SOUTHERN VOICE

Southern perspectives. Global debates.

Harnessing Innovative Data and Technology to Measure Development Effectiveness

Emmanuel Letouzé
Micol Stock
Francesca De Chiara
Alberto Lizzi
Carlos Mazariegos

Harnessing Innovative Data and Technology to Measure Development Effectiveness

Emmanuel Letouzé
Micol Stock
Francesca De Chiara
Alberto Lizzi
Carlos Mazariegos

Preface

With the advent of the Sustainable Development Goals (SDGs), discussions on development finance have been revitalised. Mobilising sufficient financial support to meet the resource gap in SDG implementation is a critical challenge for developing countries.

Traditional aid flows to these countries have been restrained by both supply-side limits and demand-side pulls. However, new actors and innovative financial instruments create opportunities for additional funding. In this context, improving the quality of development cooperation (including financial flows) and assessing its effectiveness have become more pertinent than ever.

Economic and political factors aggravate the challenge of effective development cooperation. The current global development finance architecture lacks necessary political ownership and momentum. Further, the discourse suffers from an obvious lack of credible knowledge that reflects realities on the ground. Demand is thus high for Southern perspectives so as to embed them in future reforms.

That is what Southern Voice, a network of over 50 think tanks from Africa, Asia, and Latin America, is facilitating. It provides structured inputs from the Global South for debates on the 2030 Agenda for Sustainable Development. With capacity gained through the successful execution of various research programmes, Southern Voice aims to contribute to the global discussion on the effectiveness of development cooperation in the era of SDGs.

The new initiative, "Rethinking Development Effectiveness: Perspectives from the Global South," is being carried out in partnership with the Centre for Policy Dialogue (CPD) in Dhaka, Bangladesh and with support from the Bill & Melinda Gates Foundation. The present study is the fourth in a series of nine occasional papers on rethinking development effectiveness. It explores the different analytics methods, digital data and tools that contribute to the measurement of development effectiveness.

Debapriya Bhattacharya, PhD
Chair, Southern Voice and Distinguished Fellow, CPD
Dhaka, Bangladesh

Acknowledgement

Authors

Emmanuel Letouzé is the director and co-founder of Data-Pop Alliance, a visiting scholar at MIT Media Lab, a research associate at ODI, and the executive director and co-founder of the OPAL (Open Algorithms) project. He holds a BA in political science, an MA in applied economics, an MA from Columbia University School of international and public affairs, where he was a Fulbright fellow, and a PhD from the University of California, Berkeley. He can be reached at: eletouze@datapopalliance.org.

Micol Stock is project coordinator at Data-Pop Alliance, where she works on pilot and research programmes focusing on how new technologies and alternative data sources can help measure SDGs. She holds an MA in Middle East studies from the Hebrew University in Jerusalem and a second MA in human rights and humanitarian policy from Columbia University. She can be reached at mstock@datapopalliance.org.

Francesca De Chiara is researcher at Fondazione Bruno Kessler in Trento, Italy, where she works in the Digital Commons Lab, within the ICT Research Centre. She's also visiting fellow at the Governance Lab, New York University. She holds a PhD in Sociology. She can be reached at dechiara@fbk.eu or francesca.dechiara@gmail.com.

Alberto Lizzi is the policy specialist for Knowledge Solutions at the United Nations Development Programme where he helps development programmes learn and adapt from failures and opportunities. He can be reached at albertolizzi@gmail.com.

Carlos Mazariegos is the technical programme manager for OPAL (Open Algorithms) project, where he drives the technical workstream liaising MIT Media Lab, Imperial College of London, Orange Labs and Telefonica for launching the initial OPAL pilots in Colombia and Senegal. He holds a M.S. in data analytics, policy and management from Carnegie Mellon University. He can be reached at cmazariegos@opalproject.org or cmazarie@alumni.cmu.edu.

Abstract

In this study, the authors discuss and show how new kinds of digital data and analytics methods and tools falling under the umbrella term of Big Data, including Artificial Intelligence (AI) systems, can help measure development effectiveness. Selected case studies provide examples of assessments of the effectiveness of ODA-funded policies and programmes. They use different data and techniques. For example, analysis of mobile phone data and satellite images: to estimate poverty and inequality, traffic congestion, social cohesion or machine learning approaches to social media analysis to understand social interactions and networks, and natural language processing to study changes in public awareness. A toolkit contains resources and suggestions on key steps and considerations, including legal and ethical, when designing and implementing projects aimed at measuring development effectiveness through new digital data and tools. The chapter closes by describing the core principles and requirements of a vision of a 'Human AI', which would reflect and leverage the key features of current narrow AI systems that are able to identify and reinforce the neurons that help them reach their goals. A Human AI would be a data and machine-enabled human system (such as a society) that would seek to continuously learn and adjust to improve—rather than prove after the facts—the effectiveness of its collective actions, including development programming and public policies.

Content

List of tables

List of figures

Acronyms and abbreviations

AI	Artificial Intelligence
CODEs	Councils for the Orientation of Development and Ethics
DHS	Demographic and Wealth Survey
IDO	International Development Organization
MDGs	Millennium Development Goals
ODA	Official Development Assistance
OPAL	Open Algorithms
OECD	Organisation for Economic Co-operation and Development
RCT	Randomised Control Trial
SDGs	Sustainable Development Goals
UNDP	United Nations Development Programme

Harnessing Innovative Data and Technology to Measure Development Effectiveness

Emmanuel Letouzé, Micol Stock, Francesca De Chiara, Alberto Lizzi and Carlos Mazariegos

Introduction

The development effectiveness landscape has been significantly transformed following two major milestones, namely the Paris Declaration on Aid Effectiveness of 2005 and Accra Agenda for Action of 2008. On the political economy side, the transition from the Millennium Development Goals (MDGs) to the Sustainable Development Goals (SDGs) and the recent rise of political polarisation and populism are reflecting and fuelling new concerns. On the technological and scientific side, major drivers of change have been the spread of randomised control trials (RCTs), and—of course—the growing use of digital devices, services, and data.

It has become commonplace to say that the unfolding Big Data (or fourth industrial) revolution has created both challenges and opportunities, as all previous techno-political revolutions have. Notably, there are opportunities for more agile decisions and targeted information. Challenges, for example, are widening divides as well as distrust and overconfidence in the power of technological fixes. In this context, whether and how Big Data can help donors, policymakers, and development professionals get a better, finer, faster sense of the effectiveness of development—addressed here as official development assistance (ODA), South-South cooperation, and blended finance—to increase it have received significant attention in recent years.

The improvement of predictive machine learning models has given a new impetus to an old debate about the balance between formative and summative evaluations and calls for a greater focus on "improving" over "proving." This improvement has happened while evaluation experts have increasingly recognised the implications of the growing complexity of both the interventions to be evaluated and the contexts within which they are deployed. One important insight is that Big Data has contrasting effects on the "evaluability challenge." In other words, the extent to and ways in which causality can be credibly assigned between an intervention funded by aid and observed outcomes— like the impacts of a new transportation system on economic opportunities and citizen security. Positive effects occur through new insights on human processes and experiences that these new data and tools can yield (such as fine-grained mobility or poverty estimates), including from quasi-natural experiments they can conceal. Negative effects

occur through the many feedback loops that they create, which may further complicate causal inference, while providing a temptation—with so much data to crunch—to bypass careful scientific design.

In response, a consensus has emerged on the use of mixed methods, also called the "RCT+" approach, including qualitative analysis and for them to be more embedded in daily processes to make them more adaptive (Bamberger, Tarsilla and Hesse–Biber, 2016). These mixed methods are expected to allow for more dynamic and richer sets of indicators and enable feedback that may identify unintended consequences. Guidelines have also been developed to integrate Big Data into monitoring and evaluation (M&E) of development programmes. Other new approaches can help with causality and counterfactuals. They can strengthen organisational and multi-stakeholder learning, including information value stream mapping.

> **Artificial intelligence is also poised to affect aid effectiveness in the medium and long terms.**

The new normative landscape is also testing standard practices and opening up new avenues. For example, some experts question if the standard criteria of the Organisation for Co-operation and Development's Development Assistance Committee can adequately capture the new values and objectives embedded in the SDGs and 2030 Agenda for Sustainable Development (to which the Principles for Digital Development[1] could be added) such as social inclusiveness and environmental sustainability, which challenge the praxis, metrics, and timeframes commonly used to determine development effectiveness. An area where new goals and tools meet is Tier III SDG indicators, for which no methodology has yet been formalised, as in the case of a recent project by Data-Pop Alliance with the United Nations Development Programme (UNDP) to estimate the "proportion of population satisfied with their last experience of public services" (SDG Indicator 16.6.2) through social media analysis in Botswana.

Artificial intelligence (AI) is also poised to affect aid effectiveness in the medium and long terms. Already, AI applications are used to analyse and categorise large amounts of text and images to assist in producing and connecting relevant data sets, conducting simple

---

1 Principles for Digital Development: https://digitalprinciples.org.

differentiation between images and objects, and broadly identifying people and groups. Other examples include using machine learning models to improve the effectiveness and fairness of social programmes (such as Progresa in Mexico) by predicting false positives and false negatives. The prospect of using AI approaches at scale points to core, growing needs during the assessment of development effectiveness in a digital world: having access to appropriate data and having such access in a reliable, predictable, and ethical manner. Meeting these needs will take efforts with which development effectiveness experts are familiar—building trust, partnerships, data systems, and baselines—but alongside new stakeholders and with new incentives.

A project that may facilitate the transition towards greater reliance on "private" data to assess and enhance aid effectiveness in the short term is the Open Algorithms (OPAL) project, currently being implemented by a consortium of partners in collaboration with two leading telecommunications operators in Colombia and Senegal. OPAL aims to enable the computation of key indicators (such as population density, poverty, and diversity) from data collected and controlled by private companies through a secured open source platform and open algorithms running on the servers of companies, behind their firewalls, with governance standards ensuring the security, auditability, inclusivity, and relevance of the algorithms and use cases[2].

AI applications can also provide a useful example, inspiration, or "aspirational analogy" to make future humans more effective. What makes current (narrow) AIs so impressively good at their jobs (such as predicting travel times, translating texts, and driving a car) is the credit assignment (or reward) function: the ability of algorithms to identify and reinforce the artificial neural networks that most contribute to coming up with the "right" result (such as whether what an autonomous car detects in front of it is a road or a tree) through many iterations and data-fuelled feedback loops, which allow (machine) learning. If the principles and tools of AI were applied to human systems, which we call "Human AI ecosystems, or ecologies," organisations (governments, companies, or the aid sector, for example) would both be inspired by and using AI to be set up and aim to identify and reinforce what contributes to "good results," including intended outcomes of aid programmes. They could also better understand whether these intended effects are desirable in the long run through feedback.

Building such Human AI ecosystems or ecologies (Pentland, 2017) will require and spur a few key social and technological features: a healthy data culture, with widespread data literacy, as well as incentives and means for civil society organisations, researchers, and regulators, etc., to demand that the effectiveness of taxpayer-financed policies

---

2 See the OPAL Project website: https://www.opalproject.org/home-en.

and programmes (including by aid) be assessed using the best available data and methodologies, adjust future iterations, and contribute over time to a body of evidence regarding what works, why, and how following the main historical tenets of the scientific method, social progress, and human evolution (Letouzé and Pentland, 2018). These developments may seem like a long-term dystopian perspective, but they seem inevitable and desirable to enable development aid systems to be more effective at meeting their single objective: improving the overall state of the (increasingly digital) world.

How has the development effectiveness landscape changed in the data era?

The rise of and obstacles for the aid and development effectiveness agenda

In the early 2000s, donors recognised that they were partly to blame for the often discouraging results of aid. The need to understand why aid was not yielding expected results, as well as the will to step up efforts to meet ambitious targets set by the MDGs, contributed to new debates and initiatives in the field of international development. "Effectiveness" became a central notion in the lexicon of the aid sector. The donor community and aid recipients became increasingly committed to improving aid effectiveness through better coordination mechanisms, as proven at successive High Level Forums on Aid Effectiveness. The Paris Declaration on Aid Effectiveness (OECD, 2006) established a roadmap to improve aid quality through specific implementation measures, performance indicators for assessing progress, and a mutual accountability framework between donors and recipients. The five principles of aid effectiveness—ownership, alignment, harmonisation, managing for results, mutual accountability—were then strengthened by the Accra Agenda for Action and have served as the foundation for other commitments tailored to specific contexts, such as the Bogotá Statement (concentrating on effective aid principles in South-South cooperation), Istanbul Principles (on the role of civil society), and Dili Declaration (on effective aid in fragile and conflict-affected states). As a result, the international community broadened the agenda significantly with the Busan Partnership for Effective Development Co-operation of 2011: from "aid effectiveness," as pertaining principally to the Development Assistance Committee providers of official development assistance, to "development effectiveness," encompassing the myriad actors and partnerships involved in South-South, triangular, philanthropic, and private sector cooperation.

Despite major progress in reducing poverty and increasing aid transparency, the overall consensus is that more should be done better: "Donors have never actually fulfilled their commitments, and the evidence on the effectiveness of finance delivered through Paris- and Busan-style mechanisms has been mixed" (OECD, 2011). Various factors, including structural weakness of aid recipients' economies, institutions, and capacities, will make it difficult to eradicate extreme poverty by 2030 (Kharas, Hamel and Hofer, 2018). The latest projections show that meeting the SDGs' deadline of 2030 will require increasing the number and effectiveness of development programmes (United Nations, 2017).

The bureaucratic models adopted following the Second World War—and the corresponding incentives that they generate—were very well suited to addressing specific development issues and technical problems and could be standardised, reduced to a best practice. Yet, they have shown severe limitations when responding to complex and ambiguous challenges in the context-specific settings that characterise the 2030 Agenda (Lant, Woolcock and Andrews, 2010). Much development and humanitarian work is still being done using a paradigm of predictable, linear causality and maintained by mind-sets that are focused on accountability and top-down authority, while donors and the international community advocate for aid effectiveness, institutional reform, participation, local ownership, and empowerment. In the meantime, complexity science has explored and articulated a contrasting world of understanding, which helps to explain complex dynamic phenomena in widely diverse settings using concepts like non-linearity, edge of chaos, self-organisation, emergence, and coevolution.

Increasing complexity: beyond one-size-fits-all

Both the MDGs and SDGs set ambitious targets that reflect the multifaceted nature of human development though differ in several respects (UNDP and the World Bank, 2016). While the MDGs managed to focus global attention on the plight of the poorest, official development assistance commitments, and a framework within which countries can plan their social and economic development while donors may provide effective aid, the SDGs are conceptually different and much more complex. The 2030 Agenda emphasises the need for approaches that take into account the integrated nature of the three dimensions of sustainable development—economic, social, and environmental—where decisions in one domain affect outcomes across the others. It is a profound shift in thinking about development as a desirable future that is equitable, inclusive, peaceful, and environmentally sustainable. The complexity of this vision demands interventions that go beyond the typical linear and sectoral ones of earlier decades and recognises the increasing complexity of the world in a globalising world.

When they committed to the achievement of the MDGs by 2015, governments agreed to a compact that put the onus on developing countries. Donors had a particular responsibility to increase both the quantity and quality of aid to developing countries—in other words, to increase aid effectiveness. The context in which to achieve global goals has also shifted profoundly and over the past decade a plethora of new public and private sector actors have become engaged in development. Their contributions are considered critical to the success of the 2030 Agenda, though tracking and understanding the impacts of various types of financial flows have become much more complicated than they are for traditional official development assistance.

Need for data

The SDGs require the identification of new pathways to sustainability in order to navigate the interplay of research and development, public and private investments, politics, planning, and complex decision making by many stakeholders. They also require unprecedented mobilisation of global knowledge across many sectors and regions as well as accurate, timely (not several years out of date) information. Time lags were inevitable when data were obtained through household surveys filled out by hand, but in the age of mobile phones, wireless broadband, remote sensing, and AI, data collection and dissemination should be considerably quicker. The adaptive nature of the SDGs' approach calls for technology, data collection and analysis. For example, real-time data, AI and machine learning that enable hypothesis testing and providing meaning to the ever-evolving complexity at hand. Governments and international development organisations (IDOs) should consciously invest in a quasi-real-time reporting systems for the SDGs—and in support of programmes—to produce reliable data on no more than a yearly, or ideally quarterly, basis.

To succeed, the focus should shift towards encouraging better collection and dissemination of performance data for more informed decision making. Similarly, ensuring that learning happens rapidly and lessons about what works (or does not) are analysed and shared both in a timely manner and widely helps speed up improvements. The open source and evidence-based policy movements have been championing these agendas for some years now, but these kinds of approaches remain somewhat at odds with the prevailing mindsets within many government institutions.

Making investment spending visible to constituencies can no longer be the yardstick to measure effectiveness. The new ways of working require innovative approaches that are less focused on linear reporting and measuring of aggregated results and more responsive by leveraging smarter, faster, and disaggregated data. This supports adaptive approaches and learning from development on the ground. A growing number of

innovative approaches, such as the Global Delivery Initiative and Global Learning for Adaptive Management Initiative, are promoting adaptive development as a way of Doing Development Differently and tackling complexity through more problem-driven and context-specific approaches. By leveraging agile methodology and modern technology, like machine learning and real-time data acquisition, to harness large data sets, these approaches may prove to be more effective in addressing the complexity of the SDGs and increasing development effectiveness.

> **The SDGs require unprecedented mobilisation of global knowledge across sectors and regions.**

Defining Big Data

What was initially referred to as "the industrial revolution of data" in 2008 (Hellerstein, 2008) has since been simply called Big Data. The starting point and central feature of Big Data as a phenomenon is the unprecedented growth in the volume and variety of high-frequency digital data—structured and unstructured—being passively generated by people during the course of their activities. These data have been described as non-sampled data characterised by the creation of databases from electronic sources whose primary purpose is something other than statistical inference (Horrigan, 2013) or otherwise as "data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand" (Scannapieco, Virgilito, and Zardetto, 2013). But in the words of one observer, "Big Data is not about the data" (King, 2013). Rather, the key features are captured by the three Cs of Big Data: the first C stands for "crumbs" (Letouzé et al., 2013; Letouzé, 2014; Pentland, 2012)—identifying Big Data as new kinds of passively generated individual and networked "traces of human actions picked up by digital devices" (Letouzé, 2013). These "digital breadcrumbs" have the potential to paint a picture of some aspects of the social world with unprecedented levels of detail and shades (Pentland, 2012) and their fundamental, revolutionary nature is qualitative. The second C stands for "capacities," or the intent and capacity to yield and convey what are routinely and vaguely referred to as "insights" from these new kinds of qualitative data (Toyama, 2012). This capacity involves advanced storage and computing capacities as well as advanced quantitative and computer science methods and tools—primarily statistical machine-learning techniques and algorithms. The third C stands for "community" or culture, since Big Data must also be considered as referring to the

people and groups who make use of crumbs and capacities. Many of these new actors have embraced and indeed spurred the open source movement as well as new ways of working based on the lessons of agile software development. Others in the private sector or intelligence communities operate in a highly supervised and secretive manner for obvious commercial and political reasons, respectively. At the moment, most of the data that make up Big Data are held by the private sector—especially telecommunication companies and financial institutions—and only a handful of data sets are in the public domain, most of which are unstructured and hard to work with.

Importantly, for discussions and initiatives on the applications and implications of Big Data for development and aid effectiveness to be meaningful, Big Data must be approached and conceived beyond its core raw material. Big Data must be looked at as being made up of new kinds of qualitative data on people's actions and interactions, new types of methods and tools, and new actors, incentives, and systems.

Growing awareness of and expectations for new technologies, AI, and machine learning with increased data availability and accountability pressures

The excitement over Big Data has stemmed from two factors: ever-increasing supply of data and analytics capacities as well as demand for better, faster, and cheaper information. The availability of reliable and up-to-date data has been improving significantly over time, though gaps remain in many instances. The lack of reliable data has inspired the call for a "data revolution" that led to the publication of a report by an expert group appointed by the United Nations (IEAG, 2014). The basic and somewhat simplistic rationale is that governments in the age of Big Data should be steered by policymakers relying on better navigation instruments and indicators that let them design and implement more agile and better targeted policies and programmes (Letouzé, 2015). However, most of the new analytical methods used by data scientists are good at prediction but not very good at understanding causality, which is what social scientists (and evaluators) are most often interested in. Today, different applications are routinely used for implementation of early-warning systems, emergency relief, and dissemination of information, identifying and collecting feedback from marginalised and vulnerable groups, and enabling rapid analysis of poverty. Data analytics are also implemented to create integrated databases that synthesise all of the information on topics as diverse as national water resources, human trafficking, conflict zones, and climate change (Bamberger, 2017). Given the rapid expansion of Big Data in international development, there will likely be a move towards integrated programme information systems, which will begin to generate, analyse, and synthesise data for programme selection, design, management, monitoring, evaluation, and dissemination. The United Nations Development Programme (UNDP) has been

implementing some projects based on Big Data and data-driven adaptive management (Watson and Lizzi, 2018), as described in Box 1.

Box 1. Innovation in data analytics at UNDP

Uptake of Big Data analytics is accelerating across the United Nations system with a growing number of agencies, funds, and programmes implementing and scaling operational applications for development and humanitarian use. UNDP is deploying AI and data science in many initiatives, creating efficiencies for human development and humanitarian assistance through an interesting portfolio of innovative projects aimed at improving data collection and analysis for decision making.

To strengthen programme effectiveness, UNDP partnered with the Global Delivery Initiative[3] and World Bank to leverage the power of machine learning and gather insights from historical data found in project reports, evaluations, and corporate results reporting. Grounded in the principles of being problem-driven, attentive to context and adaptive to problem solving, the Global Delivery Initiative helps development practitioners and organisations learn from collective knowledge on implementation failures in order to design better quality programmes, plus it uses predictive analytics to inform decision making.

In order to provide better policy advise on SDGs to governments, UNDP is working with IBM to automate UNDP's Rapid Integrated Assessment – a tool aiming to support the mainstreaming of SDGs into national planning by assessing the readiness for SDG implementation while determining their relevance to the country context. The first pilots have demonstrated large gains in efficiency.

In 2019, the upcoming Arab Human Development Report will expand UNDP's reach to voices from the region using Premise[4] data gathered from mobile signals, which is an alternative data collection method. UNDP and Premise will collect and analyse lean data that will inform ongoing research on the role of citizenship in achieving the SDGs in Arab countries. Repeated, short-term feedback loops will complement available official statistics for a better understanding of socio-economic and social cohesion dynamics. While samples will be limited to people with smartphones (and therefore cannot be stratified as in traditional surveys), the pervasion of mobile technology already allows diverse populations in diverse locations to be reached. This approach enables the reduction of financial costs as well as the project's carbon footprint (compared to consultative workshops), while reaching out to a broader audience for prolonged and repeated interactions. Data collection will be tailored to the languages and specificities of each country, but data visualisation tools will make responses available in English and Arabic in real time.

---

3 See: www.globaldeliveryinitiative.org

4 See: https://www.premise.com

This function is critical for the data to inform and shape the analysis as it develops, instead of being retrofitted or treated as an add-on as often happens with consultative processes.

In Iraq, UNDP is gearing up to leverage real-time crowdsourced data via mobile signals to test data-driven adaptive management and improve evidence-based decision making to support stabilisation efforts under the Iraq Crisis Response and Resilience Programme.

The portfolio also includes remote sensing for improved decision making on environmental protection in Mongolia and using drones to facilitate disaster preparedness in the Maldives. In Uganda, UNDP and the United Nations Refugee Agency are using drones to map the Oruchinga Refugee Settlement. The data are used to engage refugee and host communities in jointly developing camp and host community infrastructure.

In Lebanon in 2017, UNDP conducted a survey via WhatsApp that collected narrative data from both host community members and Syrian refugees to understand local conflict dynamics and the impact of assistance. The survey is at the core of an Innovation Project called "Speak your Mind to Prevent Conflict in Lebanon," which is funded by the UNDP innovation facility.

UNDP and partners, including Data-Pop Alliance, are testing alternative methods for harnessing data such as geospatial, social media, sensor, and mobile phone data to address official data gaps. "Measuring the Unmeasured", described in Box 2, is a cross-regional initiative to help countries tackle the challenges posed by the many indicators for measuring progress on the SDGs that are currently unmeasurable (Tier III SDG indicators).

Box 2. Measuring Tier III SDG indicators in developing countries

Given the variety of data generated and collected about human actions and interactions, particularly in middle-income countries, and the concomitant development of powerful analytical methods and capacities, developing and diffusing methodologies and standards are within reach, provided sufficient focused resources are put to the task. Great efforts are being made to fill the gap on Tier III SDG indicators. Data-Pop Alliance aims to contribute to advancing the state of knowledge, awareness, and capacities to effectively implement new methodologies in pilot countries. Together with UNDP, Data-Pop Alliance has been focusing on several countries in different regions of the world. The project seeks to enable governments to identify, collect, and use non-traditional sources of data to measure Tier III SDG indicators at the national level.

In Botswana, for example, the project focused on SDG indicator 16.6.2: "Percentage of the population satisfied with their last experience of public services." As the indicator implies, the main

source of data is perception-based surveys, which are often collected by national statistical offices that, given different capabilities, structural constraints, and limited technical capacities, ultimately produce low-quality data. Together with Fondazione Bruno Kessler, Data-Pop Alliance created a user-feedback mechanism that collects data in real time from Facebook and Twitter to measure public satisfaction with government services in a development sector and can focus analysis on a specific group, allowing for easy and anonymous sharing with government institutions through an application. Work that has been traditionally conducted through costly and time-consuming manual data collection can now capture sentiments in real time for the effective measurement of indicator 16.6.2 and potential progress towards SDG 16 on peace, justice, and strong institutions.

Paul, Jolley, and Anthony (2018) illustrate the growing awareness of AI and machine learning in the field of international development. In general, machine learning algorithms are geared towards making predictions, making them useful in early-warning systems by monitoring whether conditions are similar to those that have preceded a crisis in the past, especially concerning conflict, food security, and health. This application of machine learning allows attention and resources to be directed towards rapid response. However, not all early-warning systems rely on machine learning. It is common also to look at geospatial, economic, or health data and make predictions about what might happen. One major difference is that while human analysts tend to make predictions based on a small number of strong signals, such as anticipating a famine if rainfall is low and food prices are high, machine learning methods excel at combining a large number of weaker signals, which might have escaped human notice. These methods give early-warning systems the potential to spot emerging problems faster than traditional methods and obtain more complete results for an evaluation of development effectiveness.

Fears of techno-utopian fixes that may run counter to the principles and objectives of aid effectiveness

While the potential impact of Big Data is vast, it is not always guaranteed to be positive. For example, relying on innovative systems without understanding their strengths and limitations may lead to risks of unfairly targeting or excluding people. Algorithmic decisions may be flawed and accountability for results may be not possible.

Although many factors contribute to harm, the blind trust that may be placed in technological tools poses outsized risks. In studying AI and machine learning in development, USAID (2018) describes how excessive trust can be dangerous when it leads to unquestioning acceptance of modelled results, which can lead to misinformed choices in cases where models reach the wrong conclusions. For instance, models may be: fair

but inaccurate (though still useful); less precise for minority groups than for the majority population; unevenly balanced in terms of errors (granting more false positives to one segment of a population and more false negatives to another, which creates an uneven playing field and systematically disadvantages one group); or reproducing existing inequities, doubling down on bias, and drifting due to changes in the real world.

Technological innovation in data automation, including AI, could bring both threats and opportunities to SDG implementation. UNDP (2018) highlights the danger that AI poses to progress on at least nine SDGs if large-scale automation is unchecked. Recognising the well-documented dangers posed by technological advances, this study focuses on the promises of innovative data automation for development effectiveness, specifically how data innovation can help development practitioners and organisations obtain high-quality, reliable intelligence faster for more effective decision making and improved development financial flows.
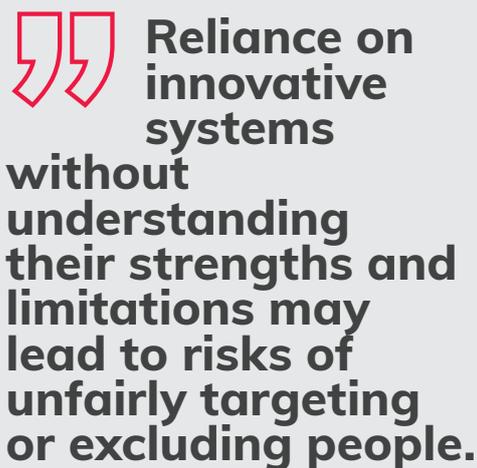
> **Reliance on innovative systems without understanding their strengths and limitations may lead to risks of unfairly targeting or excluding people.**

Defining M&E today

In recent years, there has been a shift from a narrow concept of monitoring to a more encompassing approach that recognises the complexity of development projects.

The UN Global Pulse for example, describes a monitoring programme as usually including: producing data for a results framework; accountability: did the programme achieve its outcomes in a timely manner and within budget? Actionable information on problems detected during project implementation; identifying negative outcomes or groups who are not receiving programme benefits and services; providing data inputs to the programme evaluation; providing inputs for the evaluation of complex programmes (UN Global Pulse, 2016, p. 93).

Monitoring and evaluation are often understood as being part of the same process. The main difference is their timing and focus assessment, but the two are integrally linked—monitoring usually provides data for evaluation (or assessment) and elements of evaluation occur while monitoring.

The UN Global Pulse also differentiates among four different types of M&E for development: policy and broad-based programme evaluation usually "conducted retrospectively at the end of a country programme cycle (typically lasting four to five years)"; formative evaluation, "providing regular feedback to management and other stakeholders to help strengthen the implementation of programmes and projects", and often based on a mixed-method approach; developmental evaluation, "focused on innovative programmes and those that operate in complex environments where an adaptive approach to design and implementation must be used"; and summative evaluation, "to assess the extent to which observed changes in outcome variables (the intended project goals) can be attributed to the effects of the project intervention" (United Nations Global Pulse, 2016, p.51-52).

For the purpose of this study, we will look at a combination of these categories to assess the effectiveness of development programmes retrospectively. Since development programmes have been increasingly recognised as complex in nature, seriously challenging the validity of conventional evaluation designs that assume a linear relationship between programme inputs and outcomes with the introduction of "complexity-responsive" evaluation designs is often required. In the near future, many of the data used for evaluation will likely be generated passively through the use of new technologies rather than collected through M&E processes that are commonly implemented today. Therefore, future M&E systems should be closely linked to broader systems that encompass programme identification, design, and management.

From proving to improving results

With the concept of the "Big Stuck," Lant, Woolcock, and Andrews (2010) denounce the long-term development stagnation caused by the "capability trap" of many countries and the systemic issues that do not allow development organisations and partners to make progress towards development objectives. The same authors also argue that reaching development objectives requires not only good policies and innovative ideas, but also going beyond transplanting best practices and overcoming implementation challenges that hinder the success of even the best-formulated policies. Implementation bottlenecks in development programmes are either accepted or go unnoticed, which contributes to a complacent culture that promotes success in terms of mere output reporting where the appearance of development activity masks the lack of impact (Lant et al., 2010).

Traditional M&E approaches are part of the problem. As seen in the previous section, they are instruments for the implementation of theories of change and policies informed by best practices and fall short of grasping the realities of more complex, less linear, and ever-evolving situations. Traditional approaches present data only after implementation

is complete, when these data are only useful for validating theories of change and designing new programmes instead of improving existing ones. With programme durations spanning up to five years (or more), losses in efficacy are greatly compounded and implementers rarely know where they stand relative to development objectives. Data, in this sense, are regularly used to prove results, but rarely to foster them. Learning, both contextually and from programme implementation, is essential to improve development outcomes and should be integrated into programming to strengthen traditional M&E approaches (Pritchett, Samji, and Hammer, 2013).

Big Data and other new data applications such as real-time data can increase the speed, volume, and quality of information to—if adequately supported by data literacy and organisational culture—enable development practitioners to learn and refine programmatic interventions faster and more precisely. Pairing adaptive development and Big Data holds the promise of increasing the effectiveness of development interventions by making them tactically more focused, timelier, and more efficient. Their respective capacities to tackle implementation bottlenecks and move the needle on the effectiveness of aid programmes can be considered in a number of ways:

- Better measures: Incorporating new methodologies will significantly advance processes that monitor development. Crowdsourcing feedback from marginalised groups, analysing call detail records (CDRs) to measure population mobility, and studying electronic transaction records to rapidly understand poverty are merely a few examples of data-based applications that offer clear advantages over traditional methods, such as faster data collection, real-time evaluation, and a higher granularity of information. Implementing new technologies could also help measure progress on those Tier III SDG indicators for which no methodologies have been formalised.

- Better processes: Big Data offers development professionals the ability to gain real-time insights into a population's well-being and target aid interventions to vulnerable groups in a manner that is more tailored to their specific needs. Satellite images, mobile phone records, Global Positioning System (GPS) tracking, and analysis of social media are examples of sources that can be used to structure integrated databases in order to synthesise information on diverse development topics (such as national water resources, human trafficking, conflict, climate change, epidemics, and food security) by refining learning processes and adapting programmes to work in complex contexts. The intersection of methodologies will "open up possibilities for more dynamic metrics and sets of indicators, which could foster organisational and multi-stakeholder learning" (Giller, Bell, Mock, Hijmans, Sayer and Serraj, 2014).

Over the next two sections, we will provide an overview of existing methodologies and current debates concerning related advantages and limitations and then propose an approach and a toolkit that are useful when including Big Data and new technologies in the assessment of development programmes and effectiveness.

How much is actually done about Big Data for development effectiveness?

Frameworks and best practices for incorporating Big Data in complexity-responsive M&E processes

Development organisations have been stressing the potential role of Big Data in reconfiguring and integrating existing M&E frameworks. The most important known limitations of monitoring systems could be mitigated by the availability of up-to-date accounts and real-time data. In this context, innovative methods are "unlikely to replace established methodologies" but "might rather complement them, thus adding to the already existing methodological heterogeneity" (Scheel and Ustek-Spilda, 2018). New mixed intelligent methods, based on the intersection of systems, open up possibilities to move from conventional approaches, which use linear techniques and focus on input and output, to the measurement of results at any stage of an evaluation cycle, as in the examples identified by Bamberger (United Nations Global Pulse, 2016).

The implementation of impact assessments to understand the effectiveness of an intervention, as well as whether it has a direct relationship with a specific outcome, is subject to debate among evaluators (Dinshaw, McGrayl, Rai, and Schaar, 2014). What has been recognised as a major challenge is the need to understand counterfactuality. Science and technology may provide some support towards this end.

Evaluators may use approaches based on RCTs, differentiation models, statistics, and economic research to understand counterfactuality. Although RCTs have for long been the most widely used tool for summative evaluation, this approach has also been the most challenged by Big Data analysts. An exclusive focus on RCTs is largely criticised by evaluators due to, among other reasons, a narrow focus on one or a small number of usually quantitative outcomes, a lack of attention to the process of project implementation (evidence often being available only after years have passed), high costs, and the context within which programmes are designed, implemented, and evaluated. Rights-based evaluators also stress the need to listen to multiple voices and argue that there is no one way to identify or assess programme outcomes.

Several attempts have been made to respond to these and other challenges. For example, there is a call for new approaches such as: a "nimble evaluation", focusing on cheaper access to data and on the implementation stage rather than outcomes (Apolitical, 2018 October 29) or "RCT+" (Bamberger et al., 2016) which combines experimental evaluation designs with qualitative approaches. Most evaluation designs are intended to determine whether there is credible evidence (statistical, theory-based, or narrative) that programmes have achieved their intended objectives. The logic of many evaluation processes, even those that are considered the most rigorous, does not allow for the identification of outcomes that were not specified in programme designs. Enriching RCTs with a mixed-methods approach can strengthen credibility as well as capture important unintended consequences.

To understand the efficacy of development projects, not only are counterfactuals required, but there also needs to be an appropriate match between the change of impact over time and the deployment of corresponding adequate tools which will enable an empirical understanding of such a transformation. (Woolcock, 2009). One step includes setting and establishing baselines, eventually reconstructing them and using shifting or rolling baselines, in contexts that are constantly changing. An Organisation for Economic Co-operation and Development paper on monitoring and evaluation of climate change adaptation (Dinshaw et al., 2014) emphasises the risk of a biased correlation of factors that may have not been predicted in an original impact assessment model. Although there are techniques that can enable an evaluator to separate and quantify the impacts of different influencing factors on an outcome, the capacity to generate complex computational modelling and interpret information is required to apply these methods.

Challenges

Some of the concrete challenges in integrating Big Data and satellite images into evaluation systems relate to the characteristics of Big Data, analysis methods, and models, while others are created by data processing systems (Jin, Wah, Cheng, and Wang, 2015). Akerkar (2014) and Zicari (2014) have grouped the challenges related to data-driven methodologies into three main categories: data challenges related to the characteristics of data (e.g., data volume, variety, velocity, veracity, volatility, quality, discovery, and dogmatism); process challenges related to a series of techniques: how to capture data, how to integrate data, how to transform data, how to select the right model for analysis, and how to provide results; management challenges covering privacy, security, governance, and ethical aspects. Sivarajah, Muhammad, Zahir, and Vishanth (2016) adapted and summarised this conceptual classification based on the data life cycle. Sophisticated data mining techniques can help address the constraints related to the enormous volumes of information that make it difficult to determine, retrieve, process,

integrate, and infer data on the physical world (e.g., environmental data, business data, medical data, and surveillance data) as well as the variety and multiple formats of structured and unstructured data.

An additional challenge is ensuring the quality of data and satisfaction of standards while following appropriate methodological steps. For example, the use of Twitter application programme interface (API) data can suffer from questionable quality and serious methodological constraints such as samples of unknown representativeness, lack of one-to-one correspondence between accounts and users, and proliferation of tweets created by bots (Boyd and Crawford, 2012). Whereas transparency is essential to ensure reliability and validity, data created through the use of social media are in fact often produced by companies' closed structures (Driscoll and Walker, 2014). Further awareness is needed in this regard. In addition, according to Boyd and Crawford (2012) the difference in availability, appropriateness, and effectiveness of Big Data has caused a new form of digital divide, especially in low-income countries. Indeed, the types of data that can be found for most developing countries are often limited.

Of course, there are also important privacy and ethical concerns that have to be addressed since "any data on human subjects inevitably raises privacy issues" (Nature, 2007, p.638). In the context of development effectiveness, the design of data-driven evaluation systems dramatically increases ethical concerns related to the vulnerability of target populations. For example, the combination of different databases can lead to serious privacy violations (exposing sensitive information, such as individuals' social security numbers). Moreover, in countries affected by conflict, crisis, and weak law enforcement, privacy challenges may represent serious security risks (Letouzé, 2012). In the next subchapter, we discuss certain existing frameworks to overcome these challenges.

Privacy

In some places, data protection laws, such as the General Data Protection Regulation in the Europe Union and the Federal Trade Commission's Fair Information Practice Principles in the United States, cover the matter of privacy related to personal data. However, many countries remain largely unregulated. The risks of identification and high levels of vulnerability connected to anonymisation are still high and can especially affect people with low technology literacy, plus low-cost tools are able to collect sensitive data. Among African countries, relevant domestic legislation has been introduced by some governments. In South Africa, for example, the Protection of Personal Information Act of 2013 is stringent in requiring companies to re-seek consent if data are used for a new, unexpected purpose (Mann, 2017). Evidence demonstrates the need for regulatory frameworks. These will likely facilitate the creation of personal data formal markets. Key actors in the field are trying to develop a set of ethical guidelines.

The United Nations Global Pulse and International Association of Privacy Partners (2018) jointly released a report outlining operational steps. These include: building a multidisciplinary team to practice ethics on the ground; conducting ethics assessments, including in consultation with external working groups, for new big data projects to consider the personal and societal impacts; implementing programmes that are scalable and flexible, depending on factors such as the societal context and the organisational structure of the implementing entity. Notably, the General Data Protection Regulation in the European Union is likely to affect many legal and regulatory frameworks, with its key principles of opt in and informed consent. It also requires all data controllers to keep track of the use of data for which they are legally responsible. It must be noted, however, that the regulation only applies to personal data, meaning that data not considered as personally identifiable information (e.g., pseudonymised aggregated data) can be accessed and analysed for research and statistical purposes.

Toolkit and methodology for a data-driven assessment

When considering the inclusion of Big Data in the assessment of development effectiveness in a particular context, it is advisable to keep in mind the definition of Big Data and the many components. A detailed, though not exhaustive (and constantly expanding), list can be found below.

The three Cs
Crumbs[5]

1. Exhaust data (collected electronically as a function of some other transaction)
   - Mobile-based: mobile GPS data; CDRs
   - Online traces: browser cookies, Internet Protocol addresses; search history

2. Digital content
   - Social media: content of posts; social graphs based on user connections; metadata from posts
   - Crowdsourcing: calls on social media using hashtags; mapping (OpenStreetMap, Google Maps, Yelp), monitoring/reporting apps (uReport); structured and unstructured online content

---

5 For a visualisation of the diversity of internal and external data sources (in this case based on the 3V definition of Big Data: variety, velocity and volume), see: http://staging.kapowsoftware.com/resources/ infographics/intelligence-by-variety-where-to-find-and-access-big-data.php. In addition a useful list of public data sets is included in: Almås, I., Attanasio, O., Jalan, J., Orteiza, F. & Vigneri, M. (2018) Using data differently and using different data. Journal of Development Effectiveness (pp. 30-35)

3. Sensing
- Physical: smart meters; United States Geological Survey seismographs; speed/weight/mail trackers
- Remote: unmanned aerial vehicles; satellite imagery (National Aeronautics and Space Administration [NASA], Transmit/Receive Integrated Microwave Module [TRIMM], Landsat)

Capacities

- Network analysis and social physics
- Machine learning techniques and AI
- Text mining tools (sentiment analysis, topic models)
- Spatial analysis and geographic information systems
- Crowdsourcing
- Data set disaggregation and unification
- Distributed networks of devices
- Cloud computing and storage

Community

- Academia and research organisations
- Civil society/non-governmental organisations
- Multilateral institutions
- Governments/national statistical offices
- Private sector (e.g., telecommunications companies and financial institutions)

As mentioned, Big Data can help produce evidence and thereby strengthen the value of evaluation. The United Nations Global Pulse (2016) stresses how a dynamic evaluation system based on Big Data has the potential to: evaluate the outcomes of policy, programme, and project interventions; provide robust estimates of the extent to which observed changes in outcomes can be attributed to programme interventions; assess the outcomes of complex programmes operating in complex contexts; design evaluations operating under real-world budget, time, and data constraints; allow rapid feedback on outcomes; and provide predictive as well as retrospective analyses.

In general, when developing an evaluation plan, one of the hardest challenges is to choose the kind of information that best answers the defined questions. It is therefore important to have an agreement among stakeholders about how success is defined. The logic model recommended by the United Nations Global Pulse (2016) is an excellent tool for designing performance measurement systems. The framework enables capturing the

important elements of a programme, identifying which evaluation questions should be asked and why, and determining key performance measures.

Stages in using Big Data

An overview is provided of the various stages that could be part of a data-driven external evaluation programme once the required information has been identified. Typically, stages do not need to occur in a specific order and can be repeated, be omitted, or occur simultaneously (as illustrated in Figure 1). Therefore, much of an evaluation programme design involves choosing the appropriate combination or sequence of stages to carry out a project that is feasible, ethical, effective, and scalable, while taking into account the constraints, needs, and capabilities of various stakeholders and beneficiaries in developing countries.

Figure 1. The different stages of data use
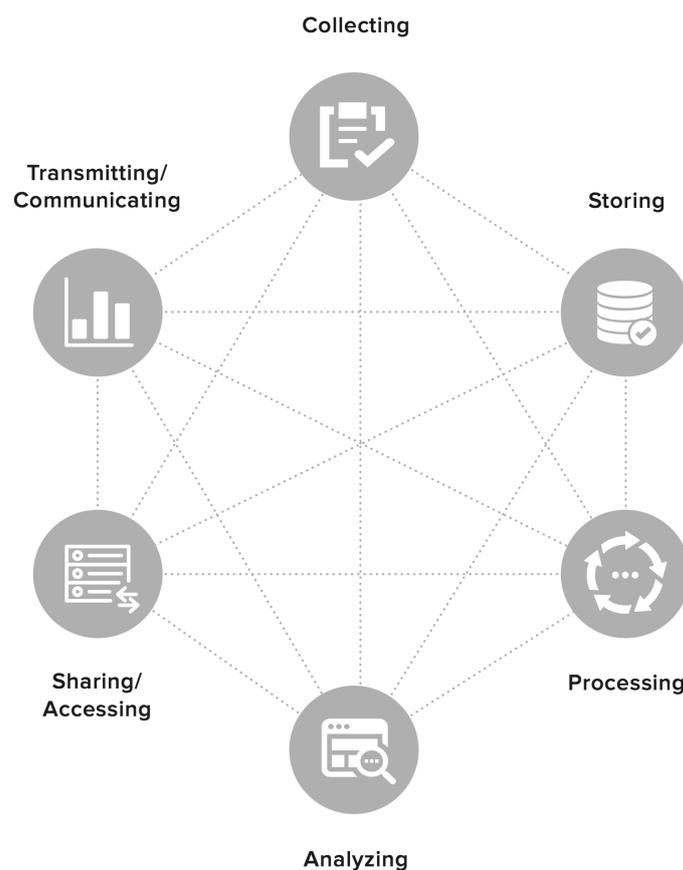
Source: Data-Pop Alliance (2017).

Depending on the context and needs, an external evaluator may wish to choose the best sources for each stage of an assessment programme from the toolbox presented in Box 3 below.

Box 3. Toolbox

1. Collecting: gathering data from people or devices
   Tools for collecting data from people:
   - Crowdsourcing tools: uReport, OpenStreetMap, Ushahidi
   - Survey tools: Kobo Toolbox
   Tools for collecting data from devices:
   - Sensors on smartphones: Bandicoot, Funf library
   - Remote sensing: Google Earth
2. Storing: keeping data for future use
   Tools for storing large volumes of data/information:
   - Database: Hadoop, MySQL, MongoDB
   - Cloud storage solutions: Amazon S3, Dropbox, Google Drive, OneDrive
   - Version control systems: Git, SVN
   Tools for storing sensitive data securely:
   - Secure data storage: OpenPDS
   - Trustless/decentralised storage: blockchain (e.g., Bitcoin)
3. Processing: turning raw data into intermediate data
   Tools for machines to execute instructions:
   - Cloud computing: Amazon Web Services, Microsoft Azure, Google Cloud
   - Decentralised computing: MIT Enigma
   - Batch automation: IFTTT
   Tools for people to make sense of raw data:
   - Data wrangling tools: OpenRefine, DataBasic.io
4. Analysing: moving from data to insights
   Tools for programming and statistical analysis:
   - Programming languages: MatLab, SPSS, Stata, Python, R, Julia
   - Open source libraries: Scipy, Numpy, Pandas, MatPlotLib, Plyr
   - Software and utilities: Rstudio, Jupyter, Anaconda scientific stack
   Tools for applying models and methodologies:
   - Machine learning tools: Scikit-learn, Spark, Accord.NET
   - Natural language processing: Apache OpenNLP, Stanford CoreNLP, LingPipe
   - Geographic information systems: ArcGIS, QGIS, Google Earth Engine
   - Causal inference: Tetrad
   - Propensity score matching: Twang R package
5. Sharing/accessing: making stored data/information available to others
   Tools for sharing/accessing data and code:
   - Code hosting platforms: GitHub, BitBucket, CDNJS

- Data pooling platforms: United Nations' HDX
- Application programming interfaces: Twitter Streaming API, Bit.ly Social Data API

Tools for sharing/accessing sensitive data:

- Privacy architecture: OpenPDS
- Governance structures: OPAL project

6. Transmitting/communicating: making stored data/information available to others or receiving data/information shared by others

Tools for transferring data:

- Standard formats: csv, json, geojson, xml, shp, xls
- Encryption tools: AES, RSA, OpenPGP

Tools for communicating information:

- Information and communications technology infrastructure: World Wide Web
- Visualisation tools: D3.js, R Shiny, Tableau

Causal modelling requires specific software. In Box 4 we describe Tetrad and Mediation, two of the most common ones.

Box 4. Causal modelling software

Tetrad [6]

Tetrad is a software that creates and simulates data from, estimates, tests, and predicts with, and searches for causal and statistical models. The programme aims to provide sophisticated methods with a friendly interface that requires a user to have very little statistical or programming knowledge. It is not intended to replace flexible statistical programming systems such as Matlab, Splus, or R. Tetrad is freeware that performs many functions in commercial programmes such as Netica, Hugin, LISREL, and EQS as well as many discovery functions that these commercial programmes do not perform.

Mediation [7]

Mediation is a publicly available R package that enables both parametric and nonparametric causal mediation analysis in applied empirical research. In many scientific disciplines, researchers' main objective is not only estimating causal effects of a treatment but also understanding the

---

6 See: http://www.phil.cmu.edu/projects/tetrad

7 See: https://imai.fas.harvard.edu/software/mediation.html

process in which the treatment causally affects the outcome. Causal mediation analysis is frequently used to assess potential causal mechanisms. Mediation implements a comprehensive suite of statistical tools for conducting such an analysis.

Risks and challenges: questions to consider

A list of contextual factors, adapted from the work of the Humanitarian OpenStreetMap Team (2018) is useful to plan and analyse the political climate, development area, and technology present:

- Infrastructure: presence of a foundation of technology, such as hardware, software, networks, data centers, or electricity, supporting an organisation.
- Imagery: availability of good quality, high-resolution imagery from satellites or other sources.
- Internet: Internet connection and amount of bandwidth or level of connectivity.
- Usage of smartphones.
- Physical accessibility of the development programme areas.
- Literacy levels: community's ability to read and write and awareness of the basic conventions of maps.
- Cultural sensitivity.
- Partnerships and possible collaborations on the ground.
- Gender: best way to promote gender equality in operations while ensuring that no one is left behind.
- Characteristics of urban vs. rural settings.

Risks and challenges in Big Data projects

We list the factors to be considered in relation to the different stages of data use, based on the work of Data-Pop Alliance.

1. Collecting

- Privacy and legal concerns
  - Data ownership/stewardship: Who owns or is the custodian of the data?
  - What legal responsibilities and protections exist in that regard?

- Privacy protection in legal context
  - Do the project's data collection methods meet privacy protection guidelines?
  - What frameworks govern the use of personal and group data relevant to this project?

- Regional privacy protection laws
  - If multiple organisations are providing data, what is the best way to balance the laws/policies that apply to each data source?

- Risks in data governance
  - Latent group discovery: What risks are associated with the revelation of latent groups in using and aggregating user data with available open data resources?
  - Do people know how their data are being used? What do people know about how their data are being grouped or aggregated?
  - If the project is deployed across multiple countries, what is the process for balancing data collection/storage with the laws of each location? What is the balance of crowdsourced, collected (e.g., sensor, survey), and official data used in the project?
  - If multiple organisations are involved, what is the process to ensure that each data provider follows relevant guidelines and data use complies with collective guidelines?

- Challenges in representation and participation
  - Selection bias: Who is contributing to the data? Does everyone have the same access to the platform to contribute?
  - Sample bias assessment/correction: Does the project account for bias in where the data comes from (e.g., market share for mobile operators) or attempt to assess/communicate it?

- Challenges to data protection and consent
  - What are the data collection policies? How are people informed about them?
  - Informed consent: Are users made aware of when and why their data are being collected and how the data will be used?
  - Is consent affected by the usefulness of the project?
  - What are the consequences of denying consent for the user? Do users have reasonable alternatives if they opt out?
  - If using data sets that were previously collected, is this in line with the uses that were originally consented?
  - If the project involves data on individuals and groups, what safeguards can be put in place to protect them? And how can consent be ensured?

- Challenges in transparency
  - Is it clear to users which product/use they are contributing data to? Do they get a service/reward in return for their contribution?
  - Process transparency: Is information collected about users for analytical purposes (now or in the future)? Are users informed in an intelligible way about current and future uses (and implications) of any collected data?

2. Storing

- Privacy and legal concerns
  - Would the information that was collected expose users (or others) to risks if it were seen or found by unauthorised people?

- Risks in data governance
  - What are the relevant laws and policies that should inform how data will be stored?

- Challenges in representation and participation
  - Invisible populations: When working with existing data sets or records, is there good representation of who exists in the data? How is the coverage of the data discerned?
  - Data breaches and leaks: How do you protect stored data that are not intended to be accessed at all?

- Challenges to data protection and consent
  - Data retention: Are users aware of how and how long their data will be stored?
  - How long can data be retained? Can users opt out?

- Challenges in transparency
  - Collected metadata: Are metadata on methodologies, user consent, and acceptable uses recorded alongside data sets?

3. Processing/analysing

- Privacy and legal concerns
  - Does the analysis require using/exposing sensitive data and does this create risks to individual and group privacy?
  - Are there individual and group privacy risks that emerge when bringing together a broad range of data sets?
  - Are there individual and group privacy risks that emerge from studying and

using the data (particularly when bringing together a broad range of data sets)?

- Risks in data governance
  - Are the data being merged with other data sets? How can new risks created by merging multiple rich data sets be assessed?

- Challenges in representation and participation
  - Consistency/quality of the data: Do all users report in the same way?
  - Community engagement: Are the intended beneficiaries involved in the insights/decision-making process?
  - Is there a risk of the analysis producing outputs (e.g., results, recommendations, or pricing) that disproportionately affect certain individuals or groups (e.g., algorithmic discrimination)?

- Challenges to data protection and consent
  - How do you ensure protection of and consent for data from a public space (e.g., traffic cameras, transit, or urban labs)?

- Challenges in transparency
  - Replicability: Are the methodologies replicable/open source? Or is there a mechanism for community review, approval, or validation?

4. Sharing/accessing

- Privacy and legal concerns
  - What are the safeguards to prevent unauthorised access during sharing of sensitive/proprietary data in order to protect individual and group privacy?

- Risks in data governance
  - What are the safeguards to prevent unauthorised access during sharing of sensitive/proprietary data in order to protect individual and group privacy?

- Challenges in representation and participation
  - Access bias: If models/insights are proprietary, how is access determined? Do any people/groups face disproportionate barriers to access?

- Challenges to data protection and consent
  - User expectations and intended use: Does the distribution of the data for external research and development conform with users' expectations? Is there

informed consent?

- Challenges in transparency
  - Are the data inputs/products made available externally (for replication/validation)?

5. Transmitting/communicating

- Privacy and legal concerns
  - How can it be ensured that the guidelines/laws of all entities (e.g., organisations, countries) involved are taken into account and balanced/reflected in the resulting framework?
  - What legal and policy frameworks exist to guide how data and information can/should be shared (e.g., balancing confidentiality, proprietary, and sensitivity with free expression as well as journalistic integrity and responsibility)?

- Risks in data governance
  - How does the framework act as a tool for making decisions about the applications and implications of data collection, storage, analysis, and access?

- Challenges in representation and participation
  - Public benefit rationale: Does the project help raise awareness of issues and build capacities? Does the project have an engagement component?
  - What mechanisms could be used to disseminate insights to potential beneficiaries? (e.g., should an API be made?)
  - Representation: Do the data provide an unbiased picture of what is going on? Are there ways in which the data could be misused/misinterpreted?
  - How can it be ensured that the framework's metrics/processes/checklists can be generalised or adapted to a specific context in order to avoid oversimplification, edge cases, invisibility, and the false appearance of objectivity?
  - How can individuals/communities use the data product for engagement/advocacy? Are there foreseeable risks in making these tools available (e.g., unintended uses)?

- Challenges to data protection and consent
  - Are the terms of consent accessible or digestible by producers of information? What are the best modalities for consent?
  - Are alternative or future uses of data effectively communicated?

- Challenges in transparency
  - Freedom of information laws: Do any of the results become subject to freedom

of information laws? What risks might that create?
- What are the trust mechanisms for validating and reviewing the decisions that underpin a framework?
- If a tool is intended to convey a specific message, is that message communicated transparently and does it have the guise of objectivity/neutrality?

Case studies

We now turn to case studies to understand how Big Data and new technologies can be applied in practice in development programmes. A summary of the different development focus areas and applications is provided in Table 1 below, followed by a more detailed description.

Table 1. A summary of case studies

Focus area	Objective	Data sources	Metrics	Tools and methods	Privacy approach
Poverty	Mapping socio-economic levels (poverty) in Bangladesh using predictive maps	Remote sensing and geographic information system data Mobile operator CDRs	Demographic and Health Survey Wealth Index Household expenditures (Progress out of Poverty Index) Household income	Voronoi polygons Bayesian geostatistical models	Data aggregated at the level of physical cell towers to preserve the privacy of individual subscribers
Transportation	Analyse travel demand to increase system efficiency in Andorra	CDRs Traffic counts Road network infrastructure	User profile Realised trips Idealised trips Acceptable congestion	Origin-destination matrix Matrix factorisation (latent factor model) Recurrent neural network	Data aggregated by individual movements from one cell tower to another cell tower

				Network analysis Passive sensing (calls, SMS, and Bluetooth interaction) Classification algorithms Cross-validation	
Social engagement	Measuring interpersonal interactions at a university in the United States	User interactions: calls, SMS, and Bluetooth logs	Interpersonal social behaviour: closeness, trust, strength of ties, and friendship		Hashed (encrypted) personal identifiers and secured storage
Finance	Loan repayment prediction and assessment in a middle-income South American country	Raw transaction country records (calls and SMS), where 34% of adults had bank accounts and 89% had mobile phones	5,500 behavioural indicators that have some intuitive link to repayment	Random decision forests Logistic regression	Use of anonymised data Possible opt in (consent) from users to enhance privacy
Women's and children's health	Evaluate the impact of the Every Woman Every Child campaign and determine whether there was any change in general public awareness of issues related to women's and children's health	Fourteen million public tweets related to women's and children's health over a three-year period	Keyword counts	Machine learning classification algorithm	Not applicable

Poverty: Poverty maps using mobile phone and satellite data

Traditional approaches for measuring and targeting poverty rely heavily on census data, which are unavailable or outdated in most low-and middle-income countries. Alternate metrics can be used to complement such information. Research shows that it is possible to estimate and continuously monitor poverty rates at high spatial resolution in countries with limited capacities to support traditional methods of data collection (Steele, Sundsøy, Pezzulo, Alegana, Bird, Blumenstock, Bjelland, Engø-Monsen, De Montjoye, Iqbal, Hadiuzzaman, Lu, Wetter, Tatem and Bengtsson, 2017). The use of innovative data sources

(like satellite data and CDRs) and analytical methods can contribute to measuring the effectiveness of country-wide interventions, such as electricity infrastructure expansion and its indirect impacts on poverty levels.

Data sources

Remote sensing and geographic information system data can be used to measure distances to roads and cities reflecting access to markets and information.

Mobile operator CDRs can be used to track movements of mobile phones at an aggregate level. The structure/geographical reach of the calling networks of individuals can be correlated with remittance flows and economic opportunities.

Remote sensing and CDR data capture distinct and complementary features that describe relevant aspects of human living conditions and behaviour. Moreover, since these timely data are collected without interruption, the ability to use them to produce accurate maps offers the promise of ongoing sub-national monitoring required for the SDGs.

Approach

Steele et al. (2017, p.2) describe how "different approaches exist to calculate indicators of living standards, including the construction of one-dimensional and multidimensional indices, as well as the use of monetary and non-monetary metrics[...] monetary-based metrics identify poverty as a shortfall in consumption (or income) and measure whether households or individuals fall above or below a defined poverty line. By contrast, asset-based indicators define household welfare based on asset ownership (e.g., a refrigerator, radio, or bicycle), dwelling characteristics, and access to basic services like clean water and electricity". Steele et al. (2017) chose

> **Timely data, collected without interruption, offers the promise of ongoing sub-national monitoring required for the SDGs.**

Bangladesh to use different sources, such as remote sensing, CDRs, and traditional survey-based data, to understand the level of accuracy that different sources can reach when

estimating different measures of poverty. The methodology that is recommended for the evaluation presents an approach for poverty modelling which is flexible, and represents the first attempt to build maps using a combination of remote sensing and CDR data. Maps (in Figure 2) were generated using CDR features, remote sensing data, and Bayesian geostatistical models. The darker colour indicates poorer areas in prediction maps and higher error in uncertainty maps.

Figure 2. National prediction maps for mean Wealth Index in Bangladesh
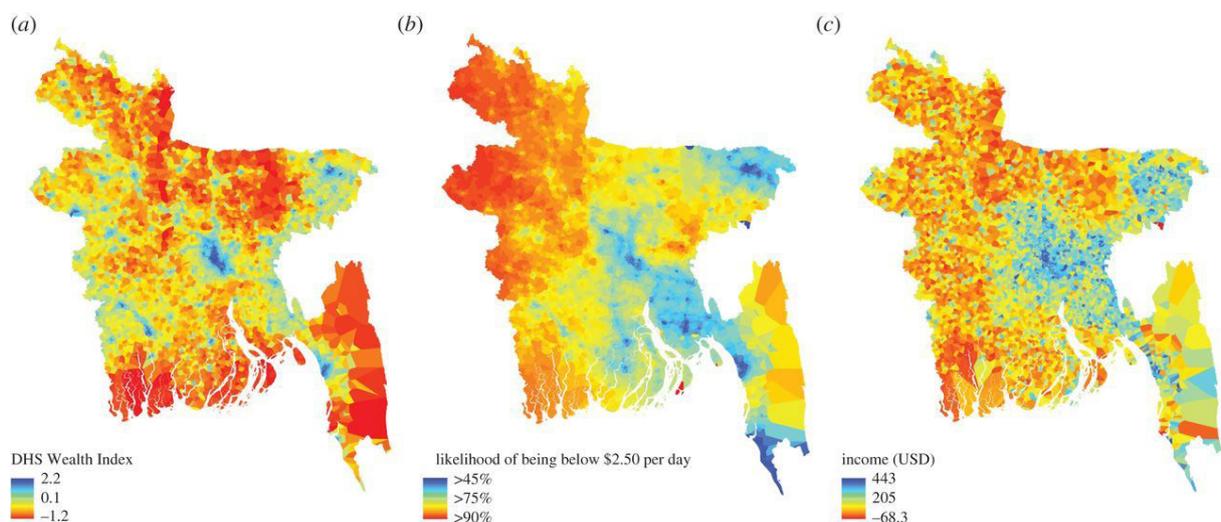
Source: Steele et al. (2017).

Furthermore, differentiating the estimations for urban and rural regions highlights the importance of specific data in distinct contexts. For instance, night-time lights and transportation time to the closest urban settlement were important nationally and in rural models for Bangladesh, while distances to roads and waterways were significant in urban and rural strata. The methodology provides additional support for existing evidence about correlations between socio-economic measures and night-time light intensity, access to roads and cities, entropy of contacts, and mobility

> ,, **Differentiating the estimations for urban and rural regions highlights the importance of specific data in distinct contexts.**

features. Such proof includes metrics of the brightness of artificial night-time lights data (Noor, Alegana, Gething, Tatem, and Snow, 2008). The methodology offers a robust and inexpensive alternative to asset-based poverty indices derived from survey data and information derived from night-time satellite imagery (see Ghosh, Anderson, Elvidge, and Sutton, 2013) and enables developing various globally consistent proxy measures of human well-being at the gridded, sub-national, and national levels.

Transportation: CDRs for traffic management

Uncoordinated travel behaviour can lead to higher traffic congestion and longer average travel delays. Mobile phone records—CDRs—have the capacity to not only reveal human mobility patterns but also enable the estimation of travel demand for system efficiency. Leng, Rudolph, Pentland, Zhao, and Kousopulous (2016) highlight the benefits of synergies among types of collective behaviour to increase system efficiency in Andorra, Spain.

Data sources

Different data sources can be combined to understand travel demand patterns and transportation system performance. For example, CDRs-Logs generated by telecommunications operators to reflect the use and attributes of telecommunication services, like calls or SMS. Table 2 presents an example of how information from CDR looks like.

Table 2. Sample fields in CDRs

interaction	direction	correspondent_id	datatime	call_duration	antenna_id
call	in	8f8ad28de134	2012-05-20 20:30:37	137	13084
call	out	fe01d67aeccd	2012-05-20 20:31:42	542	13084
text	in	c8f538f1ccb2	2012-05-20 21:10:31		13084

Source: Leng et al. (2016).

Geolocation information, such as antenna location (in Figure 3), is usually available in CDRs.

Figure 3. Example of cell tower distribution in Andorra by city
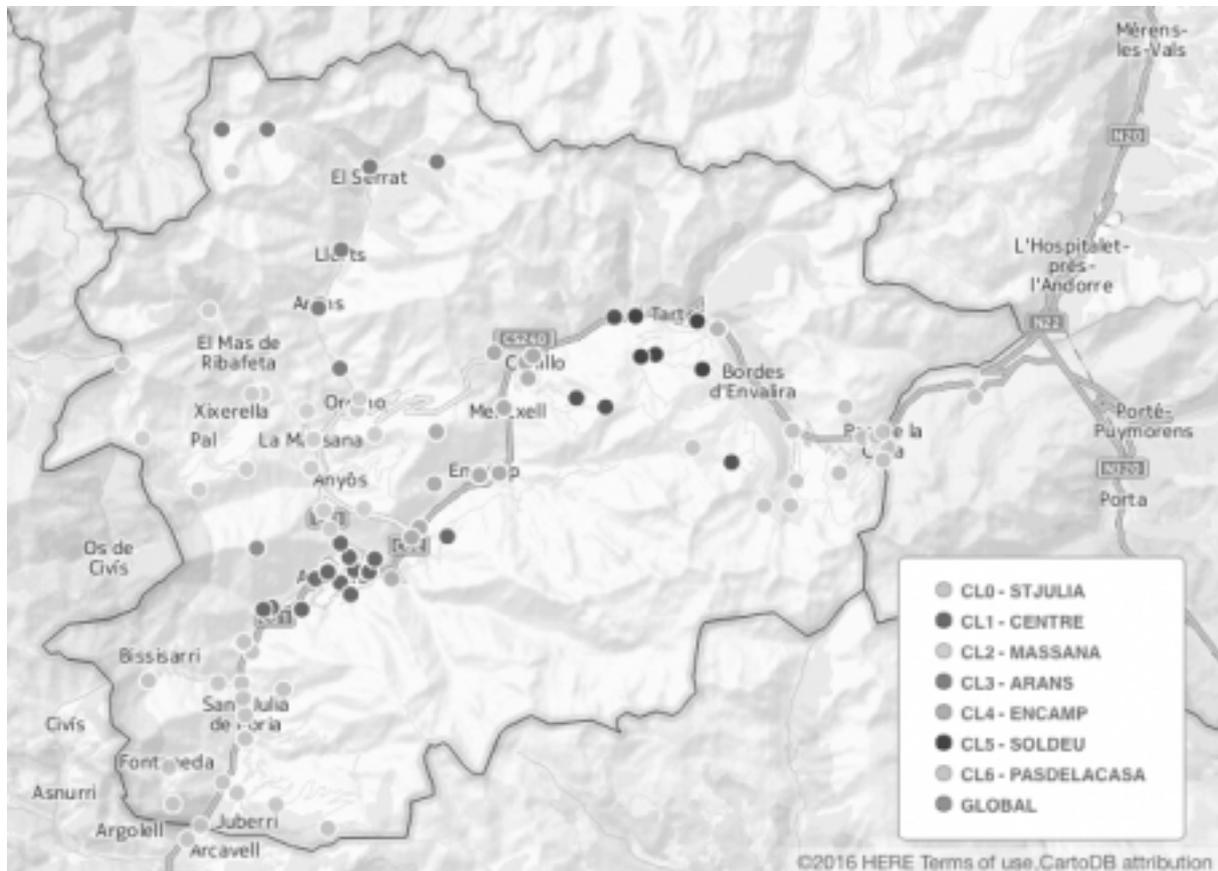
Source: Leng et al. (2016).

Indicators calculated based on CDRs and antenna records from telcommunications operators will be accessible via the OPAL platform currently being deployed in Africa and Latin America. Similar propriety solutions from these operators might exist.

Traffic counts and road network infrastructure: Actual traffic flows can be estimated from CDRs assuming that the number of travellers using their mobile phones while on the road is a constant fraction of the number of vehicles. Traffic counts can be complemented by metrics available from proprietary city traffic management systems, including data from cameras at key locations and geographic information system shape files.

Other relevant data attributes include connecting cities, number of lanes, road capacities, and optimistic/no-traffic travel time (available through Google Maps API).

Approach

Individual travel decisions made without accounting for system efficiency can lead to traffic congestion. Careful coordination of the travel behaviour of irregular travellers may reduce congestion and improve travel experiences, the quality of the environment, and life for local populations. Leng et al. (2016) provide a proven framework that can be used to assess traffic congestion and related time saving. It should also be applicable to internal fleet management by city transportation authorities. The framework is divided into three main steps: travel demand inference about vehicle trips along road links based on CDRs; personal location preferences (needs) based on location traces and next location estimation based on historical traces and recommended location comparison.

> **Different data sources can be combined to understand travel demand patterns and transportation system performance.**

The objective is to optimise travellers' location preferences with the constraint of acceptable congestion. Transportation authorities will have the freedom to trade-off between these two factors. The model can be applied to other cases with different capacity constraints. One could further establish a comprehensive framework for application in real-life situations as well as detail the distribution channels, spatial and temporal frequencies, target markets, and needs that should be studied further from a communication/marketing perspective. Integrating CDRs with other data, such as WiFi, Bluetooth, and geolocated social media data could supplement and enhance the proposed framework to provide better spatial and temporal resolutions.

Social engagement: Using social systems for measuring social interactions

Studies show that there is a strong correlation between the diversity of relationships among individuals and the economic development of communities. This correlation can be translated into an opportunity for city planners to introduce innovative social systems in order to assess the economy of communities (Nishikata, Hardjono, and Pentland, 2018). Networks of relationships, computational methods, and passive sensors can be implemented for detecting, understanding, and shaping human behaviour. Trust—an

important component—can be quantified, sensed, and applied towards this end (Shmueli, Singh, Lepri, and Pentland, 2014).

Data sources

Calls, SMS, and Bluetooth logs each capture a different feature of human interaction.

Approach

Traditional approaches adopted in social psychology to understand human behaviour include surveys. While theoretically sound, surveys can be expensive, unsuitable for long-term longitudinal analysis, subjective, and affected by perception bias (Fogg, 1999).

As listed in Table 3, sensing and understanding behaviour can be categorised according to two dimensions: personal versus interpersonal and short-term versus long-term behaviour.

Table 3. Sensing and understanding behaviour

	Personal	Interpersonal
Short-term	Emotional states Actions, poses, and gestures	Roles in meetings Outcomes of short-term group interactions
Long-term	Personality Health and wellness Financial behaviour	Community structure Organisational effectiveness

Source: Shmueli et al. (2014).

Interpersonal and long-term behaviour can be used to understand and measure community-related aspects like closeness, trust, strength of ties, and friendship. Wi-Fi and GPS logs are used for coarser analysis (but often with wider coverage) of people's mobility and co-location patterns. Shmueli et al. (2014) show how calls, SMS, and Bluetooth-based social interaction signals can be used to predict trust in relationships and quantify its role in mediating persuasion. As demonstrated through rich and dense sampling of the lives of over 100 participants living in a single community for a year it is possible to predict trusted relationships using computational methods based on passive sensing and network analysis. Trust can have a significant impact on social persuasion—trust is significantly more effective in determining behavioural change than the closeness of ties.

Finance: Mobile phone behaviour in loan repayment

Summary

Many households in developing countries lack formal financial histories, which makes it difficult for banks to extend loans and potential borrowers to receive them. Behavioural signatures in mobile phone data can be analysed to estimate loan repayment using call records (Bjorkegren and Grissen, 2018). Mobile phone behaviour can be used in new forms of lending to reach unbanked and underserved populations as well as in assessing the effectiveness of related programmes.

Data sources

The main source included raw transaction records (calls and SMS): 5,500 behavioural indicators with intuitive links to repayment. Performance is expected to increase with richer data and larger samples that can be observed over longer periods of time. Bjorkegren and Grissen (2018) use data from a telecommunications operator in a middle-income South American country, where 34% of households had bank accounts and 89% of households had mobile phones.

Approach

Two billion people across the world lack bank accounts. It can often be costly to introduce loans or replicate approaches from other regions for improving access to finance through bank branches and credit bureaus. Traditional approaches aggregate information on an individual's history of credit management. However, few households in developing countries have the formal records that generate such information. Still, many have maintained an interaction with a formal institution over a long period of time and can be traced via mobile phone activity recorded by a telecommunications operator.

As Bjorkegren and Grissen (2018) describe, it is possible to use low-cost methods to identify profitable investments, such as information on potential borrowers that is already being collected by mobile phone networks. The authors demonstrate that indicators of behaviour derived from mobile phone transaction records are predictive of loan repayment. This method has potential to achieve useful predictive accuracy, ensure stability over time, and provide indicators which are very similar to the information gathered by credit bureaus. Bjorkegren and Grissen describe how their findings "suggest that nuances captured in the use of mobile phones themselves can reduce information

asymmetries, and thus can form the basis of new forms of low cost lending. Together with mobile money, these tools are enabling a new ecosystem of digital financial services. This ecosystem is leading to what appears to be a revolution in access to finance in the developing world."

Women's and children's health: Understanding public awareness through social media

Online campaigns are an important component of advocacy and change in general public awareness can be determined by analysing social media. For example, the United Nations Global Pulse (2013) has measured the impact of the Every Woman Every Child campaign started by the United Nations in 2010 to mobilise and intensify global action towards saving and improving the lives of women and children around the world. Three years after the campaign was launched, an analysis of the number of Twitter conversations on relevant topics was conducted to understand the change in public awareness.

> **A change in general public awareness can be determined by analysing social media.**

Data sources

Using Crimson Hexagon's analytical tool ForSight, evaluators were able to access and analyse an archive of all public tweets from September 2009 to July 2013.

Approach

A taxonomy of relevant keywords was developed to identify messages related to women's and children's health (such as maternal health, breastfeeding, and vaccination of children). Following, a machine learning was leveraged applying an algorithm to recognise relevant tweets. Fourteen million public tweets were analysed to identify spikes, trends, and possible connections with real-life events and campaigns.

The findings showed that, over three years, the number of relevant tweets increased

tenfold, with the majority being on children's health. Some peaks in activities could be matched with special events or the launch of reports.

Looking forward: How can Big Data and AI revolutionise development?

The next technological frontier is AI and approaches that leverage machine learning and new data sources to make predictions, suggestions, and recommendations. AI is poised to affect aid effectiveness significantly in the next few years. Using AI for development effectiveness at scale will require having access to appropriate data to feed systems and doing so in a reliable, predictable, and ethical manner. Gaining access will take efforts that development effectiveness experts are familiar with—such as building trust, partnerships, data systems, and baselines—but with new stakeholders and incentives. The following pages are based on the recent paper Towards a Human Artificial Intelligence for Human Development written by Letouzé and Pentland in 2018.

Vision and pillars of Human AI

Over the past decade, the rise of Big Data and the recent emergence of AI have stirred hopes and increasingly fears about the fate of humankind. Are we heading towards brighter or darker times? Do Big Data and AI pose existential threats to democracy? What might be the impacts of data, algorithms, and AI on increasingly digital societies?

The first issue that should be discussed is the lack of a clear understanding of what is really happening and looming with Big Data and AI. The next is the lack of a long-term vision for how humans and machines can cooperate in the future. Yet another is the lack of a clear roadmap for mobilising and coordinating scarce resources, including both human and technological resources, towards that end. The last issue is the dominance of personal agendas favouring naive embrace or systematic fear-mongering of all things related to AI. In the final section of this study, we sketch an ambitious, optimistic vision and offer reflections on how human societies can shape their future, particularly how they could leverage AI by not just using it but also applying some of its key principles to build Human AI reflecting and serving the objectives and drivers of human development.

The gist and "good magic" of current AI

Big Data and current AI run on personal data generated by people using digital devices and services for their daily actions and interactions. Such use yields digital

signatures or data "crumbs" in the forms of mobile phone records, bank transactions, Web and social media content, geolocation data, photos, and videos. The resulting large data sets can be analysed by algorithms to unveil patterns and correlations. Most people already choose or are incentivised to rely on digital devices and services in decisions about which roads to drive, articles to read, clothes to buy, content to like, flights to book, and people to connect with. Doctors will soon use the same types of tools to diagnose diseases and suggest treatment plans. Current and future AI is and will be what was called Big Data a few years ago—computational analytics models fed and trained on large quantities of data crunched by computers to reach an objective, namely predict, optimise, suggest, recognise, or in some cases power sophisticated machines to implement decisions in a more or less autonomous manner.

The gist of Big Data and current AI is machine learning. Through many iterations and data-fuelled feedback loops, algorithms are able to identify and learn the features or combinations of features that most contribute to coming up with the "right" results. Of course, there are many caveats and complexities, but for most intents and purposes it suffices to understand that current "narrow" AI is about getting lots of data as inputs and learning how to connect them to output data (desirable or observed outcomes considered as the "right" results) through training, testing, and learning based on past cases. The reward (or credit assignment) function and learning through iterations lead to reinforcement of the combination of features to look for and use. In contrast, those features that lead to the "wrong" results will be weakened—a machine will gain an incentive not to use them.

The applications and implications of machine learning are already far-reaching. How are and should those inferences, estimations, projections, predictions, suggestions, and recommendations be used, by whom, and when, if at all? These concerns and risks are real and need to be known and addressed to limit the worst common side effects of technological change, at least in the short term, including widening inequalities. Big Data and AI are not "black magic" and the algorithms running them are not "black boxes." Given their ubiquity and power, it is important to understand how they do what they do and what insights could be gleaned from them to promote positive social change.

Applying the principles of AI to human systems towards Human AI

We call such a system Human AI. The basic principle is that, like with current narrow AI, what works to "get it right"—policies, programmes, behaviours, and actions—would get rewarded and reinforced. What "does not work" would be penalised and weakened. Both would be enabled by data-fed feedback loops. With time, we would have human systems (societies, governments, and organisations) with a good sense of what works,

in other words which sets of policies, programmes, behaviours, and actions yield good results. In addition to providing the core analogy, AI would be a central part of this system by generating and crunching data, taking over tasks, and assisting decision making under general human oversight. With Human AI, governments, companies, or the aid sector could apply AI tools to identify and reinforce what contributes to the "right results," including outcomes of aid programmes. Through feedback, they could also better understand whether these effects are desirable in the long term. What is key is learning and agreeing on what yields good versus bad results and acting accordingly the next time around. Over time, what helps to reach good results will gain importance over what does not, and become prevalent, ideally turning into norms that need less enforcement. Human systems would be better off—safer, fairer, more civil, and more sustainable—because the opposite results do not "work" for most people.

Human AI is not an Orwellian vision where citizens would be digitally monitored and rated in real time. This proposal would merit further attention, but a few points can be noted. First, societies have lots of systems—such as taxes and credit scores—in place to influence individual behaviours in ways deemed desirable. Second, the focus of Human AI is on collective rather than individual actions—instilling a culture and setting up the necessary systems and standards to improve collective decisions.

For this proposal to work, there needs to be a general agreement that decisions and outcomes ought to be evaluated on the basis of data. It may not be easy to agree on the features and factors of "good results," but assessing them on the basis of facts should at least be agreed. A preconceived agreement on the targeted level of income inequality is not necessary, but a starting point should be agreeing that and how inequality should be measured. From there, what contributes to different levels of inequality and their associated results can be understood in turn. Human AI requires a general agreement that facts should matter because otherwise systems cannot learn and improve.

Challenges and impediments to designing Human AI

This sketch of the vision of Human AI has left out many challenges, the biggest of which we can only briefly discuss. Most importantly, some of the examples mentioned above are contentious since there is no consensus on them. There is also ample evidence that facts alone do not change people's minds and even recognising facts may be getting harder in this data-rich era. Tensions between data, information, and facts are not new, but there is a sense that it becomes harder for facts to be recognised and agreed upon in a world awash in data. While so far we treated data and facts as synonyms for simplicity, there are differences between data and facts. For Human AI to work, there ought to be something more, like "connective tissue" that allows learning to happen, information to

flow, and facts to matter. Key elements for this seem to include greater trust, empathy or rational compassion, shared experiences, and mingling among and between individuals and groups.

Another basic challenge is to know what actually "works" and how, when, and where it "works." There are broad areas of agreement, but no consensus on the right sets of policies. This challenge applies to almost all domains of social life because assigning causality or credit is difficult in complex systems where so many variables (and values) interplay both as inputs and outputs. A major challenge is agreeing on what the good results ought to be. With current AI, the final result is largely given—this is not the case with Human AI. Should societies aim for perfect income equality? Should economic policies aim to raise gross domestic product, with all its limitations? Should prolonging life be the objective of any treatment? Soon values come into play. Opponents of hitting children, torture, or the death penalty will also argue along moral lines, irrespective of efficiency and outcomes. There is a need for rational, outcome-based arguments in many of these debates to promote gradual adjustments and improvements.

Further, an important challenge is access to the data that power current AI. These sensitive data hold most keys to figuring out and advocating convincingly for what "works," showing what does not, and pointing towards means of improvement. For instance, assessing whether a new transportation system may result in increased economic opportunities and lower criminality would be significantly improved by having access to fine-grained mobility data from mobile phones. The vast majority of data that power AI are collected and stored by private companies that legally act as data controllers. There have been many proposals for and discussions about data sharing projects and agreements, but to date there are no systemic standards and norms for accessing the data ethically and safely at scale in ways that could power Human AI.

Lastly, there is the privacy imperative—privacy is a fundamental human right. The vision of Human AI is not about looking into individual records or targeting specific individuals and groups. In any case, looking and targeting would not work. Recent societal reactions and legal trends suggest that while people's attitudes towards privacy may be changing, the demolition of privacy is not a marker or driver of human development. There is no need to encroach on privacy for Human AI to work since aggregated anonymised indicators suffice.

Requirements and priorities for designing Human AI

What is required for Human AI? It will take several key elements. For example, nurturing a strong, healthy data culture, including widespread data literacy, with more

trust and interest in evidence-informed debates in societies. Further, building better public governance for the systems that provide the data that can power Human AI—including private sector data systems—and allow essential data to be tapped into ethically and safely. These elements are among the key objectives of the aforementioned OPAL project. OPAL aims to allow accredited users to query private sector data through open algorithms running on the servers of private partner companies, behind their firewalls, to extract aggregated data for indicators of interest, from mobile phone activity, bank transactions, hospital records, police data, and more. With OPAL, no sensitive data ever leaves the servers of data partner companies. All queries are logged and auditable, while all algorithms are open and subject to scrutiny and redress. OPAL also aims to develop governance standards and processes that will allow data subjects to weigh in on the kinds of analyses done using data about themselves, including through local oversight bodies referred to as Councils for the Orientation of Development and Ethics (CODEs). Currently piloted in Colombia and Senegal by national statistical offices with two leading telecommunications operators, OPAL is the first real-world attempt at setting up technological systems and governance standards for building Human AI.

Notably, Human AI also requires developing incentives and means for civil society organisations, researchers, and regulators to demand evidence-based policies. There should be incentives to request that the effectiveness of publicly financed programmes be assessed using the best available data and methodologies. Data for transparency and rational compassion are a required for dealing with fake news and demagoguery. The promotion of Human AI to improve society is not the promotion of a technological utopia but rather an "aspirational analogy" that places good data sources and rational discussion frameworks at the core of a new social contract between humans as well as between humans and machines in 21st-century societies. This vision intends to improve rather than prove and requires building a body of evidence that demonstrates or at least suggests that some sets of actions yield better outcomes. The focus is on instilling such a culture and setting up the necessary tools and systems for it to work in the future— whether the future is in five, 10, 20, or 50 years.

Conclusion

In complex development contexts, it is especially hard to assign causality and draw conclusions about the effectiveness of large financial flows, tasks made even more difficult when there are many degrees of separation between them, as in the case of assessing official development assistance flows. As great opportunities come from the triangulation of different methods, combining traditional approaches with the use of Big

Data and new technologies leads to some interesting new opportunities. In the fast-changing field of AI, its relevance to M&E will likely grow. The vision of Human AI, a body of evidence confirming or at least suggesting that some sets of actions yield better outcomes, outlined in this study promises to improve rather than prove and could be part of development programming in the future.

Overall, data innovation can support people-centric assessments by making results and challenges clearer as well as helping understand the truth on the ground—localisation—while diversifying an approach away from being a one-size-fits-all approach. Data analysis tools and techniques, including machine learning, can complement conventional evaluation methodologies by providing cheap, quick, complexity-sensitive, longitudinal, and easily analysable data. Using multiple data sources can overcome the scarcity and unavailability of relevant information and provide new insights on human

Using multiple data sources can overcome the scarcity and unavailability of relevant information.

processes and experiences. The presented case studies offer insights into the advantages related to the analysis of data, such as social media content, which enable the provision of near real-time and fine-grained mobility or poverty estimates. Big Data and satellite images can also increase the possibility of creating feedback loops and lead to the temptation to bypass scientific design and other standard considerations, like ethics and political repercussions. The toolkit provided in this study should be used to harness innovative data sources in assessment programmes designed to respond to needs in specific contexts. It also contains a list of important questions to be considered when designing in-country pilots, mainly related to privacy and legal concerns.

Our proposed methodology to harness innovative data and technology to measure development effectiveness includes a combination of stages. These stages involve collecting data from people and devices, storing data for future use, data processing, analysing data to generate useful insights, sharing data with stakeholders, and transmitting information to the outside world. Special attention to specific, contextual factors will be required in the individual countries chosen for possible pilots. For example, technological infrastructure, the current political climate, and different cultural sensitivities of local communities may be salient factors. It is recommended to invest early in developing mutually beneficial partnerships, including with the private sector, to get access to complementary expertise, capacities, and data sources.

As technologies and capacities keep evolving, it is inevitable and largely desirable that development aid systems will adapt, as they have done for decades. Learning how to use new data sources and technologies to meet objectives must continue. Based on experience, these systems will also, in turn, contribute to shaping next generations' data and technological systems and standards to ensure that they do not inadvertently harm vulnerable groups. This exchange of knowledge and continuous learning through feedback are the hallmarks of agile systems and AI. Overall, the tools, methodologies, and culture among professionals in the field of development could put them in a unique position to be drivers of larger positive systemic changes. By taking inspiration from and using AI approaches to develop and be part of Human AI, there would be a striving to be better at finding out and doing more of what "works" and less of the opposite.

References

Almås, I., Attanasio, O., Jalan, J., Orteiza, F. & Vigneri, M. (2018). *Using data differently and using different data.* Journal of Development Effectiveness. Retrieved from: https://cedilprogramme.org/wp-content/uploads/2017/12/EDePo-Using-Data-Differently-DRAFT.pdf

Apolitical (2018, October 29) *Policy evaluations fail too often. Here's how to make them more nimble.* Retrieved from: https://apolitical.co/solution_article/policy-evaluations-fail-too-often-heres-how-to-make-them-more-nimble/

Bamberger, M., Vaessen, J. & Raimondo, E. (2016) *Dealing with Complexity in Development Evaluation: A Practical Approach.*

Bamberger, M., Tarsilla, M. & Hesse-Biber, S. (2016). *Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute.* Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/26874234

Bamberger, M. (2017). T*he Future of Development Evaluation in the Age of Big Data.* Retrieved from: http://merltech.org/the-future-of-development-evaluation-in-the-age-of-big-data/

Bjorkegren D. & Grissen D. (2018). *Behavior revealed in mobile phone usage predicts loan repayment.* Brown University. Retrieved from: https://dan.bjorkegren.com/danbjork_grissen_creditscoring.pdf

Boyd, D. and Crawford, K. (2012). *Critical questions for Big Data.* Information, Communication & Society, 15:5, 662-679

Data-Pop Alliance (2017). *Leveraging Big Data for Sustainable Development: Toolkit.* Unpublished document.

Dinshaw, A., McGrayl, H., Rai, N. and Schaar, J. (2014). *Monitoring and Evaluation of Climate Change Adaptation: Methodological Approaches.* Organisation for Economic Co-operation and Development, Environment Working Papers 74, No. 74. Paris: OECD Publishing.

Driscoll, K. & Walker, S. (2014). *Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data.* IJoC, University of California

Fogg, B. (1999). *Persuasive technology: Using computers to change what we think and do.* Morgan Kaufmann.

Giller, K., Bell, S., Mock, Nancy B., Hijmans, R., Sayer, J. & Serraj. (2014). *Data, metrics and monitoring in CGIAR – a strategic study.* Rome: CGIAR Independent Science and Partnership Council (ISPC)

Ghosh, T., Anderson, S. J., Elvidge, C. D. & Sutton PC. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. *Sustainability* 5, 4988 – 5019.

Hellerstein, J. (2008, November 19). *The commoditization of massive data analysis.* O'Reilly Radar. Retrieved from: http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html

Horrigan, M.W. (2013, January 1). *Big Data: A perspective from the BLS.* AMStatNews, Retrieved from: http://magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/.

Humanitarian OpenStreetMap Team. (2018). P*articipatory Mapping Toolkit: A Guide for Refugee Contexts.* Retrieved from: https://www.hotosm.org/downloads/Toolkit-for-Participatory-Mapping.pdf

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). *Significance and challenges of big data research.* Big Data Research, 2(2), 59–64.

Kharas, H., Hamel, K. & Hofer, M. (2018, June 19). *The Start of a New Poverty Narrative.* Brookings Institute. Retrieved from: https://www.brookings.edu/blog/future-development/2018/06/19/the-start-of-a-new-poverty-narrative/

King, G. (2013, May 14). *Big Data is Not About the Data!* Retrieved from: http://gking.harvard.edu/presentations/big-data-not-about-data-1

Leng, Y., Rudolph, L., Pentland A., Zhao, J. & Kousopulous, H. (2016). *Managing travel demand: Location recommendation for system efficiency based on mobile phone data.* Cornell University Library

Lant, P., Woolcock, M., & Andrews, M. (2010). *Capability Traps? The Mechanisms of Persistent Implementation Failure.* Retrieved from: https://www.cgdev.org/sites/default/files/1424651_file_Pritchett_Capability_FINAL.pdf

Letouzé, E. (2012). *Big Data for Development: Challenges and Opportunities.* UN Global Pulse. Retrieved from: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

Letouzé, E. (2013, June 11). C*ould Big Data provide alternative measures of poverty and welfare?* Development Progress Blog. Retrieved from: www.developmentprogress.org/blog/2013/06/11/could-big-data-provide-alternative-measures-poverty-and-welfare;

Letouzé, E. (2014, November 25). *Reflecting on the Future of M&E in the Age of Big Data.* Retrieved from: http://datapopalliance.org/reflecting-on-the-future-of-me-in-the-age-of-big-data/

Letouzé, E. (2015) *Big Data and Development: General Overview Primer.* Data-Pop Alliance White Paper Series. Data-Pop Alliance, World Bank Group, Harvard Humanitarian Initiative. Retrieved from: http://datapopalliance.org/wp-content/uploads/2015/12/Big-Data-Dev-Overview.pdf

Letouzé, E., Meier, P., & Vinck, P. (2013). Big data for conflict prevention: New oil and old fires. In Mancini, F. (Ed.), *New Technology and the Prevention of Violence and Conflict,* International Peace Institute, New York. Retrieved from: http://www.ipinst.org/wp content/uploads/publications/ipi_epub_new_technology_final.pdf;

Letouzé, E. & Pentland, A. (2018, December 6) *Towards a Human Artificial Intelligence for Human Development.* ITU Journal: ICT Discoveries, Special Issue No.2

Mann, L. (2017) *Left to Other People's' Devices? A Political Economy Perspective on the Big Data Revolution in Development.* Development and Change 0(0): 1–34.

Nature. (2007, October 11). *A matter of Trust.* Nature volume 449, pages 637–638

Nishikata, T., Hardjono, T., & Pentland, S. (2018). *Social Capital Accounting.* Retrieved from: https://www.media.mit.edu/publications/social-capital-accounting

Noor A. M., Alegana V. A., Gething, P. W., Tatem, A.J. & Snow, R.W. (2008). *Using remotely sensed night-time light as a proxy for poverty in Africa.* Popul. Health Metr. 6, 5.

Organisation for Economic Co-operation and Development. (2006). *Aid Effectiveness: Three Good Reasons Why the Paris Declaration Will Make a Difference.* In: *Development Co-operation Report 2005: Efforts and Policies of the Members of the Development Assistance Committee.* Paris: OECD.

Organisation for Economic Co-operation and Development. (2011). *Aid Effectiveness 2011: progress in implementing the Paris Declaration.* Paris: OECD.

Paul, A., Jolley, C. & Anthony, A. (2018) *Reflecting the Past, Shaping the Future: Making AI work for International Development.* USAID. Retrieved from: https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf

Pentland, A. (2012, August 30). *Reinventing society in the wake of big data.* Edge.org. Retrieved from: www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data.

Pentland, A. (2017, November 1). *Human-AI Decision Systems.* Retrieved from: http://

thehumanstrategy.mit.edu/blog/human-ai-decision-systems.

Pritchett, L., Samji, S., & Hammer, J. (2013). *It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space.* Retrieved from: https://www.cgdev.org/publication/its-all-about-mee-using-structured-experiential-learning-e-crawl-design-space

Scannapieco, M., Virgilito, A., & Zardetto, D. (2013). *Placing Big Data in Official Statistics: A Big Challenge?* 2013 New Techniques and Technologies for Statistics conference, Brussels

Scheel, S. & Ustek-Spilda, F. (2018). *Big Data, Big Promises: Revisiting Migration Statistics in Context of the Datafication of Everything.* Retrieved from: https://www.law.ox.ac.uk/research-subject-groups/centre-criminology/centreborder-criminologies/blog/2018/05/big-data-big

Shmueli, E., Singh, V. K., Lepri, B., & Pentland, A. (2014). *Sensing, understanding, and shaping social behavior.* IEEE Transactions on Computational Social Systems, 1(1), 22–34. doi: 10.1109/TCSS.2014.2307438

Sivarajah, U., Muhammad M., Zahir I. & Vishanth, W. (2016). *Critical analysis of Big Data challenges and analytical methods.* Journal of Business Research 70, 263-286.

Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y.A., Iqbal, A. M., Hadiuzzaman, K.N., Lu, X., Wetter, E., Tatem, A.J., & Bengtsson, L. (2017). *Mapping poverty using mobile phone and satellite data.* J. R. Soc. Interface. doi https://doi.org/10.1098/rsif.2016.0690

Toyama, K. (2010, November 1). *Can technology end poverty?* Boston Review. Retrieved from: http://bostonreview.net/forum/can-technology-end-poverty

United Nations. (2017). *The Sustainable Development Goals Report.* New York. Retrieved from: https://www.un.org/sustainabledevelopment/blog/2017/07/pace-of-progress-must-accelerate-to-achieve-the-sdgs-finds-latest-un-progress-report/

United Nations Development Programme. (2018). *Development 4.0: Opportunities and Challenges for Accelerating Progress towards the Sustainable Development Goals in Asia and the Pacific.* Retrieved from: http://www.asia-pacific.undp.org/content/rbap/en/home/library/sustainable-development/Asia-Pacific-Development-40.html

United Nations Development Programme and the World Bank. (2016). *Transitioning from the MDGs to the SDGs.* Retrieved from: http://www.undp.org/content/dam/undp/library/SDGs/English/Transitioning%20from%20the%20MDGs%20to%20the%20SDGs.pdf

United Nations Global Pulse. (2013). *Advocacy Monitoring through Social Data: Women's and Children's Health.* Retrieved from: https://www.unglobalpulse.org/projects/ EWEC-social-data-analysis

United Nations Global Pulse. (2016). *Integrating Big Data into the Monitoring and Evaluation of Development Programmes.* Retrieved from:http://unglobalpulse.org/ sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

United Nations Global Pulse and International Association of Privacy Partners. (2018). Retrieved from: https://iapp.org/resources/article/building-ethics-into-privacy-frameworks-for-big-data-and-ai/

United Nations Secretary General Independent Expert Advisory Group. (2014, November). *A world that counts: mobilizing the data revolution for sustainable development.* Retrieved from: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

United Nations Secretary General. (2018). *United Nations Secretary General's Strategy on New Technologies. (2018).* Retrieved from:http://www.un.org/en/newtechnologies/ images/pdf/SGs-Strategy-on-New-Technologies.pdf

Watson, C. & Lizzi, A. (2018). *Opinion: Data-driven adaptive management for international development, Devex World: Sponsored by Premise.* Retrieved from: https://www. devex.com/news/sponsored/opinion-data-driven-adaptive-management-for-international-development-92994

Woolcock, M. (2009). *Toward a plurality of methods in project evaluation: a contextualized approach to understanding impact trajectories and efficacy.* Journal of Development Effectiveness, 1:1, 1-14

Zicari, R. V. (2014). *Big Data: Challenges and Opportunities.* Big data computing (pp. 103–128). Florida, USA: CRC Press, Taylor & Francis Group.