DATA-POP ALLIANCE

Inputs to the Big Data and SDGs Chapter of the 2015 Global Sustainable Development Report

Reflections on Big Data and the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era

February 2015

Working Note

DATA-POP ALLIANCE

Reflections on Big Data and the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era

February 2015

Working Note

Reflections on Big Data & the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era

Input to the Big Data and SDGs Chapter of the 2015 Global Sustainable Development Report

(Working document—Not to be quoted without permission)

Emmanuel Letouzé

February 18th, 2015

Whether and how Big Data could loosely "contribute to the SDGs"—in other words, and in broad terms, the collision and intersection of these two hot topics in the public development discourse—has received significant attention in recent months. Much of this attention focuses on ways in which Big Data may help monitor the SDGs—with examples of such uses being routinely put forward, in the area of poverty monitoring using cell-phone activity using Call Detail Records (CDRs) analytics, for instance. By and large, the discussions are almost exclusively about *measurement*, and much less attention is paid to issues of *achievement*.

Quantifying and tracking development processes using 'new kinds of data' emitted by humans, the bulk of which corresponds to what Sandy Pentland has referred to as 'digital breadcrumbs', i.e. passively emitted structured data like credit card of phone transactions, and new analytical techniques falling under the umbrella of Big Data analytics, indeed holds real promise. It is now largely undisputed that "Big Data can provide alternative measures of poverty and welfare".[i]

As is now well known but is worth restating, dosing so it also carries significant challenges and uncertainties, methodological, technical, institutional, as well as areal risks to individual and group privacy and safety. Further, the measurement approach alone does not cover the whole spectrum of ways in and channels through which Big Data as an entirely new ecosystem could impact—contribute to or hamper—human progress as called for and measured by the SDGs.

To get a better picture of the full scope of the question, it is useful to unpack its terms and expose and discuss its underlying assumptions first, and limitations then. As mentioned above, for the most part the 'Big Data and SDG' question is unfortunately framed as a measurement or monitoring issue. It reflects the argument that new kinds of digital data, structured (such as CDRs, or tweets metadata) and unstructured (like the content of tweets) could be analyzed to 'say something', 'yield insights on'—complement, substitute—other more traditional ways of measuring facets of human reality, be it mortality, violence, hunger, etc. As hinted above, there is already a large and growing body of evidence that Big Data indeed holds this potential—as the examples listed in the attached table summarizes quite well.

First, some sectors and related SDGs seem more amenable than others to being monitored with and by Big Data—where Big Data here is used not as big data alone, nor as big data + analytics, but as the ecosystem of new data, new tools and new actors, which I have called the 3Cs of Big Data for Crumbs, Capacities and Community. These 'low hanging fruits' will generally be sectors and processes (and related SDGs) that are

1. correlated with trends and patterns in data production of some kind (we all pay greater attention to our phone of electricity consumption when facing financial constraints) and
2. that are currently monitored through traditional means (providing ground truth).

It is also likely that SDGs deemed more 'important' (e.g. poverty rather than overfishing) by public opinions will receive greater attention.

But Big Data could also be quite applicable to 'new' kinds of sectors and goals. It is clear that social media chatter and use hold potential in the realm of social cohesion analysis (falling under Goal 16 right now, for instance), as does criminality prediction. It is also worth noting that other uses—such as the tracking of Illegal, Unreported and Unregulated (IUU) fishing through satellite imagery analysis[ii], falling under Goal 14—is already a reality, and could become standard practice in the next few years in some regions.

The simple measurement approach—using Big Data to monitor the SDGs, fill in gaps etc.—is more than less explicitly assumes that one can only manage what one can measure, and that measuring goes a long way towards impacting. This double assertion obviously comes with lots of caveats in terms of what is necessary vs. sufficient, but the SGDs, as the MDGs before them, would not have existed in the absence of a broad consensus on the fact that monitoring a variable (measuring and tracking) mattered, in a Przeworskien sense[iii], i.e. had a causal even if indirect effect on what was being measured. The underlying argument is that doing so can be used for advocacy purposes, shapes incentives, inform policies etc. A case in point is GDP, which, since its invention in the 1930s to track the effect of post-Great Depression macro economic policies, has become the alpha of omega of economic policies—making its presence so central it need not be mentioned (rather we talk about 'growth', or 'the economy'). Governments have since the designed policies primarily meant to increase a variable created because it was easy to measure in an industrial era—even as it has had nefarious effects on the environment.

The role of measurement is indeed also evidenced by what happens to outcomes that cannot or are not measured—which are by definition or design statistically invisible: as suggested above the ecologically detrimental effects of pro (GDP) growth policies are now well known, and led to the development of sustainability indicators, some of which capturing fiscal sustainability that take into account national savings and investment. And so it is hard to deny that measuring does or can matter. But it is neither a sufficient—which everybody agrees with—nor a necessary condition (which I will come back to), and it also very much depends on the quality of the measurement—where 'quality' refers to the 'qualitative nature' of the measurement, including but beyond its being accurate vs. inaccurate, considering process and agent issues.

If we stay for a moment in this realm of 'SDG monitoring through Big Data', a few technical and institutional considerations are worth making. First, not everybody has a cell phone or uses Twitter, such that Big Data streams and sets, even with huge N, are typically non-representatives samples of the entire populations of interest. Sample bias correction methods are being developed, which typically require ground truthing data to be tested.

This stream of work will use a blending of statistical hypothesis based approaches (we can assume that Twitter data is even less representative in Niger than in Norway) and machine-learning techniques (we may find that geographic elevation plays a significant role in determining the size and sign of sample bias when using cell-phone data activity to infer population density for instance).

It should be noted however that no-representative samples can yield representative outcomes; surveying millionaires to know whether they would prefer a good meal with friends vs. being sent to jail would probably yield responses that would not differ much from those of the general population. But in most cases that will not be the case.

A lot of investments and work will need to go into developing rigorous and robust methods of that kind if Big Data, building on existing attempts[iv], if it is to be used widely for monitoring purposes on a sustained basis. Simple machine-learning approaches on which most pilots, proofs-of-concepts and fancy dashboards are based tend to have 2 main flaws that greatly limits their scaliblity and validity over time:

1. High data requirements
2. Limited external validity

Another related challenge, and a precondition, is accessing the data—to the 'big data' o Big Data—in a sustainable and stable manner. At the minute, a large chunk of what is commonly referred to as Big Data (or rather, big data) are held and often legally owned by the private sector companies that collect them—and analyze them for their own purposes. The data provided to outside organizations and individuals are often aggregated and 'anonymized'– although we also know the concept itself is more than problematic[v], either as part of data challenges or through personal connections; in some cases they can only be accessed on site.

There are of course good reasons not to share all personal data publically, but the point is that the current way data sharing is done is ad hoc, unstable (in the sense that it offers little predictability on future data access) and unsustainable (it will not last). The development and research communities are already eager to develop and test innovative approaches to SGD monitoring through partnerships, via one-shot pilots in most cases.

It will be important to ensure that issues of data access and stability are 'internalized' in developing and planning for the uptake and scaling up of these pilots.

Another important and also related consideration that veers into political territories is that of the rights to the data. Holding and owning are different concepts; a bank holds our money but does not own it. It can use and generate value from it by investing it, but it stays ours (hopefully). The analogy is not water-tight but there is currently a great deal of thinking around the future of the global data system's legal framework, starting with a greater recognition that in principle and in the future people should have greater control of the rights to their data—the data they emit. What a 'New Deal on Data' may exactly look like is yet unclear, but what is certain is that it won't look like today's patchy and blurry landscape. Either because 'we' have collectively found ways to collect, store and share people's data in ways they are broadly happy with, or because the system will shut down, or implode.[vi]

Let's reiterate the initial question as a way to turn our attention to its other side, or the other piece of the puzzle: How can/could Big Data *contribute* to the SDGs; i.e. how could Big Data improve the outcomes measured by the SDGs. One way, as we have just discussed, is by contributing to monitoring these SDGs. But as mentioned too, monitoring alone is not sufficient, and the specific question of how and by whom measurements can be leveraged to influence policies is still largely open, and part of the larger question.

For one, there will or could still be tensions between SDGs that simply monitoring all of them in the best of ways won't address. Groups will also denigrate some of the goals and play on these tensions—arguing that environmental protection comes at the expense of poverty eradication for instance. So, politics will remain the elephant on the picture, in ways that will be even more complex if and as 'we' have more goals and (possibly) better measures of progress towards them. Arbitrating will not go smoothly, for sure.

There is also the added and centrally important fact that—as noted above already: Big Data is not just big data—but also tools and techniques that are largely developed and mastered outside the reach and realm of traditional policymaking. I have argued that Big Data is best captured through its 3Cs of Crumbs (data), Capacities (tools and methods) and Community (that of emitters, analysts and users—the human element), constituting a complex system in its own right. Considering Big Data as such an ecosystem made up of these interlaced interacting elements adds depth and breath to the question; it simply brings up other questions.

A first subsequent question is who will be 'using' (leveraging, shaping, the ecosystem of) Big Data for the purposes of monitoring? Will it be primarily governments? Which ones? Democracies or autocracies? UN agencies? With which track-record and means? Private companies? For what prupose? Specialized NGOs? With which checks and balances and accountability and replicatility mechanisms? A combination of all under new Public-Partner (or Public-Partner-People) Partnerships? How, when? This clearly gets to the institutional shift that may arise with the rise of Big Data as an ecosystem; not just making is possible and/or easier to monitor better using the same actors and channels as before, the same system, but by bringing out new dynamics.

The subsequent question is: how could Big Data contribute *directly* to the SDGs—irrespective of their near-exogenous effect via monitoring—endogenously, so to speak? It is implicitly or explicitly argued that poverty can be to a good extent causally attributed to poor poverty data (the "statistical tragedy" and pilots-with-no-instrument arguments[vii]), such that getting better poverty data would necessarily contribute to reducing the object or outcome it measures—poverty—through some causal mechanisms. But—and this is where the non-necessity argument comes in, this has 2 caveats:

1. Measuring is certainly not the only way to incentivize and spur change, and in many cases it can be argued it has no effect if responsive social and institutional mechanisms aren't in place;
2. The correlation between good data and good outcomes also obviously reflects reverse causation and spurious correlations. In the hypothetical case that Norway, which tops the HDI ranking, decided to stop measuring children mortality rates for 20 years, it is doubtful that it would cause it to rise significantly. Norway has good data and low poverty because of confounding factors—chief of which its being a democratic society with lots of oil.

There are several direct ways in which Big Data—both as data and more importantly as an ecosystem—could contribute to socioeconomic changes that would positively affect the outcomes captured in the SDGs. One is by leading to changes in traffic laws and behaviors that would curb congestion for instance. This example shows that it is not all about policies but can affect change through people's behaviors directly. Although hard if not impossible to quantify, it is likely that a good share of this *direct positive effect* of Big Data on outcomes measured by the SDGs will be attributable to non-policy actions—simply by people using insights and suggestions derived from Big Data, such as Google Maps estimates, algorithmic recommendations of when to see a doctor, etc. These are largely unrelated to policies but remain in the realm of 'applications'—ways in which Big Data helps 'do stuff, concrete tasks, more effectively etc.

Another means (and a precondition for this to happen) through which Big Data will impact outcomes is through its effect on people's empowerment. Big Data can definitely disempower people—with the technocrato-technological notion that it will provide a 36,000-foot and 360-degree solution to all of the world's problems, and that 'people' cannot and need not understand Big Data, which enlightened leaders will best know how to use for the greater good (a real issue in the 'data for development' movement). This approach is quite prevalent in the current 'Data Revolution' discourse, where the light is ironically expected to come from UN Member States and agencies, despite their mixed track record when it comes to bringing about positive social change—for a host of reasons ranging from elite capture and corruption to basic inefficiency and lack of agility. The truth is that the Data Revolution has also provided a terrific excuse for development professionals and professional politicians: we have failed because we didn't have the data.

But Big Data can and must also be a force for greater political empowerment and societal change. For one, as it diffuses to the masses, governments and industry leaders will increasingly know or fear they can and will be held to greater accounts, that lying, stealing or slacking will come at a greater expected price; also if or rather as people get greater control over the rights to their data, changes in balances of power can be expected happen.

In the future Big Data-based or –related policies will have an important impact on societies (and the SDGs), but I suspect that Big Data-related or –informed laws may have an even greater impact. Clearly a world where every individual is the ultimate owner of the rights to their data will be a very different world than the one we know today, and whether and how the SDGs are monitored, met or not met will be consequently changed significantly.

Most of this will require a data-educated citizenry; societies of sophisticated users. Data literacy has emerged as the new buzzword. What it means remain unclear; but one thing must be clear: it is *not* the ability to use data—being able to do so is neither a necessary nor a sufficient condition of being data literate. By that standard indeed, a low-level NSA analysis is highly data literate—but this is probably not the only skills a thriving democratic and developed society would want to promote and monitor. Actually the concept of data literacy may be flawed in 2 ways:

1. It should not be defined in isolation of the purpose it is expected to serve—spurring social progress and development and
2. It places to much of an emphasis on technical skills and overlooks the need for citizens to understand what is done with their data, and when and how data literacy is about refusing to use data.

For these reasons I prefer thinking about the importance of and requirements for promoting *literacy in the age of Big Data* and see this as a building block to contribute to the SDGs through Big Data. This will require much more sustained and strategic engagement and investments and be much more disruptive of current decision-making processes and political structures than developing Big Data and SDG pilots, which may nonetheless serve a real purpose, but if and only if they are part of a broader and more complex vision.

Annex: Uses of Big Data for SDG monitoring

SDGs adopted by the OWG	Big data examples	What is monitored	How is monitored	Country(ies)	Year	Advantages of using big data
1. Poverty eradication	Satellite data to estimate poverty[viii]	Poverty	Satellite images, night-lights	Global map	2009	International comparable data, which can be updated more frequently
	Estimating poverty maps with cell-phone records[ix]	Poverty	Cell phone records	Cote d'Ivoire	2013-4	
	Internet-based data to estimate consumer price index and poverty rates[x]	Price indexes	Online prices at retailers websites	Argentina	2013	Cheaper data available at higher frequencies
	Cell-phone records to predict socio-economic levels[xi]	Socio-economic levels	Cell phone records	"Major city in Latin America" (Actually Mexico-City)	2011	Data available more regularly and cheaper than official data; informal economy better reflected
2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture	Mining Indonesian Tweets to understand food price crises[xii]	Food price crises	Tweets	Indonesia	2014	
	Uses indicators derived from mobile phone data as a proxy for food security indicators[xiii]	Food security	Cell phone data and airtime credit purchases	A country in Central Africa	2014	
	Use of remote-sensing data for drought assessment and monitoring	Drought	Remote sensing	Afghanistan, India, Pakistan[xiv]	2004	
				China[xv]	2008	
3. Health	Internet-based data to identify influenza breakouts[xvi]	Influenza	Google search queries	US	2009	Real-time data; captures disease cases not officially recorded; data available earlier than official data
	Data from online searches to monitor influenza epidemics[xvii]	Influenza	Online searches data	China	2013	
	Detecting influenza epidemics using twitter[xviii]	Influenza	Twitter	Japan	2011	
	Monitoring influenza outbreaks using twitter[xix]	Influenza	Twitter	US	2013	
	Systems to monitor the activity of influenza-like-illness with the aid of volunteers via the internet[xx,xxi]	Influenza	Voluntary reporting through the internet	Belgium, Italy, Netherlands, Portugal, United Kingdom, United States	ongoing	
	Cell-phone data to model malaria spread[xxii]	Malaria	Cell-phone data	Kenya	2012	
	Using social and news media to monitor cholera outbreaks[xxiii]	Cholera	Social and news media	Haiti	2012	
	Google dengue trends[xxiv,xxv]	Dengue	Web search queries	Argentina, Bolivia, Brazil,India, Indonesia, Mexico, Philippines, Singapore, Thailand, Venezuela	ongoing	
	Monitoring vaccine concerns to help tailor immunization programs[xxvi]	Vaccine concerns	media reports (e.g., online articles, blogs, government reports)	144 countries	2013	Data not available otherwise; expensive to collect data through survey
	Monitoring vaccine concerns[xxvii]	Vaccine concerns	Twitter	US	2011	
	Analysis of Twitter used to track HIV incidence and drug-related behaviours[xxviii]	HIV, drugs use	Twitter	US	2014	
7. Energy	Satellite data to estimate electric power consumption[xxix]	Electric power consumptio n	Satellite images	21 countries	1997	Regular updates
8. Economy and macroeconomic stability	Light emissions picked up by satellites to estimate GDP growth[xxx]	GDP growth	Satellite images	30 countries;	2012	Informal economy better reflected; information available at sub-national level; improves estimates for countries with poor national accounts data
	Using night-lights to estimate GDP at sub-national levels[xxxi]	GDP at sub-national levels	Satellite images	China, India, Turkey, US	2007	
	Internet-based data to monitor inflation in real time[x]	Inflation	Prices from online retailers	Argentina, Brazil, Chile, Colombia, Venezuela	2012	Cheaper data available at higher frequencies
9. Build resilient infrastructure, promote inclusive and sustainable	Map showing internet devices which could be logged in using default passwords or no passwords. Despite biases towards unsecure devices, the	Map with internet devices by location	Internet tools to scan all addresses of the fourth version of the internet protocol	World	2012	Easier, cheaper, quicker than internet use surveys. Disadvantages: illegal and may not be able to be

industrialization and foster innovation	map may reflect online usage around the world.[xxxii]					reproduced with the newest internet protocols
10. Reduce inequality within and among countries	Mapping socio-economic status by analysing airtime credit and mobile phone datasets[xxxiii]	Wealth and inequality	Airtime credit purchases	Cote d'Ivoire	2013	Disadvantage: no ground truth data to compare it with (last censuses unreliable)
11. Make cities and human settlements inclusive, safe, resilient and sustainable	Light emissions picked up by satellites to estimate urban extent[xxxiv]	Urban extent	Satellite images	Global	2005	Globally consistent way to map urban extent; more regular updates
	Use of data from transport cards to construct a picture of individual journeys and how the bus and train networks are used by the public[xxxv]	Transport use and journeys	Transport cards data	London, UK		More detailed and more frequent than survey data
	Times series of satellite images of flooded areas are used to identify flood risk areas[xxxvi]	Flood hazard and risk	Satellite images	Namibia	2014	Data available frequently
	Analysis of the temporal evolution of nightlights along the river network to obtain a global map of human exposure to floods[xxxvii]	Night lights as a proxy for population/ infrastructure along the river network	Satellite images	Global	1992-2012	
	Using satellite imagery, GIS and precipitation data to produce a flood risk map along the Niger-Benue river[xxxviii]	Flood risk	Satellite images	Nigeria, Niger-Benue River	2014	
	Using satellite remote sensing and GIS techniques for flood hazard and risk assessment in Chamoli district, Uttarakhand, India[xxxix]	Flood hazard and risk	Satellite images	Chamoli district, Uttarakhand, India	2014	
	Assessing flood impact with cell phone records[xl]	Flood impact	Cell phone records	Mexico	2014	
	Analysis of Twitter data during hurricane Sandy to identify which data may be useful in disaster response[xli]	Tweets about the hurricane	Twitter	New York, US	2012	
13. Climate change	Satellite scan to monitor population and energy related greenhouse gas emissions[xlii]					Separate emissions of urban populations from other sources; more regular updates
	Satellite images to measure net primary production[xliii,xliv]					Regular updates
	Methane observations made from space combined with Earth-based remote sensing column measurements[xlv,xlvi]	Methane	Satellite measurements	US	2014	
16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels	Use of mobile phone and demographic data to predict crime in London[xlvii]	Crime	Mobile phone and demographic data	London, UK		
	See also http://www.ipinst.org/media/pdf/ publications/ipi_epub_new_technology _final.pdf					
	Using the 'Global Data on Events, Location and Tone (GDELT)', a news stories dataset, to crunch the numbers of violent events in a conflict[xlviii]	Violent events	News stories database	Syria	2013/4	
Measures beyond GDP	Cell-phone records to predict socio-economic levels[xlix]					Data available more regularly and cheaper than official data; informal economy better reflected

Source: Martinho and Letouzé (2015)

Endnotes

[i] Letouzé, 2013. http://www.developmentprogress.org/blog/2013/06/11/could-big-data-provide-alternative-measures-poverty-and-welfare
[ii] http://www.economist.com/news/leaders/21640350-big-data-allow-fish-be-protected-never-governments-should-take-advantage
[iii] http://politics.as.nyu.edu/docs/IO/2800/go_2004.pdf
[iv] Zagheni and Weber, 2012, Letouzé, Zagheni et al, 2015
[v] http://newsoffice.mit.edu/2015/identify-from-credit-card-metadata-0129
[vi] On this see http://www.datapopalliance.org/blog/2015/1/29/big-data-and-human-rights-ema-new-and-sometimes-awkward-relationshipem-to-be-further-explored and http://www.datapopalliance.org/blog/2014/10/16/harvard-business-review-interview-with-alex-sandy-pentland-on-big-data-the-internet-of-things-privacy-and-the-new-deal-on-data and http://www.datapopalliance.org/s/WhitePaperCDRsEthicFrameworkDec10-2014Draft-2.pdf
[vii] Giugale, 2012
[viii] Christopher D. Elvidgea, Paul C. Suttonb, Tilottama Ghoshc, Benjamin T. Tuttlec, Kimberly E. Baughc, Budhendra Bhadurid, Edward Brightd (2009). A global poverty map derived from satellite data, Computers & Geosciences, vol 35, issue 8, pp. 1652-60.
[ix] http://www0.cs.ucl.ac.uk/staff/l.capra/publications/chi14-chris.pdf and http://www.cities.io/wp-content/uploads/2012/12/d4d-chris-submitted.pdf
[x] Cavallo, A., (2013). Online and official price indexes: Measuring Argentina's inflation. Journal of Monetary Economics, 60, pp 152-165.
[xi] Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E., (2011). Prediction of Socioeconomic Levels using Cell Phone Records. Lecture Notes in Computer Science, 6787, pp 377-388.
[xii] http://www.unglobalpulse.org/sites/default/files/Global-Pulse-Mining-Indonesian-Tweets-Food-Price-Crises%20copy.pdf
[xiii] A Decuyper, A Rutherford, A Wadhwa, JM Bauer, G Krings, T Gutierrez, ... (2014). Estimating Food Consumption and Poverty Indices with Mobile Phone Data. http://arxiv.org/pdf/1412.2595.pdf
[xiv] Thenkabail,P.S., Gamage,M.S.D.N.and Smakhtin,V.U. (2004). The Use of Remote Sensing Data for Drought Assessment and Monitoring in Southwest Asia. International Water Management Institute, Research Report 85. http://www.iwmi.cgiar.org/Publications/IWMI_Research_Reports/PDF/pub085/RR85.pdf
[xv] Renhua Zhang, Hongbo Su, Jing Tian, Zhaoliang Li, Shaohui Chen, Jinyan Zhan, Xiangzheng Deng, Xiaomin Sun, and Jianjun Wu (2008). Drought Monitoring in Northern China based on Remote Sensing Data and Land Surface Modeling. Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, Volume: 3. https://www.researchgate.net/publication/224383482_Drought_Monitoring_in_Northern_China_based_on_Remote_Sensing_Data_and_Land_Surface_Modeling
[xvi] Ginsberg, J. , Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature, 457, pp 1012-1014.
[xvii] Yuan, Q., et al. (2013). "Monitoring influenza epidemics in china with search query from baidu." PLoS One 8(5): e64323.
[xviii] Aramaki E, Maskawa S, Morita M. Proceedings of the conference on empirical methods in natural language processing. Edinburgh: Association for Computational Linguistics; 2011. Twitter catches the flu: detecting influenza epidemics using Twitter. Available at: http://www.aclweb.org/anthology/D/D11/D11-1145.pdf
[xix] Lamb, A., et al. (2013). "Separating Fact from Fear: Tracking Flu Infections on Twitter." North American Chapter of the Association for Computational Linguistics (NAACL), 2013.
[xx] www.influenzanet.eu
[xxi] www.flunearyou.org
[xxii] Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., Buckee, C. O., (2012). Quantifying the impact of human mobility on malaria. Science, 338(6104), pp 267-270.
[xxiii] Chunara, R., et al. (2012). "Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak." Am J Trop Med Hyg 86(1): 39-45.
[xxiv] http://www.google.org/denguetrends/intl/en_us/
[xxv] Emily H. Chan, Vikram Sahai, Corrie Conrad, John S. Brownstein (2011). Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLOS Neglected Tropical Diseases, DOI: 10.1371/journal.pntd.0001206.
[xxvi] Larson, H. J., et al. (2013). Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyse public concerns about vaccines. Lancet Infect Dis 13(7): 606-613.
[xxvii] Salathe, M. and S. Khandelwal (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 7(10): e1002199.
[xxviii] Sean D. Young, Caitlin Rivers, Bryan Lewis (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes, Preventive Medicine, Volume 63, Pages 112-115.
[xxix] Elvidge, C. et al., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing, 18(6), pp. 1373-1379.

[xxx] Henderson, V., Storeygard, A. & Weil, D., 2012. Measuring Economic Growth from Outer Space. American Economic Review, 102(2), pp. 994-1028.
[xxxi] Sutton, P., Elvidge, C. & Ghosh, T., 2007. Estimation of Gross Domestic Product at Sub-National Scales using Nighttime Satellite Imagery. International Journal of Ecological Economics & Statistics, 8, pp. 5-21.
[xxxii] Anonymous (2012). Internet Census 2012 Port scanning /0 using insecure embedded devices. Available at: http://internetcensus2012.bitbucket.org/paper.html
[xxxiii] http://arxiv.org/pdf/1309.4496.pdf
[xxxiv] Small, C., Pozzi, F. & Elvidge, C., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. Remote Sensing of Environment, Volume 96, pp. 277-291.
[xxxv] Mark Say (2014). Transport for London builds customer profiles using RFID data analytics. RFID Data Analytics. Available at: http://www.computerweekly.com/feature/Transport-for-London-builds-customer-profiles-using-RFID-data-analytics
[xxxvi] Skakun S1, Kussul N, Shelestov A, Kussul O. (2014). Flood hazard and flood risk assessment using a time series of satellite images: a case study in Namibia. Risk Anal. 2014 Aug;34(8):1521-37. doi: 10.1111/risa.12156.
[xxxvii] Serena Ceola, Francesco Laio and Alberto Montanari (2014). Satellite nighttime lights reveal increasing human exposure to floods worldwide. Geophysical Research Letters
Volume 41, Issue 20, pages 7184–7190.
[xxxviii] A. Akinbobola, E. C. Okogbue and O. Olajiire (2014). A GIS-based flood risk mapping along the Niger-Benue river basin in Nigeria using watershed approach. 20th International Congress of Biometeorology (28 September–2 October 2014). Available at: https://ams.confex.com/ams/ICB2014/webprogram/Paper251796.html
[xxxix] G.D. Bhatt, Komal Sinha, P.K. Deka, Ajay Kumar (2014). Flood hazard and risk assessment in Chamoli District, Uttarakhand using satellite remote sensing and GIS techniques. I
Nternational Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 8. Available at: http://www.ijirset.com/upload/2014/august/39_Flood.pdf
[xl] D Pastor-Escuredo, A Morales-Guzmán, Y Torres-Fernández, JM Bauer, ... (2014) Flooding through the lens of mobile phone activity. Global Humanitarian Technology Conference (GHTC), 2014 IEEE, 279-286. http://arxiv.org/ftp/arxiv/papers/1411/1411.6574.pdf
[xli] Statistics Without Borders and Humanity Road (2013). Analysis of Twitter data during hurricane Sandy. Available at: http://www.slideshare.net/CatGraham/swb-hr-hurricane-sandy-twitter-analysis
[xlii] D.E. Elvidge et al., "Satellite inventory of human settlements using nocturnal radiation emissions: a contribution for the global toolchest", Global Change Biology, vol. 3 (1997), pp. 387–395. Available from http://www.as.wvu.edu/biology/bio463/Elvidge%20et%20al%201997%20satellite%20night%20pictures.pdf
[xliii] Net primary production refers to how much carbon dioxide vegetation takes in during photosynthesis minus how much carbon dioxide the plants release during respiration.
[xliv] See section 5.3, figure 16.
[xlv] Eric A. Kort, Christian Frankenberg, Keeley R. Costigan, Rodica Lindenmaier, Manvendra K. Dubey and Debra Wunch (2014). Four corners: The largest US methane anomaly viewed from space. Geophysical Research Letters, Volume 41, Issue 19, pages 6898–6903.
[xlvi] Oliver Schneising, John P. Burrows, Russell R. Dickerson, Michael Buchwitz, Maximilian Reuter and Heinrich Bovensmann (2014). Remote sensing of fugitive methane emissions from oil and gas production in North American tight geologic formations. Earth's Future, Volume 2, Issue 10, pages 548–558.
[xlvii] Andrey Bogomolovy, Bruno Leprix, Jacopo Staianoy, Nuria Oliverz, Fabio Pianesix, Alex Pentland (2014). Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data. ACM International Conference on Multimodal Interaction (ICMI 2014). Available at: http://arxiv.org/abs/1409.2983
[xlviii] http://syria.newscientistapps.com/index.html and https://dartthrowingchimp.wordpress.com/2013/05/16/challenges-in-measuring-violent-conflict-syria-edition/
[xlix] Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E., 2011. Prediction of Socioeconomic Levels using Cell Phone Records. Lecture Notes in Computer Science, 6787, pp 377-388.