WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Data-Driven Development
Pathways for Progress

January 2015

© World Economic Forum
2015 - All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

The views expressed are those of certain participants in the discussion and do not necessarily reflect the views of all participants or of the World Economic Forum.

REF 060115

Contents

4 Executive Summary

6 The Evolving Data Landscape

8 Addressing the Data Deficit

12 Strengthening Data Governance

14 The Role of Communities and the Individual: Digital Identity and Human-Centred Design

16 Call to Action

17 Acknowledgements

Executive Summary

> **Never again should it be possible to say 'we didn't know.' No one should be invisible. This is the world we want – a world that counts.**
>
> UN Secretary-General's Independent Expert
> Advisory Group on a Data Revolution for Sustainable Development, 2014

When it comes to the opportunity for data analytics to provide new insights and visibility on the challenges of sustainable development, the world is in a pre-dawn fog. Defined currently by its ambiguity and complexity, the global dialogue on the use of data is shaped by both hope and anxiety. While using data for positive socio-economic gain holds great promise, it is not a certainty. It is imperative to chart a course through this complexity to capture the development benefits that Big Data can bring.

However, the global narrative on the use of data for development is conflated and incoherent. Competing tensions on data control and ownership, limited technical understanding, the lack of coordination, shifting power dynamics and a lack of effective governance frameworks have conspired to hinder clarity of integrated goals and principles across different communities of interest. This report aims to help clarify some of these complexities to provide a more coherent path forward. Based on insights from a cross-disciplinary community of policy-makers and experts from industry, academia and civil society convened by the World Economic Forum, this report hopes to help clear some of the fog that is currently obscuring the vision of how Big Data might be used to address the challenges of sustainable development.

Priorities for Sustainable Data-Driven Development

- Commercial incentives and trusted agreements need to be established to enable access to data streams held by private actors.
- Shared policy frameworks, legal protections and impact assessments need to be developed to strengthen trustworthy data flows.
- Capacity building at the institutional, community, local and individual level needs to be strengthened.
- Individuals must be recognized as both producers and consumers of data with unique digital identities.

With a lens focused on identifying various stakeholder relationships, the hope is that new approaches for balancing competing tensions can be advanced. This report suggests that those tensions reflect underlying inequalities in access to data (and the resources, capacity and infrastructure to use the data), and a growing trust deficit that assumes the misuse of data.

To achieve the goals of sustainable development, it is important to build the legal, cultural, technological and economic infrastructures necessary to enable the balancing of competing interests. Balance will require addressing multiple concerns about the use of data and issues such as privacy, human rights, property rights, climate change and national security. The approaches will need to be meaningful, pragmatic, adaptive and proportional. With so much uncertainty, the need for continuous experimentation, learning and sharing is paramount. Investing in small-scale pilots that bring together the private sector, regulators, civil society and local communities will provide the insights and local knowledge critical for long-term resilience and adaptation.

So where to begin? How can all stakeholders work through the complexity and identify key points of focus for collective progress? Three areas are emerging as important starting points: addressing the data deficit, establishing resilient governance, and strengthening local capacities and knowledge regarding the individual and the community. As the report by the UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development states, "There is a need for showing how resources, actors, forms of collaboration and institutions can evolve, be managed and be deployed to make the data revolution a force for progress and for enhancing possibilities.[1]" The intent of this report is to support those aims.

From Complicated to Complex

Understanding the dynamics of complexity is critical for addressing the convoluted set of political, commercial, civil and technical issues surrounding the issue of Big Data for sustainable development.

Complex challenges differ from complicated ones. For example, an automobile is complicated. It has a variety of interconnected parts which deliver very reliable and predictable outcomes. Step on the gas and it moves. Hit the brakes and it stops.

But driving the car in traffic is complex. It requires constant adjustments and application of a dense body of collective norms, rules and coordinated actions among strangers. Social coordination renders complex environments safe and resilient.

When looking at the challenges confronting Data for Development, it is helpful to discern those challenges which are generally complicated from those which are complex. Many of the technical challenges surrounding the standards and technological tools of Big Data are complicated. They are difficult to address, but tractable and known.

The challenges surrounding the policies and rules of how data can and should be used are complex. Contextual and nuanced, the uncertainties of what defines an acceptable use of data are often fuelled by power dynamics, outdated policy frameworks, unclear ethics, cross-jurisdictional legal uncertainties and an underlying lack of trust among stakeholders.

Resolving these competing commercial, political and civil interests will require deep commitment and recognition of the right moment in time for enabling positive change.

The Evolving Data Landscape

It goes without saying that a unique technology-fuelled global transformation is underway. The worldwide increase in digital connectivity, the global scale of highly personalized communications services, and advances in data analytics have coalesced to create a powerful platform for change. In this networked world, people, objects and connections are producing data at unprecedented rates, both actively and passively.

"Big Data" is the engine of this growth. A concept central to the Data Revolution, it is a term with multiple and varied definitions. For the purposes of this report, Big Data will be defined by the so-called "Four Vs": volume (massive and passively generated); variety (originating from both individuals and institutions at multiple points in the data value chain); velocity (generally operating in real time); and veracity (referring to the uncertainty due to bias, noise or abnormality in data).

Against this backdrop, the ways in which data can be leveraged to positively impact the lives of the most vulnerable are just beginning to emerge. Because of its detail, timeliness, ability to be utilized for multiple purposes at scale and to make large portions of low-income populations visible, the potential for data-driven development is unprecedented.

At the same time, it is a domain that provides keen insights into people's lives, behaviours, health, prosperity, needs and aspirations[2]. To address these concerns, it is critical to clearly understand the context and nature of the local development challenges that individuals face before initiating data-driven interventions.

At present, three general criteria determine the appropriateness of using data: ethics (the underlying principles for using data); accountability (how effectively the principles are implemented and enforced); and veracity (the accuracy and completeness of the underlying data sets). The absence of broadly shared processes, paradigms and measurements that can help dissipate these tensions is an area requiring much additional work. Easy answers do not work because they simply mask the deeper complexity of interrelated challenges which will need to be continuously managed and rebalanced.

An example of these complex challenges can be seen in the 2014 Ebola crisis in West Africa. Despite months of talks between health officials, UN agencies, mobile network operators and governments, getting access to mobile network operators' data on population movement was problematic.

> **"By having granular data that captures the experiences of poor communities, along with the analytical techniques needed to decipher that data, researchers and development practitioners can improve the accuracy, effectiveness and reach of their initiatives. Practitioners in the field of economic and social development can better monitor and track the progress of their programmes in almost real time, bring projects to scale at a lower cost, gather rapid feedback from the field, collaborate more effectively with stakeholders, and demonstrate impactful outcomes."**
>
> Jake Kendall, Bill and Melinda Gates Foundation

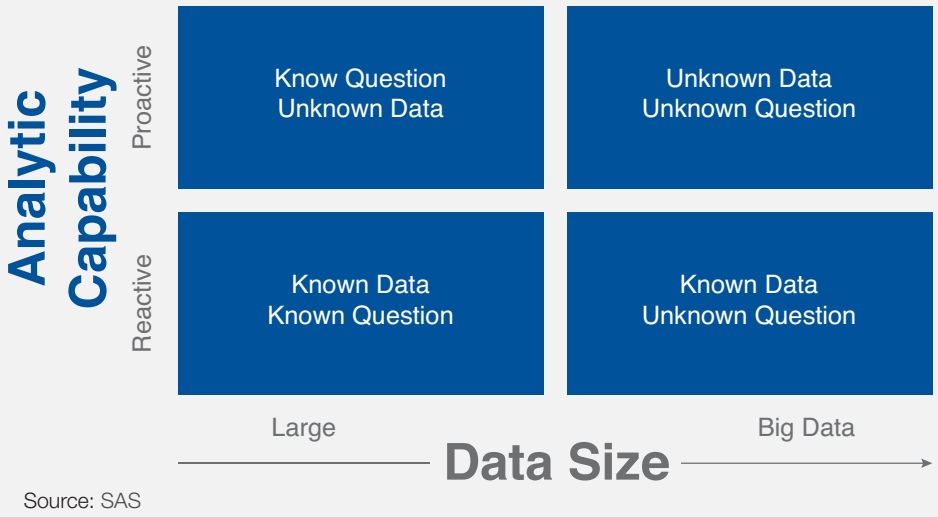
A number of factors created the entanglement: commercial interests (brand reputation risks, fear of having operating licenses revoked and disclosure of proprietary information); ethical concerns (privacy); national security concerns (releasing population movement details to third parties); regulatory uncertainty (vague legal liabilities); and knowledge and leadership gaps (lack of organizational prioritization) were just some of factors contributing to the stalemate. As stated in *The Economist* magazine, "Because there was no precedent for using call detail records in an emergency like Ebola, it was hard to bring the parties together at a high-enough political or management level to make decisions.[3]" There was no meta-institutional narrative of data sharing habits to help bring stakeholders together.

The consequence of this uncertainty is that the global dialogue on data for development is polarized. The optimists are advancing somewhat utopian views of the vast potential of using data for the common good. Advocates (often supported by well-funded public relations campaigns positioning technology executives as leaders) argue that with meaningful controls in place, a whole new range of digital insights can be applied to help track the outbreak of infectious diseases, strengthen resilience following natural disasters, enhance access to financial services for the poor and understand migration patterns of vulnerable populations.

Likewise, the pessimists with equally strong voices are pointing to dystopic futures dominated by "digital extractive industries", which leverage incumbent power asymmetries that are enabled by governments and industry alike. Headlines over how data have been used for private sector and government surveillance, identity theft, discrimination against minorities and a host of other harms have made this a non-academic debate. Underlying this view is the notion that the trust, transparency and control that individuals have regarding the use of data about them is significantly constrained and will need to be addressed for an ecosystem that is sustainable over the long term.

Big Data vs Smart Questions

One perspective on the public/private bureaucracy preventing the sharing of population movement data to stop Ebola is that the framework for delineating the types of analysis needed and the appropriate safeguards to prevent data abuse are confused.



Source: SAS

- ***Known Data and Known Question (lower left):*** This quadrant is for optimizing data for standardized processes and procedures. The questions are known and so are the data sources. Data quality, accuracy and timeliness are critical. Many of the issues for strengthening the capacities of national statistics offices fall into this quadrant. The challenges are "known knowns" and operational in nature.

- ***Known Data and Unknown Question (lower right):*** This quadrant is for domain experts to discover questions "they didn't know to ask." In this area data sets are known but their combined value to discover new correlations is unique. The key outcome is for experts to discover new knowledge to build sophisticated data models.

- ***Known Question and Unknown Data (upper left):*** This quadrant is about providing existing data models with access to specific data resources. In this quadrant, the data doesn't need to move or be pooled. Innovative data models just need access to tightly controlled data sets. Much of the confusion and inertia within the Data Revolution is occurring in this quadrant. Combining large data sets for discovering new insights isn't needed (quadrant 2). Rather, data models just need access to data that can then be turned into actionable information (quadrant 3).

- ***Unknown Question and Unknown Data (upper right):*** Quadrant 4 is about "unknown unknowns" where machine learning, massive/passive data sets and real time, personalized feedback loops come into play. This is where explorative, predictive and sentient computing comes into play that can account for the dynamic complexity of the world and stay ahead of human decision-making which is often slow and uninformed. These can feed into new types of data-driven decision support tools.

Addressing the Data Deficit

In broad terms, the Global North is experiencing a "data deluge", which provides many positive opportunities for socio-economic change. While the governance, access and quality of this data remain problematic (despite impressive technological advances), constraints on the amount of available data are not a concern. In the Global South, however, there is a relative "data deficit" because of significant constraints on the creation, collection and use of data.

Accurate, timely, disaggregated and accessible data are essential for governments to deliver services efficiently, fairly and transparently. As noted in a 2014 report by the Center for Global Development:

"*Basic data like births and deaths, the size of the labor force, and the number of children in school are fundamental to governments' ability to serve their countries to the fullest. And good data that are reliable and publicly available are a catalyst for democratic accountability. Data allow citizens to hold governments to their commitments. They allow governments and donors to allocate their resources in a way that maximizes the impact on people's lives. And they allow us all to see the results.*"[4]

One of the main challenges the poor face is being "under known". They are often not referenced in databases that are generated from interactions with official organizations, as their interactions are infrequent. They may not even exist in government voter or birth registries, let alone in credit registries or other data sources. This "administrative invisibility" prevents them from gaining access to government services to which they are entitled and often prevents them from being offered life-enhancing services (e.g. credit) by the private sector as well.

A number of dimensions of the "gaps" contribute to the data deficit. The first is a "data gap", which refers to data needed by the development community but that simply does not exist yet. The second is an "access gap", which includes data that exist but that development actors and other stakeholders cannot access due to a lack of capacity, finances or agreements. The third is a "governance gap", or the absence of legal, ethical and regulatory frameworks that enable and regulate the use of new data sources for international development.

Finally, there is a "usability gap". Often, data are collected, made available and governed by clear rules but are not put to effective use because of a range of limitations affecting usability. For example, domain expertise to guide appropriate use of data and to enable users to locate data "needles" in data system "haystacks" is often missing. Additionally, the absence of communities of practice to serve as engines of learning, innovation and diffusion is a challenge. These gaps must be bridged to give decision-makers in every sector the tools to understand the impact of the policy and business decisions they make in a data-driven world.

Yet, despite these gaps, progress is being made. During the past 15 years (starting in 2000 with the UN's Millennium Development Goals) major strides have been taken towards the monitoring and evaluation of core development goals within programme silos. But despite these gains, huge gaps remain that limit the ability of governments, donors, civil society and individuals to act and communicate fully and honestly in an integrated fashion.[5] Addressing the data deficit can strengthen the dialogue between service providers and beneficiaries, taxpayers and governments, companies and employees and among the private sector, government and civil society regarding development issues.[6]

There are multiple dimensions to the data deficit challenge. They include resource constraints at the country level, misaligned incentives for accurate collection at the local level, competing agendas between outside donors and nation states, and the lack of incentives and agreements to encourage the sharing of private-sector data.[7] Each of these dimensions presents a separate set of challenges which collectively can seem overwhelming. Nonetheless, there is no alternative to addressing these issues. Without access to data, ethical, fair and accountable governance is undermined.

Fortunately, in the new networked data landscape, sets of shared needs are emerging across stakeholder groups in both the digital North and South that can provide a basis for helping set the data-driven development agenda. With that said, constant attention must also be paid to those needs that are under-represented and for individuals who are currently "digitally invisible" and unaware of the potential impact of Big Data and its uses.

The "open data" movement presents a strategy to help address the data deficit challenge. By treating data as a common resource that can be freely used, shared and accessed by anyone, participants of the open data movement help inform global development efforts, strengthen civil engagement and improve donor decisions and policy-making. Incorporating open data sources into policy monitoring and evaluation processes enables governments to complement the mission-focused curating of relevant data sets that are accessible, interoperable and compatible with agreed-upon indicators. At the same time, it helps avoid costly and time-consuming data collection activities.

Likewise, improving access to private-sector data can serve to address the data deficit by augmenting and enhancing the statistical capacities of governments and development organizations. Many private-sector data streams have been identified as being directly relevant to governments and policy-makers. Location data, geospatial images from satellites, financial transaction data, logistics, supply chain details and the internal process data from enterprises are routinely identified as being of potentially high value in data development contexts. To date, the structures for sharing such proprietary data have not been standardized, so private data have been available only on an ad hoc basis.

To achieve the full potential of data-driven development, it will be necessary for all stakeholders – government, the private sector, development organizations, and the public – to work in coordination to bridge these gaps. However, successful collaboration has been elusive in the past. Reports on a wide range of development issues have concluded that public-private partnerships (PPPs) are only part of the answer. To be successful, stakeholders must each independently conclude that the sum is greater than the parts and that they can only achieve shared goals through collaboration.

The pervasive culture of not sharing data retards development. The private sector's reluctance to share data is due largely to a utilitarian calculus of proprietary and competitive concerns that pervade market-based economies. They are not alone. Even within the UN system or among non-profits, a proprietary default position can make it difficult to get agencies to share programme data. Many concerns about sharing data are based on a lack of trust, a fear of incurring liabilities or a loss of institutional information control and arbitrage advantages (which create and maintain power differentials both within and between organizations).

Figure 1: Classes of Uses of Mobile Phone Data

Categories of Mobile Data	
Location and Movement Data	– User location and movement patterns during active as well as passive usage of mobile services – Degree of accuracy dependent on network and device generation (few yards to several miles)
Financial and Economic Data	– Information regarding spending and specific purchases (e.g., VAS such as types of ringstones, etc.) – Personal financial information such as transaction history patterns, prepaid voucher purchases, top-up amounts and periodicity, etc. help undertsand subscriber economic capacity and income levels
Identity and Demographic Data	– Information on user's name, age, gender, employment status, ethnicity, cohort, etc. – Used for advanced segmentation and product/ service design
Social/ Browser Data	– User's social patterns and social network analysis emerging from call records, call, metadata and browser/ app usage – Helps uncover subscribers with high relative influence
Usage	– Subscriber's usage of mobile services (voice, text, data, VAS), used by operators for developing tailored products and services
Sentiment and Trends	– Language and tone of anonymized SMS messages enables understanding of sentiment – Analysis of VAS purchasing patterns helps in the understanding of popular cultural trends
Diagnostic/ Ambient Conditions	– Studying signal interference and reflection patterns from water bodies or large buildings helps develop insights into civil construction and urbanization – Measurement of ambient conditions via sensors embedded into, or connected to smartphone

Source: Bill and Melinda Gates Foundation

Table A: Perceived Risks of Data Sharing and Implications for Development Community

Perceived Risks of Sharing Data	Mitigating Actions
Weak data security systems in the development community	Create secure and trusted third-party data labs where development-community experts can access data resources Foster support for affordable data security and vulnerability assessments for humanitarian NGOs and civil society organizations Invest in affordable data security technologies, capacities and policies
Uncertain legal and regulatory environment	Establish ongoing dialogue with industry, policy-makers and civil society on the impact of robust data sharing Support ongoing efforts to create, pilot and scale standard agreements for cross-sector data sharing Identify examples of scalable use of data for socio-economic impact Manage shared risks by transferring liabilities to third-party institutions
Damage to customer or shareholder relationships or public image	Establish external ethical review boards that oversee use of shared data resources Create clear legal agreements that stipulate risks and responsibilities among stakeholders Compose data queries so that the insights extracted respect individuals' privacy Promote model regulatory terms for adoption across jurisdictions to support data flows and enhance legal interoperability across borders
Loss of competitive advantage	Identify and align on clearly articulated use cases which demonstrate a compelling pro-social impact for pooling of shared data resources Operate with clear agreements in place which define intended uses Engage in research quantifying the "cost of not sharing" and the pro-competitive impact of co-management of data across market competitors
Operational barriers	Support private-sector "Big Data R&D labs" with personnel and other resources to make data more accessible for social use Share and leverage underlying technical infrastructure for reduced operational costs

Overcoming some of the issues that limit data sharing may require more direct forms of external compulsion in the form of improved oversight and enforcement, where such an authority is available (either by government regulation or enforceable stakeholder self-regulation). Table A (above) summarizes some of the private sector's perceived risks and the suggestions for how they might be collaboratively mitigated.

The challenge for governments and the development community has been, and will continue to be, how to incentivize and/or compel the private sector to share certain data while ensuring legitimate and appropriate use. The perceived risks associated with sharing proprietary data are quite strong, mostly because they are not easily measured. The generic concerns are that sharing data costs time, money and personnel resources. It can impact competitive relationships, and misuse of the data can cause immense damage to a company's reputation with a direct loss of earnings and market value.

Of all the various private actors holding unique and valuable data sets that could support sustainable development, the most widely recognized are mobile network operators. The precision, timeliness and depth of the behavioural insights that can be gleaned from these data streams make them extremely valuable and difficult to replicate by traditional data-collection procedures. While there are still many who are "digitally invisible", mobile phone use is common among those living in even extreme poverty in many developing economies.[8] Even without access to call content (which is typically prohibited), the granular data generated by the use of phones can provide richer information and understanding in the areas of health, finance, education and a variety of other needs.

Much has been written about the power of mobile data as a tool for gaining insights on individual behaviour.[9] From a data deficit perspective, the big question is: why isn't this data being used?

The reasons are a complex set of political, commercial, ethical, regulatory and reputational factors which reflect conflicts among multiple interests interacting in the ecosystem. These dimensions play out on multiple levels touching many varied stakeholders: business leaders, government policy-makers, donors, national security leaders, civil society, research organizations, shareholders and citizens. Each has their own incentives, priorities, real and perceived institutional constraints, and agendas. Few shared metrics and guidelines are available to enable "cross talk" among these stakeholders.

Some suggest that the hesitancy on the part of actors to remove barriers to data sharing is due to their discomfort with technology – older people have more limited understanding of advanced information and

communications technology (ICT) capabilities, for instance, and this discomfort is deeply embedded in institutions. Also, as organizational representatives, people tend to act in conformity with often anachronistic policy constraints.

Achieving a balanced ecosystem – which is not to dismiss the strong incentives for incumbents to maintain power differentials and imbalance – is essential to realizing the full benefits of the data revolution. In some instances, this will require political will and leadership. More substantially, it will require the establishment of structures that can normalize risk and create unique risk reduction and leverage benefits. Addressing the challenges of the data deficit will require innovation in hybrid technology and policy architectures by the development community, private enterprises, governments, legal experts, ethicists and citizens. Stakeholders must enter new forms of dialogue and coordination (both informal and formal and across multiple sectors), and create new policy frameworks, supplemental institutional structures (such as public-private partnerships) and incentive structures.

> **Right now we have a rare chance to take the dull world of data and make it headline news. As ONE's Ebola tracker shows, bad data is killing people in West Africa right now. Bad data is helping corrupt leaders stay in power. Lack of data is making it hard for the poorest to hold leaders accountable. Data delivered by technology can't alone guarantee transformation - but in the hands of actively engaged global citizens, a free media and responsible accountable governments, good data is the fuel of progress and a data revolution delivered in 2015 is the time to fill the tank .**
>
> Jamie Drummond, Co-Founder ONE

Strengthening the Ability to Listen: The Question Box Use Case

As the UN Data Revolution report highlights, "Too many items that need to be known remain unknown." One key concern is the growing number of "Data Invisibles" – individuals who are not counted within the formal or digital economy. These often include women, the elderly, children, migrants, indigenous populations and slum dwellers. Not being visible can mean that the ability to address domestic violence, the vulnerability of children, human trafficking and many other social concerns are absent or at best uninformed. During crisis events in particular, not listing individuals accurately in data records can have catastrophic consequences.

Question Box is an innovative ICT-for-development approach to addressing the data gap. Through the deployment of networks of free community call boxes in remote and rural areas, individuals can access information and services using their local language. Question Box networks are designed to attract specific subsets of a community through targeted service offerings. This approach promotes community-based reciprocity – information is exchanged in two directions – and local resilience, while producing unique data streams on "data invisible" populations.

Strengthening Data Governance

As noted in previous World Economic Forum reports, the global anxiety over the governance of data stems from the fact that everyone is somewhat in the dark. According to author David Brin: "We're in a fog of data ignorance." To unleash the potential of data for development, a number of governance issues must be addressed.

While the data economy of emerging economies is just beginning to form, it seems safe to say that it will most likely resemble the ongoing evolution seen in developed economies in terms of its structure. Global-scale standards for data technology infrastructures will force consideration of matching global-scale policy and economic data standards for at least the subset of shared data-use concerns. Moving fluidly between jurisdictions, organizations and functions, data flows will be constantly shifting, with firms entering and exiting, new analytical tools coming into use and value chains becoming increasingly complex. The growth of passive data collected from billions of sensors will add further scale and complicate efforts to manage, monitor and audit data flows.

Currently, a set of interrelated questions is evolving over the legal requirements, policies, ethics and norms that guide the use of data in both developed and developing economies. Many of the existing approaches that guide the creation, collection, storage and use of data were based upon decades-old policies of developed economies first established in the era of mainframe (un-networked) computing. While many of the underlying Fair Information Practice Principles that currently guide stakeholders in various jurisdictions and sectors are still relevant and important, some need to be updated and refreshed to address the new challenges of networked systems and also to suit the unique needs of emerging economies (such as engagement of the individual, use limitations and purpose specification).[10]

Additionally, new governance systems and institutions must earn and maintain the trust of individuals and organizations alike regarding their data use and handling practices. Power dynamics exacerbate lack of trust that results from a lack of transparency in current data approaches. A steady stream of media reports worldwide reminds consumers that data often flow in ways that can be intrusive on individual rights either because they are outside the traditional rule of law or are under the broad umbrella of state national security interests.

In every region of the world, strengthening meaningful transparency in the ways that data are collected, stored and used has been widely recognized as a shared global priority.[11a] It has also been noted that transparency is in many ways a paradox. Greater transparency, without controls and education, can overwhelm individuals with too much information.

Additionally, the construct of transparency is generally oriented towards strengthening externally-facing "front door" relationships with individuals. When it comes to the "back door" ways that data flow from these entities through their suppliers within the "data-industrial complex", the transparency and incentives for sharing are diminished or eliminated. Much greater visibility and auditability of both the public and private data supply chains are needed to avoid "transparency-washing."

It is also important to anticipate that the proportion of personal data that is either passively observed about individuals or computationally inferred about them is growing at an ever-increasing rate. By 2020, an estimated 50 billion devices will be wirelessly connected to the internet. Because of this global change, the guidelines and protection mechanisms for governing the use of high-frequency and high-resolution data in both the Global South and North need to adapt.

Legacy privacy guidelines and data protection mechanisms currently in effect were based on an earlier presumption that data are actively collected from the individual with some level of their direct awareness. As billions of sensors come online and passively collect data (without individuals' awareness), and as computer analytics generate and synthesize more "bits about bits" (or "meta-data"), understanding how data are generated and how engaged the individual is in their creation and collection will be essential to balance interests for effective data governance. As African Studies scholar Laura Mann notes: "In Kenya, for example, the government has awarded the telecommunications company Safaricom a lucrative security and anti-terrorism contract while Kenya Revenue Authority has begun to mine mobile transaction data to identify noncompliant taxpayers. While the sale, sharing, or indeed interception of digital data may improve states' developmental capacities and lead to more targeted social policy, it also raises important ethical implications about privacy and the political manipulation of data by powerful groups."

Balancing the trade-offs between the public good that can be achieved with data and the potential harm to individuals and communities is central for effective data governance.

Trust is an important variable in evaluating such trade-offs, such as in cases where the degree to which data have been anonymized before transfer is balanced with the trust placed in the recipient and their processes to avoid unauthorized access. When the issue of anonymity is discussed, it is generally rare to also hear the level of *trust in the recipient* referenced in discussion. The recognition of trusted third parties and systems to manage anonymized datasets, enable detailed audits and control the use of data could enable greater sharing of data among multiple parties while serving to manage and mitigate identified risks.[11b] While much more research is needed in computational privacy, the widespread adoption of existing techniques could enable this trend of sharing data in a privacy-conscious way.

Overcoming these challenges will require a comprehensive revision of policy frameworks that were based upon legacy information flows within hierarchical, industrialized institutions relying on centralized information distribution systems in which data and their applications were defined and limited. The internet and its global data flows are fully distributed, challenging traditional institutional and sovereign borders.

The pervasiveness of hierarchical institutions raises a question regarding the institutional appetite for a genuinely transformative "data revolution". Reliance by stakeholders on existing hierarchical institutions is understandable but it is not clear that the mere combination of existing public and private institutions (with their centralized power structures) will capture the benefits and have aligned incentive structures for change.

Current problems, like highly centralized institutions, are artefacts of current power structures. New levels of thinking about data governance will reveal new potential governance structures. The super-structure of shared information systems among institutions compels this analysis.

Another issue shaping the governance of data is the lack of a shared taxonomy of impacts (both benefits and harms). Shared taxonomies can drive meaningful near-term progress. As Linnet Taylor of the University of Amsterdam writes on the issue of data taxonomies: "A new taxonomy of data is badly needed. Industry, government and citizens are too frequently in disagreement as to what exactly constitutes personal data and what does not – and without an understanding of how data gets positioned in each category, or flows between them, it is impossible to have a discussion about how to govern and regulate those flows."[12] Many existing privacy regulatory frameworks do not take this into account. The effect is that they indiscriminately apply the same rules to different types of data, resulting in an inefficient and less-than-trustworthy ecosystem fraught with unintended consequences that undermine reliability and predictability.

Local context is another critical governance issue. Ulrich Mans at the Peace Informatics Lab of Leiden University (Campus The Hague) comments, "We need to create and make visible a growing number of data-driven initiatives across developing economies that have a clear benefit for those living in extreme poverty." To do this, taking account of the local context is key. Attitudes and tolerance for how data are used and what is legitimate, fair or ethical vary greatly among different geographic and social groups. While incorporating context-related nuances into regulation is difficult, it is clear that universal data use policies that treat all data equally will face significant challenges to remaining relevant in all contexts and over time.[13]

Emerging Principles of the Data Revolution

The UN Secretary General's Data Revolution Independent Experts' Advisory Group has advanced 10 principles. A preliminary digest is provided below.

Data quality and integrity
Poor quality data can mislead.

Data disaggregation
To the extent possible and with due safeguards for individual privacy and data quality, disaggregated data can provide a better comparative picture of what works and help inform and promote evidence-based policy-making.

Data timeliness
Data delayed is data denied. The data cycle must match the decision cycle.

Data transparency
Publicly-funded datasets, as well as data on public spending, should be available to other public ministries or the general public. Underlying data design and sampling, methods, tools and datasets should be explained and published alongside findings to enable greater scrutiny, understanding and independent analysis.

Data openness
Data should be made public in ways that encourage greater use and be complete, machine-readable, freely available for reuse without restrictions, and transparent about underlying assumptions.

Data usability and curation
Data architecture should place great emphasis on user-centered design and user-friendly interfaces. Communities should be fostered to develop new tools that can translate raw data into something meaningful to a broader constituency of non-technical potential users.

Data protection and privacy
Clear international norms and robust national policy and legal frameworks must be developed.

Data governance and independence
Data quality and NSOs should be protected and improved, to ensure they are functionally autonomous, and independent of political influence.

Data resources and capacity
National statistical systems should be established that are capable of producing high quality statistics in line with global standards and expectations.

Data rights
Rights include (but are not limited to) the right to be counted, the right to an identity, the right to privacy and shared control, the right to due process, the right to freedom of expression, the right to participation, the right to non-discrimination and equality, and the right to principles of consent.

The Role of Communities and the Individual: Digital Identity and Human-Centred Design

With a general understanding that addressing the "data deficit" will require new forms of public/private data sharing agreements and that the governance of these shared commonly-managed resources will require contextually-based policies with granular levels of control, the question becomes: where should these new technology-based tools be applied? What challenges should be prioritized and how should they be implemented?

If the data deficit is characterized as a public/private commercial innovation challenge and governance is an ethical and policy challenge, the issue of "how well do we understand the local context" can be seen as a design challenge. Establishing an ecosystem that is sustainable, balanced and principled will require approaches that account for the complex and dynamic relationships and movement of data and information among multiple entities (i.e. infrastructure and tool providers, producers, consumers, processors, curators, auditors, etc.).[14]

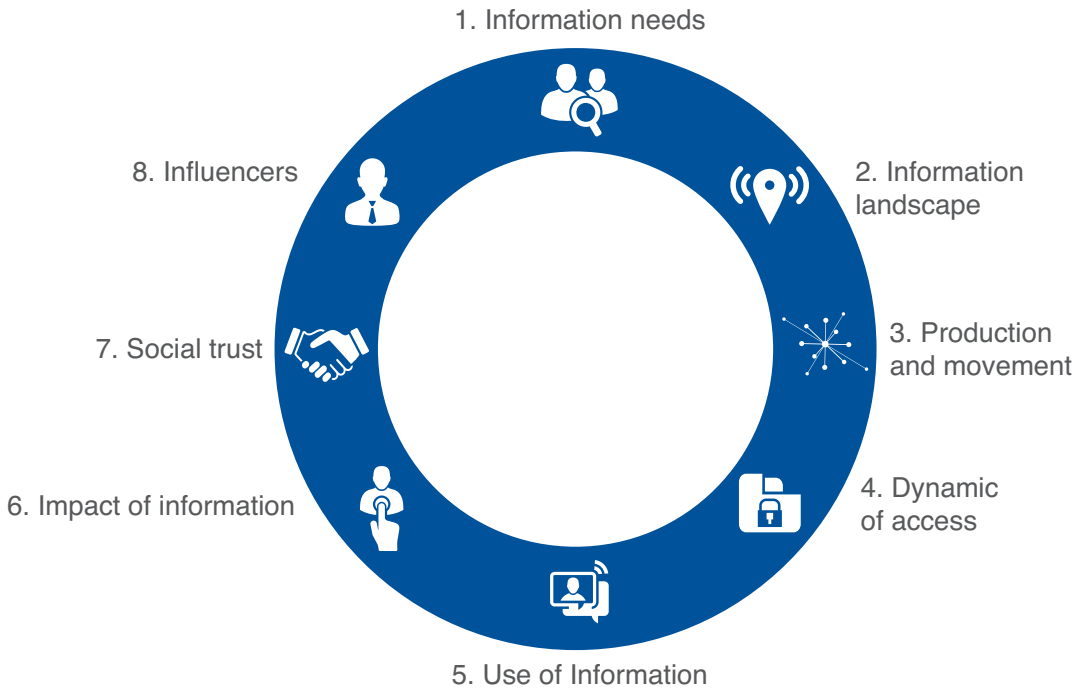
As a first step, many stakeholders must be educated. In particular, at the start the focus should be on understanding the local utility and economic value creation that can accrue to individuals through a more trusted, transparent and accountable data ecosystem. Additionally, awareness building within the technology sector would help recognize how little private actors know about the depth and complexity of delivering on the goals of sustainable development. There have been decades of failed ICT-for-development interventions led by technologists providing "solutions" to problems they did not fully understand.

Holistic approaches are needed, focused on in-depth learnings of the local needs, wants, aspirations and social context, as well as the economic and political environment of the needs. A deeper appreciation for "dark data", which is not yet known but is needed, is a requirement.

By addressing the issue of how *all* stakeholders can more effectively listen, learn and adapt as a design challenge, new ways of thinking, seeing and behaving can emerge to help address the significant power dynamics, velocity of change and trust. Incorporating an appreciation of social relationships, human context and dynamic networks of control and influence are critical to a richer understanding of the impact of data-driven technologies on communities, cities and larger socio-ecological systems.[15] Examining how data collection, storage and usage systems are actually being built in different parts of the world, how they are being contested and negotiated by different stakeholders and what impact these power struggles are having on the subsequent form of their governance are complex yet important questions to ask.[16]

> **Through the lens of human-centered design, information needs, creation and distribution systems can be seen as fluid systems that adapt and regenerate according to the obstacles, challenges and needs of a given situation and community. Combining macro-level analysis (i.e. media landscape, information infrastructures, and political/regulatory environments), granular observations (i.e. information availability, needs, and distribution), with human and social insights (i.e. identifying information disseminators and influencers) can be viewed as an important way for policymakers and practitioners to design the most appropriate and effective strategies for individuals, communities and societies.**
>
> Internews, "Why Information Matters: A Foundation for Resilience" (December 2014)

Figure 2: Elements of an Information Ecosystem



Source: Internews, "Why Information Matters: A Foundation for Resilience" (December 2014).

In that light, one of the central tensions where the dynamics of system design can be seen is in the "centralized versus distributed" debate. There are multiple and overlapping sets of conversations that pivot around the "centralized vs distributed" differences (internet governance, the impact of "blockchain" technologies on the financial services sector, open governance, etc.).

At its core, the internet is a highly distributed system that was designed to be highly resilient against attack and is therefore beyond the control of any single actor. As it matures, it is evolving in ways that are "spontaneous, autonomous, self-healing and wholly distributed."[17] A new ecology of data assets is emerging that provides the means for secure and trustworthy communications and for entirely new solution-sets related to the digital identity of people, devices and institutions. Blockchain technologies (e.g. those applied by crypto currencies such as Bitcoin), digital exchanges and "self-signing" ledgers and contracts are early examples of how the creation and exchange of value could reliably occur in a highly distributed manner.[18]

From an innovation perspective, these highly distributed technologies directly challenge conventional and centralized commercial, legal and governmental power structures. "What once required the authority of a central bank or a sovereign authority can now be achieved through open, distributed crypto-algorithms without regard to borders or human intervention," says John H. Clippinger, Executive Director of the technology non-profit IDcubed.org. "We are seeing a new kind of highly distributed, self-deploying, self-healing infrastructure that profoundly alters one of the most fundamental precepts of human social and economic organization – the issuance and management of identity, access rights and risk."

The power dynamics within the "centralized versus distributed" debate can be starkly reflected in the dialogue on identity systems for low- and middle-income economies. Few deny the critical importance of national identity and civil registration services as essential for delivering on the promise of sustainable development. Millions of individuals are denied, and excluded from, basic health, education, social protection and humanitarian response services because of insufficient national identification systems. These capabilities are also vital for policy planning, monitoring and evaluation at the national and local level. Digital identity services can help eliminate inequalities and establish new efficiencies, foster innovation and extend the reach of service delivery.

Yet the question of how these systems are implemented is an area of growing debate. From the perspective of centralized sovereign states, national identity systems provide a means of delivering services to citizens and protecting their interests. From the perspective of those advancing more fluid and distributed systems, the importance of sovereign individuals (and/or groups) is emphasized, with a focus on their capacities to self-organize and deliver collective action in transnational ways beyond the traditional governance institutions such as the nation state and standard democratic institutions. A more structured debate on the comparative risks and benefits of centralized versus distributed identity systems is critically important to genuinely advance transformative social change.

Call to Action

To strengthen the data revolution, governments, the private sector and the development community must take action in three priority areas:

- Addressing the data deficit
- Improving governance systems to ensure proper and ethical use of data
- Institutionalizing a bottom-up, human-centred approach to understanding local information ecosystem dynamics

To support these goals, strengthening capacities in technology innovation for data sharing, global data literacy, and infrastructure building and sharing are key priorities. Additionally, establishing new funding streams for indicator monitoring systems and best-practice sharing on governance and publicly-available data analytics tools are top concerns.

In particular, the following specific areas of action are noted.[20]

- Establish new funding streams and innovative financing mechanisms to address the data deficit by supporting governments, businesses and other national and sub-national actors to implement country-level plans which focus on collecting critical information – in a trustworthy and principled manner – on the hardest to reach and most vulnerable people. Enhancing global data literacy is of special importance in this process.

- Establish a global initiative on trust and transparency to strengthen governance, ethics, norms and values on the trusted flow and balanced use of data to accelerate sustainable development while minimizing risks to vulnerable individuals and communities. This will require bringing together the public and private sectors as well as civil society in order to promote the adoption of specific principles and to accelerate the development of interoperable data standards.[21]

- Recruit and train an “Army of Factivists” a group of individuals across the developing world who would be paid to collect and interpret data on an ongoing basis. One of their initial tasks would be to undertake a baseline study for the Sustainable Development Goals.

Advancing this agenda requires a technologically savvy, multi-sector coalition to overcome systemic barriers to an inclusive, sustainable digital economy. Such a coalition must collaboratively develop digital data access platforms and a policy framework for sharing data between data-for-development deployments. Regulatory frameworks, commercial best practices, and new forms of decentralized, autonomous organizations must also be designed, tested and adopted to enable the exchange of digital assets and personal data.

Acknowledgements

The World Economic Forum would like to acknowledge the support of all those who contributed to this initiative in 2013 and 2014. The global dialogue series included sessions in New York, USA; Dublin, Ireland; London, UK; Dalian, People's Republic of China; Brussels, Belgium; Abu Dhabi, United Arab Emirates; and Davos, Switzerland.

The project engaged a multistakeholder community across government, the private sector, civil society, academia and others throughout the dialogue series and through its Working Group.

Special thanks are extended to all those who joined regular calls and supported these events with their insights, including: A.T. Kearney, American Red Cross, Annenberg Center on Communication Leadership, BBVA, Brielle Marketing, Data-Pop Alliance, the Engine Room, Gates Foundation, GSMA, Impact Institute, Internews, Leiden University (Campus The Hague) Peace Informatics Lab, MasterCard, Microsoft, NYU GovLab, ONE, Orange, Overseas Development Institute, Palantir, PopTech, Rockefeller Foundation, SAS, Stanford University, World Bank, UN Global Pulse, University of Washington, Ushahidi, Vital Wave and WITNESS.

Special thanks for the editorial input to this report from Vital Wave Consulting and the Global Agenda Council on Data-Driven Development.

Members of the Global Agenda Council on Data-Driven Development are:

Sandy Pentland, MIT (Council Chair)

Mikael Hagstrom, SAS (Council Vice-Chair)

Anastasia Ailamaki, Ecole Polytechnique Fédérale de Lausanne

Seth Ayers, World Bank Institute

Mitchell Baker, Mozilla Foundation

Adam Bly, Seed

John H. Clippinger, IDcubed.org

Nicolas de Cordes, Orange

Kate Crawford, Microsoft

Scott David, University of Washington

Jamie Drummond, ONE

Nathan Eagle, Jana Mobile

Jeremy Howard, Strategic Data Science

Jake Kendall, Gates Foundation

Robert Kirkpatrick, UN Global Pulse

Ali Al Khouri, Emirates Identity Authority

Marc A. Levy, Columbia University

Guo Ping, Huawei Technologies

Juliana Rotich, Ushahidi

Linnet Taylor, University of Amsterdam

Mihaela Ulieru, Impact Institute

At the World Economic Forum, William Hoffman leads the Data-Driven Development initiative. The Peace Informatics Lab at Leiden University (Campus The Hague) under the co-leadership of Ulrich Mans and Gideon Shimshon served as content adviser. From the Peace Informatics Lab, Jos Berens liaised with the World Economic Forum as research analyst.

Contact:
William Hoffman, Head, Data-Driven Development
World Economic Forum
william.hoffman@weforum.org

Endnotes

1. “A World That Counts”, UN Data Revolution Independent Experts Advisory Group, November 2014.

2. “Using Mobile Data for Development”, Bill and Melinda Gates Foundation, 2013 (http://www.cartesian.com/wp_content/upload/Using-Mobile-Data-for-Development.pdf).

3. “Waiting on Hold”, The Economist, 25 October 2014 (http://www.economist.com/news/science-and-technology/21627557-mobile-phone-records-would-help-combat-ebola-epidemic-getting-look).

4. “Delivering on the Data Revolution in Sub-Saharan Africa – Final Report of the Data for African Development Working Group”, Center for Global Development & The African Population and Health Research Center, 2014.

5. “A World that Counts”, UN Data Revolution Independent Experts Advisory Group, 2014.

6. Ibid.

7. “Delivering on the Data Revolution for Sub-Saharan Africa”, Center for Global Development, 2014.

8. “Using Mobile Data for Development”, Bill and Melinda Gates Foundation, 2013. Approximately 89 active subscriptions per 100 people are cited for developing economies. Given that people can own and share multiple subscriptions, these numbers are only roughly indicative of the overall rate of mobile ownership.

9. “Cell-Phone Data might help predict Ebola spread”, David Talbot, MIT Technology Review, 22 August 2014, and “Mobile Phone Network Data for Development”, United Nations Global Pulse, October 2013.

10. OECD Privacy Principle Update, 2013.

11a. Rethinking Personal Data: A New Lens for Strengthening Trust, World Economic Forum, 2014

11b. de Montjoye, Kendall and Kerry, Enabling Humanitarian Use of Mobile Phone Data, Brookings Institute, 2014

12. Rethinking Personal Data: A New Lens for Strengthening Trust, World Economic Forum, 2014

13. Rethinking Personal Data: A New Lens for Strengthening Trust, World Economic Forum, 2014

14. Internews, “Why Information Matters: A Foundation for Resilience” (forthcoming).

15. Ibid.

16. Laura Mann, African Studies Centre, Leiden University, 2014.

17. John H. Clippinger, IDcubed.org, “The Ecology of Digital Assets: Identity, Trust and Data”, 2014.

18. Ibid.

19. Ibid.

20. These agenda items are directly supported by the ONE Campaign as well as multiple institutions within the World Economic Forum’s Global Agenda Council initiative.

21. See also: “A World That Counts”, UN Data Revolution Independent Experts Advisory Group, November 2014.

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum is an international institution committed to improving the state of the world through public-private cooperation in the spirit of global citizenship. It engages with business, political, academic and other leaders of society to shape global, regional and industry agendas.

Incorporated as a not-for-profit foundation in 1971 and headquartered in Geneva, Switzerland, the Forum is independent, impartial and not tied to any interests. It cooperates closely with all leading international organizations.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744

contact@weforum.org
www.weforum.org