy relevant international affairs research as any national official statistical agency.

The implication is that these companies may be far more important than multilateral data clearing houses such as the World Bank, OECD or UN bodies, on whose highly questionable traditional data—in terms of quality, coverage, granularity and timeliness—much of the current research and analysis in international affairs and development depends. While academics have often pontificated about new and alternative measures of progress, such as the "happiness index," lesser known firms than Google or Facebook, like Jana, are experimenting with SMS based surveys on a global scale that are able to deliver a real-time snapshot of societal well-being.[24]

Humanitarian crises and disaster relief 2.0

The tragic earthquake off the coast of Haiti's capital in January 2010 marked a watershed moment for the impact of big data and open data on disaster relief. The earthquake "created a chasm between what the international humanitarian community knew about Haiti prior to the quake and the reality it faced in the immediate aftermath."[25] The response in Haiti demonstrated an important change in how the huge information gap between damage assessment and response planning was filled. For the first time, two new data inflows were added to the typical crisis response data: one from volunteer and technical communities around the world (principally open source mapping communities like OpenStreetMap, Sahana, CrisisMappers and Ushahidi), and one directly from the affected community of Haitians.[26]

The experience in Haiti showed that the international humanitarian community was not equipped to handle these new information channels, in terms of both speed and complexity.[27] The volunteer technical communities approached the problems in ways that fundamentally challenged the status quo of large humanitarian agencies leading the recovery efforts while the smaller groups follow.

Criticism of the Haiti experience revolves around the overflowing information pipeline.[28] Yet, this is a far better problem than the opposite situation. The ability to learn and rapidly apply lessons in future crises, as discussed below, demonstrates the benefit of having "too much" information. Before focusing on the lessons, it is important to emphasize that the Haitian response proved the rise of big data and open data is not simply about data or technical sophistication. One of the most useful roles played by volunteers was language translation of a huge volume of SMS and other messaging through social media channels. The disruptive innovation was that a highly networked and highly technical, yet contextually aware, virtual community emerged organically. Arguably, the creation of such a community may not have been possible, no matter how many pilot projects were funded by well-meaning donor agencies.[29] One reason is that the problem-solving, transparency-driven, open source mindset that underpins much of the virtual community is not always shared by big bureaucracies and senior

policymakers.[30]

Lessons from the Haitian earthquake have been applied in other contexts. User generated crisis maps have saved lives in subsequent disasters.[31] Volunteers involved in the Haiti mapping project have supported other crowd-sourced mapping initiatives, including projects that emerged in the wake of the earthquake in Chile, floods in Pakistan, the crisis in Libya, the earthquake and tsunami in Japan, and the typhoon in the Philippines. With each experience, the work has gotten better as lessons are rapidly shared within a likeminded, highly motivated, and well organized community. The process of interlinking real-time, geo-spatial crisis data with other relevant data feeds, such as traditional media, has grown exponentially in the past few years. The time taken between crisis impact and information generation has shrunk dramatically compared to historical response times. In the case of Japan, within two hours after the earthquake and tsunami, real-time witness reports were being mapped and shared. In a context where seconds and minutes can determine the difference between life and death, the rise of big and open data and their associated communities has disrupted how society plans humanitarian responses, ensuring such tools will be leveraged in future crises.

The poverty of poverty measures

At the other end of the velocity spectrum are data on typically slow moving measures like poverty. Not only are poverty trends relatively slow moving, at least in comparison to the examples discussed above, but the reporting lags are enormous. The significant lag time of the data bears repeating: When the World Bank announced that 22 percent of the world's population lived on less than $1.25 a day in 2012—and, consequently, the first Millennium Development Goal had been achieved—that data was four years old when reported, dating from 2008.[32]

The data is the poorest where it matters the most. Recent analysis of the state of widely used economic indicators, such as GDP in sub-Saharan Africa, raises serious issues. While international databases like the World Bank report time-series data for many countries, the countries themselves were found to have not published their own data for many of the years covered. Many countries in the region have or are in the process of updating their national income accounts methodology, making these more consistent with what most countries use. In so doing, many are finding a very different picture than they had been led to believe.

For instance, Ghana's 2010 revision showed that GDP was 60 percent higher than expected, instantly catapulting a low income country to middle income status. Research comparing GDP data from country sources with GDP data from the World Bank is alarming. GDP estimates according to national sources in some countries like Burundi (2007) were found to be 32 percent higher than the same reported by the World Bank. However, in other cases the reverse was true, and for Guinea-Bissau in 2006, the World Bank's estimate was 43 percent higher than that of the national authority.[33]

It is important to understand that the problems underpinning these data challenges are not merely an issue of technical capacity, competence, or cost of collection. A far greater problem is perceived or actual interference, whether from political authorities, donors, or other actors. These issues have been aptly termed "the political economy of bad data," which neatly describes the situation in many developing countries.[34] Huge incentives to misreport plague administrative data systems on many levels. For example, when Kenya decided to abolish fees in primary school, this radically changed the incentives for reporting by school administrators, as schools are allocated more teachers and funding if they attract more students. While administrative data from the Ministry of Education shows a steady increase in primary school enrollment rates,

demographic survey and national statistical data fails to confirm the trend and instead indicates enrollment rates have been flat over the same time period.[35]

These findings, while extremely troubling, are made worse by added issues that complicate incentives. For instance, a fast growing trend among donors is cash-on-delivery or performance-based aid, a trend based on the idea of paying for results instead of paying for inputs. Whatever one may think about this conceptually as an aid modality, the fact is that these approaches greatly increase the data burden. In this approach donors pay for development results or outcomes such as increased educational enrollment and improved performance. For performance-based measures to work, organizations need better, more timely, and more granular data. The more ingenuity society can throw at the problem the better.

How are big data and open data disrupting this landscape? Given the context described above, tapping into passively generated and proxy data, if only to triangulate results or provide baseline referential information, could be a welcome innovation. Big data approaches have thrown up three interesting possibilities. The first is analysis of anonymized call detail records (CDRs). A recent project in Cote d'Ivoire, using five million anonymized CDRs from Orange telecommunications customers collected over a five month period, analyzed both the level and location of activity. The analysis indicated that a wider range of calls and longer durations were good proxies for wealth. Using this data, researchers were able to create a granular geospatial estimate of poverty in Cote d'Ivoire—the first data pertaining to a full survey of the country that has been available since the late 1990s due to political strife and economic turmoil in recent years, which have hampered traditional methods.[36]

Another interesting innovation in small scale poverty measurement and prediction is an approach using night light illumination. This approach rests on the assumption that poorer places are quite literally in the dark. Using geospatial, night light, and census data for Bangladesh in 2001 and 2005, researchers showed that a regression model combining the data was able to predict poverty at a granular level. The cost effective and non-intrusive nature of this approach makes it a useful source of proxy poverty data, and makes up for potentially lower accuracy.[37] The concept is also being extended with application to other geographic regions, including in Africa.

A third avenue is high-frequency micro-surveys conducted using mobile phones and other platforms. The World Bank's Listening to Latin America or L2LAC project was launched out of a frustration among policymakers looking for information on the impact of the 2008 economic crisis in Latin America. Typically this sort of analysis depends on household survey data collected and reported over years—and at a high cost. The L2LAC pilot covered nationally representative samples in Peru and Honduras and demonstrated that by using mobile platforms, small versions of wider household surveys can be conducted on a monthly basis and at a fraction of the cost. This provides much closer to "real-time" insights into poverty, employment, inequality, and other trends essential for effective responses to fast moving crises. L2LAC also provides a useful gauge of poverty dynamics and trends between official reporting periods, which can be years apart.[38] The model has since been extended to pilot projects in Africa.[39]

Anonymized CDR analysis, proxy light source data, and mobile phone based micro-surveys are big data innovations that are disrupting how we measure and respond to poverty at various levels. Aspects of each approach have the potential to be "mainstreamed," which would have been unthinkable just a few years ago.

Making the Most of Disruption Requires a New Kind of Policy Entrepreneur

Simon Maxwell, formerly of the Overseas Development Institute, popularized the term "policy entrepreneur" when he proposed a simple self-assessment questionnaire that categorized analysts into four types: *story-tellers* who are steeped in powerful grand narratives that often inform policy; *networkers* who rely mostly on their connections with policymakers; *engineers* who are grounded in testing ideas expected to have policy import; and *fixers* who wield expert power around specific problems.[40]

In an age where big data and open data are fast becoming part of the mainstream in several domains of international affairs, we need to add a new, fifth type of policy entrepreneur: the *disrupter*. The policy entrepreneur who focuses on making the most of disruptive innovation opportunities ushered in by big data and open data typifies the role of the disrupter.

A set of key trends underpin our discussion and make international affairs ripe for disruptive innovation. They are worth repeating. The majority of analyses in international affairs rely on clean, composite, largely linear, and mostly backward-looking indicators that remain firmly within the closed-loop single source of truth paradigm. This, combined with the asymmetric influence of a handful of influential senior advisors privy to the inner circle of senior policymakers, makes international affairs fertile ground for disruptive innovation fueled by the rapid mainstreaming of big data tools.

A key trend underlying the examples discussed in this paper is the dramatically abbreviated learning, development and deployment curves associated with these innovations, and their shrinking marginal costs. A result of these improvements is that sophisticated analytical capacity is becoming more available to an ever wider range of users and analyzers, with a much wider set of applications. The scope of who—whether individual or institution—can be a data generator, aggregator, analyzer, and synthesizer is expanding rapidly.

The rise of big data and open data has shrunk the time lag between the start of a trend, when responders have access to essential information needed to respond to the trend, and the feedback loop generated by those who are affected by the response.[41] Traces of emerging trends show up faster in machine-level exchanges across online data platforms—or in the data stream generated by the nearly 35,000 Facebook likes that brands and organizations receive every minute—than they do in official statistics.[42] As the level of digital activity grows, as social networks become more ossified, and as literacy and awareness around tools increase across the developed and the developing worlds, traditional barriers to adoption of new technologies will fall rapidly. These trends create new analytical and engagement opportunities.

The rise of big data is disrupting clean, composite, backward-looking indicators and the result is often messy, probabilistic, but real-time and forward looking dashboards. We are just at the beginning of a shift in the nature of policy analysis in international affairs, from in-depth research to on the fly analytics. What do these trends mean for a new generation of policy entrepreneurs? How can the new policy entrepreneur make the most of the disruptive innovation potential fueled by big data and open data?

The successful policy entrepreneur will first and foremost invest in staying on top of these rapidly moving trends and their associated tools and technologies. The landscape is evolving fast. Take for instance crowd sourced mapping, which only emerged at scale less than five years ago and is already a mainstream tool in crisis response. To make the most of the disruptive potential, the successful policy entrepreneur will work to break down perceived dichotomies and distinctions, and get comfortable working in polarized terrains. Or consider the

distinction between "public" and "private" sources of information. As we have seen, Google and Twitter can be as important a data and analytical resource as the World Bank or any national statistical agency. Policy entrepreneurs that succeed in blurring and breaking these dichotomies will be best placed to capitalize on the disruptive potential.

Harnessing disparate sources requires investing in tools that help drive interoperability. A lot of potentially disruptive data exists across disparate levels and domains—for example, at the national, local, regional or even neighbourhood levels. Potentially disruptive data also exists in both actual and virtual communities and networks, as well as across government departments and agencies. Furthermore, disruptive data can be found across different data types that may not intuitively work well together—for instance, well-structured data (such as numerical relational tables) and unstructured data (such as Twitter hashtags or search data with multiple and complex interrelationships). There are currently over 300 known open data sites hosted at various levels of government across the world.[43] As a starting point, there remains significant unexploited potential to better link and leverage this fast growing public sector open data, not only to increase efficiency, but to drive entirely new business and service-delivery models.[44]

In addition to tools that stimulate interoperability across disparate sources, types, and levels, the successful policy entrepreneurs will distinguish themselves by having a keen eye on which combination of tools and information source work best for which policy questions. The greatest potential impact may be in areas where experimental alternatives, like triangulating multiple big or open data sources, either significantly improves or fundamentally alters the view presented by standard single-source closed-loop indicators.

With the expanded scope of who can be a data generator, aggregator, or analyzer come inherent problems. These challenges will not only be a major preoccupation of the new policy entrepreneur but will also represent a significant opportunity. A key issue is validation of new and innovative approaches. Here, successful policy entrepreneurs will make the most of their "analog" skillsets and deep contextual awareness. Big data, no matter how powerful, can easily go astray without a domain expertise. Maintaining contextual awareness requires a great deal of offline investment. The importance of building agile teams and maintaining loose but motivated networks with diverse backgrounds and skillsets cannot be overemphasized. The new breed of policy entrepreneur, while versed in data science, cannot be restricted to data. Big data policy entrepreneurs will need to be obsessed with developing innovative validation and verification solutions, with the aim of disrupting the incumbent closed-loop single source of truth paradigm, still emblematic of most international affairs research and policy analysis.

Notes

[1] Gartner, Inc. Research Methodologies, "Hype Cycle Research Methodology," *Gartner, Inc.*, http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp.

[2]Kate Crawford, "The Hidden Biases in Big Data," *HBR Blog Network*, posted April 1, 2013, http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

[3] IBM, "The Four V's of Big Data" *IBM, Inc*., http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg As recently as 2000, less than a quarter of the world's stored information was digital. Now, around 98 percent is digitally stored. There are several similar statistics. A good reference for the scale of the rise of big data is WikiBon, found at http://wikibon.org/blog/big-data-statistics/.

[4] There are several big data evangelists and enthusiasts; see in particular: Kenneth Cukier and Viktor Mayer-Schoenberger, "How It's Changing the Way We Think About the World," *Foreign*

Affairs 92, no. 3 (2013): 27-40; World Economic Forum 2012 Report, "Big Data, Big Impact: New Possibilities for International Development," *World Economic Forum*, http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development; Harvard Business Review October 2012, "Getting Control of Big Data," *Harvard Business Review*; Aniket Bhushan, "Big Data, Democratized Analytics and Deep Context Will Change How We Think and Do Development," *The North-South Institute*, http://www.nsi-ins.ca/wp-content/uploads/2013/04/2012-Big-Data-Democratized-Analytics-and-Deep-Context-will-Change-How-We-Think-and-Do-Development.pdf.

[5] Kate Crawford, "The Hidden Biases in Big Data," *HBR Blog Network*, posted April 1, 2013, http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

[6] Marc Bellemare, "Big Dumb Data," *Marc F. Bellemare Blog*, posted May 16, 2013, http://marcfbellemare.com/wordpress/2013/05/big-dumb-data/. The point is interesting, but questionable. Much of the artificial intelligence and machine learning at the core of big data is in fact underpinned by sophisticated and causal, if only iterative, research designs and models. In fact, there is no reason one couldn't apply more or bigger data to various social science research designs. The greatest potential from a methodological and academic perspective may be found by overcoming some of these gaps.

[7] Konrad Yakabuski, "Big Data should inspire humility, not hype," *The Globe and Mail*, March 4, 2013, http://www.theglobeandmail.com/globe-debate/big-data-should-inspire-humility-not-hype/article9234569/.

[8] Bernholz describes how big data hype has captured the imagination of the commercial and to some extent the public policy domains, but virtually ignores civil society, a sector that generates a great deal of significantly untapped data. See: Lucy Bernholz, "Civil Society and Big Data," *Philanthropy 2173 Blog*, posted February 24, 2014, http://philanthropy.blogspot.ca/2014/02/civil-society-and-big-data.html?m=1.

[9] Michael Schrage, "Big Data's Dangerous New Era of Discrimination," *HBR Blog Network*, posted January 29, 2014, http://blogs.hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination/.

[10] Clayton Christensen, "Disruptive Innovation," Clayton Christensen's personal website, http://www.claytonchristensen.com/key-concepts/.

[11] Paul Avey and Michael Desch, "What Do Policymakers Want From Us? Results of a Survey of Current and Former Senior National Security Decision-makers," *International Studies Quarterly* 58, no. 4 (2014) Forthcoming.

[12] Paul Avey and Michael Desch, "What Do Policymakers Want From Us? Results of a Survey of Current and Former Senior National Security Decision-makers," *International Studies Quarterly* 58, no. 4 (2014): 34. Forthcoming. (Page number may differ in final edition.)

[13] Several examples come to mind, but think of highly composite measures like the Human Development Index (HDI) or the World Governance Indicators (WGI) that are popular in international affairs research and rely on such untimely data that they are more a historical caricature than an insightful tool for understanding coming trends.

[14] Similarly, inflation data is published monthly, but with a three week lag after the reporting month. See: Nii Ayi Armah, "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf.

[15] The Rachel Maddow Show, "The Excel Error Heard Around the World," *MSNBC*, http://www.msnbc.com/rachel-maddow-show/the-excel-error-heard-round-the-world. See also: Thomas Herndon, Michael Ash and Robert Pollin, "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff," (working paper no. 322, PERI University of Massachusetts, Amherst).

[16] See: Mitsuru Obe, "Japan Looks to Big Data for Timely Economic Indicator," *The Wall Street Journal*, September 24, 2013; L. Einav and J. D. Levin, "The Data Revolution and Economic Analysis," (working paper no. 19035, *National Bureau of Economic Research*); A. Binette and J. Chang, "CSI: A Model for Tracking Short-Term Growth in Canadian Real GDP," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-binette.pdf; Nii Ayi Armah, "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf.

[17] Binette and J. Chang, "CSI: A Model for Tracking Short-Term Growth in Canadian Real GDP," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-binette.pdf; Nii Ayi Armah, "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf.

[18] Mitsuru Obe, "Japan Looks to Big Data for Timely Economic Indicator," *The Wall Street Journal*, September 24, 2013.

[19] Nii Ayi Armah, "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf.

[20] A counterargument often made is that focusing on frequency tends to overestimate both inflation and the variability of inflation. This is particularly the case in other measures such as the Everyday Price Index, which measures inflation from the perspective of goods that people purchase frequently. However, these critiques can be easily countered by statistically correcting for additional variability inherent in higher frequency data. For more information, see: Matthew Yglesias, "Here's a Deliberately Inaccurate Inflation Index," April 9, 2013, http://www.slate.com/blogs/moneybox/2013/04/09/aier_s_everyday_inflation_index_is_terrible.html

[21] UN Global Pulse, PriceStats, and the Billion Prices Project, "Daily Tracking of Commodity Prices: the eBread Index," http://www.unglobalpulse.org/projects/comparing-global-prices-local-products-real-time-e-pricing-bread.

[22] Hyunyoung Choi and Hal Varian, "Predicting Initial Claims for Unemployment Benefits," Google Inc., http://static.googleusercontent.com/media/research.google.com/en/us/archive/papers/initialclaimsUS.pdf; Hyunyoung Choi and Hal Varian, "Predicting the Present with Google Trends," Google Inc., http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf; Paul Cheung, "Big Data, Official Statistics and Social Science Research: Emerging Data Challenges," (presentation at the World Bank, December 12, 2012).

[23] Mining Twitter data on food related conversations has also been shown to be strongly correlated with food price inflation. See: World Economic Forum 2012 Report, "Big Data, Big Impact: New Possibilities for International Development," *World Economic Forum*,

http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development.

[24] JANA and UN Global Pulse, "Global Snapshot of Well Being - Mobile Survey," *United Nations*, http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf.

[25] UNOCHA et al., "Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies," *United Nations Foundation*, http://issuu.com/unfoundation/docs/disaster_relief20_report.

[26] UNOCHA et al., "Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies," *United Nations Foundation*, http://issuu.com/unfoundation/docs/disaster_relief20_report; See also: Kim Rose, "The humanitarian power of big data," *Hortonworks Blog*, posted May 6, 2013, http://hortonworks.com/big-data-insights/the-humanitarian-power-of-big-data/; Maja Bott, Björn-Sören Gigler, and Gregor Young, "The Role of Crowdsourcing for Better Governance in Fragile State Contexts," International Bank for Reconstruction and Development/World Bank, https://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/crowdsourcing_final_0.pdf; Maja Bott and Gregor Young, "The Role of Crowdsourcing for Better Governance in International Development," *The Fletcher Journal of Human Security* 27 (2012): 47-70.

[27] During the first week, volunteers mapped some 1,600 reports from affected Haitians based on information from Twitter, Facebook, and online news. Over 30,000 SMS messages from affected Haitians were sent through the Ushahidi led Project 4636 in the first month alone.

[28] Maja Bott, Björn-Sören Gigler, and Gregor Young, "The Role of Crowdsourcing for Better Governance in Fragile State Contexts," *International Bank for Reconstruction and Development/ World Bank*, https://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/crowdsourcing_final_0.pdf.

[29] It is worth noting that once innovations like open source crowd mapping prove their utility, donors and others see value in supporting and facilitating them from the "inside." A great example of this is USAID's recent MapGive initiative, which encourages volunteers to learn and get involved in crowd mapping using the OpenStreetMap platform: http://mapgive.state.gov/index.html

[30] The Standby Volunteer Task Force for Live Mapping (SBTF), an online volunteer initiative for crisis mapping that was founded as a consequence of the various loosely connected projects for Haiti's recovery.

[31] Jing Guo, "How User Generated Crisis Maps Save Lives in Disasters," *The World Bank Blog*, posted February 26, 2014, http://blogs.worldbank.org/publicsphere/how-user-generated-crisis-maps-save-lives-disasters.

[32] Johan Mistiaen, "What will it take to improve poverty data?" *The World Bank: MDGs and Beyond 2015* (2012): 2.

[33] Jerven Morton, "African Growth Miracle or Statistical Tragedy? Interpreting trends in the data over the past two decades," (paper presented at UNU-WIDER conference on Inclusive Growth in Africa, September 20-21, 2013). More worryingly, as the paper shows, in some cases base year revisions are nearly three decades apart, implying that we should not be surprised if we see a number of countries go through a statistical inflation in fundamental data like GDP, as in the recent experience in Ghana.

[34] Justin Sandefur and Amanda Glassman, "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics," (paper presented at UNU-WIDER conference on Inclusive Growth in Africa, September 20-21, 2013).
[35] Justin Sandefur and Amanda Glassman, "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics," (paper presented at UNU-WIDER conference on Inclusive Growth in Africa, September 20-21, 2013). We should also note that analyses such as these are made far easier by the spread of open data including administrative level data. Kenya is viewed as a leader in this regard among developing countries.
[36] Christopher Smith, Afra Mashhadi, and Licia Capra, "Ubiquitous Sensing for Mapping Poverty in Developing Countries," (paper submitted to the D4D session of the NetMob 2013 conference). UN Global Pulse, "Mobile Phone Network Data for Development," October 2013, http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf.
[37] Prasanna Lal Das, "Scenes from the DC big data dive- the final report," *The World Bank Blog*, posted May 28, 2013, http://blogs.worldbank.org/opendata/scenes-dc-big-data-dive-final-report.
[38] A. Ballivian and J. Azevedo, "Listening to LAC: Using Mobile Phones for High Frequency Data Collection," *The World Bank*.
[39] An example at the metropolis level is the Listening to Dar project in Dar es Salaam, Tanzania.
[40] Simon Maxwell, "Policy Influence: Policy entrepreneurs," *Overseas Development Institute Blog*, posted January 2009, http://www.odi.org.uk/publications/5896-simon-maxwell-engineer-networker-fixer-storyteller-policy-entrpreneurship.
[41] World Economic Forum 2012 Report, "Big Data, Big Impact: New Possibilities for International Development," *World Economic Forum*, http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development.
[42] Nii Ayi Armah, "Big Data Analysis: The Next Frontier," *Bank of Canada Review*, Summer 2013, http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf.
[43] For more see Open Data Sites: http://opendatasites.com/
[44] A good example is the emergence of public sector open data based companies and business models. In the United States, real estate sector companies Trulia and Zillow come to mind. There are also innovative business models emerging around searching, indexing, meta-tagging, and visualizing open public data, like Evision.io.