



Data for development: What's next?

Concepts, trends and recommendations for German development cooperation



Published by:

As a federally owned enterprise, GIZ supports the German Government in achieving its objectives in the field of international cooperation for sustainable development.

Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices:
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Germany
T +49 (0) 228 4460 - 0
F +49 (0) 228 4460 - 17 66

Dag-Hammarskjöld-Weg 1-5
65760 Eschborn, Germany
T +49 (0) 61 96 79 - 0
F +49 (0) 61 96 79 - 11 15

E info@giz.de
I www.giz.de

On behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ)
Division 303, Education and the digital world
Dr Katrin Bornemann
Dahlmannstraße 4, 53113 Bonn, Germany
T +49 (0) 228 995 35 - 0
F +49 (0) 228 995 35 - 35 00

Programme/project description:
Sector Project Digital Development
Strategic Project Digital Change

Authors:
Andreas Pawelke, Michael Cañares, World Wide Web Foundation
Kevin Hernandez, Pedro Prieto Martin, Institute of Development Studies

Supported by Ben Ramalingam, Institute of Development Studies

Responsible:
Björn Richter, Sector Project Digital Development
Lea Gimpel, Strategic Project Digital Change

Editor:
Balthas Seibold, Elisabeth Hobl, Sector Project Digital Development

Design/layout, etc.:
Crolla Lewis, Aachen

GIZ is responsible for the content of this publication.
The views expressed in this paper are those of the authors and do not necessarily represent those of the German Federal Ministry for Economic Cooperation and Development, the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, the Institute of Development Studies, or the World Wide Web Foundation.

URL links:
This publication contains links to external websites. Responsibility for the content of the listed external sites always lies with their respective publishers.
When the links to these sites were first posted, GIZ checked the third-party content to establish whether it could give rise to civil or criminal liability.
However, the constant review of the links to external sites cannot reasonably be expected without concrete indication of a violation of rights. If GIZ itself becomes aware or is notified by a third party that an external site it has provided a link to gives rise to civil or criminal liability, it will remove the link to this site immediately. GIZ expressly dissociates itself from such content.

Maps:
The maps printed here are intended only for information purposes and in no way constitute recognition under international law of boundaries and territories. GIZ accepts no responsibility for these maps being entirely up to date, correct or complete. All liability for any damage, direct or indirect, resulting from their use is excluded.

Printing and distribution:
bonnprint.com GmbH, Siegburg

Printed on 100% recycled paper, certified to FSC standards.

November 2017

This work is licensed under a Creative Commons Attribution ShareAlike 4.0 International License.

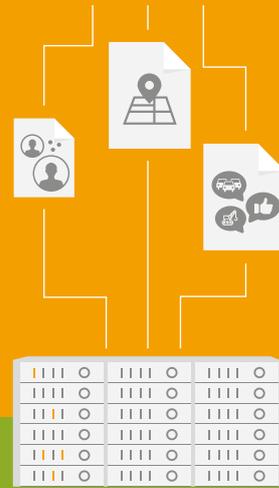
Data for development: What's next?

Concepts, trends and recommendations for German
development cooperation

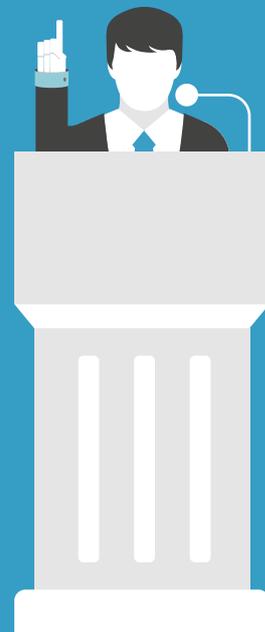
	Executive summary	6
1	Introduction	16
2	The state of affairs	18
	2.1. Big data	20
	2.2. Open data	26
	2.3. Citizen-generated data	32
	2.4. Real-time data	36
	2.5. Summary	41
3	Trends	42
	3.1. More data and new data sources create opportunities	44
	3.2. New technologies improve data analysis	47
	3.3. New types of actors, partnerships and ways of working emerge	50
	3.4. Balancing access to personal data and privacy	53
	3.5. The value of contextual, granular information	56
	3.6. Information inequality is likely to persist	58
	3.7. Summary	61
4	Recommendations	62
	4.1. Maximise the potential of data, but do not treat it as a 'silver bullet'	66
	4.2. Build internal data capacity	67
	4.3. Leverage partnerships for a strategic advantage	68
	4.4. Support strong technological and legal data privacy frameworks	71
	4.5. Be experimental and focus on a few sectors and geographies	73
	4.6. Address data inequalities	74
	4.7. Summary	76
5	The way forward	78
6	Annex	80
	About this report	82

Executive summary

The state of affairs



Trends



Recommendations

‘Inequality and poverty are about more than income – they are also about information.’

Web inventor Sir Tim Berners-Lee

Ninety per cent of all data ever created has been produced in the past two years. This includes data generated by people as well as the billions of sensors all over the world that are creating data every second and communicating with servers over the internet, creating what is called the Internet of Things.

Enabled by the rapid spread of technology and a more affordable access to the internet and mobile networks, the availability of digital data, in new and different forms, has grown at massive scale. Several initiatives have been launched in recent years that have explored diverse ways to leverage these new types of digital data for more targeted, effective and efficient development interventions.

In this report, we look at the state of digital data for development and emerging trends. We aim to support German development cooperation in integrating and prioritising data approaches and investments in their work. In this study we focus on four data categories: big data, open data, citizen-generated data and real-time data. The selection of these categories considered two key dimensions: (1) the growing use in development-related policy discussions, and (2) the ability to capture key characteristics of interest, including size, access, source, and timeliness of data. We believe these categories provide a good starting point to explore how digital data production and use might lead to better development outcomes.

	Basic definition	Value	Example
	<p>Big data</p> <p>High-volume, high-variety and high-velocity data that can contain both structured datasets and flows of unstructured data, often created and collected by the private sector.</p>	<p>Quickly analyse situations, diagnose and understand underlying problems, predict likely scenarios and even prescribe the most beneficial potential actions.</p>	<p>Tackling the spread of malaria using mobile phone data.</p> <p>(see Box 2.a. on p. 23)</p>
	<p>Open data</p> <p>Data that can be freely used, reused, and redistributed by anyone.</p>	<p>Strengthen (government) transparency, accountability and responsiveness, and spur social and business innovation.</p>	<p>Identifying lower-priced generic medicines through open data in South Africa.</p> <p>(see Box 2.e. on p. 31)</p>
	<p>Citizen-generated data</p> <p>Data actively produced by citizens with a specific purpose in mind.</p>	<p>Understand on-the-ground conditions, document and map incidents and sentiment, especially when data from other sources is unavailable, inaccessible or of poor quality.</p>	<p>Using drones to collect data for land rights advocacy in Indonesia.</p> <p>(see Box 2.f. on p. 35)</p>
	<p>Real-time data</p> <p>Data that is gathered and analysed quickly enough to make timely decisions.</p>	<p>Quickly make data-based decisions, for example in health or humanitarian crises.</p>	<p>Using real-time data on birth registration in Tanzania to improve policy-making.</p> <p>(see Box 2.h. on p. 40)</p>

What is the current state of data for development?

Big data

Generally, big data increases an organisation's capacity to describe and visualise situations, diagnose and understand underlying problems, predict likely scenarios and even prescribe beneficial potential actions. In the data for development space, actors like UN Global Pulse and the Data-Pop Alliance seek to promote a people-centred approach to big data, undertaking research and projects that use big data to solve persistent development problems, such as combating malaria.

To date, big data for development projects have mainly been pilots and proof of concepts showing the potential of using big data to tackle development problems. While these pilots have helped garner interest and hype around big data, so far there is no clear indication that they have led to data-based decision-making. However, there is reason to remain hopeful as these pilots enable the international development sector to start integrating long-term big data applications into development programming.

Open data

The global open data movement began by advocating for openly available government data. For open data to make a real impact, however, there are at least three key building blocks: the publication of open data by governments, the conversion of data to actionable information by intermediaries, and the use of this information by government officials, citizens, and others.

Case studies of open data initiatives in developing countries revealed at least three emerging impacts: increased government transparency and accountability; innovation and economic development; and greater inclusion and citizen empowerment. These impacts, however, are dependent on a number of factors, the availability of open data being just one. It has become clear that open data impacts depend on partnerships and collaboration, the quality of public open data infrastructure, clear open data policies, and government responsiveness.

Citizen-generated data

The idea behind citizen-generated data is for citizens to be not only passive consumers of information, but also active data producers, users and intermediaries. There is evidence of the value of citizen-generated data for improved delivery of development projects. Crowdmapping initiatives have contributed to a better understanding of the nature and scale of social issues such as corruption, harassment and voter turnout. Granular and real-time data provided by citizens have enabled governments to take more evidence-based decisions, for example in regard to urban infrastructure. Moreover, research has documented the impact of citizen-generated data projects on government policy and practice in areas including measuring education outcomes and improving community water supply. However, citizen-generated data initiatives often face significant barriers to scaling, including the requirement for a robust data collection infrastructure and ensuring that data production is sustainable and citizens are committed to contributing data.

Real-time data

Digital real-time data requires employing digital technology to enable and accelerate collection, sharing, management, analysis, and reporting of data to enable timely decision-making. The expected impacts of using real-time data in development programmes are related to increasing their capacity to respond to changes in their operating contexts and to learn from constant evaluations of the effectiveness of their actions. Real-time data has the potential to give development actors the means to uncover anomalies, respond to issues as they arise, improve internal coordination, optimise resource allocation, react to citizen feedback, and anticipate trends and future events.

There is still little evidence that real-time data leads to adaptive decision-making, as there are many barriers to ensuring that data is used by decision-makers. Some of the hindrances include competing data sources, distrust in data quality, lack of awareness of available data, data not adequately turned into information, lack of visualisation, and information that is not tailored to the end user's needs.

What emerging trends do we see?

Our research has uncovered six main trends impacting the digital data for development space that could shape or magnify the role and impact of data for development.

1 More data and new data sources

The quantity of data available is increasing at an exponential rate, a trend likely to continue in years to come. By 2021, 8.3 billion mobile phones are expected to be in use with over 50 per cent being smartphones.

The Internet of Things might increase from 8.4 billion objects in 2017 to 20 billion in 2020. All these devices will generate an ever growing flow of data. If adopted effectively by the development sector, new data sources and technologies could help overcome data gaps and allow development practitioners to gain a better understanding of contexts before, during and after interventions. However, digital data over-represents those already connected, and more data does not automatically translate to use. The challenge is to ensure inclusiveness, especially in triggering data demand and use.

2 New powerful technology to analyse data

An increase in data availability and the speed at which it is collected, processed and visualised with analytical tools is providing us with new ways to make sense of development problems and opportunities. Machine learning algorithms can identify patterns in observed data, build explanatory models and make predictions quicker and with more accuracy than humans do. These developments enable practitioners to better understand problems through data and have encouraged some agencies to introduce agile and adaptive programming. However, algorithmic transparency remains a challenge.

3 **New types of actors and partnerships**

The complexity of challenges in development requires not only new data sources and tools, but also new actors and partnerships. In data for development, at least two kinds of partnerships have emerged: data philanthropy and data collaboratives. Data philanthropy encompasses corporations that share their data for the public good, while data collaboratives share data assets and combine their expertise and/or tools to solve specific public problems. Non-traditional actors have also emerged, such as data and innovation labs, and data analytics companies, which increasingly play a role in the development sector.

4 **Balancing access to data and privacy**

Privacy protection is likely to receive growing attention in the coming years. Key players in the data for development domain will realise that protecting individual privacy while using personal data to tackle development challenges is virtually impossible. Meanwhile, the protection of personal privacy may be used as an excuse to withhold public sector data that could be made open for citizens to advocate for better public services, hold governments accountable and tackle corruption in the public sector. The hope is that an inclusive multi-stakeholder approach will emerge to identify workable solutions to privacy challenges.

5 **The importance of contextual, granular information**

The growing recognition of the importance of contextual information and of the need to integrate local knowledge into data for development initiatives has multiple implications. Future data initiatives are likely to move from current technology-driven top-down approaches towards becoming more bottom-up, problem-driven interventions. These initiatives will increasingly be ‘human-centric’ and aim for long-term empowerment of local communities, governments and civil society organisations, in ways that enable them to cooperate in using data to improve their situations.

6 **Information inequality is likely to persist**

The current system of data production and ownership rests primarily in the hands of private sector players, and to a lesser extent governments. Existing information inequalities are likely to be exacerbated in the coming years as some actors are in a better position than others to harness the positive developments arising from data. Unless there is a drive to diminish this data inequality, those with the capacity to handle large volumes of data will be in an advantageous position, while others will lose out.

What role could German development cooperation play?

There are negative and positive potential scenarios for these six trends as they unfold within the next five years and beyond. More sources of data can make the development sector more agile, but may open opportunities for surveillance and threats to individual privacy. Artificial intelligence offers novel ways of tackling development problems, but in some cases, algorithms have proven to be biased, opaque and out of reach of scrutiny. While new partnerships emerge between development and private sector actors to facilitate collaborative data-driven approaches, these partnerships could create silos, excluding other actors. Finally, while opening and sharing data between partners can serve the public good, this can also threaten citizen privacy due to the technical challenges of fully anonymising data.

These opportunities and challenges are not set in stone. The actions taken by those working in the sector will shape whether we amplify the positive trends, and mitigate the negative. The following **recommendations** on how to harness data – big, open, citizen-generated, or real-time – can contribute to global development goals and affirm the role of German development cooperation in international development.

1 **Maximise the potential of data, but do not treat it as a ‘silver bullet’**

German development actors should embrace the potential of digital data to drive development, while being conscious that data is only a means to an end, and acknowledging that the political economy of decision-making and the demand for data are as important as the supply of data. When establishing data for development initiatives, they should ensure that an enabling environment to incorporate data into decision-making already exists, or that one will be developed alongside the initiative.

2 **Build internal data capacity**

Organisations will need to invest in internal capacity building to develop the data literacy needed to leverage data-driven approaches. They should promote the emergence of multi-disciplinary teams to work in these initiatives. An assessment of current capacities vis-à-vis the role German development organisations would like to play is necessary to determine how the demand for domain experts, social scientists, data scientists and technologists can be met.

3 **Leverage partnerships for a strategic advantage**

There is a need to consider actions to reinforce Germany’s commitment to data in achieving the Sustainable Development Goals (SDGs), such as participating in the Global Partnership for Sustainable Development Data and endorsing the International Open Data Charter. German development actors should review their engagement strategy with other sectoral partners, especially the ‘development mutants’ – those emerging actors that are revolutionising the way development is done.

4 **Support strong legal and technological data privacy frameworks**

We suggest German development organisations support partner governments in strengthening their data privacy frameworks, work with business and civic organisations to better understand the risks of data collection and use, and champion responsible data approaches in global initiatives. Through existing networks, these organisations could help mobilise voices from the global South to contribute to global debates on data privacy issues, conduct targeted capacity building for actors in partner countries to mainstream responsible and secure data practices, and help shape the debate on the role of artificial intelligence and algorithms in data analysis.

5 **Be experimental and focus on a few sectors and geographies**

German development actors should be taking an experimental, iterative and flexible approach to data for development, targeted at specific problems, sectors and geographies instead of cross-cutting data initiatives. They could combine their strong sectoral expertise with the opportunities that new types of digital data provide, and help nurture data ecosystems in sectors like procurement, extractives and health.

6 **Tackle data inequalities**

German development organisations should help reduce data and information inequalities through targeted research and empowerment of local stakeholders given their strong track record of empowerment through capacity building and the financing of social and economic infrastructure. With a reputation of being able to work across government, civil society and the private sector, and with expertise in a range of thematic areas such as food security, financial inclusion and disaster preparedness, they are well positioned to occupy the niche of human-centred, privacy-aware data skills development. This also applies to the financing of the necessary investments into software, hardware and up-to-date telecommunication infrastructure required for wide-spread data access and use. Public institutions and citizens, including marginalised people and those living in remote areas in developing countries, need to have access to technology and the internet to avail of the benefits derived from using data for improved decision-making, strengthened transparency and accountability, and better public service delivery.

Digital data offers manifold opportunities to development by enabling development actors to strengthen decision-making processes, improve service delivery, elicit meaningful citizen participation, and increase responsiveness in humanitarian services, among others. At the same time, it generates new forms of exclusion, new methods of surveillance and threatens individual privacy.

Promising that digital data will solve development challenges is an overstatement. Data is just one of the ingredients that can help solve a given problem. There are people, partners, communities, gatekeepers and leaders that make up the whole enabling mechanism for digital data to work best in development.

This is why the recommendations identified in this paper emphasise the role of stakeholders along the development value chain – building individual and organisational capacities, protecting personal privacy, building collaborations across sectors, and empowering the powerless. This is in recognition that data is useless in development without people, that technology is useless without civic engagement, and that efforts to engage technology with no regard to communities and their culture will gain no ground.

Examples of data usage in international development cooperation



Nigeria: The German Federal Ministry for Economic Cooperation and Development supported the development of the open source platform 'Surveillance, Outbreak Response Management and Analysis System' that uses real-time data. The platform serves as a disease outbreak early warning and management system and was used during the West African Ebola outbreak in 2014/2015.



Tanzania: UNICEF and the Tanzanian government use real-time data for birth registration in Tanzania to improve policy-making and planning.



Indonesia: Incomplete data about land type, usage and ownership has led to frequent conflicts between businesses and local communities over territorial rights in Indonesia. Now citizen-generated data via drones is used to create cartographic material that can be used to solve disputes over land use.



Kenya: A Kenyan initiative is using citizen-generated data via the platform 'Una Hakika?' – Swahili for 'Are you sure?'. People can report potentially harmful rumours, which are then quickly validated to mitigate the escalation of ethnic conflicts based on misinformation.



Global: 'FinTechs' and 'InsureTechs' tap into data to develop financial products and innovative ways of making financial services available. Big data can help with removing traditional bottlenecks for financial inclusion by taking advantage of new technology, analytical tools and data access.

Global: Actors like the Humanitarian OpenStreetMap Team and Missing Maps work across the globe to enable data collection and provide detailed and accurate maps for humanitarian response and economic development.

1

Introduction

Data has been used extensively in international development for many years. It has helped development organisations define the most appropriate and relevant intervention, assess progress of projects over time and evaluate the effectiveness of the global development agenda, among others.

Enabled by the rapid spread of technology and a more affordable access to the internet and mobile networks, the availability of data, in new and different forms, has grown at massive scale. A number of initiatives have been launched in recent years that explore diverse ways to leverage these new types of digital data coming from cell phones, social media or sensors for more targeted, effective and efficient development interventions.

In this report, we look at the current state of data for development. We review the state of affairs and explore emerging trends to indicate potential areas of interest that various actors involved in German development cooperation may want to explore in making choices on priorities, approaches and investments in their future development agenda. By presenting key concepts, actors and achievements and by identifying key trends that we believe will shape and influence the work of development organisations in the next five years, we hope to provide a foundation for decision-makers of German development actors to help them explore in more depth the various opportunities and risks.

The report is structured as follows. **The ‘state of affairs’ chapter explores four digital data for development concepts**, namely big data, open data, citizen-generated data and real-time data. In this report, we focus on these relatively new and emerging data for development concepts¹ rather than more traditional development data that includes household surveys and censuses. For

¹ Unless otherwise noted, in this report the term ‘data’ refers to digitally generated or digitally processed data.

each of the concepts, we discuss key actors and initiatives as well as progress made in interventions and programmes over recent years.

The ‘trends’ chapter presents six trends that we believe will shape the data for development space over the next five years. Through these trends – identified through literature review, interviews with experts, and workshop discussions with staff from German development organisations – we seek to capture not only technical advancements, but also social, political and legal aspects that will impact future data for development work.

The ‘recommendations’ chapter seeks to respond to the identified trends. While we have made an effort to tailor the recommendations to the German government’s specific development cooperation agenda, they should be understood as suggestions for a general direction of travel rather than as immediate, concrete and actionable. This is largely due to the variety of actors involved in German development cooperation, which include funders, implementing organisations working in financial and technical cooperation, media development and faith-based institutions as well as political foundations. Spelling out customised, context-specific advice would require in-depth organisational assessments, review of priorities and analyses of project portfolios, all of which was beyond the scope of this research. Nonetheless, we are confident that the exploration of the trends and the presentation of key recommendations, albeit broad, will help German development actors identify a niche in the data for development field.

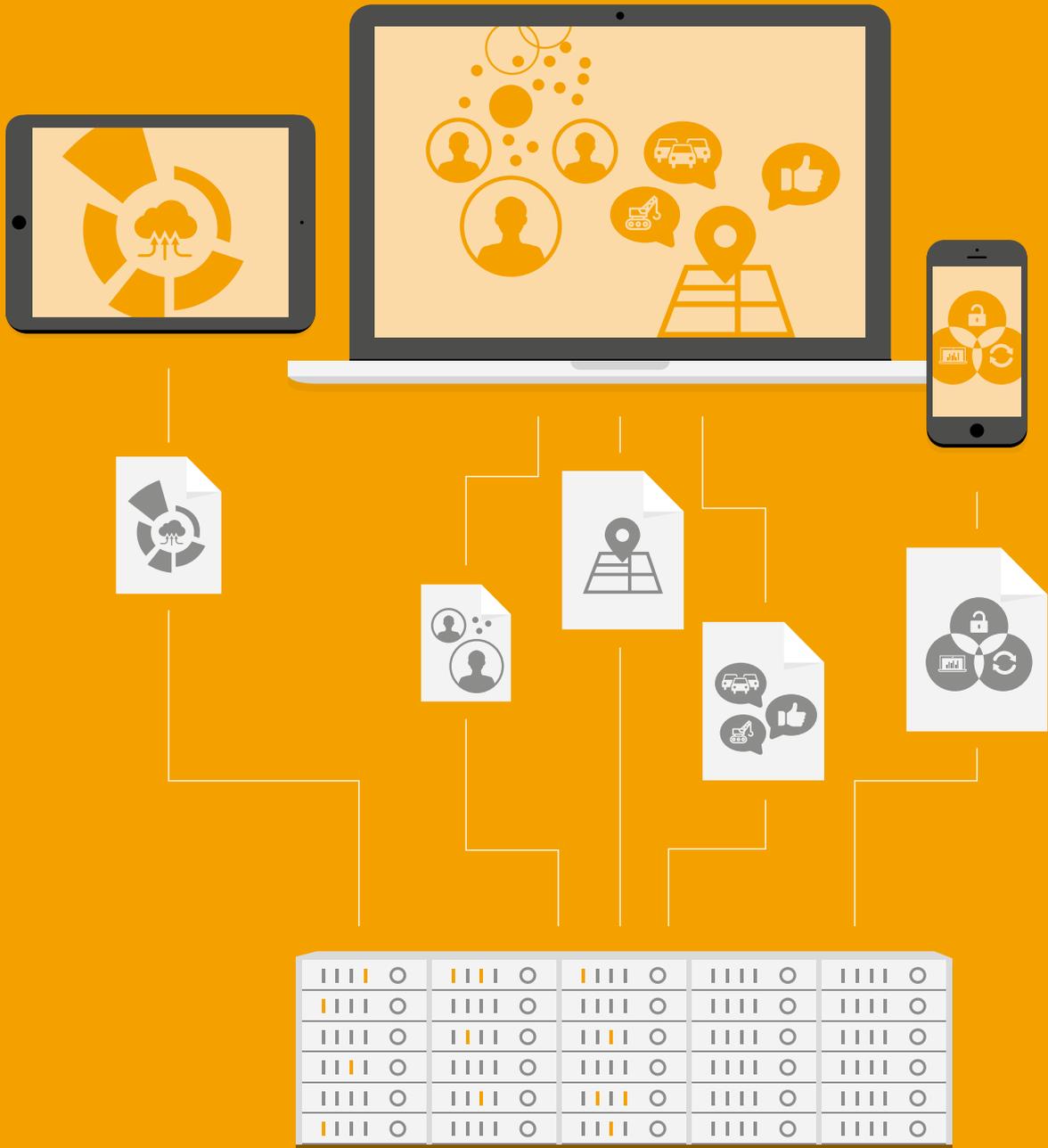
2

The state of affairs

Over the last decade or so, several data concepts, ways of categorising data, and patterns of data collection and use have emerged. While for a long time data collection and use debates have focused on quantitative versus qualitative data or paper-based versus electronic, new forms of comparisons have dominated the data discourse recently. This is also thanks to new types of digital data finding their way into the development sector: closed versus open, big versus small, thick versus thin, real-time versus right-time, machine-readable versus human-readable, downstream versus upstream, and so on.

While these categories have enriched and shaped our understanding and appreciation of data, there are also significant overlaps between them. For example, fiscal budget data is quantitative, non-personal and public, and can be closed or open depending on publication processes, while citizen-generated data can also be shared as open data, and real-time data can be threshed out from big data sources. The boundaries among these categories are wide and selecting which of these categorisations to study is a daunting task.

In this study, we focus on four data categories: **big data, open data, citizen-generated data and real-time data**. The selection of these categories, while conditioned by the practicalities of the research, considered at least two dimensions: (1) a growing evidence of their use in development-related discourses, and (2) the ability to capture key characteristics of interest, namely size, access, source, and timeliness of data. We believe these categories are good starting points to explore the relationship of digital data production and use for better development outcomes.



2.1. Big data

Big data could be described as data that is too big to fit into a standard computer drive and too complex to be analysed using Excel sheets.

Informally, big data could be described as data that is too big to fit into a standard computer drive and too complex to be analysed using Excel sheets. In fact, big data includes both structured datasets and flows of different unstructured data, like emails, social media, photos, videos and data streams produced by sensors and devices. Data from these sources is often collected, analysed, and turned into actionable information in real time, to provide enhanced awareness, understanding and forecasting capabilities.² The insights provided by big data have enabled transformations in many domains like health care, transportation and financial services.

The most recognised definition is the one coined in 2012 by the Gartner corporation that defines big data as ‘high-volume, high-velocity, and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making, and process automation’.³ Big data thus requires not only access to large data assets, but also the competence and infrastructure to process them in a timely manner, and the capacity to realise the valuable insights extracted from it.⁴ The three V’s that characterise the data assets are:



Volume: the sheer volume of data available nowadays is enormous and growing. 90 per cent of all data ever created has been produced in the past two years.⁵ This includes data generated by people as well as the billions of sensors all over the world that are creating data every second and communicating with servers over the internet, creating what is called the Internet of Things.



Velocity: the speed at which the data is created, stored, processed, analysed, visualised and acted upon has increased to up to real time. Big data usually involves collating data generated at various speeds and moments, and accommodating bursts of activity.



Variety: there are a growing variety of data assets. These include structured and unstructured data from databases, devices and sensors, logs, social media, websites and posts, images, email communications, and audio and video streams (like radio and television), among others.

Beyond the technical aspects of big data

From a development perspective, however, it is important to look not only at the technical aspects of big data, but also consider its positive and negative potential impacts. Big data has thus been described as the ‘growing ability to generate, manage, analyse and synthesise data to create or destroy different forms of value in the economic, environmental, rights and human development domains’.⁶

2 Coppola, A.; Calvo-Gonzalez, O.; Sabet, E.; Arjomand, N.; Siegel, R.; Freeman, C. and Massarrat, N. (2014) Big Data in Action for Development, Washington DC: World Bank, understandrisk.org/wp-content/uploads/Big-Data-for-Development-Report_final-version.pdf

3 research.gartner.com/definition-what-is-big-data

4 Ali, A.; Dadir, J.; Rasool, R. ur; Sathiaseelan, A.; Zwitter, A. and Crowcroft, J. (2016) Big Data for Development: Applications and Techniques, Big Data Analytics 1: 2

5 IBM (2017) 10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations, IBM, ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

6 Spratt, S. and Baker, J. (2015) Big Data and International Development: Impacts, Scenarios and Policy Options, IDS Evidence Report 163, ids.ac.uk/publication/big-data-and-international-development-im-

The Data-Pop Alliance, a global coalition bringing together researchers, practitioners, and activists to promote a people-centred approach to big data, considers **big data for development as an ecosystem** with three main elements:⁷



Crumbs of data: the new digital information from ubiquitous data sources that become available and enable alternative approaches to development challenges. In the development context what matters is not so much the sheer size, speed or variety, but the nature of the new data, which is passively emitted by people as they interact with digital technology.



Capacities: the expansion of tools and methods to collect, aggregate, analyse, and visualise data by development actors. They include things like algorithms and parallel computing, which are able to analyse in seconds what would in the past have taken years.



Communities: it is extremely important to look not just at the technology, but at all organisations and communities taking part in data generation, governance, and usage. This includes those producing the data, policy-makers, privacy advocates and end users.

Evolving actors and emerging initiatives

Two prominent actors in the big data for development space on a global level are UN Global Pulse, a flagship UN Secretary-General initiative, and the Data-Pop Alliance. However, **several smaller and new actors and partnerships have emerged in recent years.**

The UN Statistical Commission has created the Global Working Group on Big Data for Official Statistics,⁸ which looks into the potential benefits and challenges of big data to complement and improve traditional statistical sources and to monitor progress around the Sustainable Development Goals. Moreover, the International Telecommunications Union has recently launched the Artificial Intelligence for Good Global Summit,⁹ which every year will gather government officials, UN agencies, civil society organisations, industry leaders and artificial intelligence experts to look into how big data and artificial intelligence can be applied for social good and help achieve the SDGs. Some other actors and initiatives worth mentioning include: The World Bank's big data innovation challenge,¹⁰ the Data2X initiative on partnering for a gender data revolution,¹¹ the platform for big data in agriculture from the International Center for Tropical Agriculture,¹² LIRNEasia's work including their big data initiative analysing land usage patterns from space,¹³ and Paris21's programme on the data revolution.¹⁴

pacts-scenarios-and-policy-options

7 Letouzé, E. and Sangokoya, D. (2015) Leveraging Algorithms for Positive Disruption: On Data, Democracy, Society and Statistics, Data-Pop Alliance, datapopalliance.org/wp-content/uploads/2016/03/DataPopAlliance_LeveragingAlgorithms.pdf

8 unstats.un.org/bigdata

9 itu.int/en/ITU-T/AI

10 bigdatainnovationchallenge.org

11 data2x.org

12 blog.ciat.cgiar.org/cgiar-platform-for-big-data-in-agriculture

13 lirneasia.net/2017/05/big-data-igf-presentation-athuraliya

14 datarevolution.paris21.org/

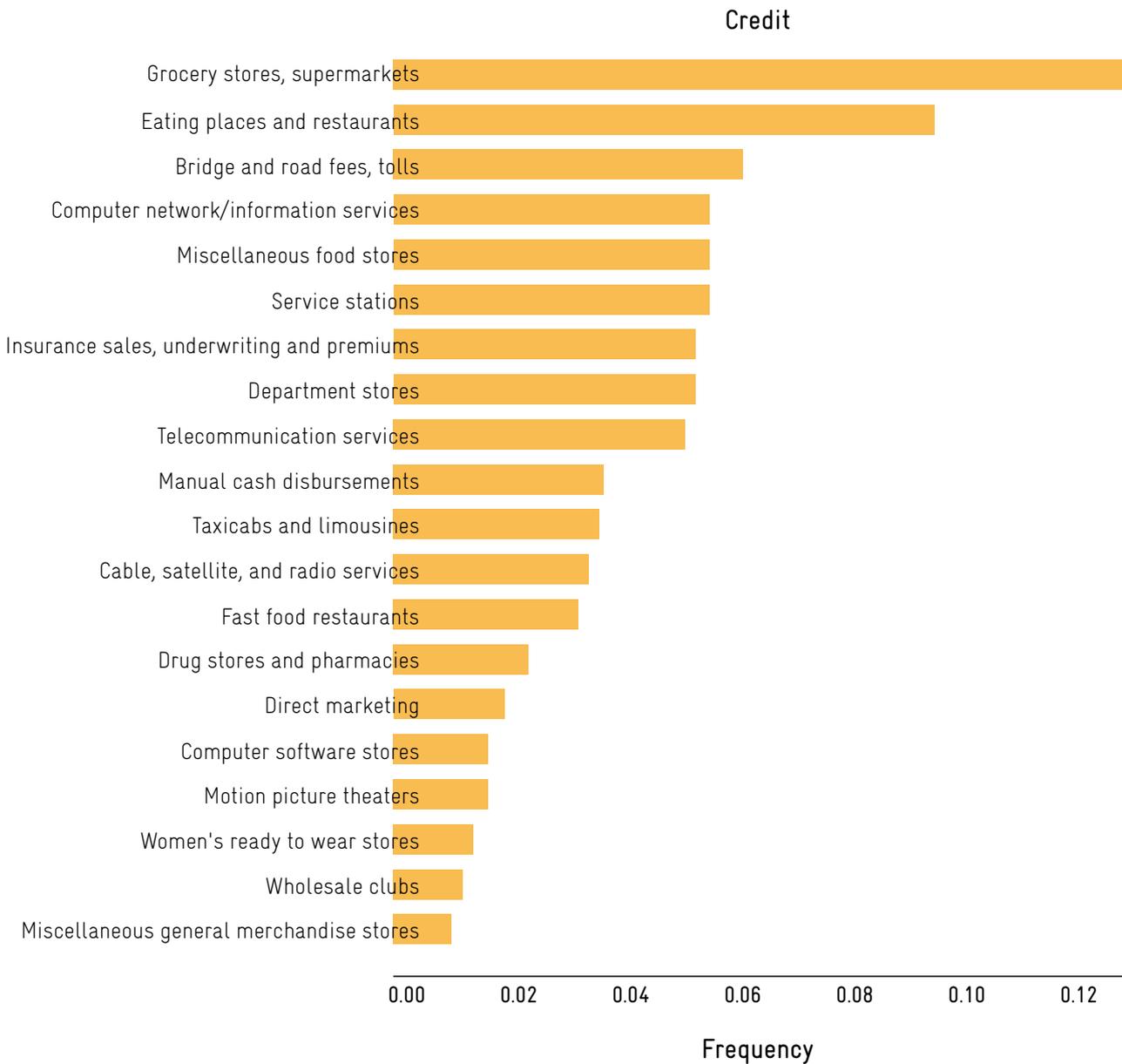


Figure 2.a. Looking deeper into big data and gender

The report 'Big Data and the Well-Being of Women and Girls Applications on the Social Scientific Frontier'¹⁵ by the organisation Data2X¹⁶ describes various approaches of using different kinds of big data in order to fill the global gender data gap.

One example from the study is shown in the diagram above: The diagram shows the frequency of transactions in different expenditure categories, as assessed by anonymized credit card data and cell phone data of women living in a major Latin-American metropolis. Thus behavior patterns are portrayed and illustrate needs and priorities.

¹⁵ <http://data2x.org/wp-content/uploads/2017/03/Big-Data-and-the-Well-Being-of-Women-and-Girls.pdf>

¹⁶ Data2X is a collaborative technical and advocacy platform dedicated to improving the quality, availability, and use of gender data in order to make a practical difference in the lives of women and girls worldwide: www.data2x.org

The value of big data

The ability to quickly analyse very large and diverse data streams potentially enables the extraction of insights that would have been unattainable using traditional methods. Generally, big data increases an organisation's capacity to describe and visualise situations, diagnose and understand underlying problems, predict likely scenarios and even prescribe the most beneficial potential actions.

This disruptive capacity also affects the development sector, where actors are figuring out how to incorporate big data into their strategy, programming and operations. Some proponents of the use of big data in development argue that **big data and artificial intelligence provide an opportunity to shift from biased, inefficient and unjust decision-making towards more objective and evidence-based decisions.**¹⁷ However, that does not mean that new insights will be used as it is well documented that evidence does not necessarily lead to action,¹⁸ especially for controversial issues.

Others argue that big data's largest contribution is its provision of new ways to visualise development and thus guide – and to some extent also limit – its responses.¹⁹ Traditionally, survey data including censuses provided static visualisations of data which was updated infrequently. Because big data is collected continuously, information can be disseminated through more fluid visualisation techniques²⁰ such as dashboards providing panoptic views of operations, word clouds capturing citizen sentiment and exposure to shocks, and maps that capture movements of people and the intensity of these movements.

Box 2.a. Precision farming using big data and algorithmic decision-making for more targeted manuring and soil irrigation

Algorithmic decision-making refers to instances where algorithms make automatic decisions based on the analysis of data, or at least present decision choices to humans. One such example is **precision farming**, which mixes data from weather forecasts with data from sensors measuring the humidity and temperature of air and soil, aerial photography from drones or satellites measuring crop maturity, and data on the nearness of equipment and human labour. This data is then used to generate predictive models that enable farmers to apply inputs to specific areas of land as needed rather than applying inputs uniformly across the farm as done in traditional farming.

Big data increases an organisation's capacity to describe and visualise situations, diagnose and understand problems, predict scenarios and prescribe potential actions.

- 17 Letouzé, E. and Sangokoya, D. (2015) Leveraging Algorithms for Positive Disruption: On Data, Democracy, Society and Statistics, Data-Pop Alliance, datapopalliance.org/wp-content/uploads/2016/03/DataPopAlliance_LeveragingAlgorithms.pdf
- 18 Georgalakis, J.; Jessani, N.; Oranje, R. and Ramalingam, B. (Eds) (2017) *The Social Realities of Knowledge for Development*, Brighton: IDS, www.ids.ac.uk/publication/the-social-realities-of-knowledge-for-development-sharing-lessons-of-improving-development-processes-with-evidence
- 19 Flyverbom, M.; Madsen, A.K. and Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities, *The Information Society: An International Journal* 33.1: 35–42
- 20 Robinson, I. (2016) *Data Visualisation: Contributions to Evidence-Based Decision-Making*, SciDev.Net, www.scidev.net/filemanager/root/site_assets/learning_reports/data_visualisation_learning_report_scidevnet.pdf

Being able to visualise problems in new ways allows development actors to tackle them differently. Big data is promoting a shift in what is considered actionable knowledge, that gives more prominence to correlations analysis that unearths patterns even if these are not fully understood.²¹ The rationale is that sometimes knowing ‘what’ is happening is enough to take action, even if the ‘why’ remains undiscovered. Besides, **big data analysis can provide the ‘digital smoke signals’ that uncover anomalies, allowing actors to follow up quickly and investigate why it is happening.** Diverse tests, including methods from other research disciplines, can be carried out to validate the correlations before taking action in cases where knowing ‘why’ is important.

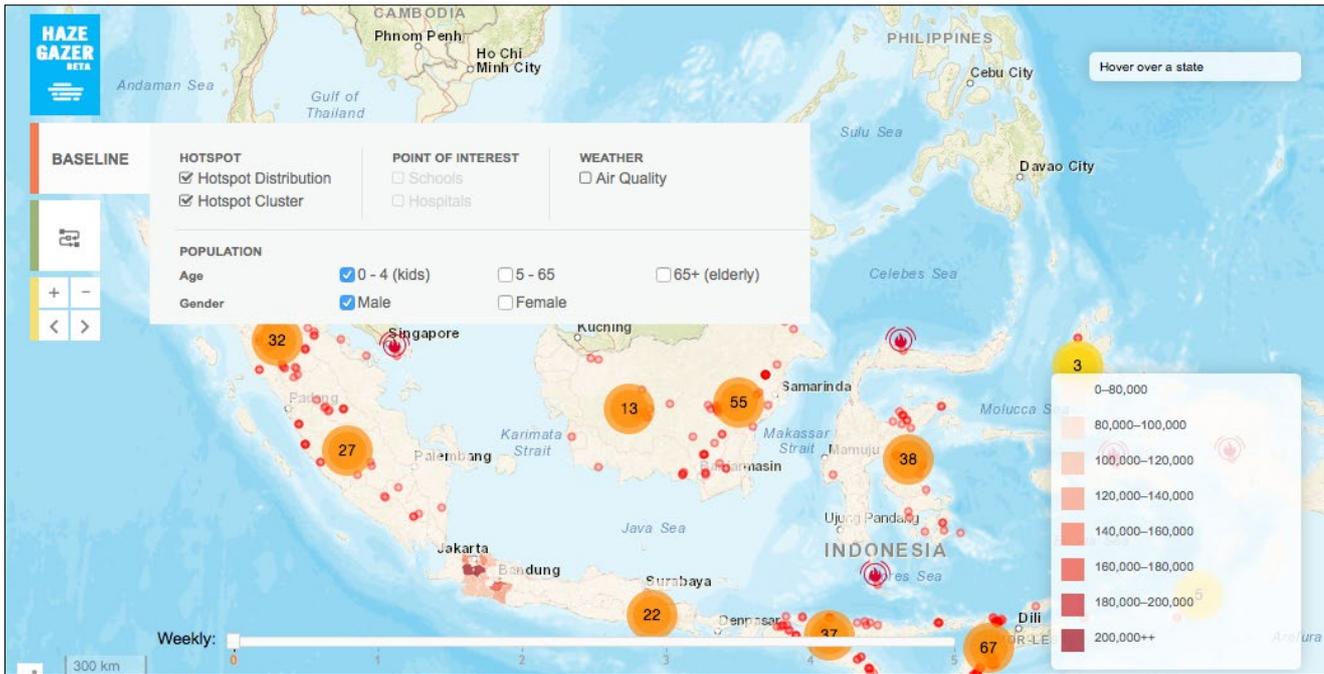


Figure 2.b. Pulse Lab Jakarta is enabling public access to Haze Gazer, which is a crisis analysis and visualization tool that provides real-time information on fire and haze hotspots in Indonesia.²²

Big data algorithms are very much limited depending on which data is available.

Big data’s potential is yet to be realised

Although big data algorithms could generate new ways of looking at development and tackling problems, they are very much limited depending on which data is available. For example, algorithms that rely on mobile phone data will not be inclusive of people living in areas with low mobile phone penetration or poor network connectivity/availability as well as those less likely to own mobile phones (e.g. the poor, women, the elderly). Moreover, using Twitter data can help to gain useful insights about what is currently happening,²³ but could also be limiting in other situations: Twitter only makes a small subset of its data available and tweets are generally short outbursts devoid of background contextual information, making it difficult to tell complete stories.²⁴ Thus, taking a

21 Flyverbom, M.; Madsen, A.K. and Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities, *The Information Society: An International Journal* 33.1: 35-42
 22 hazegazer.org
 23 fastcompany.com/3007178/how-uns-new-data-lab-indonesia-uses-twitter-preempt-disaster
 24 Flyverbom, M.; Madsen, A.K. and Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities, *The Information Society: An International Journal* 33.1: 35-42

problem- and user-centric approach to big data for development – which is complemented by traditional data sources when possible – is essential to ensure that the right data from the right people is analysed to tackle the right problem.

To date, **big data for development projects have mainly been pilots and proof of concepts** showing the potential to use big data to tackle development problems. These pilots have helped garner interest and hype around big data for development. Although there is no clear indication that any of these pilots have resulted in significant impacts, there is reason to remain hopeful. Several data for development experts have told us that thanks to pilots and proof of concepts, the development sector is ready to start integrating long-term big data applications.

Box 2.b. Tackling the spread of malaria using mobile phone data

The World Health Organization estimates that nearly half the world's population is at risk of getting malaria, with roughly 212 million cases and 429,000 deaths from the disease in 2015. Children under five are especially vulnerable, making up 70 per cent of all deaths.²⁵ The mosquito species responsible for spreading the disease does not innately produce malaria, but acquires it by biting already infected humans and then passes it to humans through subsequent bites. Since one or two weeks can pass without symptoms showing, people often carry the disease and unknowingly contribute to the infection of others. When people become infected while travelling, it can lead to new outbreaks.

Wesolowski et al.²⁶ combined malaria incidence data collected by health officials (pre-existing data) with anonymised call record data from 15 million Kenyan mobile subscribers disaggregated at the cell tower level, to understand how travel patterns contribute to spreading malaria. Using this data allowed them to uncover 'sources' from which humans typically picked up the disease and 'sinks' to which they were likely to travel and put others at risk. They were able to predict on a daily basis which communities were likely to import the disease and the probability that visitors to a given community would get sick. Such an understanding provides insights on where and when prevention efforts should be focused, at a much more granular level. However, past experience shows that current 'efforts' to tackle malaria such as distributing bed nets do not automatically lead to the behavioural change that would make them effective.²⁷ Thus, it is uncertain that big data-backed efforts will lead to take-up either.

There are other more recent **initiatives that use big data and machine learning to tackle malaria and other infectious diseases such as Zika, Dengue or Ebola**. These initiatives illustrate the wide diversity of approaches available: from integrating infection cases data with satellite, environmental and climatological variables to produce disease outbreak risk maps and recommendations for action,²⁸ to using drones that autonomously locate mosquito hotspots, robotic traps to identify and collect interesting specimens, and machine-learning algorithms to search for pathogens in the genetic material obtained from mosquitoes.²⁹

Researchers were able to predict which communities were likely to import malaria and the probability that visitors to a given community would get sick.

25 WHO (2016) World Malaria Report 2016, Geneva: World Health Organization, who.int/malaria/publications/world-malaria-report-2016/report/en

26 Wesolowski, A.; Buckee, C.O.; Pindolia, D.K.; Eagle, N.; Smith, D.L.; Garcia, A.J. and Tatem, A.J. (2013) The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales, *PLoS One* 8.1

27 nyti.ms/2kqTm8M

28 disarm.io

29 microsoft.com/en-us/research/project/project-premonition

2.2. Open data

Open data is data that can be freely used, reused, and redistributed by anyone.

The earliest widely accepted definition characterises open data as ‘data that can be freely used, reused, and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike’.³⁰ This definition points to at least three characteristics, namely accessibility, machine-readability, and reusability, but hinges on the use of technology not only to produce and publish data, but also to access it.

Box 2.c. Eight characteristics of open government data

The Sebastopol Principles emphasise eight characteristics of open government data:³¹

- 1 **Complete** – all public data not subject to valid privacy, security or other limitations is made available;
- 2 **Primary** – at source, not aggregated, and granular;
- 3 **Timely** – available at the earliest possible time;
- 4 **Accessible** – available to widest range of users and purposes;
- 5 **Machine-processable** – structured to allow automated processing;
- 6 **Non-discriminatory** – available to anyone, without registration;
- 7 **Non-proprietary** – in a format over which no entity has exclusive control;
- 8 **Licence free** – not subject to any copyright, patent, trademark or trade secret regulation.

Acknowledging that **moving towards this level of data openness is a process rather than an end state**, a graduated system of deployment has been formulated³² so that it will capture the progression from closed to open data, and acknowledge the efforts of different actors in opening up datasets. The Five Star Deployment Scheme for Open Data considers, for example, an openly licensed PDF file containing structured data with sufficient level of granularity and published online as a move towards data openness (see also Figure 2.c.).

30 opendefinition.org/od/2.1/en

31 public.resource.org/8_principles.html

32 5stardata.info



Figure 2.c. Progressing from closed to open data: The Five Star Deployment Scheme for Open Data. OL: open license, RE: machine readable, OF: open format, URI: uniform resource identifier, LD: linked data
PDF, XLS, CSV, RDF and LOD are file formats and related standards. LOD: linked open data, RDF: resource description format, CSV: comma-separated values

From government-driven to multi-stakeholder collaboration

The global open data movement started off focusing on advocating for openly available government data more than any other type of data. Country governments, more particularly the US and the UK, led the early pack of reformers in open government data, with the US launching its open data portal in 2009 and the UK in 2010. A wave of similar initiatives was launched in the developing world, including in Kenya, Moldova and the Philippines, many of which were supported by the World Bank which also made public previously internal data on its platform in 2010.

Open data became a primary advocacy target of the Open Government Partnership (OGP), ‘a multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance’.³³ The OGP Open Data Working Group, composed of representatives from government and civil society, coordinated the process of promoting open data practices globally that led to the eventual launch of the International Open Data Charter³⁴ in 2015, a manifesto and a commitment on how governments should publish government data.

Currently, the open data landscape is made up of a very diverse set of actors from across the globe. It includes non-profits, philanthropic organisations, multilaterals, think tanks and regional organisations (see Box 2.d. for examples). These actors are members of different networks that promote open data globally, more notably, the Open Government Partnership, the International Open Data Charter, the Open Data for Development Network,³⁵ the Global Open Data for Agriculture and Nutrition,³⁶ and loose alliances of researchers like the Open Data Research Network.³⁷

A diverse set of actors promotes open data globally – non-profits, philanthropic organisations, multilaterals, think tanks and regional organisations.

33 opengovpartnership.org/about/about-ogp

34 opendatacharter.net

35 od4d.net

36 godan.info

37 opendataresearch.org

Box 2.d. Open data actors

- » Non-profit organisations such as Open Knowledge, the Open Data Institute, and the World Wide Web Foundation;
- » Philanthropic organisations like the Omidyar Network;
- » Multilateral organisations including UNDESA and the World Bank;
- » Think tanks like NYU GovLab and the International Development Research Centre;
- » Regional organisations like Iniciativa Latinoamericana por los Datos Abiertos and Code for Africa.

A review of open data initiatives across the globe³⁸ found that several of the open data initiatives focused on capacity building and training, advocacy and lobbying for data disclosure, and data dissemination and reuse. This indicates not only the nascent nature of the field, but also the theory of change that those engaged in the initiatives propose (see discussion on theorised impacts below). To date, major initiatives work on standard-setting (e.g. the Open Contracting Data Standard), implementation and practice (e.g. Open Data Lab Jakarta³⁹), and research and measurement (e.g. the Open Data Barometer, see also Figure 2.e.).

Thematically, open data is used largely in programmes or projects in mapping, business, research and consulting, governance, energy and climate, finance and insurance (see Figure 2.d.). **Emerging fields, in terms of open data use are health, transportation and logistics, and education.**⁴⁰

38 Boyera, S. and Iglesias, C. (2014) Open Data in Developing Countries: State of the Art, London: Open Data Institute, https://docs.google.com/document/d/1FMylLu-jouL7j7Pw0kEwUn_B07aZ9IX3vIFGqP00gX0/edit

39 labs.webfoundation.org

40 opendataimpactmap.org/map.html

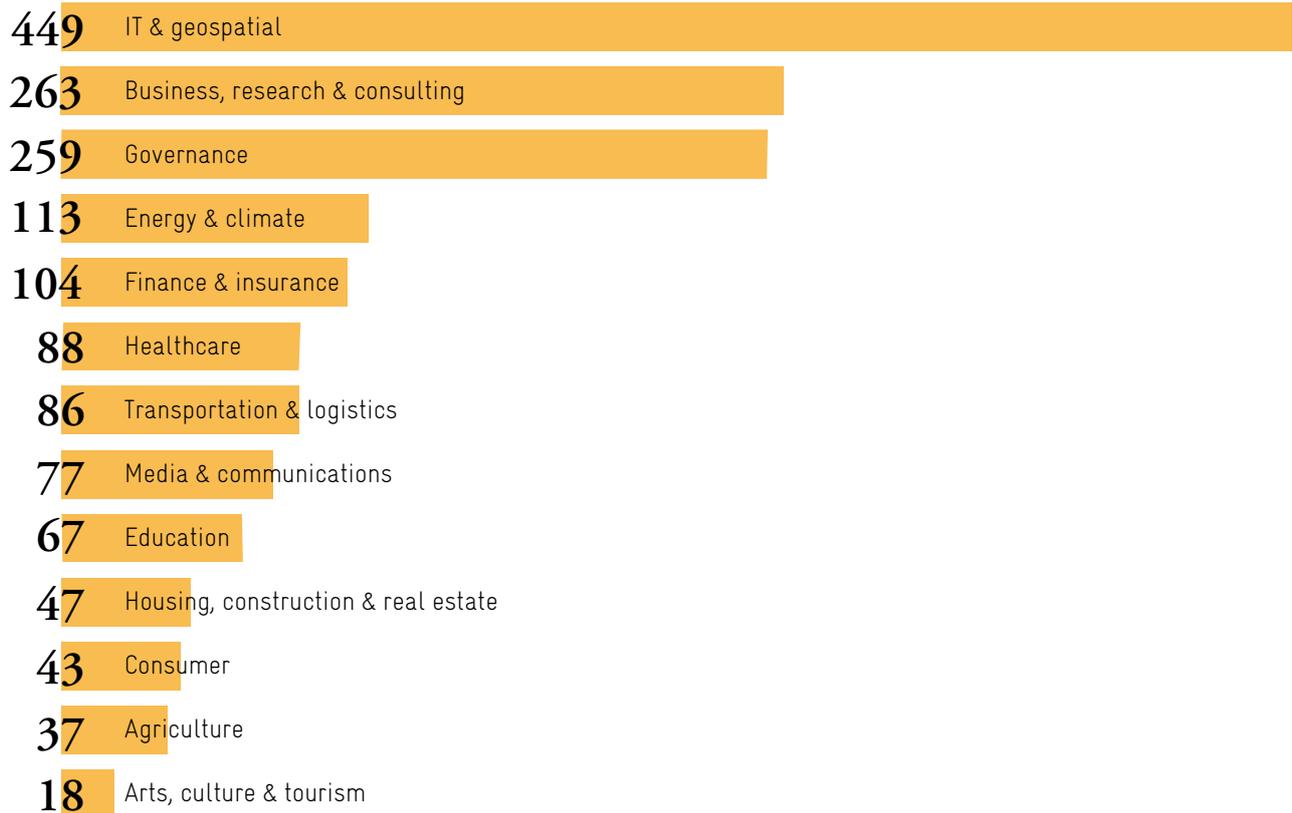


Figure 2.d. Number of projects using open data across the globe according to sector.⁴¹

Open data cannot achieve impact alone

For open data to result in actual impact, there needs to be at least three building blocks in place: (1) the publication of open data by governments, (2) the conversion of data to actionable information by intermediaries, and (3) the use of data by citizens, government officials and other stakeholders to achieve development outcomes.

In a review of more than 15 case studies of open data initiatives in developing countries, Davies and Perini⁴² argue that at least three emerging impacts can be identified: increasing government transparency and accountability, innovation and economic development, as well as greater inclusion and citizen empowerment. Verhulst and Young⁴³ arrived at a similar conclusion on the positive effect that open data can have on governments, citizens, public problems, and economic opportunities. The evidence, however, is limited to a few cases only. Nevertheless, the impacts, while they are achieved, are dependent on different factors, the availability of open data being just one of them. What has become clear is that open data impacts depend on partnerships and collaborations, the

Emerging impacts of open data include government transparency and accountability, innovation and economic development, greater inclusion and citizen empowerment.

⁴¹ opendataimpactmap.org

⁴² Davies, T. and Perini, F. (2016) Researching the Emerging Impacts of Open Data: Revisiting the ODDC Conceptual Framework, *The Journal of Community Informatics* 12.2, ci-journal.net/index.php/ciej/article/view/1281

⁴³ Verhulst, S. and Young, A. (2016) *Open Data Impact: When Demand and Supply Meet*, New York: GobLab, thegovlab.org/static/files/publications/open-data-impact-key-findings.pdf

The provision of open data does not automatically result in impact. A conducive political, economic and social context influences the disclosure and use of data.

quality of open data public infrastructure, clear open data policies, and the degree to which a government responds to a problem or issue.^{44,45}

This realisation is not new. The Open Data Barometer, a comprehensive assessment of the state of open data globally, acknowledges that data is just one part of the equation to make open data result in positive outcomes. Among these are the conducive political, economic, and social context that largely influences the disclosure and use of data. The provision of open data does not automatically result in impact, indicating that the journey from open data to impact needs several intermediary steps, and intermediary stakeholders.⁴⁶

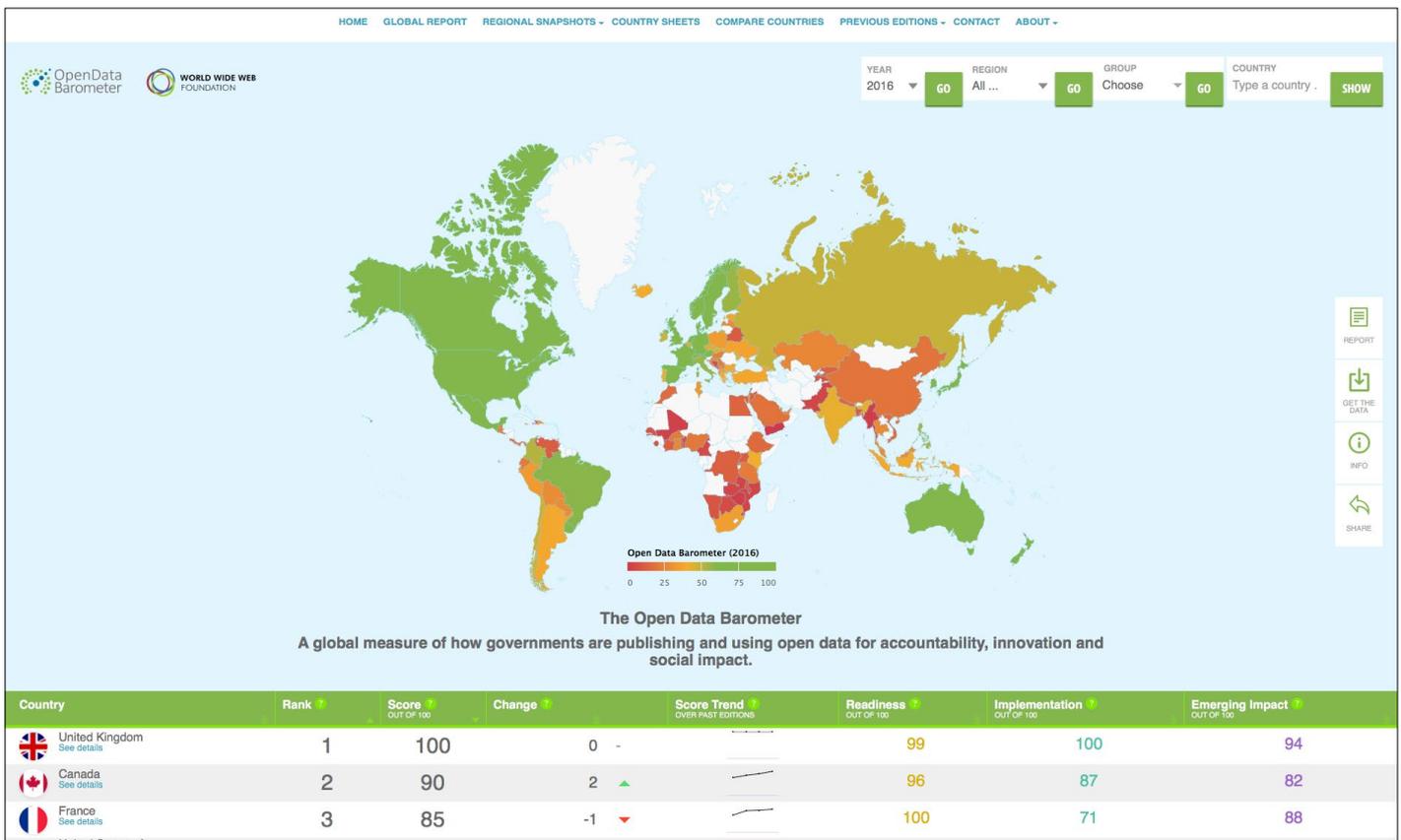


Figure 2.e. The Open Data Barometer Rank measures how governments around the world are publishing and using open data for accountability, innovation and social impact.

44 Verhulst, S. and Young, A. (2016) Open Data Impact: When Demand and Supply Meet, New York: GobLab, thegovlab.org/static/files/publications/open-data-impact-key-findings.pdf

45 Heusser, F. (2012) Understanding Open Government Data and Addressing its Impact, Background Papers, ODDC

46 Davies, T. (2014) Open Data in Developing Countries: Emerging Insights from Phase 1, World Wide Web Foundation, opendataresearch.org/sites/default/files/publications/Phase 1 - Synthesis - Full Report-print.pdf

Open Data Barometer Rank



Figure 2.f. Top five developing partner countries of German development cooperation ranked according to the Open Data Barometer Rank (115 countries and jurisdictions ranked in 2016).⁴⁷

Box 2.e. Identifying lower-priced generic medicines through open data in South Africa

In South Africa, the prices of medicines are regulated by the government, which requires that pharmacists offer private patients the lower-priced generic medicine for prescriptions issued by doctors. However, this does not always happen, despite the fact that generic drugs are 40 per cent cheaper than branded ones. While data about medicines and their prices are published regularly by the country's health ministry, it is not easy to find the data and difficult to understand and use it. A little-known open dataset from the website of the Department of Health shows a registry of medicines, their generic names and their costs.

Using this data, in 2014 Adi Eyal of Open Up (formerly Code for South Africa)⁴⁸ developed an application called the Medicine Price Registry Application⁴⁹ and an online database⁵⁰ of medicines. This app enables patients to identify and request lower-priced generic medicines and check that they are not being overcharged by their local pharmacies.

This online tool is now used by citizens to source cheaper alternatives of prescription medicine as well as by doctors who work in mixed-income communities to provide the best medical advice to patients with due consideration of their health budget or income. The initiative has been widely appreciated by its users for its cost-saving effects without compromising efficacy.

47 opendatabarometer.org

48 openup.org.za/articles/cheaper-medicines-unexpected-user.html

49 mpr.code4sa.org

50 data.code4sa.org/Government/Database-Of-Medicine-Prices/mba4-xngh

2.3. Citizen-generated data

The idea behind citizen-generated data is for citizens to be not only passive consumers of information, but also active data producers, users and intermediaries.

The idea behind citizen-generated data is for citizens to be not only passive consumers of information, but also active data producers, users and intermediaries. Employing a citizen-generated data approach can be effective in situations where no data from other sources is available to affected stakeholders as it is either not collected by government or the private sector, or not made accessible. Likewise, data actively generated by citizens can be useful when the available data is incomplete, inaccurate or not timely enough. What is more, in rural areas where limited internet access makes accessing official data difficult, citizen-generated data collected at the local level and shared with relevant target groups through offline channels could be an effective response to the lack of accessible information.⁵¹ While the citizen-generated data concept can be used to describe the use of both digital and traditional data sources, this section focuses on digital data sources and digital data collection and management tools.

How citizens shape data collection

Proponents of the concept argue that data actively produced by citizens with a specific purpose in mind can provide a direct representation of citizens and enable them to ‘directly monitor, demand or drive change on issues that affect them’.⁵² **Citizen-generated data approaches give people agency** over decisions on what data is collected, what it is used for, and how it is used. Initiatives like I Paid A Bribe,⁵³ an online corruption-reporting platform run by an Indian civil society group, or HarassMap,⁵⁴ a collection of crowdsourced SMS and online reports on sexual harassment in Egypt, were launched to tackle issues that affected citizens, for which no public data was available.

What distinguishes the citizen-generated data concept from the more traditional citizen participation approaches is the strong emphasis on the role of citizens as data producers compared to mere consumers of information.⁵⁵ Moreover, the concept highlights the importance of the data in question being collected with a specific usage in mind. Thus, unlike data that is passively generated through the use of technology devices, often referred to in the literature as ‘data breadcrumbs’,⁵⁶ citizen-generated data is consciously produced for a specific purpose by a group of citizens who normally retain ownership of the data collected. In many cases the data is made publicly available with no restrictions in terms of usage.

As is the case with the other three data for development concepts discussed in this report, definitions of citizen-generated data can be broad and sometimes include cases where data is produced by citizens but collected and owned by third parties. Private firms like Premise or Findyr have pioneered remote data collection through cell phones, largely for commercial interests. Other

51 data4sdgs.org/guide-making-use-of-citizen-generated-data

52 Wilson, C. and Rahman, Z. (2015) Citizen-Generated Data and Governments: Towards a Collaborative Model, DataShift, civicus.org/thedatashift/wp-content/uploads/2015/07/statistical-perspectives-on-cgd-web_single-page.pdf

53 ipaidabribe.com

54 harassmap.org/en

55 Rebecca, S.; Rono-Bett, K. and Kenei, S. (2017) Citizen-Generated Data and Sustainable Development: Evidence from Case Studies in Kenya and Uganda, Nairobi: DevInit and DRT, makingallvoicescount.org/publication/citizen-generated-data-sustainable-development

56 Letouzé, E. and Sangokoya, D. (2015) Leveraging Algorithms for Positive Disruption: On Data, Democracy, Society and Statistics, Data-Pop Alliance, datapopalliance.org/wp-content/uploads/2016/03/DataPopAlliance_LeveragingAlgorithms.pdf

initiatives under the broad umbrella of citizen-generated data use terms like ‘user-generated data’⁵⁷ to refer to any digital data created by individuals online, such as social media platforms.

Many actors, but little coordination

Unlike in the open data and big data spaces, there are no large coordinated international networks or initiatives that promote the citizen-generated data approach in development. The only exception is DataShift,⁵⁸ which seeks to explore and promote the concept through research and capacity building of civic groups to engage in citizen-generated data projects.

Key players in the field include Ushahidi⁵⁹ whose open source software powers a great number of crowdmapping initiatives. Actors like the Humanitarian OpenStreetMap Team⁶⁰ and Missing Maps⁶¹ that work across the globe to provide detailed and accurate maps for humanitarian response and economic development, have focused on maps as a tool in enabling citizens to provide data. In addition, several smaller initiatives have contributed to the advancement of the citizen-generated data field through practical experimentation with various forms of data collection techniques, tools and platforms. PetaJakarta,⁶² for instance, has been exploring the combination of crowdsourced data through social media networks with government data to tackle the problem of flooding in the Indonesian capital of Jakarta.

Other projects include crowdmapping projects like the aforementioned I Paid A Bribe⁶³ and HarassMap.⁶⁴ Depending on the conceptual definition of citizen-generated data, initiatives such as U-Report,⁶⁵ a reporting and polling portal for social issues, and Wefarm,⁶⁶ an information-sharing platform for small-scale farmers, could be subsumed under citizen-generated data initiatives.

Open source software powers a great number of crowdmapping initiatives.

57 devpolicy.org/bridging-data-gaps-policy-making-crowdsourcing-big-data-development-20160708

58 civicus.org/thedatashift

59 ushahidi.com

60 hotosm.org

61 missingmaps.org

62 petajakarta.org

63 ipaidabribe.com

64 harassmap.org/en

65 ureport.in

66 wefarm.org

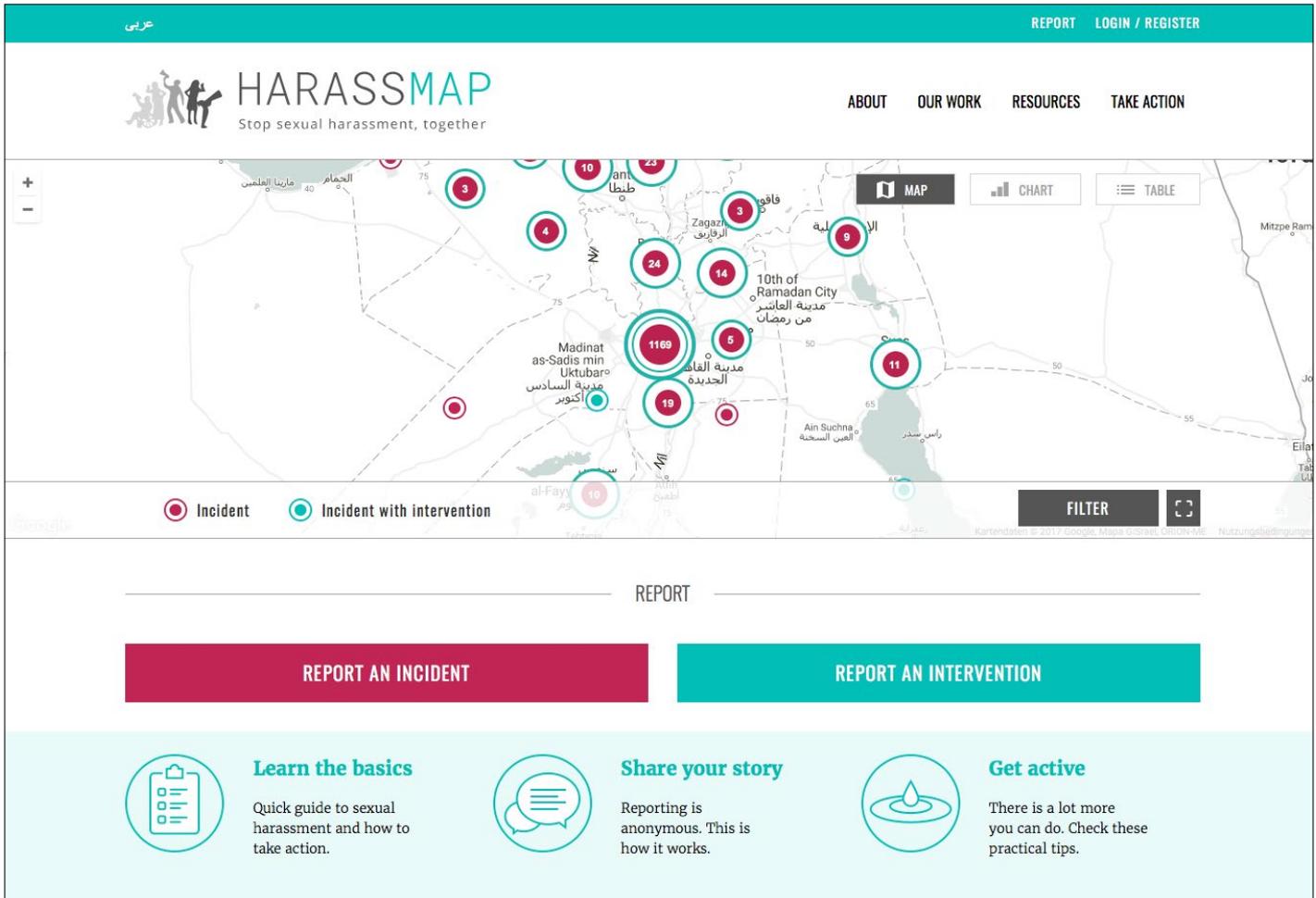


Figure 2.g. HarassMap is a collection of crowdsourced SMS and online reports on sexual harassment in Egypt.⁶⁷

Crowdmapping initiatives have contributed to a better understanding of corruption, harassment, or voter turnout.

Scaling remains a challenge

While much of it is anecdotal in nature, there is evidence of the value of citizen-generated data for improved delivery of development projects.⁶⁸ Crowdmapping initiatives have contributed to a better understanding of the nature and scale of social issues such as corruption, harassment, or voter turnout.⁶⁹ Granular and real-time data provided by citizens has enabled government authorities to take more evidence-based decisions about urban infrastructure such as in the case of PetaJakarta. Moreover, some research has documented the impact of citizen-generated data projects on government policy and practice such as adjustments of measuring education outcomes in Uganda and improvements in community water supply in Malawi.⁷⁰

67 harassmap.org/en

68 Rebecca, S., Rono-Bett, K. and Kenei, S. (2017) Citizen-Generated Data and Sustainable Development: Evidence from Case Studies in Kenya and Uganda, Nairobi: DevInit and DRT, www.makingallovoicescount.org/publication/citizen-generated-data-sustainable-development

69 Bailard, C.S. and Livingston, S. (2014) Crowdsourcing Accountability in a Nigerian Election, *Journal of Information Technology and Politics* 11.4: 349–367

70 Rebecca, S., Rono-Bett, K. and Kenei, S. (2017) Citizen-Generated Data and Sustainable Development: Evidence from Case Studies in Kenya and Uganda, Nairobi: DevInit and DRT, makingallovoicescount.org/publication/citizen-generated-data-sustainable-development

However, **citizen-generated data initiatives often face barriers that make a rigorous application at scale difficult.** Challenges include providing a sufficiently robust and strong data collection infrastructure, capacities to ensure sustainable data production, and incentives for citizens to contribute data and to sustain their initial commitment to data collection efforts. Moreover, when it comes to combining or complementing public sector with citizen-generated data, questions of accuracy, representation and quality assurance processes become relevant.⁷¹ Other issues that require a closer look in the design of citizen-generated data projects include privacy implications for personally identifiable information, data provenance as well as interoperability with other data.

Box 2.f. Using drones to collect data for land rights advocacy in Indonesia

Official data about land type, usage and ownership in Indonesia is often incomplete, inaccurate and even contradictory.⁷² As a result, concessions for logging or mining activities have been awarded to companies based on low-quality data, which has led to frequent conflicts between businesses and local communities over territorial rights. What is more, public access to official map data held by the government is limited, which makes it difficult for activists and civic groups to engage in an informed conversation with the government over land rights and address issues ranging from land grabbing to environmental pollution resulting from illegal exploitation of natural resources.⁷³

In a response to the lack of inaccessible and reliable data, and faced with limited financial resources, a group of activists assembled do-it-yourself drones to capture high-resolution images of forest areas in the Indonesian part of the Borneo island. Equipped with these low-cost unmanned aerial vehicles, the **activists worked with local communities to create their own cartographic material in order to better understand ecological damage** caused by private companies operating near villages.

The drone footage collected by the villagers revealed that land rights of local communities were being violated by private businesses. The maps were even used in a Constitutional Court trial to show that mineral and coal-mining companies were not complying with environmental regulations.

One of the organisations involved in the initiative, the Swandiri Institute, has built on these initial successes by setting up several ‘drone schools’ throughout Indonesia. The organisation also started testing drones in collecting data on the health of crops to allow for more effective prevention of pests and disease that could severely limit agricultural output with significant impact on the incomes of local farmers.⁷⁴

71 Piovesan, F. (2017) Statistical Perspectives on Citizen-Generated Data, DataShift, civicus.org/thedatashift/wp-content/uploads/2015/07/statistical-perspectives-on-cgd_web_single-page.pdf

72 civicus.org/thedatashift/learning-zone-2/research/changing-what-counts/community-drones

73 Radjawali, I.; Pye, O. and Flitner, M. (2017) Recognition through Reconnaissance? Using Drones for Counter-Mapping in Indonesia, *The Journal of Peasant Studies* 44.4: 1–17
tandfonline.com/doi/full/10.1080/03066150.2016.1264937

74 unglobalpulse.org/news/drones-community-planning

2.4. Real-time data

Real-time data: data on current activities that is analysed and presented quick enough to make timely decisions and change plans.

Real-time data refers to data on current activities that is analysed and presented quick enough to make timely decisions and change plans based on the data. For example, routing applications like Waze or Google Maps provide the expected time of arrival for several routes based on data collected in real time on the speed of travel of smartphone users travelling on those routes. The information displayed on the app can then be used to choose an alternative faster route in the case of an accident or heavy traffic. Although not always as technologically sophisticated or immediate, real-time data is now an emerging trend in international development.

Although real-time data systems are often associated with the use of digital technologies, in the context of development interventions information and communications technology (ICT) usage is not necessarily the most important aspect.⁷⁵ Data systems need to be contextually appropriate and therefore their intensity of ICT usage varies widely. At one end of the spectrum there are systems that rely on face-to-face surveys and in-person observation, while at the other end there are automated computational systems that rely on sensors for immediate decision-making. However, there is a clear tendency to increase the use of ICT in real-time data systems for development,⁷⁶ as well as an increase in real-time data initiatives using big data, open data and citizen-generated data as their data sources. This section focuses on digital real-time data, which can be defined as: ‘Employing digital technologies, typically involving mobile technologies and related software applications, to enable and accelerate the collection, sharing, management, analysis and reporting of data to inform more rapid and timely decision-making’.⁷⁷

The term ‘real time’ seems to imply immediacy, but in development contexts it refers to timeliness: data that is gathered, analysed and disseminated quick enough to make timely decisions, according to its intended use. For example, data published monthly may be sufficient for local civil society groups to make resource allocation decisions in the midst of dynamic contexts. However, in emergency situations, for example natural disasters, data that is several days old may not be timely enough to respond properly. Frequently, in development literature real-time data is used to refer to the increased timeliness of data sources; for example, new data that is available quarterly rather than every several years.⁷⁸

Not many development organisations have initiatives focusing on real-time data.

Few actors, public-private partnerships are common

Not many development organisations have initiatives focusing on real-time data. However, there are a number of development projects that use digital technologies and rely on real-time data in one way or another, even if they do not consider themselves to be real-time data projects. For example, **most big data projects using recently generated data streams are making use of real-time data**, but consider it as an implicit component of big data. Moreover, projects

75 Lucas, H., Batchelor, S. and Berdou, E. (2013) Real Time Monitoring and the New Information Technologies, IDS Bulletin 44.2: 31–39 bulletin.ids.ac.uk/idsbo/article/view/244

76 Akhter, M. and Chaudhuri, J. (2013) Real Time Monitoring for the Most Vulnerable: Pre-Primary Education in Bangladesh, IDS Bulletin 44.2: 97–112

77 Ramalingam, B.; Barnett, I.; Levy, A.; Oppenheimer, C. and Valters, C. (forthcoming) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

78 Lucas, H., Batchelor, S. and Berdou, E. (2013) Real Time Monitoring and the New Information Technologies, IDS Bulletin 44.2: 31–39

involving citizen-generated data, such as monitoring of election-related violence, also require immediate analysis of the incoming data in order to react in time, but these initiatives tend to emphasise the ‘crowdsourcing’ rather than the real-time element as their focus.

UNICEF⁷⁹ has supported the development and expansion of several open source real-time data systems, including RapidFTR, RapidSMS and RapidPro. One of UNICEF’s flagship initiatives is U-Report,⁸⁰ a mobile social messaging tool used in 35 countries, which allows communities to share what they care about and their sentiment towards development issues and programmes through polls. USAID has explored the potential of real-time data to support adaptive management of development interventions, encouraging them to respond to change in complex dynamic environments. On behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ), the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) has supported the development of the open source platform SORMAS (see Figure 2.f),⁸¹ a disease outbreak early warning and management system that was used during the West African Ebola outbreak in 2014/2015.

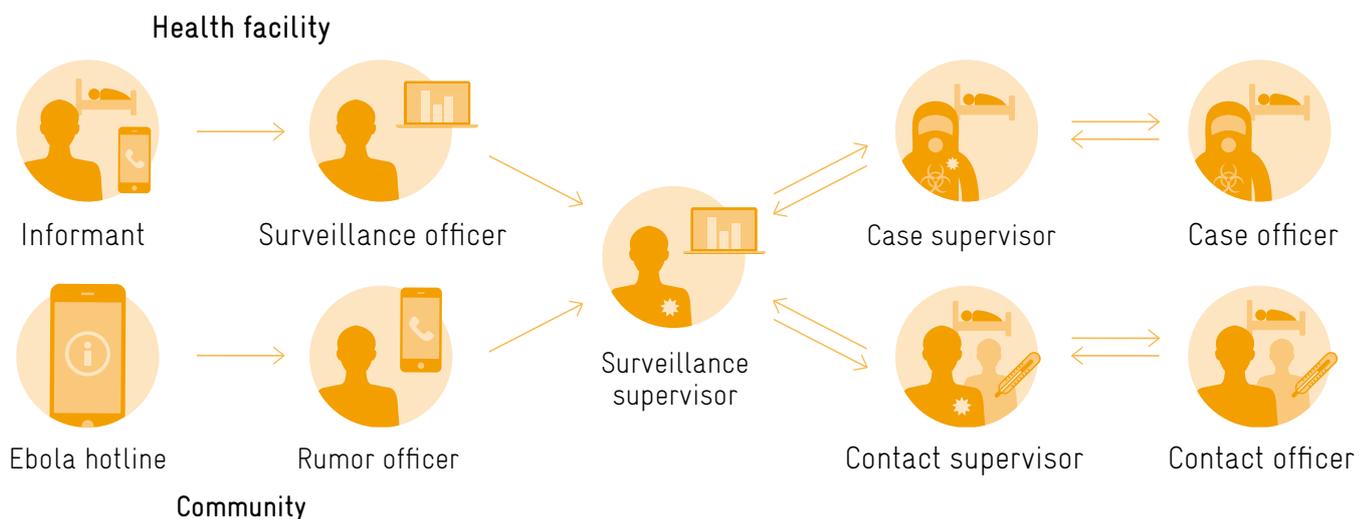


Figure 2.b. Tackling disease outbreaks in West Africa: The SORMAS computer-assisted network for real-time interaction.⁸²

Frequently, real-time data initiatives rely on partnerships between organisations tackling development problems and technical partners that provide software and tools to collect, manage and visualise the data. For example, the health organisation PATH, in partnership with the Tableau Foundation,⁸³ works with national governments to tackle malaria using real-time data.⁸⁴ Magpi⁸⁵ provides development organisations, such as the World Health Organization or Oxfam, with information systems that enable them to handle real-time data in sectors such as education, elections, microfinance or water and sanitation.

79 unicefstories.org/category/real-time-information

80 ureport.in

81 sormas.org

82 Helmholtz Centre for Infection Research, <https://www.helmholtz-hzi.de/en/>

83 tableau.com/foundation/featured-projects

84 visualizenomalaria.org

85 magpi.com

Beyond data: a new approach to development

Conventionally, development programmes are designed with a linear approach, where a series of predefined activities, outputs and outcomes, together with a timeline and a series of indicators, are established upfront to guide the programme and its evaluation. Implementation of the programmes in the field is expected to conform to logical frameworks, with little space to react to changes from the surrounding environment.

The expected impacts of using real-time data in development programmes are related to increasing their capacity to respond to changes in their operating contexts and to learn from constant evaluations of the effectiveness of their actions. This learning and adaptation ranges from the more operational, aimed at optimising the efficiency of the existing activities, to the more strategic, which would involve changing the activities and adapting the processes that guide the programme. Real-time data is considered especially relevant in the context of health or humanitarian crises, where a lot of the activities depend on the timely distribution of updated information about the conditions and capacities on the ground.

Unlike traditional survey data, which is typically collected to evaluate projects at yearly or bi-annual intervals, **the purpose of real-time data is to continuously collect data to make timely informed decisions**, which is especially important in sectors like health, agriculture, infrastructure management, or the monitoring of markets and environment. Real-time data could give development actors the means to uncover anomalies, respond to issues as they arise, improve internal coordination, optimise resource allocation, react to citizen feedback and anticipate trends and future events. Rather than solely measure success, real-time data facilitates adaptation and corrective actions within the lifespan of initiatives.

Box 2.g. Seven uses of real-time data for decision-making⁸⁶

- 1 **Responding quickly in emergencies** – Real-time data has been used to respond to humanitarian and health crises,⁸⁷ and to curb election violence. This is especially important during natural disasters in which crowdsourced digital maps and updated imagery from satellites and drones have helped to identify vulnerable and hard-hit areas.⁸⁸
- 2 **Optimising resource allocation and efficiency** – Real-time data can help to ensure that resources are being deployed where they are needed. UNICEF’s mTrac,⁸⁹ for example, uses real-time data to alert officials of stock-outs of medicine and vaccines, thus helping health clinics to stay adequately stocked.
- 3 **Managing public perceptions** – Monitoring of social networks and other real-time data sources can help to understand what people are thinking and, for example, react to negative rumours and misinformation.⁹⁰ The Kenyan initiative ‘Una Hakika’⁹¹ prevents conflict escalation by quickly collecting

86 Ramalingam, B. et al. (forthcoming) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

87 RIA (2014), Leveraging Anonymized Telecom Data to Fight the Ebola Outbreak, RIA, realimpactanalytics.com/content/data-cases/ebola-case-fighting-ebola-with-telecom-data-201411.pdf

88 Soesilo, D.; Meier, P.; Lessard-Fontaine, A.; du Plessis, J. and Stuhlberg, C. (2016) Drones in Humanitarian Action – A Guide to the Use of Airborne Systems in Humanitarian Crises, Geneva: FSD, drones.fsd.ch/en/drones-in-humanitarian-action

89 mtrac.ug

90 dw.com/en/using-social-media-for-good-during-ghanas-elections/a-36646065

91 firstdraftnews.com/how-una-hakika-helped-slow-the-spread-of-dangerous-rumors-in-kenya

rumours, verifying them and reporting back to the affected communities about whether they are true or not.

- 4 **Monitor and react to human behaviours and trends** – Online browsing history is used in real time by the private sector to personalise sales and adverts. In the development sector, target populations are less likely to be online. However, call detail records from cell phone usage can be used to monitor and react to human movements to, for example, more effectively contain the spread of diseases.⁹²
- 5 **Understand and efficiently manage whole systems** – Integrating a series of data sources, real-time data is used in fisheries to combat overfishing and by-catch rates. Thanks to fine-scale, real-time assessments of changing conditions, dynamic closures can be used, which affect smaller portions of the ocean for shorter periods of time.⁹³
- 6 **Develop and drive new business models** – Real-time data have spurred new business models to extend services to previously underserved communities. For example, M-Kopa's energy systems⁹⁴ can be remotely disabled in the case of non-payment, allowing the company to securely provide off-the-grid solar electricity to previously excluded people.
- 7 **Anticipate trends and build resilience to future events** – Data collected in real time can feed predictive analytics that aim to anticipate risks and trigger preventive actions. For example, extant data from public administrations, once digitised, can be combined with real-time data and used to predict abuse cases and improve social services interventions to prevent them.⁹⁵

Barriers to real-time data use by decision-makers

There is still little evidence that real-time data leads to adaptive decision-making, as there are many barriers in getting data taken up or used by decision-makers.⁹⁶ Some of the **hindrances include competing data sources, distrust in the quality of the data or in those inputting the data, lack of awareness of the data, data not adequately turned into information**, lack of visualisation, and information that is not tailored to the end user's needs.⁹⁷ Moreover, it is important to take into consideration the political economy of decision-making to understand why available data is not leveraged to improve decision-making processes. In many instances, personal convictions, interests and power asymmetries can motivate decisions more than evidence.

An actual demand for data from decision-makers is as important as the supply of relevant real-time data, to make sure that it is integrated into structures and processes: getting the data into the right format to the right people at the right time makes the data most impactful, rather than the pure speed or quality of

Getting the data into the right format to the right people at the right time makes the data most impactful, rather than the pure speed or quality of the data.

92 Ali, A., Qadir, J., Rasool, R. ur, Sathiaseelan, A., Zwitter, A. and Crowcroft, J. (2016) Big Data for Development: Applications and Techniques, Big Data Analytics 1: 2

93 Dunn, D.C., Maxwell, S.M., Boustany, A.M. and Halpin, P.N. (2016) Dynamic Ocean Management Increases the Efficiency and Efficacy of Fisheries Management, Proceedings of the National Academy of Sciences 113.3: 668–673

94 m-kopa.com

95 Packard, T. (2016) Literature Review: Predictive Analytics in Human Services, San Diego: SACHS

96 Lucas, H.; Batchelor, S. and Berdou, E. (2013) Real Time Monitoring and the New Information Technologies, IDS Bulletin 44.2: 31–39

97 Ramalingam, B. et al. (2017) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

the data.⁹⁸ For example, sharing relevant information with responsible actors in the form of visualisations at key moments before decisions need to be made increases the probability that the data will be acted upon, rather than sharing continuous data in raw format.

Box 2.h. Using real-time data on birth registration in Tanzania to improve policy-making

Tanzania has the second-lowest level of birth registrations in the eastern and southern Africa region at 13 per cent. The least likely children to be registered are often the most vulnerable, as birth registration is especially rare for rural children and children from poor families. **Obtaining a birth certificate can be costly and cumbersome:** parents are required to travel to the district headquarters, which could be over 100 km away, where they are required to pay a fee; and it may take up to five days to complete the paper-based registration process.

In 2012, UNICEF and the Tanzanian government piloted a mobile-based real-time birth registration information system that simplifies, decentralises and speeds up the process, while making registration less cumbersome for parents. Together with other interventions around maternal health supported by Germany development cooperation, **the programme has contributed to improve health and survival rates for newborn babies.**^{99,100} The system allows health workers to register children on the behalf of parents free of charge at local government offices, ward offices, hospitals, clinics and health dispensaries and collect a birth certificate near home.

Data is transmitted to a central system instantaneously and aggregated at the village, ward, district, regional and national level to improve policy-making. Dashboards help decision-makers identify anomalies and underperforming areas. **The data has been used in village-level budget and planning discussions and there is potential to use the data for national long-term planning,** by anticipating future needs in specific areas based on birth rates. Previously, census data was only available once every ten years.

However, the technology alone did not induce demand. In order to increase the incentive for birth registration, some services have been linked to birth certificates including passports, access to higher education and voting. Moreover, targeted public awareness campaigns were launched in several regions to enable parents to understand the importance of birth registration and welcome them to register all children. Health centres were incentivised to improve their registration rates through a friendly competition. The system's pilot showed that it was most used where there was sufficient political buy-in.¹⁰¹

98 Ramalingam, B. et al. (2017) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

99 Grimm, S.; Evans, R. and Sroka, V. (2016) Better Services for Better Health: Tanzanian-German Programme to Support Health Phase IV (2013–2016), Bonn: GIZ, health.bmz.de/where_we_work/countries/Programme_Phase_Report_Webversion_160610.pdf

100 health.bmz.de/where_we_work/countries/tanzania

101 Ramalingam, B. et al. (forthcoming) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

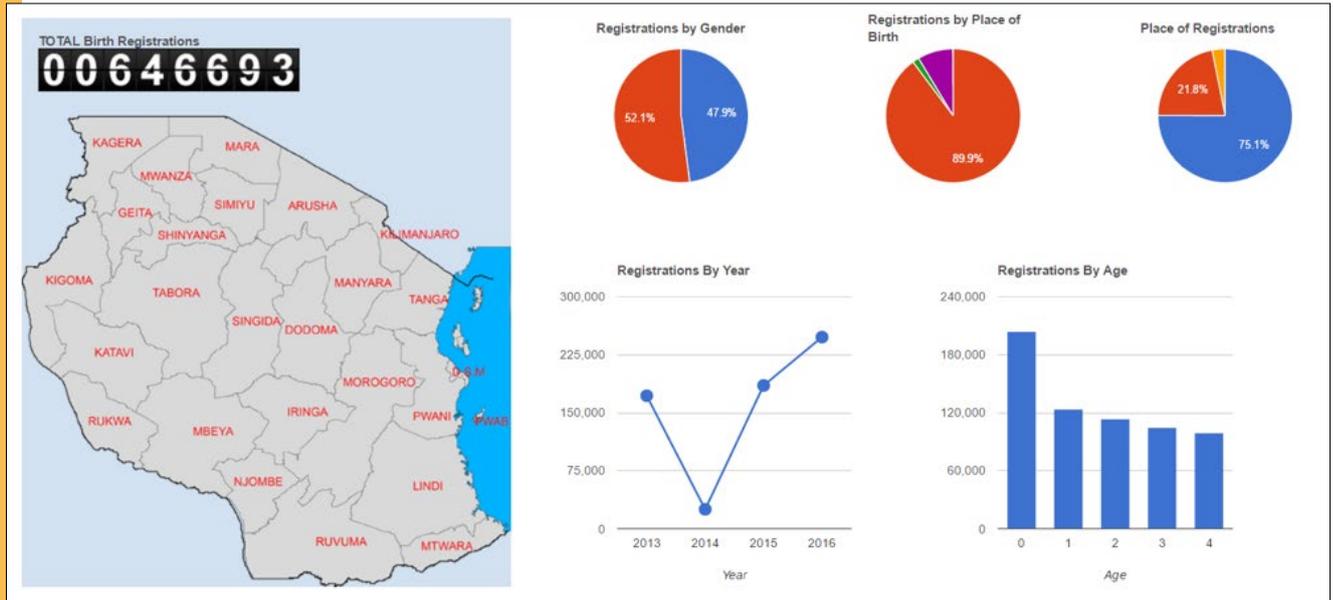


Figure 2.i. UNICEF and the Tanzanian government use real-time data for birth registration in Tanzania to improve policy-making and planning.

2.5. Summary



Big data

Basic definition

High-volume, high-variety and high-velocity data that can contain both structured datasets and flows of unstructured data, often created and collected by the private sector.

Value

Quickly analyse situations, diagnose and understand underlying problems, predict likely scenarios and even prescribe the most beneficial potential action.



Open data

Data that can be freely used, reused, and redistributed by anyone.

Strengthen (government) transparency, accountability and responsiveness, and spur social and business innovation.



Citizen-generated data

Data actively produced by citizens with a specific purpose in mind.

Understand on-the-ground conditions, document and map incidents and sentiment, especially when data from other sources is unavailable, inaccessible or of poor quality.



Real-time data

Data that is gathered and analysed quickly enough to make timely decisions.

Quickly make data-based decisions, for example in health or humanitarian crises.

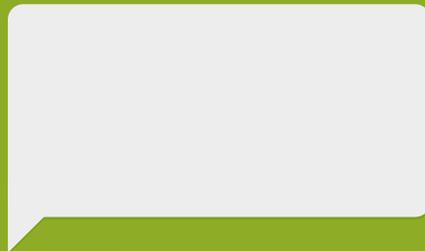
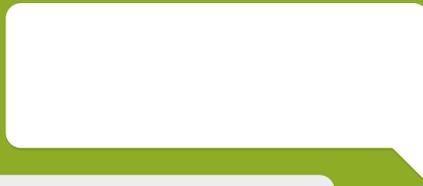
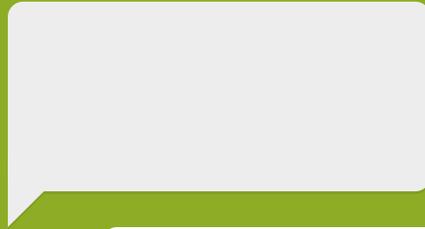
3

Trends

Our research has uncovered six main trends impacting the digital data for development space.¹⁰² For each of the trends we have analysed its main characteristics, the drivers behind them, and their potential positive and negative implications for the development field.

- 1 More data and new data sources create opportunities
- 2 New technologies improve data analysis
- 3 New types of actors, partnerships and ways of working emerge
- 4 Balancing access to personal data and privacy
- 5 The value of contextual, granular information
- 6 Information inequality is likely to persist

¹⁰² See the Annex for details on the methodology.



3.1. More data and new data sources create opportunities

The quantity of data available is increasing at an exponential rate, a trend that is likely to continue in years to come. If adopted and adapted effectively by the development sector, new data sources and technologies could help overcome data gaps and allow development practitioners to gain a better understanding of contexts before, during and after the implementation of interventions. Moreover, seeing the potential of data to provide valuable insights for development, many governments have begun to open their data for use by non-government actors.

The quantity of data available is increasing at an exponential rate, a trend that is likely to continue in the years to come.

New technologies as drivers

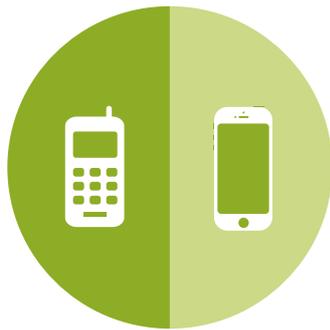
This explosion of data is driven partially by exponentially increasing computing power, lower costs and increased speed of broadband connections, but also by novel data sources such as sensors used by the Internet of Things, drones, nano-satellites and mobile phones.

The amount of computing power available per dollar has been growing exponentially for the last five decades. This has led to smaller devices with greater capabilities that are increasingly affordable. However, affordability of capable devices is only one piece of the puzzle. They must also be connected to create relevant data. Mobile broadband has helped make connectivity affordable for more people. The average price of mobile broadband is now half the price of fixed broadband globally, and a third in least developed countries. Moreover, as prices are decreasing, speeds are increasing, which means that mobile broadband users are getting faster speeds for less.

Increasing affordability of devices and connectivity has spurred the near ubiquity of mobile phones and mobile networks. As of the end of 2016, approximately 95 per cent of the global population is now living in places covered by at least a 2G mobile signal.¹⁰³ By 2021, 8.3 billion mobile phones are expected to be in use with over 50 per cent being smartphones.¹⁰⁴

103 ITU (2016) ICT Facts and Figures 2016, ITU, itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf

104 Cisco (2017) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, Cisco, cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf



By 2021:

8.3 billion

mobile phones

50%

being smartphones

Figure 3.a. By 2021, 8.3 billion mobile phones are expected to be in use with over 50 per cent being smartphones.¹⁰²

New data sources

Computers and mobile phones are not the only gadgets connected to each other. The physical world is increasingly becoming represented digitally. Objects equipped with sensors are being integrated into the **Internet of Things** at an accelerating rate. Gartner Inc. predicts that the Internet of Things will increase from 8.4 billion objects in 2017 to 20 billion in 2020.¹⁰⁵ These devices will generate even more data, which can be used to garner further insights.

Moreover, advancements in other technologies and fields allow for the integration of those technologies into the Internet of Things, providing us with new sources of data.¹⁰⁶ Some of these technologies are touted for their ability to democratise data collection, making them ideal methods to collect data by citizens. For example, **Unmanned Aerial Vehicles** (also known as drones) equipped with cameras can capture high-quality aerial images for mapping and can be bought off the shelf starting from 100 Euros. There are also **drones** small enough to fit in a backpack that can float and swim in water and allow us to collect data and samples from the ocean more efficiently. **Low-cost nano-satellites**, which constituted 48 per cent of space launches in 2015, have made near real-time satellite imagery affordable for all kinds of projects.¹⁰⁷

Along with an increased use of data gathered from sensors in objects, citizens are also increasingly becoming sensors through their mobile phones and other connected devices. This is often done voluntarily, where citizens consciously capture and share data on platforms. One such case is that of the platform ‘Una Hakika?’ – Swahili for ‘Are you sure?’ – in Kenya, which allows people to report potentially harmful rumours, which are then quickly validated to mitigate the escalation of ethnic conflicts based on misinformation. At other times, citizens crowdsource data unconsciously and indirectly, for instance when a Twitter user signals in a tweet that their child may have an infectious disease, without being aware that the tweet might be used to uncover an epidemic and inform decision-making. Similarly, data generated automatically from phone calls and texts can be used to track human movement.¹⁰⁸

Citizens are increasingly becoming sensors through their mobile phones and other connected devices.

¹⁰⁵ gartner.com/newsroom/id/3598917

¹⁰⁶ BMZ (2015) Using Sensors for Good: How the Internet of Things Can Improve Lives, Bonn: BMZ, better-place-lab.org/wp-content/uploads/giz2016-bmz-digital-IKT-studie.pdf

¹⁰⁷ Space Foundation (2016) The Space Report 2016 – The Authoritative Guide to Global Space Activity, Space Foundation, spacefoundation.org/sites/default/files/downloads/The_Space_Report_2016_OVERVIEW.pdf

¹⁰⁸ Wesolowski, A.; Buckee, C.O.; Pindolia, D.K.; Eagle, N.; Smith, D.L.; Garcia, A.J. and Tatem, A.J. (2013) The Use of Census Migration Data to Approximate Human Movement Patterns across Temporal Scales, PLoS One 8.1

More data and new data sources allow us to measure and uncover things more accurately and faster than before.

Advances in cryptography, encryption, peer-to-peer networks, and digital currency, among others, have brought about new ways of keeping records.

Distributed Ledger Technology (also referred to as blockchain) could transform the way we keep records while making more data available to the public. Transactions on the blockchain are transparent, encrypted and immutable. Assets can be tokenised and represented on the ledger to prove ownership at any given time. The blockchain is updated approximately every ten minutes and everyone with an internet connection could have an up-to-date copy of all transactions ever made. These properties make blockchain a potential contributor to the growth in publicly available data.

This is by no means an exhaustive list of data sources available. As the Internet of Things becomes the **Internet of Everything** and advancements are made in other sciences, potential data sources will become virtually endless.

The opportunities and challenges

The increasing quantities of data and novel data sources available are likely to have both positive and negative implications for development.

On the one hand, more data and data sources provide us with increasing opportunities to understand social and natural phenomena. They allow us to measure and uncover things more accurately and faster than before. Moreover, these data sources provide more granular pictures allowing development practitioners to pinpoint exactly where efforts should be focused. Some of these new sources provide timely data that, if integrated into programmes the right way, could lead to a shift from linear to adaptive management, helping facilitate better and more iterative design of solutions from the end user and back. Moreover, if opportunities to gather data from the crowd are harnessed, data used by decision-makers may become more representative, thus increasing the likelihood that solutions fit the needs of intended beneficiaries. Moreover, availability and public accessibility of this kind of data provide citizens with a tool to hold governments and service providers accountable.

On the other hand, there are also potential negative implications requiring action to mitigate. Although increased data sources and new ways to gather data can make data samples more representative, we still do not live in a world where 'digital' data can be considered representative. Because digital data generation requires digital tools, digital inequalities need to be addressed before the most vulnerable are adequately included in digital data.

Digital data over-represents those who are connected. At the most basic level, people without mobile phones will not be included in call detail records used to track human movement and the spread of diseases. Furthermore, deep qualitative insights are more likely to arise from more advanced uses of mobile phones requiring smartphones and broadband connections, such as citizens expressing themselves through social media posts. Poor and elderly people, ethnic minorities, women and other disadvantaged groups are likely to be under-represented in digital data when compared to 'traditional' data collection strategies, which often tackle inequality with counter strategies.

Box 3.a. More and better data does not guarantee use and impact

Although there is an increasing amount of data available, recent experiences have evidenced how data availability does not necessarily lead to its use or incentivises action.¹⁰⁹ The use of data is as important as its supply. Since data is often adapted into pre-existing decision-making timings and structures, getting the data into the right format to the right people at the right time makes the data most impactful, rather than its speed, quality and availability. Data initiatives increasingly incorporate specific actions to trigger demand and usage.

When data is taken up, decision-makers must ensure that solutions to problems that can be datafied are not prioritised at the expense of problems that do not show up in the existing data. Moreover, **there is a risk to invest in the collection and generation of data that will never be used or acted upon.**

Unused data is already an issue in the private sector where in many instances a grand share of the data collected is never considered.

3.2. New technologies improve data analysis

Traditional data, including censuses, have provided us with static visualisations of data that were only occasionally updated. Today, increases in the amount of data available, the speed at which data is collected, and new data sources, combined with digital analytical tools that automatically analyse the data and feed it into visualisation tools, are providing new ways to make sense of development problems and opportunities.

Although many of our interviewees stated that the sector still suffers from digital data ‘pilotitis’, development actors have also extended the limits of what can be done with digital data. This is especially true for crowdsourced data initiatives, which have contributed to the development of open source software solutions capable of being quickly replicated, adapted and scaled to facilitate crowdsourcing initiatives. Development organisations have also helped build and support open source mobile-based real-time data systems that facilitate two-way communication and real-time data insights.

New digital analytical tools that automatically analyse data and feed it into visualisation tools, are providing new ways to make sense of development problems and opportunities.

109 Ramalingam, B. et al. (forthcoming) Bridging Real-Time Data and Adaptive Management in International Development – Final Case Study Report, Brighton: IDS

Box 3.b. How algorithms and big data can aid in addressing development issues

According to Flyverbom et al.,¹¹⁰ algorithms and large quantities of digital data increase our capacity to understand and tackle development issues in four ways.

Firstly, these data concepts bring with them new techniques of visualisation. Secondly, algorithms are able to sift through development data much faster than was possible with previous methodologies. Thirdly, the ability to visualise things differently and quicker contributes to new ways of ‘rationalising knowledge claims that underlie development issues’. Finally, using digital data results in shifting power from topical experts to data scientists who are trained to see things differently.

Artificial intelligence has come of age

Artificial intelligence is becoming an important tool in an increasing number of sectors. New technology has made it easier to garner insights from troves of data at speeds previously unimaginable, offering the opportunity to see things we were never able to see, to see them faster, to open things up to be seen and scrutinised by more people, and to democratise the way they are seen and what gets seen. These new ways of analysing and visualising problems allow us to understand issues in new ways and thus respond to them differently.

Algorithms make it increasingly possible to uncover potential issues, predict when they are likely to happen, estimate their consequences, and prescribe solutions based on causal inferences.¹¹¹

The development sector is now using algorithms that monitor social media data to react to citizen sentiment.

Real-world digital data solutions and applications are now being integrated into organisations and having impacts on their performance. Much of the early algorithmic solutions and applications have been developed by the private sector. Although the development sector lags behind in the adoption and use of these tools, the sector benefits from being able to make use of existing and maturing digital data infrastructure, software, and methodologies. For example, **the development sector is now using algorithms that monitor social media data to react to citizen sentiment.** However, sentiment analysis has been used by private sector players for over five years. In 2011, the hamburger and fast food chain McDonald’s was already monitoring social media to maintain its reputation by refuting false rumours in real time.¹¹²

A shift in how data is analysed

The advent of digital data analytical tools has shifted the way we do data analysis and thus how we make sense of things. In the past, it was common to create new models based on pre-existing historical data, to get as much out of the data as possible. Today, there is so much recent data that the focal point has shifted to mining the data to figure out what it tells you. However, this also raises the **risk of ‘garbage in, garbage out’ and spurious correlations being over-emphasised.** To safeguard against these risks, insights garnered through

110 Flyverbom, M.; Madsen, A.K. and Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities, *The Information Society: An International Journal* 33.1: 35-42

111 Letouzé, E. and Sangokoya, D. (2015) Leveraging Algorithms for Positive Disruption: On Data, Democracy, Society and Statistics, Data-Pop Alliance, datapopalliance.org/wp-content/uploads/2016/03/DataPopAlliance_LeveragingAlgorithms.pdf

112 Divol, R.; Edelman, D. and Sarrazin, H. (2012) Demystifying Social Media, *McKinsey Quarterly*, April, mckinsey.com/business-functions/marketing-and-sales/our-insights/demystifying-social-media

data science are being combined with methods from behavioural sciences. The idea here is that after knowing ‘what’ is happening, the ‘why’ can be underpinned by a human behaviour analysis.

Events have driven adoption of new tools and approaches

Certain types of events have driven the adoption and usage of new digital data tools. Interviewees have identified **emergencies – including natural disasters and epidemics – as a main driver of real-time data**. In these situations, the status quo has changed dramatically, making it senseless to rely on previous knowledge. The 2014 Intergovernmental Panel on Climate Change report found that the poor tend to be the most vulnerable to emergencies and are also the least able to cope with them.¹¹³ Having real-time data on the situation of vulnerable people is vital for constructing effective responses. The shared need to respond to natural disasters between humanitarian aid actors seeking to help the vulnerable and private sector actors in pursuit of maintaining a business-enabling environment has led to data-sharing collaborations that speed up responses and generates co-benefits (see Section 3.3. New types of actors and partnerships emerge).

The pros and cons of using sophisticated technology

The increased use of new types of technologies has both potentially positive and negative implications on the development sector. For example, traditional survey data has provided us with static visualisations of data which were updated infrequently. Because digital data is collected more continuously, information can be disseminated through more fluid visualisation techniques such as dashboards providing panoptic views of operations, word clouds capturing citizen sentiment and exposure to shocks, and maps that capture movements of people and the intensity of movement. Being able to visualise problems in new ways allows development organisations to tackle them using novel methods.

Also, digital data analytic tools can be used to induce agile adaptive programming of development interventions. They provide development organisations with the ability to describe and understand situations better, predict what will happen, and prescribe improved actions based on data. This enables experimental and agile development programming approaches in which programmes adapt and react to the dynamic and complex environments they operate in, rather than hold on to assumptions based on outdated data.

However, there are potential negative effects that require attention. For instance, many of the algorithms that help provide new ways of making sense of data were developed in the private sector. Because private companies seek to protect their intellectual property, the code that underpins how algorithms process data, analyse it and come to conclusions remains out of sight and non-transparent. **A lack of algorithmic transparency can be especially worrisome when poorly constructed algorithms are underpinned by biased processes and data, which lead to biased results and decisions.** Actors in the digital data development space are well aware of potential algorithmic biases and are working on ideas to manage them, such as the Open Algorithms project supported by Data-Pop Alliance, the World Economic Forum and others.¹¹⁴

Increased use of new types of technologies has both potentially positive and negative implications on the development sector.

113 IPCC (2014) Climate Change 2014: Synthesis Report, p. 151, Geneva: IPCC

114 opalproject.org/closer-look

3.3. New types of actors, partnerships and ways of working emerge

Using data for development is complex. To take advantage of the growing number of means to collect and aggregate data while addressing the intricacies of development challenges, we need new ways of working (see Section 3.2. New technologies improve data analysis). This prompts new actors to come to the fore and hasten new partnerships and collaborations to emerge. In recent years, the emergence of new actors and partnerships in the data for development space has been observed, and this is likely to continue in the immediate future.

New and old actors collaborate

With the resurgence of data-based approaches to undertaking development, and a high regard for aid effectiveness in development aid delivery, a set of actors, referred to by experts as ‘development mutants’ have emerged.¹¹⁵ **These actors are rubbing elbows with traditional players and forging alliances to make aid delivery more effective, efficient, and sustainable**, and launching innovative ways of programme implementation. In recent years, we have seen players like innovation and data laboratories that provide new technologies for monitoring diseases or environmental quality, among others, but also data companies that collect and aggregate data from consumers, entering the development sector.

In recent years, the emergence of new actors and partnerships in the data for development space has been observed.

While we still have to see how these actors and initiatives create public good, it is important to recognise that these new players emerge with explicit development intentions. They are often able to produce evidence from large datasets, but often lack the capacity to verify their findings on the ground, where they seldomly deploy staff. A collaboration between data companies, ‘development mutants’ and traditional development cooperation actors could be beneficial, as also stated in the ICT4Refugees study commissioned by the German Federal Ministry for Economic Cooperation and Development (BMZ).¹¹⁶

Complex problems demand new ways of working

This trend is driven by the fact that the problems that development actors today face are becoming increasingly complex, or what experts would call ‘wicked’.¹¹⁷ They are difficult to define, have many contrasting explanations and require a new set of ways of ‘solving’ that is based on a better understanding of context. Of course in this case **data is important and, increasingly, data from different sources and actors is critical, with all the necessary investments required.**¹¹⁸

The trend is also fueled by a growing general consensus that solving the world’s persistent problems requires the participation and cooperation from different

115 diytoolkit.org/profiling-the-development-mutants

116 regasus.de/online/datastore?epk=74D5roYc&file=image_8_en

117 Ramalingam, B; Laric, M. and Primrose, J. (2014) From Best Practice to Best Fit: Understanding and Navigating Wicked Problems in International Development, Working Paper, ODI, odi.org/publications/8571-complexity-wiked-problems-tools-ramalingam-dfid

118 Espey, J. (2015) Data for Development – A Needs Assessment for SDG Monitoring and Statistical Capacity Development, Sustainable Development Solutions Network, unsdsn.org/wp-content/uploads/2015/04/Data-for-Development-Full-Report.pdf

sectors. From the OECD¹¹⁹ to the World Economic Forum,¹²⁰ the call is for increased collaboration of different actors so that new solutions can potentially emerge to accelerate the world's progress towards human development. DFID, for example, partnered with the global design team IDEO, to launch the Amplify programme,¹²¹ which seeks to solve the world's greatest challenges not through bureaucracy and competition but via collaboration and creativity. Several initiatives have started to spring from different parts of the globe and many open calls for solutions have been made so far in order to find new ways of doing things and at the same time increase participation of different actors in pursuing development (e.g. MAVC,¹²² UNLEASH¹²³).

Data for development brings new forms of partnerships and networks to life

The combination of specialised expertise and resources in the data for development field opens up new ways of addressing complex problems. It also challenges the current structure of the development enterprise as more stakeholders become involved in using data to achieve development outcomes. **Joint platform projects make possible productive conversations among different stakeholders** who are habitually working in isolation, with limited opportunities for data and information exchange; and this has created opportunities for future collaborative work. For example, PetaJakarta, a real-time flood-mapping application in Jakarta, is the result of a collaboration among Twitter, the Jakarta city government and University of Wollongong Australia. This collaboration opened up new forms of partnerships between the city government and other stakeholders.

However, the need for new partnerships also raises important issues. For example, one of the experts interviewed pointed to the possibility of blurring the lines between public and private initiatives that can potentially affect development outcomes. It also has the tendency to cause the unintended effect of creating silos, potentially excluding others not within these circles.

Nevertheless, new actors and new partnerships are considered intrinsically desirable, especially in a time and space where traditional ways of thinking and doing have not always delivered the desired benefits for the world's poor and marginalised. In the next years, new actors will continue to emerge and new patterns of institutional arrangements will evolve. This will change how development aid is delivered and how local actors participated in this process.

In data for development two innovative partnership arrangements have become fashionable recently: 'data philanthropy' and 'data collaboratives'. **'Data philanthropy'¹²⁴ is a term used to describe corporations that share data they collect, curate, and aggregate as part of their business operations for the public good.** For example, Waze, a 'community-based traffic and navigation app', provides transport and traffic data to city governments for better traffic management.¹²⁵ Similarly, Twitter makes their social media data accessible to the UN system to address development challenges.

The combination of specialised expertise and resources in the data for development field opens up new ways of addressing complex problems.

In data for development two partnership arrangements have become fashionable: 'data philanthropy' and 'data collaboratives'.

119 OECD (2014) Effective Development Co-operation: An Important Enabler in a Post-2015 Global Development Framework, Post-2015 Reflection series, Paris: OECD, oecd.org/dac/FINAL_POST-2015_Effective_Development_Co-operation.pdf

120 weforum.org/agenda/2017/01/realising-the-potential-of-cross-sector-partnerships

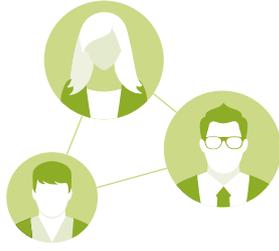
121 ideo.org/programs/amplify

122 makingallvoicescount.org

123 unleash.org

124 unglobalpulse.org/data-philanthropy-where-are-we-now

125 waze.com



Data Cooperatives or Pooling

Corporations and other important dataholders group together to create ‘data pools’ with shared data resources.



Prizes and Challenges

Corporations make data available to qualified applicants who compete to develop new apps or discover innovative uses for the data.



Research Partnerships

Corporations share data with universities and other academic organisations giving researchers access to consumer datasets and other sources of data to analyse social trends.



Intelligence Products

Shared (often aggregated) corporate data is used to build a tool, dashboard, report, app or another technical device to support a public or humanitarian objective.



Application Programming Interfaces (APIs)

APIs allow developers and others to access data for testing, product development, and data analytics.



Trusted Intermediary

Corporations share data with a limited number of known partners. Companies generally share data with these entities for data analysis and modelling, as well as other value chain activities.

Figure 3.b. Six types of data collaboratives.¹²⁶

Data collaboratives, on the other hand, refer to a group of actors who share data assets and combine their expertise and/or tools to solve particular problems of public interest. The GovLab has curated a list of data collaboratives¹²⁷ around the world that includes a data-sharing platform on agriculture in Sub-Saharan Africa,¹²⁸ and a transport bus map in Kenya.¹²⁹

In the last three years, we have also seen new networks of actors established at the global scale – the Global Partnership for Sustainable Development Data, Open Data for Development, Big Data Alliance, Open Contracting Partnership, and International Network of Crisis Mappers, among others. At the same time, we have witnessed the space being reclaimed by traditional actors such as UNICEF, and the emergence of new organisations within traditional institutions like UN Global Pulse.

¹²⁶ datacollaboratives.org/static/files/data-collaboratives-intro.pdf

¹²⁷ datacollaboratives.org/explorer.html

¹²⁸ esoko.com

¹²⁹ dataimpacts.org/project/mapping-commuters-streamlines-transit

Box 3.c. Building internal capacity versus collaborating

Expertise niches emerge as new technologies are increasingly becoming sophisticated and accessible. Data scientists are mining myriad social media data; non-profit organisations are developing disaster-preparedness maps; security companies are using drones to monitor conflict-vulnerable areas, and so on. Organisations either intentionally or consequentially have built a set of expertise that never before was available in the development sector. Whether or not building separate organisational capacities on data science is desirable remains a question because analytical capacity can also be strengthened through collaboration. Also development actors may opt to find strategic, light, smart and frugal approaches to data science that can be applied easily in partner countries with scarce resources.

3.4. Balancing access to personal data and privacy

Ethical considerations, in particular protection of privacy, are likely to receive growing attention in the coming years. While the potential negative implications of the use of data in and for development have been on the agenda of expert organisations for some time, there has been a **recent increase in investment in the guidelines, frameworks and technology to better protect personal data** in order to prevent surveillance, exclusion and stigmatisation of individuals and groups, among others.

Ethical considerations, in particular protection of privacy, are likely to receive growing attention in the coming years.

New partnerships for data privacy

New networks, initiatives and partnerships like the Responsible Data Forum,¹³⁰ the International Data Responsibility Group¹³¹ and the UN Global Pulse Privacy Advisory Group¹³² have emerged to contribute to a better understanding and protection of privacy through research and the development, testing and mainstreaming of various data privacy guidelines,¹³³ responsible data guidelines,^{134,135} or risk assessment tools.¹³⁶

While much of this work has been focused on big data for development, given the key role that personal data plays in this field, there are an **increasing number of organisations looking into privacy issues around open government data** such as Stiftung Neue Verantwortung¹³⁷ and the Berkman Klein Center.¹³⁸ This is especially important where there is a need to balance the privacy of public

130 responsibledata.io

131 responsible-data.org

132 unglobalpulse.org/data-privacy-advisory-group

133 UNDG (2017) Big Data for Achievement of the 2030 Agenda: Data Privacy, Ethics and Protection, United Nations Development Group, undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf

134 Hastie, R. and O'Donnell, A. (2017) Responsible Data Management, Oxford: Oxfam, policy-practice.oxfam.org.uk/our-approach/toolkits-and-guidelines/responsible-data-management

135 Raffree, L. (2016) How to Develop and Implement Responsible Data Policies, ICT Works, blog, 21 November, ictworks.org/2016/11/21/how-to-develop-and-implement-responsible-data-policies

136 unglobalpulse.org/privacy/tools

137 Manske, J. and Knobloch, T. (2017) Leitfaden für Datenschutz bei Open Data, Stiftung Neue Verantwortung, stiftung-nv.de/de/publikation/leitfaden-für-datenschutz-bei-open-data

138 cyber.harvard.edu/publications/2017/02/opendataprivacyplaybook

There is no common and agreed set of principles on data privacy, ethics and protection.

The future conversation might revolve around weighing the benefits against the risks of providing access to data.

officials with the public good as is the case, for example, with the fight against corruption through open government data.¹³⁹

Despite the many efforts, there is no common and agreed set of principles on data privacy, ethics and protection and the current regulatory landscape is fragmented which complicates efforts to develop standardised and scalable approaches to risk management and data access. At the same time, current techniques for anonymisation and de-identification of datasets have proved not to provide adequate safeguards to protecting privacy.¹⁴⁰ The two closely interrelated types of challenges, legal and technical, are therefore likely to be at the centre of discussions in the coming years.

Tech, legal or both: the search for adequate privacy protection

The quest for adequate legal foundations to guide and govern access, use and dissemination of data will keep researchers, policy-makers and practitioners busy for the years to come. Since **legal frameworks that govern personal data protection are non-existent, outdated or poorly implemented in most developing countries**,¹⁴¹ the challenge will be to either invest in the development or improvement of such regulatory frameworks or come up with alternative (legal) instruments. While the various responsible data and data privacy guidelines of organisations can inform and guide their individual efforts, a coordinated and inclusive approach to rule setting will be needed that involves a broad range of stakeholders to identify context-appropriate provisions.

Closely linked to the legal debate, the search for and development of technical solutions to enable a sufficient level of aggregation or anonymisation of personal or personally identifiable data will continue. However, as indicated above, given the growing availability of data from different sources on the same sector, geography or population, **experts increasingly question the possibility of fully anonymising datasets**.

Weighing the risks and benefits of data access

The future conversation might be less focused on finding ways to ensure personal privacy through sophisticated technology. Instead, it might revolve around weighing the benefits against the risks of providing access to data – preferential, public or otherwise – to those who could make use of it to advance the public good. This applies in particular to data from citizens derived from mobile phone usage, social media or consumption, commonly used in big data and real-time data initiatives; however, open data advocates will also need to determine what types of government data, for example on public officials and politically exposed persons, can be made publicly accessible in open data repositories.

More tech, more data, more potential privacy issues

The trend to look more carefully into the potential negative implications is **driven by the growing spread and use of technology, such as mobile phones, computers and wearables**. More personal devices means more personal data is being produced, which provides more opportunities to use it, which in turn results in more possible unintended negative impacts.

139 opendatacharter.net/anticorruption

140 IDRG (2017) IDRG Annual Report – People First in a Digital Age, The Hague: International Data Responsibility Group, responsible-data.org/uploads/1/5/6/9/15692298/idrg_report_2017__29052017_.pdf

141 unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=1468

Early investments into digital data for development focused on proof of concept projects to test what types and sources of big data were more promising to advance development,¹⁴² advocacy for greater supply of openly available government data, and case study research into effective strategies and approaches to data use.¹⁴³ This focus has partly shifted to efforts targeted at understanding the unintended negative consequences of data for development initiatives.

Various documented cases have contributed to the growing realisation that **providing access to anonymised personal data can put privacy at risk.**¹⁴⁴ Criticism has been voiced about the sharing of anonymised and aggregated call detail records that contain sensitive personal data, including location and communication history. This happened, for example, during the Ebola crisis in 2014 in West Africa¹⁴⁵ and for research on mobility patterns in Côte d’Ivoire,¹⁴⁶ where legal frameworks were largely absent and mobile phone users were not asked for consent to have their data shared. Likewise, activists have criticised the promotion of national ID systems by bilateral development agencies in developing countries that includes collection and storage of significant amounts of biometric and other personal data. The emergence of responsible data and privacy groups and networks has been both a result of such cases while also acting as a driver of conversations.

During the Ebola crisis in 2014 the release of sensitive personal data put privacy at risk.

The role of citizens in data for development: data farmed or data empowered?

The debate around responsible use of data has gained traction in data for development expert circles in recent years and will increasingly find its way into conversations of more traditional development organisations. It is difficult to predict the direction of travel of these developments. While there seems to be a growing consensus of the importance of putting greater emphasis on addressing privacy issues, there is a debate around just how much.

In emergency situations, some experts suggest, the harm of not using the data for the public good might outweigh the possible risks of compromising an individual’s privacy. One interviewee likened the situation to the dilemma faced in vaccination where the individual has to accept a small risk of being infected in order to protect the overall population from harmful and possibly deadly diseases.

There are at least two broad directions in which the trend could evolve over the next five years, with different implications for development organisations. The first sees the key players in the data for development domain realise that protecting individual privacy while using personal data to tackle development challenges is virtually impossible. In an extreme case, this realisation could result not only in a **negative scenario of citizens in developing countries being used as ‘data farms’** for business interests and/or enhanced government control, but also as sources of data for development initiatives that have no intent to give people agency through a fair, agreed and equitable exploitation of the data they produce.

142 unglobalpulse.org/projects

143 opendataresearch.org/emergingimpacts

144 Manske, J. and Knobloch, T. (2017) Leitfaden für Datenschutz bei Open Data, Stiftung Neue Verantwortung, stiftung-nv.de/de/publikation/leitfaden-für-datenschutz-bei-open-data

145 McDonald, S. (2016) Ebola: A Big Data Disaster, CIS Papers, Washington DC: The Centre for Internet and Society, <http://cis-india.org/papers/ebola-a-big-data-disaster>

146 [youtube.com/watch?v=FAoQHJH6l8Y](https://www.youtube.com/watch?v=FAoQHJH6l8Y)

At the same time, when it comes to openness of government data, **the protection of personal privacy might be used as an excuse to withhold relevant public sector data** which could be used by citizens to advocate for better public service provision, to hold governments accountable or to tackle corruption in the public sector.

The second, arguably more positive, direction could feature **an inclusive multi-stakeholder approach to identifying workable solutions to privacy challenges**. From mainstreaming organisational responsible data practices to institutionalising privacy guidelines and reworking entire regulatory frameworks, this work would need to be closely linked to investments on cyber security, digital rights and the protection of human rights online.¹⁴⁷

3.5. The value of contextual, granular information

There has been a recognition of the limitations of solely relying on quantitative data and the need to pay greater attention to contextual factors and local, granular information.

As several interviewees pointed out, there are only a few cases of effective and mature use and impacts. Many of the data for development projects are still considered proof of concepts, and many of the programmes are driven primarily by intent and promise rather than documented evidence.

The challenge to achieve sustained and meaningful results, along with the struggle experienced when attempting to replicate programmes, scale them, or even just sustain them over time, have led to widespread awareness among the data for development community of the need to rethink current approaches. More specifically, there has been a recognition of the limitations of solely relying on quantitative data and the need to pay greater attention to contextual factors and local, granular – often qualitative – information.

It is essential to take into account the institutional and personal contexts throughout the different phases of data for development initiatives. These go from the identification of the problems and asking the right questions,¹⁴⁸ to the selection and harvesting of good data, to carrying out proper analysis and correctly interpreting the results, to the effective communication and uptake of the insights, and even to the capacity on the ground to finally act upon the generated knowledge.

The limitations of data to explain social interactions

A key driver of this trend is the growing evidence that show the limitations of data in capturing and explaining social phenomena. Data related to social interactions always carries some bias, because of what and who is included in the datasets and, more importantly, because of what and who is left out. **Datasets are not just a reproduction of reality, but are affected by who gathered the data and how it was harvested.** Big data can exclude those who are unwilling or unable to contribute their digital traces, omitting those who, for example, do

147 See for example the digital rights work of organisations like Tactical Technology Collective (tacticaltech.org), Global Partners Digital (gp-digital.org) or the World Wide Web Foundation (webfoundation.org).

148 World Bank (2016) Big Data Innovation Challenge: Pioneering Approaches to Data-Driven Development, Washington DC: World Bank, documents.worldbank.org/curated/en/396861470905612761/pdf/107751-REVISED-PUBLIC-BigData-Publication-e-version-FINAL.pdf

not have a mobile phone.¹⁴⁹ Big data analytics and algorithmic decision-making tend to disregard the exclusionary politics that frequently underpin data collection processes, and could end up reinforcing historical inequalities.

Implications for data for development initiatives

The growing recognition of the importance of contextual and human factors, and of the need to integrate local and qualitative knowledge into data for development initiatives has multiple implications which can be summarised as affecting the what, how and why of development interventions.

Regarding the ‘what’ dimension, **the main transition is from initiatives that are technology-driven towards more sector- or problem-driven interventions.** Up to now, funding lines from several development agencies have focused on particular types of data, like open, big, or citizen-generated data, or the application of particular technologies, like artificial intelligence or the Internet of Things. However, the distinctions among data types are becoming less meaningful and more blurry as the various technologies are becoming more integrated.

Future data initiatives are likely to consider various data types as dictated by the contexts and by the problems that need to be tackled. For example, if working on data for health in a certain region, a holistic approach would aim to see where big data contributes more value, how patients’ surveys could add to rigour, reflect on the level of openness and engagement needed when handling the data, and undertake actions that increase data awareness among public workers and community organisations.

The impact of the trend on the ‘how’ of **data initiatives relates to a merging of qualitative and quantitative analysis methodologies and data sources**, where big quantitative datasets get integrated with ethnographic data. This mixed-methods approach allows for better identification of spurious correlations, obtaining more nuanced and actionable insights, and a critical evaluation of the wider contexts surrounding data initiatives. It would also help identify what ‘personalised data’ is more important to local actors: the part of data and insights which are relevant for a certain person because it is close to his or her context and needs, and useful in a given moment.

Finally, regarding the ‘why’ dimension, **data for development initiatives will increasingly be ‘human-centric’ and aim for long-term empowerment of local communities, governments and civil society groups**, in ways that enable them to cooperate and use data to collectively improve their situations.¹⁵⁰ Several of the experts we interviewed argued that organisations have been naive to expect that change can be triggered just by extracting data from people, and remotely creating models that reveal hidden correlations, without explicitly involving those affected by the problems.

A bottom-up involvement will be required to achieve sustained impacts on the ground, which frequently involve incremental, slow changes in mindsets and institutional cultures.¹⁵¹ The processes of struggle when collectively solving problems and the relationships and learning that result from them, are

149 Flyverbom, M., Madsen, A.K. and Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities, *The Information Society* 33.1: 35–42

150 Waugaman, A. (2016) *From Principle to Practice: Implementing the Principles for Digital Development*, Washington DC: Principles for Digital Development Working Group, digitalprinciples.org/from-principle-to-practice

151 GIZ (Ed.) (2014) *Cooperation Management for Practitioners: Managing Social Change with Capacity WORKS*, Wiesbaden: Springer

frequently as important as other outcomes resulting from data initiatives, like new datasets, evidence or knowledge.

3.6. Information inequality is likely to persist

Data can enrich and empower those who already have the means and resources to take advantage of its benefits.

A persistent and insightful critique of open data, which can also be applied to other data concepts, is that (open) data can enrich and empower those who already have the means and resources to take advantage of its benefits.¹⁵² As one of the experts we interviewed contended, those who collect and own the data, and those who develop the technology for production and use, including algorithmic capacities, will be able to shape the world. We can unpack further this contention by looking at who produces and owns and who uses and benefits from data.

Citizens as passive data producers

At the current state of things, several data projects take the stance of what is referred to as **extractive paradigm**,¹⁵³ where citizens are passive producers of data who hand over their data assets, even personal information, to those who analyse them behind firewalls. Motorists using the mobile application Waze, for instance, contribute locational and traffic data to a portal that aggregates this data and passes them back to app users for their use and decision-making processes. But the volume of data is at the disposal of Waze and not its users, and when Waze shares its data with governments, it does not necessarily have to seek permission from its users, as the fine print on terms of use has already taken care of it.

This **current system of data production, collection and ownership** often rests in the hands of private sector players, and to a lesser extent government, which can exacerbate existing information asymmetries, given the increasingly central role of information and data in business, society and politics. For example, as one interviewee has pointed out, technology and data giants such as Google and Facebook have a better overview of mobility and movement of people than government agencies, and they have achieved this with little dissent and protest from users.

Data ownership means control of data use

As an effect, what happens in other development sectors is mirrored in the data for development field. Web inventor Sir Tim Berners-Lee once said that **'inequality and poverty are about more than income – they are also about information.'** Because initial conditions are already unequal, the ability to fully participate in and benefit from the positive developments arising from data will be differentiated. There will surely be losers and winners in the process unless equalising conditions are present. Organisations, government agencies included, that have more resources at their disposal – expertise, financial capacity and

152 Gurstein, M.B. (2011) Open data: Empowering the empowered or effective data use for everyone?, First Monday 16.2, firstmonday.org/ojs/index.php/fm/article/view/3316

153 nesta.org.uk/blog/data-innovation-where-start-road-less-taken

sophisticated technology – will have more capacity to use data for their benefit, to the detriment of ill-resourced players and stakeholders. Users who also have more resources – access to the web, faster broadband connection, better data skills – will benefit more as compared to others.¹⁵⁴

Needless to say, data is a powerful resource. It can be put to good or bad use, and **the one who has the possession and ownership of data has the control on its utilisation**. Locational data of people is important in times of disasters, for rapid emergency evacuation procedures and swift delivery of food packages by government and early responders. But it can also be a tool of further marginalisation and intentional targeting during conflict situations, when placed in the wrong hands.

Affordable technology enables cheap data collection

A key driver of this trend is the current data ownership regime in which private companies collect vast amounts of data from customers and subscribers. They use these data assets not only to evaluate product offerings, but also to forecast new consumption patterns, define potential products and institutionalise marketing programmes based on buyer behaviour. Data collection, aggregation, and analysis systems are important investment decisions that every company makes to better manage operations and attain profitability.

Governments, on the other hand, also make these types of investments, but for an entirely different purpose. Governments collect personal information of citizens and engage in citizen identification schemes in order to manage public services, receive citizen feedback, forecast future social security spending, automate elections, deliver public goods, and plan settlements, production areas, and recreation facilities, among others.

These new uses of data are **made possible by the availability and relative affordability of technology** to collect, aggregate, and analyse personally identifiable information and mesh it with contextual variables to arrive at insights that were hard to arrive at with outdated statistical data. In almost all digital consumer goods, data is collected as a precondition of use. Likewise, more and more governments are using technology to register new births, enrol citizens in social security programmes and conditional cash transfer schemes, manage property registration, and update voter registries and national identification systems.

Empowering the empowered

Increasing customisation and personalisation of services based on data that is collected by either government or private companies have led to more responsive services, better product or service delivery, and to some extent, better appreciation of issues and concerns affecting customers and citizens. But as earlier indicated, **it can cause the empowerment of the empowered** – those who are endowed with capacity to handle volumes and large amounts of datasets will be in an advantageous position. Even cross-sector collaboration in the form of data-sharing partnerships or data collaboratives will be ruled by the party that has the competitive advantage. For example, cities with weak data capacities are likely to remain dependent on external actors to make sense of the data they own and collect.

A key driver of this trend is the current data ownership regime in which private companies collect vast amounts of data from customers and subscribers.

¹⁵⁴ This applies to women and girls in particular as a number of studies have found, including [bmz.de/en/publications/type_of_publication/weitere_materialien/study_eSkills4girls.pdf](https://www.bmz.de/en/publications/type_of_publication/weitere_materialien/study_eSkills4girls.pdf)

The danger lies in putting a premium on data in the design of development interventions. Many **data projects can fall into the trap of ignoring the fundamental institutional structure, culture and priorities**, and focus more on the data and the tools available. But these will not change incentive structure, power relations, and patterns of influence so that the poor and marginalised get real influence in agenda-setting and solution-building. For example, activists might be able to access data, have the skills to use it, and be given resources to get their message across; but if they do not have the power, the freedom, and the political leverage to bring about the change they would like to see, because the underlying political and social structure remains the same, then virtually nothing can be achieved.

With this, the danger of using data for development within the same patterns of interaction across sectors and actors will be inevitable. When data is used as a tool or as a resource for project delivery, it might have the tendency of maintaining the status quo and not change the way development intervention is designed or implemented. As such, **data for development could potentially reinforce or reproduce existing asymmetries**, and not result in empowerment of citizens or communities. The use of data in development may not necessarily result in a change in the treatment of people as passive data producers to active data owners who are able to shape the kind of future they want to achieve.

3.7. Summary

	Trend	What is this trend about?
	<p>More data and new data sources create opportunities.</p>	<p>New technologies are emerging that make data production, collection, analysis, and use faster and easier than before, such as sensors integrated into the Internet of Things, drones, nano-satellites, blockchain technologies, and many more.</p>
	<p>New technologies improve data analysis.</p>	<p>Artificial intelligence has come of age and data algorithms have improved autonomous data analytics, both of which are providing us with new ways to make sense of development problems and opportunities.</p>
	<p>New types of actors, partnerships and ways of working emerge.</p>	<p>Data for development has introduced at least two kinds of partnerships: data philanthropy and data collaboratives. New non-traditional development actors, such as data innovation labs and data analytics companies, are emerging.</p>
	<p>Balancing access to personal data and privacy.</p>	<p>Finding adequate privacy protection measures in line with rapid tech expansion and without hindering access to needed data for development initiatives will become all the more pressing in the coming years.</p>
	<p>The value of contextual, granular information.</p>	<p>Data for development initiatives will increasingly utilise various data types and become more 'human-centric', moving from current technology-driven top-down approaches towards more bottom-up, problem-driven interventions.</p>
	<p>Information inequality is likely to persist.</p>	<p>Information inequality will continue to persist or increase, with some actors in a better position than others to harness the positive developments arising from data.</p>

4

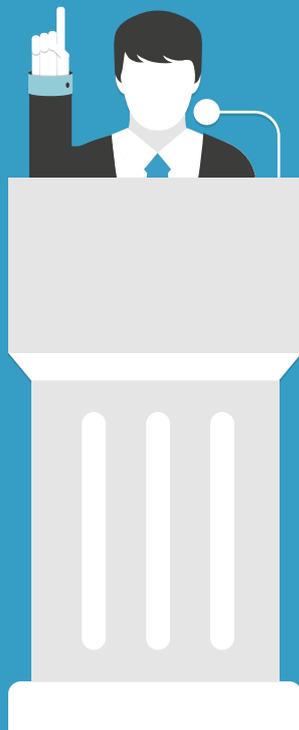
Recommendations

As presented in the preceding chapter, the increasing amount of data from various sources can make the development sector more agile but can also open more opportunities for surveillance and more threats to individual privacy. Artificial intelligence offers new ways of tackling development problems but the algorithms can be biased, opaque, and out of the reach of scrutiny.

New partnerships between development and private sector actors emerge to facilitate more collaborative data-driven approaches; however, these partnerships could create silos by excluding actors not in the circle. Also, sharing data between partners and/or opening up can serve the public good, but it can also make citizens susceptible to privacy breaches as it seems not technically possible to fully anonymise personal data.

These are just a few of the opportunities and challenges inherent in using digital data for development; their negative or positive scenarios mentioned are not set in stone. The actions taken by actors working in the data for development space and their partners will shape whether positive implications are amplified and negative ones mitigated.

This report ends with a set of recommendations on how to harness the power of data, be it big, open, citizen-generated or real-time, to contribute to global development goals and affirm Germany's role in international development. The recommendations follow from the trends identified above and take into account existing work by German development actors, as well as the general strengths and weaknesses of German development cooperation. In the boxes we highlight initiatives related to data for development that German development organisations have embarked on.



Box 4.a. The role of digital tools and data in German development cooperation

In a review of its global work on information and communication technology (ICT), which served as baseline for its new digital strategy, the German Federal Ministry for Economic Cooperation and Development (BMZ) reported that it is implementing over 390 projects across the globe in 79 countries using new technologies to address development challenges. The majority of these projects are situated in Africa, in line with the BMZ Digital Africa strategy, while others are located in Asia, the Americas and Europe.

Most of BMZ’s ICT projects are implemented by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) (202), followed by the KfW Entwicklungsbank (60), the German Academic Exchange Service (DAAD) (47), Deutsche Welle Akademie (DWA) (36), the Federal Institute for Geosciences and Natural Resources (BGR) (20), the Deutsche Investitions- und Entwicklungsgesellschaft (DEG) (18), Engagement Global (6), the Physikalisch-Technische Bundesanstalt (PTB) (2) and Goethe-Institut (1).¹⁵⁵ BMZ’s work in using ICT to address development issues touches on different sectors and themes including rural development, good governance, social development, economy and employment, sustainable infrastructure, security and reconstruction, and environment and climate.¹⁵⁶

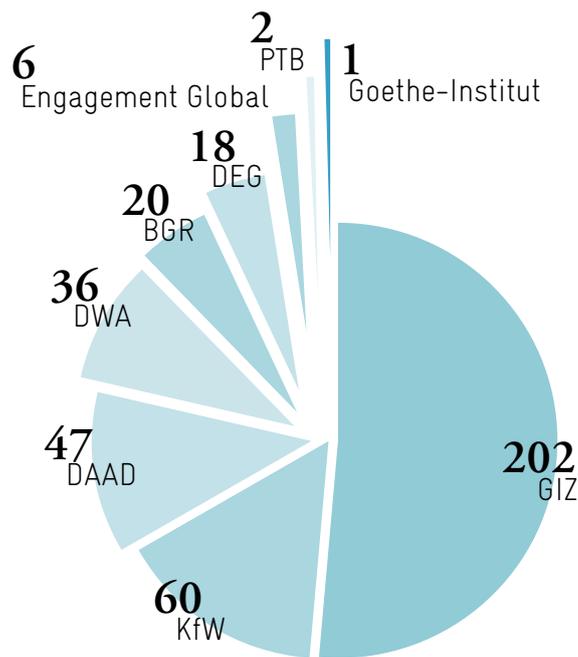


Figure 4.c. Number of projects using ICT by implementing organisation.

155 The data is based on a survey of selected recipients of German development cooperation funding and implementing organisations (collected in 2017).

156 GIZ (2016) Toolkit – Digitalisation in Development Cooperation and International Cooperation in Education, Culture and Media, Berlin: BMZ, giz.de/fachexpertise/downloads/bmz2016-en-toolkit-digitalisation.pdf

It does seem that the efforts of German development actors in this space can be commended for being experimental, bold, and ambitious. While several of their projects are also geared at collecting aid performance data alongside projects that gather sector-specific data from individuals and communities, it is not clear how the vast amount of data assets are being used internally for reflection, design, implementation and reprogramming. While a clear digital strategy is in place,¹⁵⁷ an articulation of the role of new and emerging data in German development cooperation programming needs to be defined.

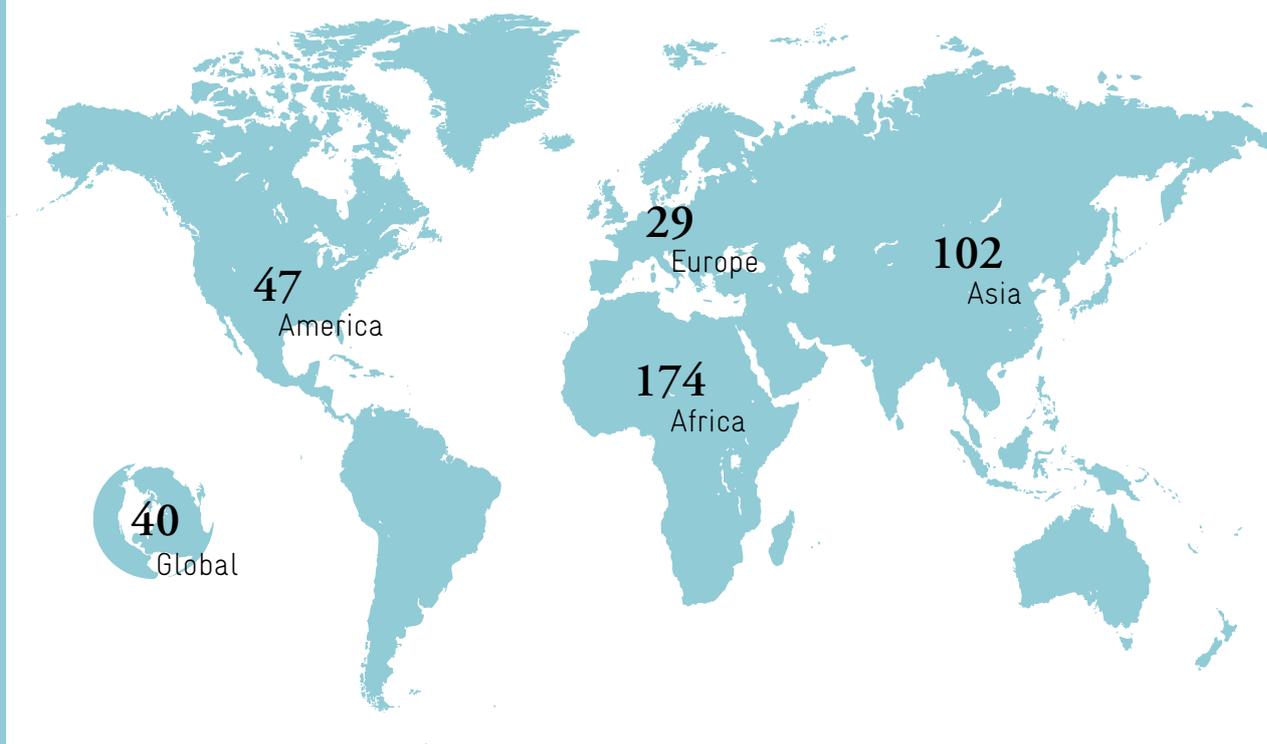


Figure 4.d. The BMZ's projects where digital technologies are playing a central role according to region.¹⁵⁴

¹⁵⁷ BMZ (2017) Harnessing the Digital Revolution for Sustainable Development – The Digital Agenda of the BMZ, Berlin: BMZ, [bmz.de/en/publications/type_of_publication/information_flyer/information_brochures/Materialie276_digitale_agenda.pdf](https://www.bmz.de/en/publications/type_of_publication/information_flyer/information_brochures/Materialie276_digitale_agenda.pdf)

4.1. Maximise the potential of data, but do not treat it as a ‘silver bullet’

German development actors should embrace the potential of data to drive and support development.

Experts interviewed for this study suggested that they should **focus on supporting demand-driven initiatives by building on existing work in a small number of sectors**. The decentralised structure of many German development organisations, with staff being placed in or working directly with government agencies, civil society organisations and business associations of partner countries, gives German organisations a very granular understanding of context, which is vital for the effective use of data to address complex challenges (see Section 3.5. The value of contextual, granular information). Specific recommendations made by the interviewees included demand-driven and sector-specific identification of data gaps in partner countries as well as funding, training and infrastructure for data use projects.

It is important to not get trapped in narrow data concepts, strategies and funding streams for big, open, real-time or other data categories, but to understand that data is only a means to an end, one of many building blocks in creating sustainable solutions to pressing development issues.

Moreover, German development actors should approach data for development with usage limitations in mind. It is not safe to assume that just because data exists, decision-makers will make use of it. **The political economy of decision-making and the demand for data are as important as the supply of data**. When seeking to fund or establish data for development initiatives, organisations should either ensure that an enabling environment to incorporate data into decision-making already exists, or that one will be developed alongside the initiative.

Such a problem-focused approach will in many cases require the combination of different types and sources of data. In these cases more data is not always better, but accuracy, timeliness and comprehensiveness of the data in question will be more important.

Box 4.b. Using multiple data sources for infrastructure planning to better understand labour market needs

Digital TVET is an initiative by BMZ that not only aims at integrating technical and vocational education and training (TVET)¹⁵⁸ as a component in the tender process for infrastructure projects financed by international development banks. It also explores the use of multiple data sources to better understand labour market needs in partner countries of German development cooperation. Using official statistics, labour market information and big data analytics, a digital platform will enable forecasting through targeted data analysis. By analysing data from scoping studies and tender documents of completed infrastructure projects, Digital TVET will provide comparative data on costs, required skills and the number of personnel needed for planned infrastructure projects. Furthermore, it will provide information about the number of qualified workers locally available and potential needs for additional training of personnel as part of the integrated TVET component.

4.2. Build internal data capacity

German development organisations will need to invest into internal capacity to build the data literacy needed to apply data-driven approaches to development work.

While partnerships with external providers will be required for more complex projects, **a basic level of expertise, knowledge and skills would be beneficial** to position German development cooperation at the forefront of the data for development field. These may include capacity building on data-driven approaches in designing, managing, and monitoring development projects and programmes, hiring personnel with expertise in digital data to scope and develop data-driven projects, as well as investing in research and experimentation capabilities to investigate and test new and novel approaches in solving persisting problems. Moreover, nurturing creative spaces within the organisations, and promoting experiments to explore and share ways of working with data, could accelerate the diffusion of data knowledge and capacities.

To do this, an assessment of current capacities vis-à-vis the role German organisations would like to play is necessary to determine how needs for domain and context experts, social scientists, data scientists and technologists can be best met – whether these have to be built from within, outsourced, or cultivated through strategic partnerships. If built within, there is a need to **facilitate the emergence and consolidation of multi-disciplinary and multi-functional teams**¹⁵⁹ where data-scientists inspect and manipulate data to determine ‘what’ is happening, while social scientists such as anthropologists can help find answers to ‘why’ things are happening.

158 TVET stands for Technical and Vocational Education and Training and is defined by UNESCO as ‘those aspects of the educational process involving, in addition to general education, the study of technologies and related sciences and the acquisition of practical skills, attitudes, understanding and knowledge relating to occupation in various sectors of economic life’; <http://www.unesco.org/new/en/newdelhi/areas-of-action/education/technical-vocational-education-and-training-tvet/>.

159 World Bank (2016) Big Data Innovation Challenge: Pioneering Approaches to Data-Driven Development, Washington DC: World Bank, documents.worldbank.org/curated/en/396861470905612761/pdf/107751-REVISED-PUBLIC-BigData-Publication-e-version-FINAL.pdf

Box 4.c. Digital skills and data capacity building within German technical cooperation

Digital transformation impacts the work of development organisations. In order to harness this potential, German development actors have launched various training measures to help their staff become familiar with digital data concepts and tools. For instance, GIZ is strengthening the skills and knowledge of its staff about open data and big data through tools like webinars, Massive Open Online Courses and reverse mentoring where specialists coach senior managers regarding digital development and solutions.

Moreover, GIZ has launched an Innovation Fund to foster the use of digital tools in development programmes. When it comes to data for development, GIZ seeks to raise awareness of the potential impacts and challenges of data, and to assess possible use in project management and evaluation.

4.3. Leverage partnerships for a strategic advantage

It is inevitable and also beneficial for German development organisations to work closely with other actors in the data for development space.

While a digital strategy exists, a data for development direction is not yet defined, and while stakeholder involvement in the digital strategy is explicitly particularised, we still have to see how German development organisations will encourage participation of other actors and build partnerships with stakeholders within the data for development space. GIZ has already initiated the Partners for Review, a multi-stakeholder network monitoring national progress in the SDGs. It may need to connect this with other similar initiatives. The following concrete recommendations are raised for consideration.

There is a need to consider reinforcing Germany's global commitment to the importance of data in achieving the SDGs. While Germany is a signatory to the SDGs and one that explores the use of data for this purpose, it should also consider joining or endorsing the Global Partnership for Sustainable Development Data to liaise with other stakeholders in ensuring that the relevant data is available to establish baselines, monitor progress, and build capacity of various stakeholders in using data to help achieve the SDG commitments. Likewise, while its digital strategy specifically mentions its commitment to open data, **the German government may consider adopting the International Open Data Charter to cement its credibility in promoting citizens' right to data**, especially in contexts where it supports partner governments to make government data available.

German development actors should review their engagement strategy with other sectoral partners and key actors, especially the 'development mutants', those new and emerging actors that are revolutionising the way development is done.¹⁶⁰ As a starting point, **German development cooperation may need to profile new actors doing work on data and development and assess how each actor's key capacities can be harnessed to contribute to Germany's development agenda**. This inventory can be used as a basis to explore data collaboratives

¹⁶⁰ diytoolkit.org/profiling-the-development-mutants

that may need to be established for key development problems of interest or priority to Germany.

Box 4.d. Germany participates in the Open Government Partnership

The Open Government Partnership (OGP) is a multilateral initiative of 75 countries that have committed to transparent, accountable and responsive governance.

Germany announced its participation at the OGP Summit in Paris in December 2016. The German government has recently published its First National Action Plan,¹⁶¹ which was co-created with civil society. The plan outlines activities in a range of sectors to strengthen transparency, accountability and public participation in government. Germany's engagement in the OGP emphasises its commitment to the principles of open government, which also guides German development organisations in working with partner countries in governance reforms and data for development initiatives.

Needless to say, the types of actors to partner with or the kinds of network to be part of will largely be dependent on the kind of role and the sector the German government would like to engage with. For example, for the use of data in agriculture, it may consider strengthening its engagement in the Global Open Data Agriculture and Nutrition network,¹⁶² a multi-stakeholder group that supports the proactive sharing of data on agriculture and nutrition to help achieve food security. If it wants to work with big data from telecommunications companies, it may consider working with the Digital Impact Alliance,¹⁶³ a partnership among USAID, the UN Foundation, the Swedish government, and other stakeholders concerned with ensuring a more inclusive digital economy, on their data for development projects.

161 bmi.bund.de/SharedDocs/Downloads/EN/Broschueren/2017/ogp-aktionsplan-en.pdf

162 godan.info

163 digitalimpactalliance.org/what-we-do/data-for-development-2

Box 4.e. Forming new partnerships

The World Poverty Clock¹⁶⁴ project shows how the German Federal Ministry for Economic Cooperation and Development is already engaging with organisations that can provide data analytics expertise. Developed by the World Data Lab, the World Poverty Clock visualises likely future economic and social development based on multiple sources of data. It aims to monitor progress against one of the UN's Sustainable Development Goals on ending extreme poverty. The project combines poverty data, demographic projections and estimates of economic growth to visualise how poverty will develop in various countries in the coming years.



Figure 4.e. Drawing on the data refinery concept, the World Poverty Clock combines and standardises poverty data, demographic projections and estimates of economic growth to understand how poverty will develop in various countries between now and 2030.

164 worldpoverty.io, screenshot taken on 11.10.2017, 17:50pm.

4.4. Support strong technological and legal data privacy frameworks

Supporting partner governments in strengthening their data privacy frameworks, working with business and civic organisations to better understand the potential negative implications of data collection and use, and championing responsible data approaches in global initiatives are suggested action areas for German development organisations to promote data privacy in development.

A recurring theme in conversations with the experts interviewed for the study was the role of **Germany as forerunner for personal data protection in the digital age**. Interviewees saw a natural fit between the reputation, expertise and perceived credibility of German development cooperation in this area with the growing need for frameworks, approaches and instruments to ensure the privacy of individuals and groups while at the same time exploiting data for the social good.

While German development actors have shown increased interest in fields like data security, data sovereignty and data protection, the impression of interviewees was that a clear profile was largely missing. Moreover, while some data protection work is under way in sectors like financial inclusion¹⁶⁵ or in the wider human rights space, **there seems to be no concerted effort yet to mainstream data privacy as part of a broader focus on digital rights**. The BMZ Digital Agenda, however, may provide a good point of departure for such work.

There are a number of roles German development actors could play in that field. Privacy regulation is currently either largely absent, inadequate or missing in developing countries – or does not apply to the digital realm. Legal regimes are copied from developed economies with little adjustment to the country-specific context. German development organisations could be working with governments, parliaments, civil society and businesses in partner countries to develop appropriate regulatory frameworks and practices that allow secure data sharing.

Likewise, through their existing networks they could **help mobilise voices from the global South to contribute to global debates on data privacy issues**, especially when it comes to possible ethical issues in the increasing number of data-sharing arrangements, or data collaboratives. In addition, targeted capacity building is required for actors in partner countries to mainstream responsible and secure data practices ranging from data collection to archiving and sharing, especially in sensitive environments.

165 Jentsch, N. (2016) Data Protection in the Context of Digital Financial Services and Big Data, GIZ Discussion Paper, www2.giz.de/wbf/4tDx9kw63gma/Datenschutz-Diskussionspapier_E_140416_Internet.pdf

Apart from regulatory frameworks, capacity and awareness-raising on the potential negative consequences of data for development work, the German government should also consider **investing into shaping the debate on the role of artificial intelligence and algorithms in data analysis**. Advances in artificial intelligence will make it even less likely to ensure anonymity because of the ability of artificial intelligence systems to cross-reference between vast quantities of data in multiple datasets. Meanwhile, poorly constructed algorithms can reproduce and reinforce existing patterns of discrimination, especially if the data that was used to train the algorithm is biased, incomplete or in other ways poor in quality.

Box 4.f. The role of German development cooperation in strengthening data privacy in the financial sector

In recent years, digitisation and the greater availability of digital data have transformed the financial sector. Traditional providers of financial services such as banks, but also new actors such as mobile network operators, ‘FinTechs’ or ‘InsureTechs’, tap into data to develop financial products and innovative ways of making financial services available. For financial inclusion, using big data means removing traditional bottlenecks by taking advantage of new technology, analytical tools and data access. In lending, for example, traditional processes to verify identity or income have been time-, money- and people-intensive with high drop-off rates. Big data applications and analytics help to make those verification processes instantaneous, automated and cheap, so that it becomes easier to obtain the necessary information on potential customers.

Although increased volumes of data facilitate a better understanding of customers, the use of big data in the financial sector raises new challenges. Data is obtained from different sources – including social media – and is assessed using algorithms that lack transparency. These practices of consumer profiling represent a risk to data protection and to financial exclusion. Access to financial services can be denied for vulnerable customers based on information gathered by insurance companies.

To avoid negative developments with regards to data protection and financial exclusion, BMZ and GIZ have worked on establishing data protection guidelines for financial services, for instance through the Global Partnership for Financial Inclusion. The Partnership provides a platform for all G20 countries, interested non-G20 countries and other stakeholders to advance financial inclusion, including implementation of the G20 Financial Inclusion Action Plan.

4.5. Be experimental and focus on a few sectors and geographies

German development organisations should be taking an experimental, iterative and flexible approach to data for development, targeted at specific problems, sectors and geographies.

Data for development is a young, dynamic and rapidly evolving field with little well-documented good practices, which calls for a step-by-step approach to test how digital data can play a role in addressing development challenges. The GIZ Innovation Fund (see Box 4.c.) and the African Union data lab (see Box 4.g.) are examples of such innovative approaches that emphasise experimentation and learning.

While data can act as a catalyst of change, a too narrow focus on the data, its supply and use rarely lead to achieving the desired impact. German development cooperation, with its strong presence in partner countries, is **well placed to combine novel data-driven approaches with a deep, local and context-specific understanding**. It can therefore connect data-driven interventions with existing demands to address real pain points as experienced by local communities.^{166,167}

Contextual knowledge is needed to assess the quality, representativeness and granularity of the data being used, as well as to interpret the patterns and the anomalies detected in the data analysis. It is necessary to complement big data insights with what has been termed as thick data: qualitative, social and ethnographic knowledge of the local context, which helps to understand the meanings and behaviours that underpin observed patterns.¹⁶⁸ Thus, purely quantitative data needs to be supplemented and critically assessed with qualitative data to better understand ‘why’ things happen, as failing to do so can lead to poor decisions, including undesired discrimination and exclusion.

German development organisations should contribute to realising the potential of data in development by investing in specific sectors instead of cross-cutting data initiatives. This will require engagement with governments and local actors, paying special attention to countries and sectors that expose positive deviance¹⁶⁹ with regard to specific development problems. German development actors could combine their strong sectoral expertise and engagement with the opportunities that new types of digital data provide, and help nurture sectoral data ecosystems in countries in sectors like procurement, extractives or health.

Take an experimental, iterative and flexible approach to data for development, targeted at specific problems, sectors and geographies.

Do not solely rely on quantitative data but complement insights with thick data i.e. qualitative, social and ethnographic knowledge of the local context.

166 Kaliati, A.; Kachieng'a, P. and de Lanerolle, I. (2017) What Data Do We Want? Understanding Demands for Open Data Among Civil Society Organisations in South Africa, MAVC Research Report, Brighton: IDS

167 medium.com/data-zetu/a-problem-driven-approach-to-promoting-data-use-978456f60e63

168 medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7

169 Pascale, R.; Sternin, J. and Sternin, M. (2010) The Power of Positive Deviance, Boston: Harvard Business Press

Box 4.g. An interactive platform and a data lab for the African Union

The African Union (AU) seeks to accelerate inclusive growth and sustainable development. ICT and broadband are recognised in the Agenda 2030 and Africa’s Vision 2063¹⁷⁰ as crucial for the development of an information society and to enable ICT services to citizens, government and the economy. Growing amounts of data are generated which calls for innovative ways to analyse and visualise it to support development efforts.

During the High-Level Conference on the Data Revolution in 2015, representatives of the AU agreed on the ‘Africa Data Consensus’, which calls for a data revolution to drive social, economic and structural transformation in Africa. Its key actions aim to create an inclusive data ecosystem involving government, private sector, academia, civil society, local communities and development partners, and to use innovative methodologies to improve data collection, analysis and data usage in African countries.

The German government has shown interest in supporting the AU in the creation of a **data lab** to gather and analyse various types of data to inform the implementation of Agenda 2030 and Africa’s Vision 2063. Moreover, an **interactive information platform** will be developed to inform citizens in AU member states about decisions, activities and impacts achieved by the AU and to provide interactive communication channels between the AU and citizens.

4.6. Address data inequalities

People and public institutions need access, skills and infrastructure in order to benefit from data for development.

The German government should help close the gap in data and information inequalities through targeted research and empowerment of local stakeholders.

There is a **scarcity of research that deals with data and power dynamics in societies and communities**, an area the German government could invest in by engaging with actors and funding initiatives that seek to build a rich understanding of the complexities that exist at the intersection of data and power. Many of the experts interviewed suggested that despite the many pilot projects and case studies, it remains largely unclear who is affected in what ways by data for development interventions and to what extent. They see a major gap in documented experiences on whether and to what extent data is actually shaping decision-making on policies and programmes and, more importantly, who benefits as a result.

Greater investments are also needed for the actual empowerment of local players with little resources and capacity to engage with data and to influence the data for development agenda. **Connected and capacitated public institutions, local civic groups, journalists and activists are needed to diffuse power that comes with the current centralisation of data and skills.** While access to data is an essential step, it requires the capacity to use it along with the agency and voice to effect change with data. There is a real risk in data for development actors promoting the empowerment of the connected and empowered while those without access, capacity and voice are further relatively disadvantaged, thus potentially exacerbating inequalities.

170 au.int/en/agenda2063

As representatives of German development cooperation actors have pointed out in interviews, there is a strong track record of German development organisations in empowerment through capacity building and through financing social and economic infrastructure. Coupled with a reputation of being able to work across government, civil society and the private sector, and with expertise in a range of thematic areas such as food security, financial inclusion and disaster preparedness, **German development organisations are very well positioned to occupy the niche of human-centred, privacy-aware data skills development.** This also applies to the financing of the necessary investments into software, hardware and up-to-date telecommunication infrastructure required for widespread data access and use. Public institutions and citizens, including marginalised people and those living in remote areas in developing countries, need to have access to technology and the internet to avail of the benefits derived from using data for improved decision-making, strengthened transparency and accountability, and better public service delivery.

Financing necessary investments into software, hardware and up-to-date telecommunication infrastructure is required for widespread data access and use that benefits everyone.

Box 4.h. Data skills training of journalists in Cambodia

In Cambodia, due to censorship of press, radio and television, journalists and human rights activists are increasingly using online media and social networks not only to share information and connect with one another, but also to shape public discourse.

In order to empower local journalists and human rights activists in using social media and digital data effectively and securely, BMZ, GIZ and Deutsche Welle Akademie, Germany's leading organisation for international media development, established a project with several local and international partners in 2013. Its aim was to assist bloggers, human rights activists and developers in setting up a shared training centre for capacity building. Media practitioners were trained in data-driven journalism. Course participants learned how to analyse and publish digital data on topics such as public service provision and rice prices. The data was then used to inform public debates and radio shows in rural communities.

4.7. Summary

Recommendation

What is this recommendation about?



Maximise the potential of data, but do not treat it as a ‘silver bullet’.

Embrace the potential of digital data to drive development, while being conscious that data is only a means to an end. Acknowledge that the political economy of decision-making and the demand for data are as important as the supply.



Build internal data capacity.

Invest in internal capacity to build the data literacy needed to leverage data-driven approaches. Nurture creative spaces to explore and share ways of working with data to accelerate the diffusion of knowledge and capacities.



Leverage partnerships for a strategic advantage.

Reinforce Germany’s global commitment to the importance of data in achieving the SDGs by endorsing the Global Partnership for Sustainable Development Data and by adopting the International Open Data Charter. Review engagement strategies with key actors in data for development, especially the ‘development mutants’.



Support strong legal and technological data privacy frameworks.

Work with partner governments, businesses and civic organisations to strengthen national data privacy frameworks. Champion responsible data approaches in global initiatives and help mobilise voices from the global South to contribute to global debates on data privacy, especially in data-sharing arrangements.



Be experimental and focus on a few sectors and geographies.

Combine strong presence in partner countries and deep, local and sectoral understanding of context with new data-driven approaches to nurture sectoral data ecosystems.



Address data inequalities.

Help close the gap in data inequalities through empowerment of local stakeholders. Support targeted research aimed at building a rich understanding of the complexities that exist at the intersection of data and power.

5

The way forward

Digital data offers manifold opportunities to development by enabling development actors to strengthen decision-making processes, improve service delivery, elicit more meaningful citizen participation, and increase responsiveness in humanitarian services, among others. At the same time, it poses threats to individual privacy, generates new forms of exclusion, and opens new ways of surveillance. Against this background, it becomes significantly important that actors are aware and cognisant of not only the benefits, but also the risks associated with working with digital data in development.

This research began by exploring different digital data concepts and how they can potentially impact development processes. The study showed that while data categories are useful in understanding how data is produced (e.g. citizen-generated) and published (e.g. open), or how data behaves (e.g. big) and gets used (e.g. real-time), development actors, including those for whom development is intended, shape the outputs and the outcomes of these processes. They drive the different contextual factors that create, maintain, or mutate the trends we have identified.

As such, attributing to digital data the promise of solving development challenges is an overstatement, if not an over-commitment. Data is just one of the ingredients that help solve a particular problem. There are people, partners and communities, regulatory frameworks, gatekeepers and leaders that make up the whole enabling mechanism for digital data to work best in development.

This is the reason why the recommendations identified in this paper emphasise the role of stakeholders along the development value chain – building individual and organisational capacities, protecting personal privacy, building collaborations across sectors and among actors, and empowering the powerless. This is a recognition of the fact that data is useless in development without people, that technology is inutile without the civic, and that efforts that over-capitalise on technology without due regard for communities and their culture will gain no ground.

6

Annex

Methodology

Our research aimed to help German development organisations inform policy and practice for the next five years regarding data for development with an emphasis on four concepts: big data, open data, citizen-generated data and real-time data. To do this the team used a mixed-methods approach including a literature review, semi-structured and unstructured interviews, and two workshops.

To start, the team carried out a literature review that consisted of peer-reviewed journal articles and grey literature. When grey literature was used, the research team did its best to triangulate sources to ensure data quality. Because members of the team already had backgrounds on the topics before the project commenced, much of the literature was uncovered by looking back at previous work. The team used literature read in previous work and ensured that their knowledge was up to date on these four concepts by using academic search engines including Google Scholar. Moreover, team members provided each other with relevant literature and sources throughout the project and uncovered further material from the interviewees. This literature review was mainly used to write the 'state of affairs' section.

Alongside the literature review the team also conducted 24 semi-structured telephone interviews with key informants in the data for development field. The team applied a snowball approach. Key informants were uncovered through contacts in our internal networks at the World Wide Web Foundation and the Institute of Development Studies. We then asked our interviewees to point us in the direction of others knowledgeable in the data for development space. The key informant interviews were mainly used to extract data for the 'trends' chapter. To do this the team asked each interviewee questions with regard to what is currently happening, what they see happening in the future, and recommendations for new development actors entering the data for development space. The team also sought input from staff working with German development organisations through interviews and email exchanges in order to get a better understanding of the current work of German development cooperation actors in data for development and to tailor the recommendations to their needs.

After conducting the literature review and most of the interviews, the team facilitated a workshop, held on 11 May 2017 in Berlin, with 13 participants from organisations involved in and knowledgeable about German development cooperation. In this workshop, the team shared their initial findings with participants using both presentations and participatory workshop methods. The research team facilitated sessions to help participants think through strengths and weaknesses of German development organisations and opportunities and threats with regard to data for development. Documentation from the workshop was shared with participants and their networks for feedback. The workshop documentation, feedback from participants and unstructured interviews, as well as some of the recommendations proposed by our interviewees, were the main resources used for the 'recommendations' chapter.

About this report

About the World Wide Web Foundation

The World Wide Web Foundation is an independent, international organisation working for digital equality – a world where everyone has the same rights and opportunities online. Established in 2009 by web inventor Sir Tim Berners-Lee, the Web Foundation works to advance a free and open web ‘for everyone’ by influencing government and corporate policies to ensure everyone can use the web freely and fully.

About the Institute of Development Studies

The Institute of Development Studies is a leading global institution for research, teaching and learning, and impact and communications, based at the University of Sussex. Our vision is of equal and sustainable societies, locally and globally, where everyone can live secure, fulfilling lives free from poverty and injustice. Since 1966 IDS has been working with partners to tackle complex development challenges and contribute the evidence, analysis, theory and facilitated learning that can help communities, practitioners and decision-makers at all levels work together for practical, positive change.

About the Sector Project Digital Development

The use of digital tools has increased dramatically over the last twenty years. This has led to fundamental changes in social, political and economic structures, particularly in developing and emerging countries. With the increasing digitalisation of society as a whole, new opportunities arise for the achievement of sustainable development objectives. The sector project supports the Federal Ministry for Economic Cooperation and Development (BMZ) in exploring new fields of action, practices and instruments for all areas of development cooperation, including education, health, economic development and good governance.

Acknowledgements

We would like to thank all the experts who provided insight and expertise that greatly assisted the research: Natalia Adler (UNICEF), José M. Alonso (Web Foundation), Kitty von Bertele (DFID), Ciro Cattuto (ISI Foundation), Adi Eyal (Open Up), Wolfgang Fengler (Word Data Lab), Manuel Garcia-Herranz (UNICEF), Robin Gravesteyn (UNCDF), William Hoffman (World Economic Forum), Charles Johnson (USAID), Johannes Jütting (OECD), Lauren Kahn (DFID), Christian Kreuz (Independent), Julia Manske (Independent), Claire Melamed (Global Partnership for Sustainable Development Data), Nnenna Nwakanma (Web Foundation), Michael Pittelkow (SAP), Giulio Quaggiotto (Innovation Expert), Zara Rahman (Engine Room), Ben Ramalingam (Institute of Development Studies), Nanjira Sambuli (Web Foundation), Enrique Santos (Independent), Claudia Schwegmann (Independent), Mamello Thinyane (UNU Computing and Society), Stefaan Verhulst (NYU GovLab).

We would also like to show our gratitude to the workshop participants for providing valuable feedback on the initial research results: Ruth Bachmann (GIZ), Marvin Boehm (GIZ), Valentina Cullmann (GFA), Rahel Dette (GIZ), Dr Philipp Fuerst (IBM), Lea Gimpel (GIZ), Christian Gmelin (GIZ), Elisabeth Hobl (GIZ), Attila Mong (Deutsche Welle Akademie), Sabine Olthof (GIZ), Michael Pittelkow (SAP), Thomas Wollnik (GIZ).

We thank Anne Doose, Judith Frickenstein, Lea Gimpel, Elisabeth Hobl, Henrike Klavert, Konstantin Pagonas, Balthas Seibold, Arlett Stojanovic, Norman Schräpel (all GIZ), and Anne Schwarz (Brot für die Welt) for providing helpful feedback on the workshop documentation and the draft report. All remaining errors are ours.

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Germany
T +49 (0) 228 44 60 - 0
F +49 (0) 228 44 60 - 17 66

Dag-Hammarskjöld-Weg 1 - 5
65760 Eschborn, Germany
T +49 (0) 61 96 79 - 0
F +49 (0) 61 96 79 - 11 15

E info@giz.de
I www.giz.de



On behalf of

**Federal Ministry
for Economic Cooperation
and Development**