# INVESTMENT CASE

# Multiplying progress through data ecosystems

United Nations

Global Partnership for Sustainable Development Data

A **DATA WITH PURPOSE** publication

## Acknowledgments

We wish to thank the following partners for their input and advice in the development of this report:

- **Global Partnership for Sustainable Development Data:** Claire Melamed, Jenna Slotin, Jennifer Oldfield, Amy Leach
- **United Nations:** Priscilla Idele, Kersten Jauer, Tomas Zak, Francesca Perucci
- **World Bank:** Haishan Fu, Craig Hammer, Keith Garrett
- **Future Advocacy:** Olly Buston, Victoria Gilbert, Iona Cable

We also acknowledge with appreciation the work of Dalberg Advisors, who served as lead authors and conducted the analysis presented in this report. The team included: Ciaran Cooper-Craig, Michele Franchetti Pardo, Henry Koster, Audrey Hanard, Annette Chau, and Matt Frazier.

We would like to acknowledge with thanks the contributions of Amanda Glassman (Center for Global Development), Johannes Jütting (PARIS21), Susan Hodgson (PARIS21), Sasha Ramirez-Hughes (PARIS21), Martin Lestra (OECD), Shaida Badiee (Open Data Watch), Lorenz Noe (Open Data Watch), Philip Cockerill (United Kingdom Foreign Commonwealth and Development Office), Paula McLeod (United Kingdom Foreign Commonwealth and Development Office), Andrea Ries (Swiss Agency for Cooperation and Development), Sarah Orzell (Bill and Melinda Gates Foundation), and Ingrid Plag (MDF) who provided technical expertise and valued feedback.

We would like to recognise with gratitude the generous funding received from the New Venture Fund to support the creation of this report.

## Executive summary

**Despite the vast impact of humanitarian, development, and domestic programs over the past 30 years, the resources these programs deploy have not been distributed with optimal efficiency or equity, in part due to information gaps that leave decision makers flying blind.** Over recent decades, low- and middle-income countries (LICs and MICs) have seen a tremendous positive change in the experiences and outcomes of individuals and communities; funding from donors and domestic governments underpins this success. However, a lack of information on the varied needs of local communities, how these needs change over time, and what "works" means that donors and domestic governments alike have missed opportunities to make their resources go as far as possible.

**Data and data ecosystems enable decision makers to improve lives and livelihoods by better understanding the world around them and acting in more effective and targeted ways.** In a time of growing crises and shrinking budgets, it is imperative that every dollar is spent in the most efficient and equitable way. Data ecosystems provide decision makers with the information needed to assess and predict challenges, identify and customize solutions, and monitor and evaluate real-time progress. Together, this enables decisions that are more collaborative, effective, efficient, equitable, timely, and transparent. And this is only getting easier—ongoing advances in our ability to harness and apply data are creating opportunities to better target resources and create even more transformative impact.

**In addition, data ecosystems can enhance dialogue between leaders and their people, strengthening accountability and the democratic process.** Increased transparency in leaders' actions and decision making processes allows citizens and civil society to be informed and make sure that decisions are taken and implemented effectively and equitably. Relatedly, improved data ecosystems provide a communication channel through which people can elevate their own views and needs.

**By enhancing the efficiency, accountability, and impact of humanitarian, development, and domestic spending, data ecosystems drive a diverse range of benefits for individuals and communities.** These benefits range across four inter-related categories:

**ECONOMIC BENEFITS –**
Data creates real value through new economic and impact opportunities, and cost savings through everything from process efficiencies to reduced corruption; analysis of past investments has shown an average return of USD 32 for every dollar invested.[1]

**SOCIAL BENEFITS –**
Data can save lives and enhance the quality and equity of living standards by increasing the effectiveness of social programs; for example, data fosters stronger and more effective health systems, higher rates of attendance and performance in schools, more proactive and anticipatory action to climate or humanitarian emergencies, and much more.

**ENVIRONMENTAL BENEFITS –**
Data can bolster the transition to sustainable development, shedding light on the drivers, rates of change, and impacts of environmental issues.

**INSTITUTIONAL BENEFITS –**
Public data systems support evidence-based decision making and provide the information needed to hold institutions accountable.

**Fully realizing these benefits requires an integrated, systems-focused approach.** Integrated data ecosystems are communities in which public, private, academic, international, and civil society actors come together to develop, validate, and use data. Investments in the foundational pillars and enablers of these ecosystems can help improve the quality, coverage, safety, and usability of all data created within the system. Moreover, because integrated ecosystems at a national level act as the building blocks of regional and global systems, efforts to develop and strengthen national data systems will further support the creation of robust regional and global data ecosystems.

1. The range of returns across the analyzed investments was USD 7 to USD 73 per dollar spent; see Chapter 2 and Annex 1 for more detail on this economic ROI analysis.

**Supporting L/MICs in building effective integrated national data systems will require governments and donors to take a more coordinated funding approach that prioritizes systemic, cross-sector issues.** A severe lack of coordination to date has contributed to fragmentation of investments, which are frequently disjointed, duplicative, and concentrated in one-off initiatives. This undermines the impact and sustainability of building data ecosystems. Pooled funding instruments and improved mechanisms for tracking investments in data offer a solution by institutionalizing a more holistic and cost-effective approach to building integrated data ecosystems. However, deploying existing resources in a more coordinated manner is not enough. Unleashing data's potential to optimize the effectiveness and impact of resources will necessitate greater investment and support from a wider pool of new donors and public and private sector actors.

**The UN Complex Risk Analytics Fund (CRAF'd) and World Bank Global Data Facility (GDF) are pooled funds that have been designed to harmonize and catalyze investment in data ecosystems.** The funds aim to transform the lives of billions by raising and deploying over USD 500 million for data and data ecosystems. While CRAF'd will coordinate investments in risk data and analytics to enable faster and more targeted, efficient, and effective programs for people in fragile and crisis-affected settings, GDF is designed to catalyze long-term domestic and international support for integrated national data ecosystems in LICs and MICs.

**Together, CRAF'd and GDF have the ability to reshape the data funding and capacity building landscapes.** Not only do these funds have the deep expertise and in-country connections required to design a systems-focused action agenda that is responsive to local needs, but they are also uniquely positioned to bring together and coordinate the diverse community of stakeholders necessary to realize this agenda and build effective integrated systems. Moreover, as pooled funds hosted by the UN and World Bank, CRAF'd and GDF enable contributors both to make more effective use of their capital (e.g., by sharing financial burdens, minimizing risks, and cutting transaction costs) and to access the extensive advocacy capabilities, financial know-how, and networks needed to identify and design investments that can catalyze additional funding. These attributes allow the funds to maximize the impact of contributors' resources and enable systems change that increases returns on future investments. In short, they help governments and donors to improve lives and livelihoods at lower cost.

**Realizing the potential of data to amplify humanitarian, development, and domestic programming will require coordinated, collaborative action by all stakeholders.** Therefore, we call on…

| Actor | |
|---|---|
| **Bilateral and philanthropic donors** | **Allocate a minimum of 0.8% of annual investment to data ecosystems;** maximizing impact by…<br>• Directing funding allocations **first and foremost through CRAF'd and GDF**<br>• Increasing the **transparency of funding for data and data ecosystems** |
| **Domestic governments** | **Lower-income (IDA-eligible) countries to allocate a minimum of 0.5%, and middle-income (IBRD-only eligible) countries 0.1%, of annual expenditure to data ecosystems** |
| **Private sector actors** | **Be a core partner in establishing and utilizing integrated data ecosystems** and leveraging industry data, technology, capacity, knowledge, and best practice to strengthen the capacity of, and ties between, ecosystem participants |
| **All ecosystem participants** | **Collaboratively drive this agenda forward,** including integrating data into the design, implementation, and evaluation of all programming and contributing to the ongoing research agenda |

**With better coordination and increased investment, we have the power to enable a step change in the speed, efficiency, and fairness of decision making—transforming lives and livelihoods around the world.** These systemic efforts will integrate support for data ecosystems into a wide array of stakeholder programming. The resulting virtuous cycle will unleash data's potential to enhance the impact and efficiency of humanitarian, development, and domestic spending, accelerating headway on shared goals and setting a course for lasting progress.

# CHAPTER 1

## Why data ecosystems, and why now?

**Over the last 30 years, humanitarian, development, and domestic resources have helped create prosperity and reduce suffering for billions.** Between 1990 and 2020, adult literacy rates rose globally by 13%, calorie consumption per capita grew by 12%, and life expectancy increased by seven years.[2,3,4] As a result, despite significant population growth, the proportion of people living in poverty fell by 75%.[5]

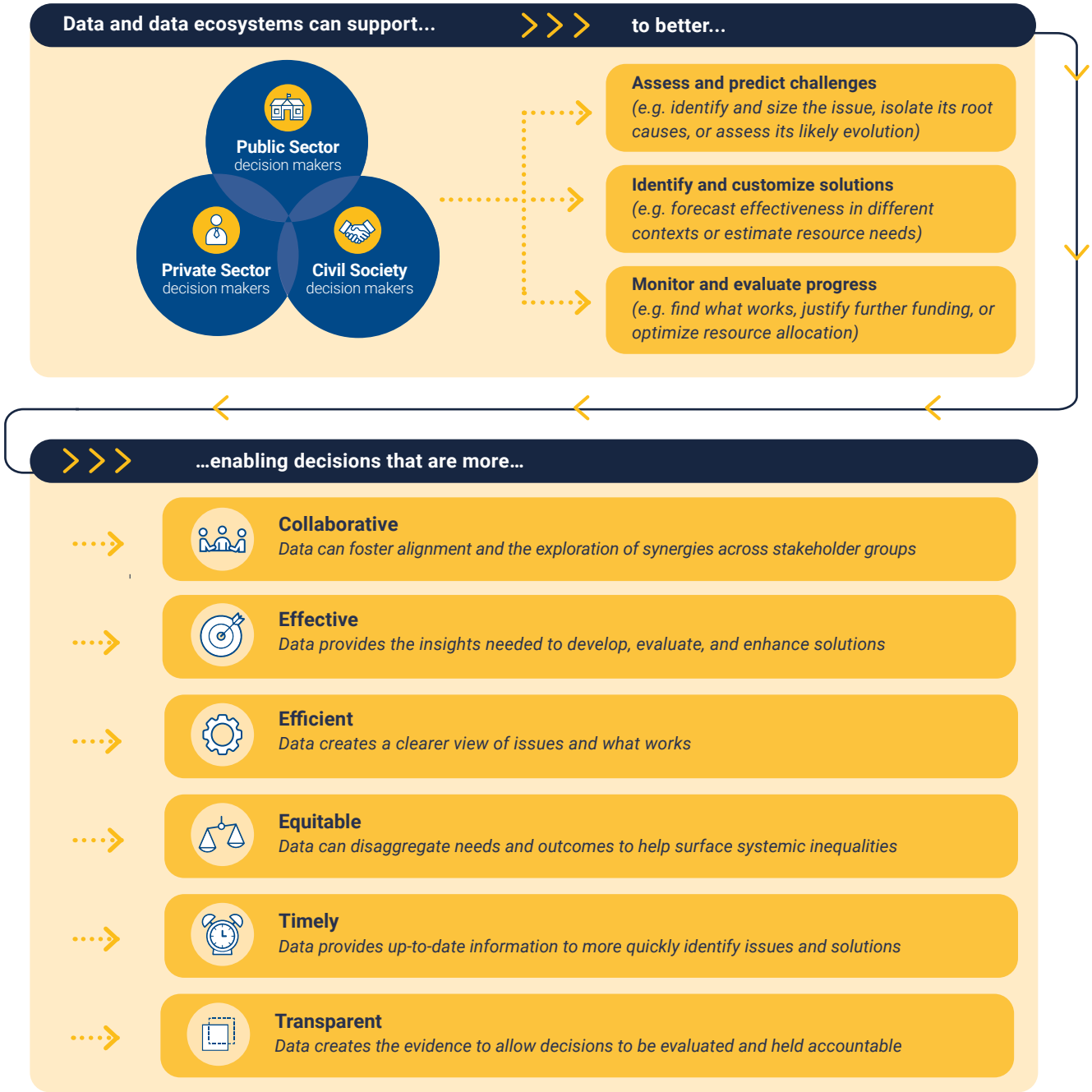
**However, the deployment of these resources has not been optimally efficient or equitable, in part due to information gaps that leave decision makers flying blind.** For example, low- and middle-income countries (LICs and MICs) often do not have the data needed to make informed decisions about how to mitigate or prepare for climate change, and lack the capacity and institutions to produce it.[6] Even in countries where these capabilities are more entrenched, data gaps remain a challenge – a recent survey of leading UK banks and insurers found that insufficient data was seen as the foremost challenge in addressing climate risk over the next five years.[7] Unfortunately, these data gaps often have a direct impact on the effectiveness and equity of outcomes. For instance, throughout the COVID-19 pandemic, a lack of testing data and persistent gaps in both coverage and segmentation (e.g., in racial and gender terms) have severely limited the ability of many governments to target those most in need or at highest risk.[8]

**In a time of growing crises and shrinking budgets, it is imperative that every dollar is spent in the most efficient and equitable way.** Every year, more than USD 30 billion of official development assistance (ODA) is allocated to respond to urgent needs in fragile and crisis settings,[9] and the need is surging—the past two years have seen a 40% increase in the number of people requiring urgent humanitarian support.[10] Development needs are likewise growing rapidly; as of 2021, the gap in funding needed to reach the UN's Sustainable Development Goals (SDGs) is set to increase by up to 70%.[11] Given this reality, even small improvements in how resources are allocated could translate into sizable cost savings and significant impact returns.

**Data ecosystems offer an outstanding opportunity to maximize resource efficiency and impact, enabling decision makers to improve lives and livelihoods by better understanding the world around them and acting in more effective and targeted ways.** Strong data ecosystems are built around a community of public, private, academic, international, and civil society actors that come together to develop, validate, and use data to inform their decisions (see *Chapter 3*).[12] These data ecosystems provide decision makers with the information needed to assess and predict challenges, identify and customize solutions, and monitor and evaluate real-time progress. In short, data ecosystems enable decisions that are more collaborative, effective, efficient, equitable, timely, and transparent (see *Figure 1*). In turn, better decisions create tangible positive impacts on lives and livelihoods, including lifting living standards in the most urgent and high-need settings (see *Spotlights 1, 2, and 3*).[13] If we are to make the best use of every dollar to reduce suffering and put the world on course for equitable and sustainable growth, we must bolster data ecosystems and make data-driven decision making the norm.

**2.** World Bank, Youth Literacy Rate, 2020.
**3.** Our World In Data, Daily supply of calories per person, 2018.
**4.** World Bank, Life expectancy at birth, 2020
**5.** World Bank, Poverty headcount ratio, 2018.
**6.** Lacuna Fund, Climate Change, 2022.
**7.** Willis Towers Watson, Net Zero strategy must use all available levers, 2021.
**8.** Science, 'Huge hole' in COVID-19 testing data makes it harder to study racial disparities, 2020.
**9.** The Global Partnership, Unlocking Data For A Better, Greener, Safer Future, 2022.
**10.** UN, UN and partners launch Complex Risk Analytics Fund, 2021.
**11.** OECD, Global Outlook on Financing for Sustainable Development, 2021.
**12.** Note, from here on, references to "data ecosystems" assume that the relevant system is effective and fit-for-purpose unless otherwise indicated—for more information on data ecosystems and their components, please refer to *Chapter 3* and the concept of Integrated National Data Systems outlined in World Bank, Data for Better Lives, 2021.
**13.** For more information on the benefits of data, please see: The Global Partnership, Value of Data, 2018; UN, Our Common Agenda, 2021; UN, Secretary-General's Data Strategy, 2020; and World Bank, Data for Better Lives, 2021.

*FIGURE 1: HOW DATA ECOSYSTEMS CAN DRIVE BETTER DECISION MAKING*




**This will only get easier – ongoing advances in our ability to harness and apply data are creating new opportunities to better target resources and create even more transformative impact.** For example, basic improvements in our ability to combine different types of data have eroded historical silos. In Kenya, merging administrative, crowdsourced, and private sector data enabled the National Police to generate 100 new variables for identifying road crash hotspots and focus its road-safety interventions on just 150 kilometers of Nairobi's 6,200-kilometer road network.[14] Analytical advances have generated similar benefits. Passive data collection through the "internet of things" (IoT) can grant access to unprecedented volumes of low-cost information, fueling analytical innovations (e.g., machine learning) to produce new and deeper insights that transcend sectoral boundaries.[15,16] These advances have the potential to drive sustainable development, with recent estimates suggesting that artificial intelligence (AI) has the capacity to positively impact 80% of the SDG targets through everything from matching electricity demand with renewable energy supplies to helping pinpoint sources of inequality.[17]

**14.** World Bank, Administrative and Monitoring Data, 2020.
**15.** Ryax, IoT and sustainable development, 2020.
**16.** McKinsey, What AI can and can't do (yet) for your business, 2018.
**17.** Nature, The role of artificial intelligence in achieving the Sustainable Development Goals, 2020.

**Fully realizing the benefits of data ecosystems will require investments in their foundational components** (i.e., governance, institutions, infrastructure, economic policies, and cross-cutting enablers—see *Chapter 3* for more detail).[18] Establishing strong pillars improves the ability of a data ecosystem to produce, process, store, share, and analyze data. The returns from investment in system foundations are huge. For instance, building data infrastructure, such as internet exchange points (IXPs) and co-location centers, can enable states to share and store data at a far lower cost, reducing the price of broadband by 85–97% and opening the floodgates to far larger data flows capable of sustaining analytical innovations like AI.[19] Similarly, investments in the creation of effective governance frameworks (e.g., open data standards or access to information rights) and institutions (e.g., data protection or antitrust authorities) can improve the quality, coverage, safety, and usability of all data outputs within their remit, creating better data that enables better decision making. Investments in the system fundamentals, therefore, pay for themselves many times over, opening up new opportunities to save and improve lives.

**To date, attempts to optimize resource distribution by addressing critical data gaps and building data ecosystems have been held back by a lack of resources and a lack of coordination between donors.** Data ecosystems offer high returns (see *Chapter 2)* and have modest needs, equating to less than 0.2% of the annual SDG funding gap.[20,21] Nonetheless, negative perceptions and misconceptions surrounding investments in data ecosystems (e.g., that data ecosystems are too removed from creating impact) have held back much-needed action. As a result, most states face serious SDG data gaps, such that no country reported data on more than 90% of the SDG indicators between 2015 and 2019 and the average country reported on only 55% of indicators.[22] Moreover, when funding is available, it often lacks coordination. For example, in Nigeria, insufficient information-sharing and coordination led seven different funders to sponsor ten separate efforts to update the country's Master Health Facility List, with each state surveyed a minimum of four times.[23]

**With better coordination and increased investment, we have the power to transform lives and livelihoods around the world, enabling a step change in the speed, efficiency, and fairness of decision making.** As described in *Chapters 2 and 3*, investment in data ecosystems can create efficiency gains large enough to cover the costs of investment many times over and multiply the impact that each subsequent humanitarian, development, and domestic dollar achieves. Filling the funding gap thus has the potential to create a virtuous cycle in which data use accelerates headway against shared goals and sets a course for lasting progress. In turn, this catalyzes further funding for data ecosystems, sustaining the transformative trajectory.

**SPOTLIGHT #1**

## HOW EARLY WARNING DATA SYSTEMS ENABLED A FASTER EMERGENCY RESPONSE TO RECORD-BREAKING FLOODS IN BANGLADESH, AT HALF THE COST

As shown by the disastrous 2022 monsoon season in Pakistan, effective responses to flooding require data-driven planning and preparation. In 2020, Bangladesh was hit by an unprecedented "triple peak" flood that came with the second-highest flood levels since 1989 and the second longest flooding since 1998. However, early warning data systems enabled the prediction of severe monsoon flooding and cyclones two weeks before they hit. This timely data enabled humanitarian organizations to get ahead of the crisis by distributing funding to emergency responders just four hours after the forecast.

By the time the water reached life-threatening levels, over 220,000 people had already received a wide variety of assistance, from cash advances and animal feed to clean birthing and post-rape treatment kits. Moreover, by facilitating coordination on the roles of each agency and allowing them to begin procurement ahead of peak needs and price spikes, the early warning data systems helped to halve the cost of previous response efforts per person reached.

**SOURCE:** ***UNOCHA, BANGLADESH MONSOON FLOODING, 2020; UNOCHA, ANTICIPATORY ACTION TOOLKIT, 2020***

**18.** Note that, while crucial and impactful, investments in data (e.g., through new sector-specific sub-systems, cutting-edge analytical software, or visualization tools) are distinct from investments in the foundational pillars (e.g., institutions, governance frameworks, or infrastructure and economic policies) and enablers (e.g., demand, funding, human capital, incentives, and trust) of data ecosystems and, therefore, do not count towards the systems-focused percentage pledge called for by *Chapter 4* of this report.
**19.** World Bank, Data for Better Lives, 2021.
**20.** OECD, Global Outlook on Financing for Sustainable Development, 2021.
**21.** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019.
**22.** World Bank, Are we there yet?, 2021.
**23.** Master Health Facility Lists create a record of all health facilities in a country and their capacity. For more information, please refer to Ayodeji Makinde et al., Duplication of effort across Development Projects, 2018.

SPOTLIGHT #2

## HOW A REAL-TIME SMS REPORTING TOOL SAVED THOUSANDS OF LIVES BY PROVIDING A CLEAR PICTURE OF NEEDS

In 2009, information flows between the Ugandan Ministry of Health (MoH) and local health facilities were limited. As a result, nearly 80% of facilities experienced shortages of life-saving malaria medications and the response time to new outbreaks averaged almost one week.

mTRAC, an SMS-based health reporting tool, filled the information gap. It created real-time data on medication shortages, disease outbreaks, and public feedback on the quality of facilities. Collecting 10,000–16,000 SMS messages every week, mTRAC provided the MoH with a clear picture of needs in every district and introduced the reporting mechanisms needed to hold facilities accountable.

Using this data, the MoH could make more timely, effective, and efficient decisions. It rooted out USD 400,000 in misappropriated funds, halved response times to disease outbreaks, and reduced malaria medication stockouts fivefold. Despite a price tag of just USD 8 million, by creating better access to timely treatment, mTRAC saved nearly 11,000 lives, creating over USD 310 million in additional lifetime earnings. Initially envisaged as a means of preventing malaria-related deaths, mTRAC's data has since saved countless more during Ebola, Marburg, and cholera outbreaks.

SOURCE: *GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA, VALUE OF DATA CASE STUDIES, 2018*

SPOTLIGHT #3

## HOW MOBILE DATA AND GEOSPATIAL ANALYTICS SAVED THOUSANDS OF LIVES IN GHANA

Early in the COVID-19 pandemic, widespread transmission created uncertainty among local communities and overwhelmed health facilities in Ghana. The government appealed for innovative and cheap ways to produce the timely information needed to shape response campaigns.

Vodafone Ghana and FRAYM (a geospatial start-up) stepped up to the challenge. By providing de-identified call records and geospatial analytics, they helped the government understand local travel patterns and assess the effectiveness of movement restrictions on disease transmission.

Insights from this public-private partnership provided the real-time data needed to identify potential COVID-19 hotspots before they arose. This enabled the government to implement and monitor lockdowns more effectively, fostering a ~50% fall in mobility across districts in Greater Accra, and helping to save thousands of lives via averted spread.

SOURCE: *GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA, GHANA MOBILE CALL DETAIL RECORDS, 2020*

Data ecosystems offer an outstanding opportunity to maximize resource efficiency and impact, **enabling decision makers to improve lives and livelihoods by better understanding the world around them and acting in more effective and targeted ways.**

# CHAPTER 2

## What is the opportunity?

**By enhancing the efficiency, accountability, and impact of humanitarian, development, and domestic spending, data ecosystems foster a diverse range of benefits for individuals and communities.** These benefits fall into four categories: economic, social, environmental, and institutional (*Figure 2*).

***FIGURE 2: BENEFITS OF DATA ECOSYSTEMS***


**The sections below outline how investments in data ecosystems contribute to each benefit,** spotlighting recent examples of the impact data ecosystems can achieve. It should be noted that while the cases below are each intended to highlight specific kinds of benefits, in reality, data ecosystem investments can yield multiple types of interrelated benefits at once.

## Economic benefits

**An analysis of recent investments in data shows an average economic benefit of USD 32 for every dollar invested.**[24] The economic benefits of these initiatives, in terms of the direct value generated, ranged from USD 7 to USD 73 per dollar spent – reflecting variations across sectors and country income levels. Accounting for the prevalence of these factors across the investments, this yields a global average of USD 32 : 1.[25] This means that fully funding data ecosystems over the next eight years could create approximately USD 1.4 trillion in additional value over the same period – equivalent to the total funding need of SDG 3 (Good Health and Well-Being) for four to six years.[26,27]

**Data creates direct financial value by capturing new economic and impact opportunities.** For example, WHO increased the lifetime earning potential of community members in remote parts of South Sudan by introducing disease surveillance systems. Each surveillance kit costs just USD 15,000 and generated over USD 340,000 in estimated lifetime earnings that would have been lost to higher morbidity and mortality.[28] Data also supports the discovery of new opportunities. By partnering with Planet, a private provider of satellite data, the New Mexico State Land Office identified over 50 cases of natural resource violations (e.g., illegal removal of material from mining sites or improper waste disposal) within just a year. As a result, it was able to generate an additional USD 800,000 in annual resource leases for the State Administration, as well as nearly USD 3 million in fines, which it reinvested in public schooling and healthcare (see *Spotlight 9*).

**24.** ROI estimate is based on modeling of discrete case examples of data investments and their impacts. See Annex 1 for information on the approach and limitations of this analysis.
**25.** Given the variation of economic returns across sectors and income groups, this weighted average accounts for the relative size of each sector in national economies and the total number of countries in each income group.
**26.** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019.
**27.** Development Initiatives, The cost of achieving SDG 3 and SDG 4, 2020.
**28.** WHO, Success stories of WHO in the African region, 2018.

**Meanwhile, data can also enable real cost savings, through everything from process efficiencies to reduced corruption.** Tangible examples of data-driven cost savings are numerous. In 2020, an early warning weather system in Bangladesh enabled a faster emergency response to record-breaking floods at half the cost, generating savings of over USD 5 million (see *Spotlight 1*). Similarly, in Uganda, data on the spread of banana wilt helped farmers avoid over USD 15 million in yield losses (see *Spotlight 4*).[29] And, in 2017, the US Atlantic City Police Department saved the criminal justice system USD 2 million by using geospatial modeling to better target and reduce the number of violent crimes (see *Spotlight 5*).

SPOTLIGHT #4

## HOW CROWDSOURCED DATA SAVED UGANDA'S BANANA INDUSTRY, AVOIDING OVER USD 15 MILLION IN PRODUCTION LOSSES

Banana bacterial wilt (BBW) is a rapidly spreading disease that can cause yield losses of up 90%, wiping out banana crops worth up to USD 360 million each year in Uganda. In 2013, the country was hit by a BBW epidemic.

In response, UNICEF's community polling project (U-Report) mobilized a network of over 330,000 volunteers, using SMS to understand the prevalence of BBW, provide information on how to control the infection, and monitor how well government initiatives were mitigating its spread.

Not only was U-Report able to provide the government with unprecedented data that enabled a more timely and effective response, but within just five days of the first SMS being sent out, U-Report had also alerted 190,000 Ugandans to the disease and how to save bananas on their farms. As a result, U-Report was able to avert over USD 15 million in potential losses at a cost of just USD 230,000 – yielding a return on investment of USD 67 per dollar spent. Given that bananas provide up to 27% of the population's calorie intake, this timely response also had an important impact on local food security.

SOURCE: *WORLD BANK, HOW YOUTH SAVED BANANAS IN UGANDA, 2014; U-REPORT, BANANA BACTERIA WILT DISEASE, 2013; THE GUARDIAN, OPEN DATA: HOW MOBILE PHONES SAVED BANANAS IN UGANDA, 2015*

SPOTLIGHT #5

## HOW GEOSPATIAL MODELING REDUCED VIOLENT CRIMES BY A THIRD IN ATLANTIC CITY, USA

Atlantic City was ranked the 8th most dangerous city in US in 2015. The city's financial difficulties and a reduced police force further exacerbated the crime rate, creating issues for citizens and businesses.

To better predict and prevent crime, the police department used "risk terrain modeling". Without targeting individuals or demographic groups, this geospatial approach uses environmental factors that can create opportunities for crime (e.g., parks or certain types of retail) to forecast crime in different parts of the city. This enables police to devise risk-prevention strategies that address relevant factors and make the environment less conducive to crime.

The model identified various environmental factors and was used to inform new citywide policing and community relation strategies. As a result, at a cost of just USD 80,000, the geospatial modelling helped to save the local criminal justice system USD 2 million between 2016 and 2017, reducing robberies by 37% and homicides and shootings by 26%. Accounting for both the justice system savings and the estimated increase in lifetime earning potential for those whose lives were saved through averted homicides, this initiative yielded an aggregate estimated return of USD 34.2 for each dollar invested.

SOURCE: *GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA, ATLANTIC CITY CASE STUDY, 2018*

**29.** Banana bacterial wilt causes plants to rot from inside out, leading to yield losses that can be as high as 90%.

**What's more, data drives additional indirect benefits that further increase the economic value of data systems investments.** Data and data systems are inherently reusable, creating numerous new business and impact opportunities and in turn leading to further downstream benefits. These ripple effects enable data and data systems to provide indirect economic benefits far beyond the directly measured impact. For example, through the Landsat program, NASA and the U.S. Geological Survey manage a series of Earth-observing satellite missions. The data created provides free information that is used to improve decision making across a variety of sectors including agriculture, climate conservation, and health. In each of these areas, the decisions and impact it enables ripple outward. For instance, Landsat is used to monitor agricultural outputs and inform decisions on fertilization, irrigation, and crop rotation. But better agricultural decisions do not only increase yields and support farmers themselves—they also lead to enhanced food security and nutritional outcomes for local communities, and improved environmental outcomes and sustainability for the land itself. In addition, it supports more accurate crop insurance evaluations and reduces fraud.[30] All of these come with additional financial and non-financial value that combined far exceed the direct changes in income for the farmers themselves (see sections below for more on non-financial benefits). Even further, every dollar going into local economies will circulate onward, creating a fiscal multiplier that further increases the total value of every dollar generated.

## Social benefits

**Data can save lives and enhance the quality and equity of living standards by increasing the effectiveness of social programs.** By providing information on people's needs and the progress of public initiatives, data can improve the effectiveness and timeliness of decision making across humanitarian and development sectors. Examples below spotlight the potential impacts in public health, education, and agriculture and food security.

**Data can increase life expectancy by providing health systems with timely and disaggregated information on beneficiaries' needs and outcomes.** This allows health authorities to launch more efficient and effective prevention and response programming that better targets their recipients' basic needs. For example, in 2010, Bangladesh's Thakurgaon district introduced a system to collect data on maternal and neonatal deaths, with the aim of informing remedial action at the community level. As a result of the information collected, local authorities were able to deploy multiple initiatives and avoid hundreds of maternal and child deaths (see *Spotlight 6*).

**Data can also improve education outcomes.** For example, in Guatemala, the Ministry of Education curbed school dropouts by using data to identify the causes of lower enrollment rates, target at-risk students, and design effective solutions. Collectively, these investments in data cost just USD 650,000 and prevented thousands of dropouts, providing an estimated economic benefit of nearly USD 14 million—or more than USD 20 for every dollar invested.[31,32] Likewise, in Pakistan, publicizing regular evaluations of school performance has proven a cost-effective means of holding educators accountable, driving better learning outcomes, increasing enrollment, and lowering private school fees.[33]

**In addition, data can help inform anticipatory action by predicting humanitarian emergencies.** Natural disasters can have devastating effects on food security (among many other consequences). Recent initiatives have proven that data ecosystems can effectively forecast extreme weather to predict natural disasters and inform anticipatory actions that mitigate damage to crops and livestock. In 2020, Somalia's food-insecure population was projected to triple to 3.5 million people due to the cumulative impact of locust infestations, floods, and COVID-19. When predetermined trigger levels for oncoming food shortage were breached, the anticipatory action framework automatically mobilized over USD 200 million for proactive response efforts, helping prevent 500,000 people from needing later emergency aid. In particular, the initiative allowed the timely expansion of coverage against the locust infestation, saving enough food to feed over 130,000 people for a year.[34]

**30.** NASA and USGS, Landsat's critical role in agriculture, 2022.
**31.** World Bank, Strengthening statistical capacities to tackle school dropout in Guatemala, 2020.
**32.** Dalberg analysis—see Annex 1 for details on the ROI estimation methodology.
**33.** Andrabi et al, Report cards: The impact of providing school and child test scores on educational markets, 2017.
**34.** UN CERF, Somalia anticipatory action against food insecurity, 2020, 2021.

SPOTLIGHT #6

## HOW SURVEILLANCE OF NEONATAL MORTALITY HELPED PREVENT 1 IN 5 MATERNAL DEATHS IN BANGLADESH

In 2010, maternal and neonatal disorders caused more than 82,000 deaths in Bangladesh. However, because no system existed for reviewing the causes of these deaths, the potential to introduce effective initiatives was limited.

In response, Bangladesh's Thakurgaon district introduced a system for monitoring and reviewing the causes of maternal and perinatal deaths. With the aim of informing action plans at the community and facility level, the system gathered, aggregated, and analyzed death reports from health providers and validated the causes of death reported from autopsies.

The monitoring and review system informed the district's remedial programming, targeting new efforts to more effectively prevent locally-prevalent causes of death. These efforts reduced maternal deaths by 20% and neonatal deaths by 8%. Due to this success, Bangladesh scaled up the program to cover over two-thirds of the population by the end of 2019.

SOURCE: *OPEN DATA WATCH, MATERNAL AND PERINATAL DEATH REVIEW, 2018; UNFPA BANGLADESH, MPDSR IN BANGLADESH, 2020; GBD, GLOBAL BURDEN OF DISEASE DATA, 2022*

SPOTLIGHT #7

## HOW INCLUSIVE DATA SYSTEMS SUPPORTED GIRLS' EDUCATION IN SIERRA LEONE

The 2014, an Ebola outbreak in Sierra Leone caused a wave of school closures. Pre-closure, the average girl already acquired less than half a year of schooling for every year acquired by boys. On reopening, female enrollment rates declined by a further 16%.

Sierra Leone's lack of gender disaggregated education data made it difficult to identify solutions to improve female enrollment. In response, the Ministry of Education included questions on disability and gender in its annual education census, generating disaggregated data at the district level.

Consequently, the Ministry was able to clearly see how female enrollment rates differed across each of the country's sixteen districts. This enabled the government to target those most in need—for example, by building a girls' school and Center of Learning and Teaching Excellence in Port Loko, the country's second most populous district, which had one of the highest teen pregnancy rates and lacked a girls' school.

SOURCE: *GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA, SUPPORTING GIRLS EDUCATION, 2021; MALALA FUND, GIRLS' EDUCATION AND COVID, 2021*

## Environmental benefits

Data can bolster the transition to sustainable development, shedding light on the drivers, rate of change, and impacts of environmental issues. In turn, this enables more targeted solutions that support communities to better mitigate environmental risks and adapt to changing conditions. For example, in Australia, Google's Environmental Insights Explorer (EIE) utilizes geospatial and satellite data along with modelling capabilities to produce estimates of the greenhouse gas emissions created by different activities in each Australian State. As a result, state governments have been able to identify and work with the larger emitters in their remit to reduce their footprints.[35] Likewise, in Brazil, firefighters have integrated a fire alert system that facilitates the monitoring of forest fires using Global Forest Watch's real-time satellite imagery data. By using this system, firefighters have been able to reduce their response time to forest fires from 36 hours to 4 hours (see *Spotlight 8*).

**35.** Google Earth, Australian councils and communities look to data for solving climate change challenges, 2020.

SPOTLIGHT #8

## REDUCING DEFORESTATION IN BRAZIL AND INDONESIA USING SATELLITE IMAGERY

Between 2001 and 2011, Brazil and Indonesia lost 45 million hectares of tree cover—the equivalent of 25 gigatons of $CO_2$ emissions.

Launched in 2011 to curb global deforestation, the Global Forest Watch (GFW) utilizes satellite imagery and crowdsourced data to monitor forest loss in real time. As an open-sourced platform, GFW enables a variety of users to access content on everything from forest fires to illegal mining.

GFW data has enabled new levels of transparency, informing multiple advocacy campaigns, contributing to legal proceedings against illicit logging, and reducing firefighter response times by nearly 90% (from 36 to 4 hours). As a result, GFW has played a considerable part in curbing deforestation, which recently decreased by 18% in Brazil and reached its lowest level in a decade in Indonesia.

SOURCE: *: GLOBAL FOREST WATCH, DASHBOARD, 2022; ODIMPACT, STOPPING DEFORESTATION IN BRAZIL AND INDONESIA, 2017*

SPOTLIGHT #9

## HOW PRIVATE SATELLITE DATA HELPED TO CURB THE DEGRADATION AND THEFT OF NATURAL RESOURCES, GENERATING MORE THAN USD 3 MILLION IN PAYMENTS

The New Mexico State Land Office (NMSLO) oversees 22 million acres of land and subsurface minerals. Given the vast and remote nature of the area under its supervision, field visits focused on discovering activities outside of lease terms are costly and challenging.

Working with Planet, a private provider of satellite imagery, NMSLO leveraged daily imagery to monitor the leased lands and achieve greater compliance. The wide coverage of Planet's satellite data allowed NMSLO to identify illegal removal of materials, activities outside of lease terms, and improper waste disposal.

Planet's high frequency and real time insights empowered quick action from NMSLO, identifying 53 theft cases within a year. Out of these, 22 have been converted into new leases, producing additional USD 800,000 in net revenues. USD 2.7 million in back payments has also been generated, which is being distributed to institutions such as public schools and hospitals.

SOURCE: *NMSTATELANDS, PROTECTING PUBLIC LANDS CASE STUDY, 2020*

## Institutional benefits

**Public data systems support evidence-based decision making and provide the information needed to hold institutions accountable.** This creates institutions that are more responsible, responsive, representative, and reliable to the societies they serve. For example, in the Philippines, the introduction of a new biometric ID system is set to create over USD 6 billion in cost savings by improving the coordination of government spending, at the same time as enabling over 10 million Filipinos to newly or more reliably access financial services (see *Spotlight 10*).

**In addition, data ecosystems can enhance dialogue between leaders and their people, strengthening accountability and the democratic process.** Increased transparency in leaders' actions and decision making processes keeps citizens and civil society informed, and ensures decisions are taken and implemented effectively and equitably. Relatedly, improved data ecosystems provide a communication channel through which people can elevate their own views and needs. For example, in Ukraine, the introduction of transparent public procurement processes in 2014 led to a 50% increase in the number of new businesses bidding for government contracts. Consequently, the government was able to save USD 1 billion while providing more reliable, inclusive, and accountable public services (see *Spotlight 11*).

**SPOTLIGHT #10**

## UNLOCKING BILLION-DOLLAR SAVINGS THROUGH AN INTEGRATED BIOMETRIC ID SYSTEM IN PHILIPPINES

Until 2018, 33 different identification cards were issued by various government agencies, resulting in duplication of efforts and corruption in social protection programs.

In 2018, the Filipino government assigned the coordination of the PhilSys national ID system to a single agency and provided every citizen with a unique identification number that is linked to a variety of datapoints, including their passport, driving license, biometrics, and health data.

By improving the coordination of government social spending, PhilSys is expected to create USD 6 billion in cost savings over the next five years, at the cost of only USD 462 million over the same period. But PhilSys's benefits are not limited to government spending. Over 25% of the 51 million unbanked Filipino cite a lack of official documents as a hurdle to opening a bank account—a challenge that PhilSys is rapidly addressing.

**SOURCE:** : *GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA, IMPROVED DATA GOVERNANCE LEADS TO BETTER ECONOMIC OUTCOMES FOR PHILIPPINE CITIZENS, 2018; BUSINESSWORLD, MORE THAN 37 MILLION RECEIVE NATIONAL ID, 2021; BANGKO SENTRAL NG PILINAS, FINANCIAL INCLUSION SURVEY, 2019*

**SPOTLIGHT #11**

## HOW A MORE TRANSPARENT ELECTRONIC PROCUREMENT PLATFORM REDUCED CORRUPTION AND SAVED THE UKRAINIAN GOVERNMENT USD 1 BILLION

In 2014, high levels of perceived corruption and abuse of power in the Ukrainian public sector led Ukraine to enact a series of reforms to increase the transparency of public procurement.

In collaboration with partners from the private sector, civil society, and government institutions, Ukraine piloted ProZorro and DoZorro—an e-procurement platform that makes all procurement documents publicly available and a channel for citizens to report violations.

At a cost of just USD 14 million, in its first two years of operation, ProZorro saved the public sector USD 1 billion by reducing the value of contracts by up to 25%. In addition, 80% of surveyed businesses witnessed reduced corruption and the number of new businesses bidding for contracts increased by 50%. At the same time, DoZorro increased the accountability of public procurement—citizens reported over 30,000 violations that, in some cases, led to sanctions and criminal charges.

**SOURCE:** *OPEN GOVERNMENT PARTNERSHIP, THROUGH THE POWER OF THE PEOPLE, EMPOWERING CITIZEN WATCHDOGS, 2021; OGP, UKRAINE'S INSPIRING JOURNEY IN THE OPEN GOVERNMENT PARTNERSHIP, 2022*

## Benefits of coordinated investments in data ecosystems

**The spotlights and figures above tend to be isolated, sector-specific initiatives that focus on data, rather than the foundational pillars of data ecosystems.** While crucial and impactful, investments in data (e.g., through new sector-specific sub-systems, cutting-edge analytical software, or visualization tools) are distinct from investments in the foundational pillars (e.g., institutions, governance frameworks, or infrastructure and economic policies) and enablers (e.g., demand, funding, human capital, incentives, and trust) of data ecosystems.[36]

**Coordinated approaches to investing in data ecosystems can provide even greater returns than investments in data alone.** Above, we estimate that investments in data alone can yield economic benefits ranging from USD 7 to USD 73 for ever dollar invested—along with a range of social, environmental, and instutitonal benefits. The returns of coordinated investments in data ecosystems are likely to be far higher, for three reasons:

1. **Investments in the foundational pillars and enablers of data ecosystems can help improve the quality, coverage, safety, and usability of all data created within the system –** For example, investments in the creation of effective economic policies (e.g., safeguards for innovation), governance frameworks (e.g., cybersecurity laws), or institutions (e.g., data protection authorities) will improve all outputs within their remit. By comparison, one-off investments in data improve only the sector, sub-system, or data gap for which they were designed.

2. **Most investments in data ecosystems provide a combination of economic, social, environmental, and institutional benefits –** Coordinated investments in data ecosystems are highly cost-effective because the same foundational data (e.g., censuses or household surveys) can be reused for many different purposes. By comparison, the USD 32 average figure above is based on the economic returns of more isolated investments, without these benefits of shared infrastructure.

3. **Taking a coordinated approach to systems investment will reduce funding inefficiencies –** As highlighted in *Chapter 1,* duplicated investments and parallel systems are common; in Nigeria, seven different funders sponsored ten separate efforts to update the country's Master Health Facility List, with each state surveyed a minimum of four times.[37,38] More coordinated spending means lower costs to achieve the same or better outcomes.

**36.** For more detail, see Chapter 3 below, as well as World Bank, Data for Better Lives, 2021 and Table 1 in Calleja and Rogerson, Financing challenges for developing statistical systems, 2019.
**37.** Ayodeji Makinde et al., Duplication of effort across Development Projects, 2018.
**38.** For more detail on the benefits of data, please see: The Global Partnership, Value of Data, 2018; UN, Our Common Agenda, 2021; UN, Secretary-General's Data Strategy, 2020; and World Bank, Data for Better Lives, 2021.

# CHAPTER 3

## How do we chart a path to progress?

**Maximizing the effectiveness of humanitarian, development, and domestic funding means supporting efforts to build the foundations and enablers of data ecosystems.** If data were produced by factories, data ecosystems would be the buildings in which public, private, academic, international, and civil society actors come together to develop, validate, and use data.[39] As shown in *Figure 4,* four foundational pillars (institutions,[40] governance frameworks,[41] infrastructure, and economic policies[42]) provide the basis for an effective and efficient system. In turn, these foundations are supported by a groundwork of enablers, which provide the inputs needed to sustain the pillars (i.e., demand, funding, human capital, incentives, and trust).

**Adequate foundations and enablers are essential to sustainably and consistently creating high-quality, high-coverage data that is safe and easy to use.**[43] Without these foundations, the whole factory (i.e., the data ecosystem) and its outputs can become unproductive (e.g., due to slow and expensive data infrastructure) and potentially even dangerous (e.g., due to a lack of data governance to protect against confidentiality breaches, bias, or access inequalities).[44]

### *FIGURE 3: PARTICIPANTS IN DATA ECOSYSTEMS*

| | |
|---|---|
| **Academia** | Academic institutions, think tanks, and research entities often produce data to guide and evaluate policy or inform and influence media and the public debate.<br><br>Crucially, academic institutions also provide important **education and training for data users and producers in government, the private sector, and civil society.** |
| **Civil Society and Individuals** | Civil Society Organizations, NGOs, youth groups, local communities, the media, and individuals play a critical role as the producers and users of data that **hold governments and the private sector accountable and highlight issues of public concern.**<br><br>This accountability function can act as a check on official data, fill gaps in coverage, or otherwise complement public intent and private intent data. |
| **Government Entities** | Government entities are the primary **producers** of public intent data (e.g., censuses or household surveys) and act as data **stewards** that are responsible for laying out and enforcing quality standards, ensuring data accessibility and protection, setting out the rules that govern data use, and ensuring secure data transactions. |
| **International Organizations** | International organizations often develop methods and tools, acting as **standard setters** that ensure the international comparability and quality of data.<br><br>In LICs and MICs, international organizations also frequently act as **donors** that support data production (e.g., through funding or by offering effective mechanisms for creating economies of scale in data and statistical capacity). |
| **Private Sector** | Firms are **prolific producers of data and drivers of innovation,** creating and housing much of the ecosystem's cutting-edge technology, capacity, and knowledge.<br><br>Much of the data produced by the private sector is **valuable to the other participants.** The relationship is reciprocal – firms routinely rely on public data to improve business decisions or to create new products and services. |

***Note:*** *All participants contribute to the ecosystem by producing and/or using data. The descriptions above are intended to highlight distinct roles and ways in which each actor's contributions to the ecosystem differ.*

**39.** For more information, please see World Bank, Data for Better Lives, 2021.

**40.** A community of local, regional, and national public, private, and civil actors that implement (i.e., collect, process, store, share, analyze, use, archive, preserve, and destroy) and govern (i.e., set policy objectives, develop supporting rules and standards, enforce compliance, and improve governance through learning and evaluation) the data lifecycle.

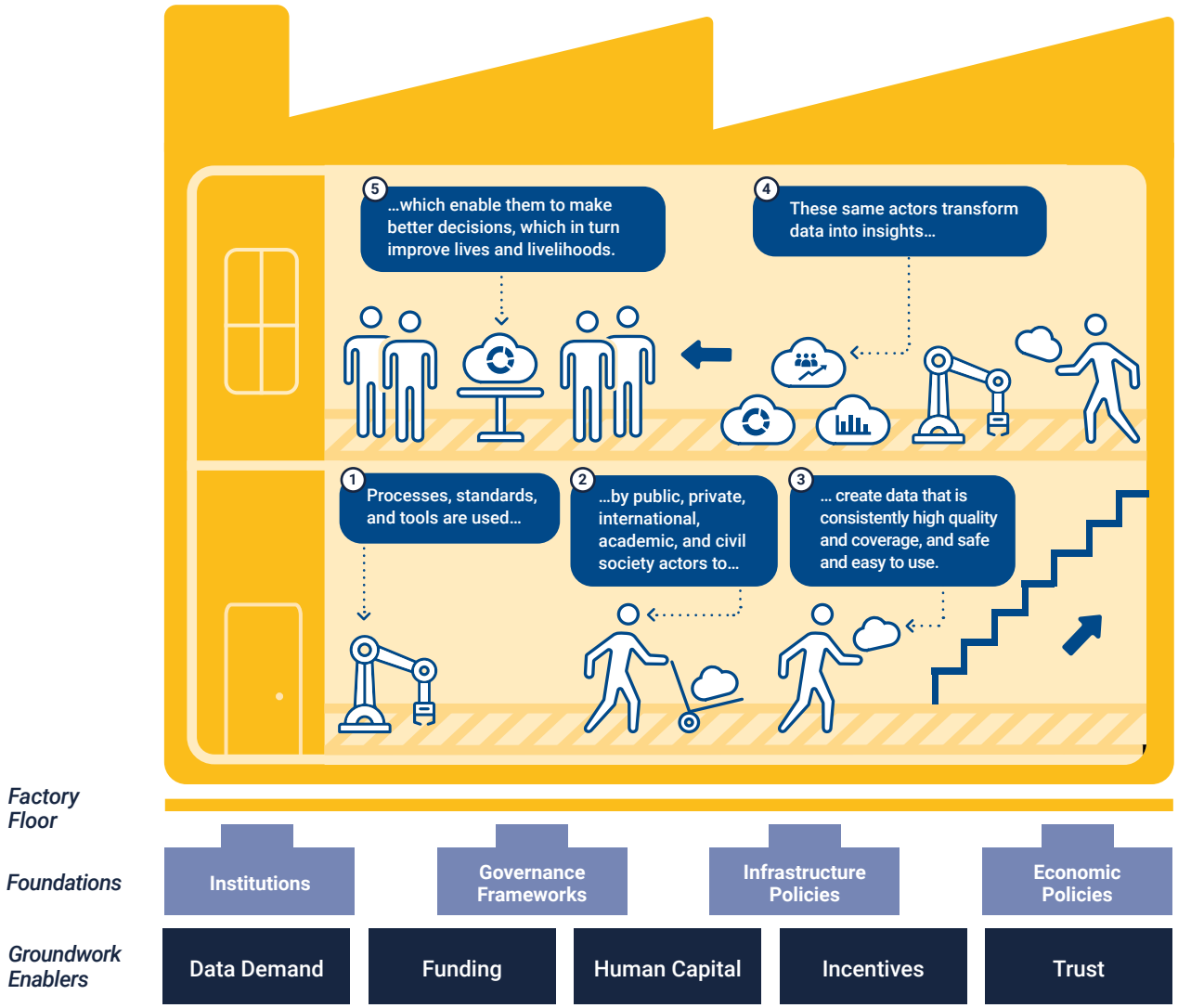
**41.** Laws and regulations that promote trust in, facilitate access to, and optimize the social and economic value provided by data.

**42.** Hard and soft infrastructure designed to enable the equitable and trustworthy production, processing, flow, and use of data, alongside economic policies that foster data-driven markets and whole-of-government approaches to data governance, management, and use.

**43.** Effective integrated data ecosystems enable the production of data that is high coverage (i.e., sufficiently complete, timely, and frequent to tackle real-time issues), high quality (i.e., sufficiently granular, accurate, and comparable to enable targeted solutions and effective MEL), easy to use (i.e., sufficiently accessible, understandable, and interoperable to enable deeper and more widespread analysis and reuse beyond its initial creators), and safe to use (i.e., sufficiently impartial, confidential, and appropriate to create the trust needed to enable its collection and use).

**44.** For more information on the enablers, components, outputs, and current state of integrated national data ecosystems, please see World Bank, Data for Better Lives, 2021.

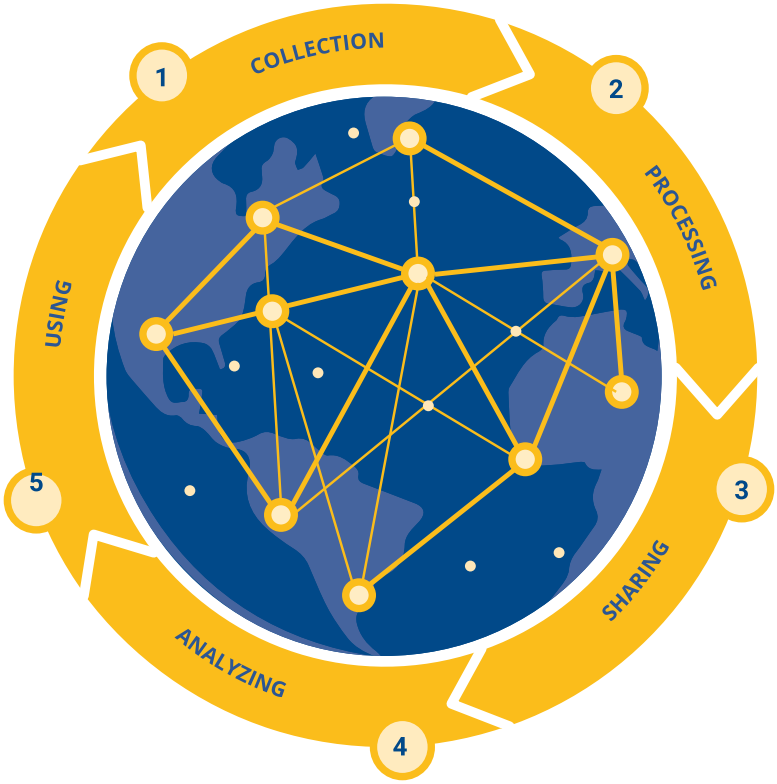
*FIGURE 4: THE ENABLERS, FOUNDATIONS AND ACTIVITIES OF DATA ECOSYSTEMS*




**To realize the full value of data, we must strengthen these foundations and enablers at a national, regional, and international level.** Integrated national data systems are the building blocks of regional and global systems.[45] As a result, efforts to strengthen their foundations and enablers will also support the creation of robust regional and global data ecosystems. Investing in national ecosystems alone, however, is not enough to create strong regional and global systems—governments, international organizations, and donors will need to reinforce these same pillars and enablers in regional and international systems. Combining these top-down and bottom-up approaches creates a web of interlinked ecosystems—akin to each national data "factory" contributing to a broader global supply chain (see *Figure 5)*. These come together to enhance data integration, interoperability, and use at scale, creating cross-cutting insights that catalyze greater improvements in lives and livelihoods.

**45.** For more information on the concept of Integrated National Data Systems see World Bank, Data for Better Lives, 2021.

*FIGURE 5: GLOBAL NETWORK OF INTERCONNECTED DATA ECOSYSTEMS*




**Most LICs and MICs are still assembling the foundational pillars or starting to initiate data flows and, as a result, have an array of opportunities to bolster their national data ecosystems.** For example, large numbers of LICs and MICs have room to greatly improve the direction, coverage, capacity, and independence of their data institutions. Nearly 90% have yet to create a standalone national data strategy and over 40% have yet to establish a cyber security agency.[46] Equally, many LICs and MICs would benefit from new elements of good governance practice, enhancing the inclusivity of data ecosystems and equity of data outcomes (e.g., 45% are yet to provide individuals with the right to rectify, or object to the use of, their personal data). Addressing these needs will also require states to bolster the enablers of integrated national data systems. Again, this will require focus in multiple areas, from increasing trust in data[47,48] (e.g., recent surveys show that 17–25% of ministry officials lack confidence in national data) to building staff capacity and capabilities (e.g., 83% of African and Asian National Statistics Offices (NSOs) stated strengthening human resources was one of their top five goals).[49,50]

**Supporting these countries in building effective data ecosystems will require governments and donors to take a more coordinated funding approach that prioritizes systemic, cross-sector issues.** Existing resource tracking mechanisms do not consistently provide a clear picture of the volume and nature of funding for data.[51] While a number of efforts underway represent important steps toward enhanced transparency,[52] much more remains to be done. This lack of visibility contributes to the fragmentation of donor investments, which are frequently disjointed, duplicative, and concentrated in one-off initiatives, such as sector-specific sub-systems, cutting-edge analytical software, and visualization tools.[53] While these may be impactful in their own right, the system fundamentals are too often left behind.[54] Under the factory analogy, this is equivalent to investing in additional floors, heavy-set machinery, and new product packaging, while forgetting to reinforce and expand the foundations to cope with the growing weight and instability of the system. As outlined above, this could lead the system to become unproductive and dangerous. Building durable and efficient systems, therefore, requires more coordinated funding to ensure that the enablers and pillars of integrated data ecosystems are in place and functioning effectively.[55]

**46.** World Bank, Data for Better Lives, 2021.
**47.** The Global Partnership, Reimagining Data and Power: A roadmap for putting values at the heart of data, 2022.
**48.** World Bank, Data for Better Lives, 2021.
**49.** Sethi and Prakash, Counting on Statistics, 2018.
**50.** PARIS21, New approaches to Capacity Development and Future Priorities, 2018.
**51.** UN Decade of Action, Data Financing Trends - Tracking aid for data and statistics, 2022.
**52.** Of particular relevance is ongoing work by: the Bern Network Clearinghouse, GPEDC, OECD Data Profiles, and PARIS21, among others
**53.** For example, Somalia is able to cover only 40% of the SDG indicators, in part, because much of its development data is held by over 20 international organizations, encased in unpublished evaluations of isolated and often parallel donor programs. For more information, see Clearinghouse, Somalia Works Towards Full Production of SDG Indicators, 2022 and Somalia Public Agenda, Who owns data in Somalia?, 2021.
**54.** Data values digest, Making pizza without dough (or the state of funding for data), 2022.
**55.** For more information on the current state of funding for data ecosystems, please see Calleja and Rogerson, Financing Challenges for Developing Statistical Systems, 2019 and World Bank, Data for Better Lives, 2021.

**Pooled funding instruments and improved mechanisms for tracking investments in data offer a solution by institutionalizing a more holistic and cost-effective approach to building integrated data ecosystems.** Pooled funds aggregate and coordinate funding and capacity building around systemic challenges. In so doing, they enable contributors to make more effective use of their resources by sharing financial burdens, minimizing risks, and cutting transaction costs. At the same time, they also establish a more coordinated, focused, and demand-driven approach to fundraising and disbursement—bringing openness to donor contributions and aligning allocations with prioritized recipient needs. Pooled funds also reduce allocative inefficiencies (e.g., duplication), prioritize solutions to create lasting change at scale, and facilitate large-scale, consistent financing flows that provide governments with the confidence and capital they need to invest in the enablers and pillars of integrated data ecosystems.

**Yet deploying existing resources in a more coordinated manner is not enough—unleashing data's potential will necessitate greater involvement from a wider pool of donors and public and private sector actors.** Globally, data ecosystems require an estimated USD 5.6 billion per annum to achieve the objectives laid out in the Cape Town Global Action Plan for Sustainable Development Data (CTGAP).[56,57] Unfortunately, existing domestic funding meets less than 40% of the needs of LICs and MICs.[58] Donor funding, meanwhile, is stagnant at just USD 700 million per annum—significantly short of filling the gap.[59] What is more, national data ecosystems often lack the integration, across the community of participant groups, needed to establish strong ecosystem foundations and enablers.[60] For example, although private sector actors frequently hold key data repositories and essential expertise in data processing and analytics, the incentives and infrastructure needed to facilitate their participation are often missing. Putting the necessary enablers in place will therefore require a significant increase in investment, as well as commitments from a wider pool of actors. Only half of all countries fully fund their national statistical plans, and five donors account for approximately two-thirds of commitments.[61,62] Crucially, however, the total funding required equates to less than 1% of total donor and public sector spending in LICs and MICs and, as outlined in Chapter 2, data ecosystems pay for themselves by optimizing the efficiency and impact of humanitarian, development, and domestic resources.[63]

**The UN Complex Risk Analytics Fund (CRAF'd) and World Bank Global Data Facility (GDF) are pooled funds that have been designed to harmonize and catalyze investment in data ecosystems.** The funds aim to transform the lives of billions by raising and deploying over USD 500 million for data and data ecosystems. While CRAF'd will coordinate investments in risk data and analytics to enable faster and more targeted, efficient, and effective programs for people in fragile and crisis-affected settings, GDF is designed to catalyze long-term domestic and international support for integrated national data ecosystems in LICs and MICs.

**Together, the funds have the ability to reshape the funding and capacity-building landscapes by:**

- **Designing a systems-focused agenda that prioritizes the solutions needed to create lasting change at scale.** CRAF'd and GDF have the expertise to understand, build, and fix rapidly evolving and technically complex data ecosystems.[64] Not only are CRAF'd and GDF directed by program managers responsible for extensive thought leadership on data ecosystems, they are also the only funds designed to work across the entire data lifecycle, helping a wide array of stakeholders to build every enabler and pillar of integrated data ecosystems in LICs and MICs. Moreover, the funds are demand-led. They have the relationships with national stakeholders needed to co-identify ecosystem priorities and are built to facilitate and respond to country participation, ownership, and leadership. Collectively, these expertise and connections uniquely position the funds to develop a systems-focused agenda capable of delivering universal data ecosystems.

---

**56.** CTGAP provides a roadmap outlining the activities needed to modernize statistical systems and expand statistical capacity to meet future needs, including the SDGs. For more information, please refer to UNDES, Cape Town Global Action Plan for Sustainable Development Data, 2017.

**57.** Note that, while crucial and impactful, investments in data (e.g., through new sector-specific sub-systems, cutting-edge analytical software, or visualization tools) are distinct from investments in the foundational pillars (e.g., institutions, governance frameworks, or infrastructure and economic policies) and enablers (e.g., demand, funding, human capital, incentives, and trust) of data ecosystems and, therefore, are not considered to contribute to the USD 5.6 billion need outlined in CTGAP—for more information, please see Table 1 in Calleja and Rogerson, Financing challenges for developing statistical systems, 2019.

**58.** Due to a lack of adequate tracking mechanisms, it is not possible to pinpoint the volume of domestic funding for data ecosystems. However, it is safe to assume that funding is significantly below the USD 2.2 billion per year that PARIS21 estimated would be required to measure the SDGs. The vast majority of countries face serious SDG data gaps, such that no country reported data on more than 90% of the SDG indicators between 2015 and 2019, 22 reported on less than 25% of the indicators, and the average country only reported on 55% of the indicators.

**59.** PARIS21, Partner Report on Support to Statistics, 2021.

**60.** Relevant actors include the public sector, private sector, civil society, academia, and international organizations. For more information, please refer to *Figure 3* and World Bank, Data for Better Lives, 2021.

**61.** World Bank, Data for Better Lives, 2021.

**62.** PARIS21, Partner Report on Support to Statistics, 2021.

**63.** For more information on the current state of funding for data ecosystems, please see Calleja and Rogerson, Financing Challenges for Developing Statistical Systems, 2019 and World Bank, Data for Better Lives, 2021.

**64.** Includes keystone reports such as World Bank, Data for Better Lives, 2021 and the UN, Secretary-General's Data Strategy, 2020.

Adequate foundations and enablers are essential to sustainably and consistently **creating high-quality, high-coverage data that is safe and easy to use.**

- **Coordinating diverse actors around a single agenda.** CRAF'd and GDF have the partnerships needed to integrate and coordinate the diverse data funding and capacity-building landscapes. For example, as institutions hosted by the UN and World Bank, they are able to leverage connections and resources in 190 countries around the world. What is more, they have close connections with a variety of broad-based initiatives in the data space, such as the Global Partnership for Sustainable Development Data, which brings together over 600 partners, ranging from technology giants to grassroots civil society groups. Likewise, their governance structures include a wide array of country ministers and thought leaders, such as the Bern Network's Clearinghouse and UN World Data Forum. Strong relationships with this diverse network of actors enable CRAF'd and GDF to break down silos and drive effective collaboration at a local, national, and global level.

- **Delivering the resources needed to realize the agenda.** Meaningfully bolstering data ecosystems requires large-scale, consistent funding flows that are capable of providing governments with the confidence and capital needed to invest in the enablers and pillars of integrated data ecosystems. CRAF'd and GDF are well positioned to catalyze these flows. Building on their UN and World Bank platforms, the Funds have developed the advocacy capabilities needed to raise awareness of integrated data ecosystems, as well as the financial capabilities and relationships needed to identify and design investments that catalyze financing from public, private, and philanthropic actors. What is more, as pooled funds, they establish a more coordinated and focused approach to resource distribution and enable contributors to make more effective use of their capital by sharing financial burdens, minimizing risks, and cutting transaction costs.

- **Supporting on-the-ground efforts to build data ecosystems, bolstering their foundations and enablers.** For example, in addition to cultivating relationships that enable it to be demand-led, GDF is highly flexible and can reshape the way it allocates resources to maximize impact in the local context, providing everything from "first mile" diagnostics and in-kind support (e.g., technical assistance), to results-based financing and fund matching.[65] CRAF'd is equally focused on delivering practical solutions, concentrating its investment in critical risk data (i.e., real-time, high-resolution, interoperable data that helps decision makers better understand fragile settings), crisis analytics (e.g., data and models capable of predicting, diagnosing, and prescribing solutions to crises), and knowledge and capacity building (e.g., data-sharing platforms that enable faster cross-pillar responses).[66] This practical approach to capacity building optimizes the impact of investments made by the funds, ensuring that investees have the knowledge, assets, and capacity to deliver stronger data ecosystems.

**From an internal perspective, the scopes, functions, and structures of CRAF'd and GDF have also been carefully designed to ensure that their work is synergistic and additional.** CRAF'd and GDF operate in tandem, with distinct sets of responsibilities and functions that build on the capabilities and expertise of their host organizations. While GDF is designed to catalyze and coordinate financing for integrated national data systems capable of driving and monitoring sustainable development, CRAF'd strengthens these integrated systems in fragile and crisis-affected settings to support diverse stakeholders better anticipate, prevent, and respond to complex risks. The structures of the funds are also designed to maximize additionality and synergies. For example, shared board members and direct links between the project management units of the funds enable day-to-day operational and strategic alignment between CRAF'd and GDF, ensuring synergies are realized and duplication is avoided.

**Crucially, the activities of the funds will enhance the return on contributors' future investments, enabling governments and donors to improve lives and livelihoods at a lower cost.** The collective endeavors of CRAF'd and GDF will help to bolster national data governance, enhance efforts to manage the data lifecycle, close data gaps, build data literacy, and connect siloed data systems. This will create stronger data ecosystems that deliver better insights to enable better decisions. For example, by breaking down international and thematic silos and fostering real-time data for real-time issues, the funds will enable the timely identification of previously unseen cross-sector or cross-border insights. As shown by the COVID-19 pandemic, in which imported infections accounted for over 10% of the incidence of the disease in more than 100 countries, these insights have the capacity to save millions of lives.[67] Moreover, by coordinating investments and filling gaps in foundational data sources (e.g., administrative or geospatial data), the funds will reduce the potential for resource inefficiencies. As a result, each government and donor dollar will go further, facilitating a faster rise in living standards in the most urgent and high-need crisis and development contexts.

**65.** GDF's priorities include, but are not limited to, modernizing data ecosystems (e.g., by enhancing the integration and cohesion of system actors and assets), strengthening and expanding inclusive data (e.g., by filling gaps related to undercounted groups), and capacity building (e.g., by bolstering the management and governance of the data lifecycle).

**66.** For more information on the role and capabilities of GDF and CRAF'd, please see The Global Partnership, Unlocking Data For A Better, Greener, Safer Future, 2022; UN, CRAF'd, 2022; and World Bank, Global Data Facility, 2021.

**67.** Russell et al., Effect of internationally imported cases on internal spread of COVID-19, 2021.

# CHAPTER 4
# Call to action

**Realizing the potential of data to amplify humanitarian, development, and domestic programming will require coordinated, collaborative action by a diverse community of stakeholders.** Public, private, academic, international, and civil society actors must unite in funding, supporting, and participating in integrated data ecosystems. Each stakeholder has a distinct role and brings different capabilities and resources (see *Figure 3*). Without contributions from every stakeholder, the vision of universal data ecosystems driving societal change cannot be achieved. Prompt action from every party is therefore crucial to moving this agenda forward.

### *FIGURE 6: OUR CALL TO ACTION*


**1** **WE CALL ON BILATERAL AND PHILANTHROPIC DONORS TO ALLOCATE A MINIMUM OF 0.8% OF THEIR ANNUAL INVESTMENT TO DATA ECOSYSTEMS.**[68] Data ecosystems multiply the efficiency and effectiveness of humanitarian, development, and domestic funding, enabling future spending to achieve a greater impact on individuals' lives. Fulfilling the vision of universal data ecosystems is therefore one of the best opportunities open to impact-focused investors, as its gains cover the cost many times over and drive a radical shift in the impact created by existing resources. What is more, this vision can be achieved with a small percentage of donor and domestic funding and represents a small portion of existing expenditure on data (e.g., best practice recommends allocating 5–10% of program budgets to monitoring, evaluation, and learning [MEL]).[69] To realize this vision, we call on bilateral and philanthropic donors to pledge 0.8% of their total annual funding to investments in the enablers (i.e., demand, funding, human capital, incentives, and trust) and foundational pillars (i.e., institutions, governance frameworks, infrastructure, and economic policies) of data ecosystems. Note investments in data (e.g., sector-specific systems, analytical software, or visualization tools) are distinct from investments in these enablers and pillars and therefore are not included under this systems-focused pledge target.

**68.** For more detail on estimating the target value of the percentage pledge, see Annex 2.
**69.** NIDOS, Monitoring, Evaluation and Learning (MEL) Guide, 2015.

**To maximize the impact and catalytic potential of this pledge, we further call on the donor community to:**

**A. DIRECT THE PERCENTAGE PLEDGE FIRST AND FOREMOST THROUGH CRAF'D AND GDF.** These mechanisms represent the highest-value channels for funding data ecosystems, for the following reasons:

- **CRAF'd and GDF are critical building blocks in global efforts to multiply progress through data ecosystems.** First, CRAF'd and GDF will strengthen the foundations and enablers of data ecosystems, ensuring that all future data is better quality, higher coverage, safer, and easier to use. Second, CRAF'd and GDF will provide future efforts to strengthen data ecosystems with the inputs to achieve their goals—for example, by filling critical risk data gaps, creating analytical models, building knowledge sharing platforms, and enhancing the capacity of value chain actors, CRAF'd provides future initiatives with the information, tools, skills, and networks needed to create and manage crisis data ecosystems.

- **As pooled funds, CRAF'd and GDF will bring coordination and economies of scale to donor efforts.** The funds will establish a more coordinated, focused, and demand-driven approach to resource distribution, thereby reducing allocative inefficiencies, prioritizing the solutions needed to create lasting change at scale, and facilitating large-scale, consistent capital flows that provide governments with the confidence and resources needed to invest in the enablers and pillars of integrated data ecosystems. Moreover, CRAF'd and GDF will enable contributors to make more effective use of their capital by sharing financial burdens, minimizing risks, and cutting transaction costs.

- **CRAF'd and GDF will unlock additional resources for data ecosystems by catalyzing bilateral, philanthropic, domestic, and private funding.** As entities hosted by the UN and World Bank, CRAF'd and GDF enable contributors to access the advocacy capabilities, financial know-how, and networks needed to identify and design investments that can catalyze additional funding. Deployments will be coordinated around integrated data ecosystems, whose costs can be sustained by national resources with limited or no donor support. As a result, contributions to the two funds are likely to be some of the most leveraged and sustainable investments in the data financing space.

Once CRAF'd and GDF are fully funded, additional flows should be directed to data ecosystems, alongside investments made by the two funds. This will ensure that funding remains coordinated, allocations focus on data ecosystems (rather than sector or program-specific data), and that the long-term needs of data ecosystems are fully met.

**B. INCREASE THE TRANSPARENCY OF FUNDING FOR DATA AND DATA ECOSYSTEMS THROUGH GREATER USE AND MORE GRANULARITY OF OPEN PROJECT DATA.** Currently donors provide varying levels of detail on the funding volumes, recipients, and objectives of their investments. Open project data would bring new clarity to where support is being provided, and where needs remain. Further, such transparency will facilitate the work of ODA and philanthropic aggregation efforts (such as the Bern Network Clearinghouse and PARIS21 for data ecosystems specifically, and OECD DAC CRS for ODA distributions generally) in turn to provide more granular overviews of funding flows. These efforts will enable a more coordinated approach to investing in data ecosystems by providing donors with the clarity needed to identify areas of opportunity and duplication.

**2** **WE CALL ON LOWER-INCOME COUNTRY GOVERNMENTS TO ALLOCATE 0.5%, AND MIDDLE-INCOME COUNTRY GOVERNMENTS 0.1%, OF ANNUAL SPENDING TO DATA ECOSYSTEMS.**[70] Public entities play a crucial role in developing integrated national data systems, acting as the primary stakeholders responsible for implementing and governing the data lifecycle and providing the resources needed to sustainably upgrade, run, and maintain national systems. By committing resources to the enablers and foundational pillars of integrated national systems, country governments will provide decision makers with the information needed to better assess and predict challenges, identify and customize solutions, and monitor and evaluate real-time progress.

**70.** For more detail on estimating the target value of the percentage pledge, see Annex 2.

Lower-income (IDA eligible) country governments, working in partnership with donor funding that is anticipated to meet half of estimated needs, can fully fund strong and sustainable data ecosystems with 0.5% of annual government expenditure. Meanwhile, with donor support projected to cover 5% of the total funding need in middle-income (IBRD-only eligible) countries, governments can fully fund the needs for strong national data ecosystems by dedicating 0.1% of annual government expenditure.

3 **WE CALL ON THE PRIVATE SECTOR TO ACT AS A CORE PARTNER IN ESTABLISHING AND UTILIZING DATA ECOSYSTEMS.** Data ecosystems foster private sector growth by facilitating the production of better data (i.e., information that is high quality and coverage, and safe and easy to use) and providing open access to data that firms can use to better understand their customers, markets, and wider economies (e.g., agriculture, health, household, or labor force surveys; geospatial data; national accounts; etc.). To amplify the benefits that businesses accrue from integrated ecosystems, private sector actors and governments will need to work together to strengthen the capacity of, and ties between, ecosystem participants (e.g., by role-modeling open data principles or spearheading system-building initiatives that bring together all types of system participants). Leveraging industry data, technology, capacity, knowledge, and best practice can accelerate this process and, in turn, contribute to an ongoing cycle of economic development and expansion that drives private sector growth.

4 **FINALLY, WE CALL ON ALL ECOSYSTEM PARTICIPANTS TO COME TOGETHER IN COLLABORATIVELY DRIVING THIS CRITICAL AGENDA FORWARD.** This means…

A. **MAXIMIZING THE IMPACT OF THEIR WORK BY INTEGRATING DATA INTO THE DESIGN, IMPLEMENTATION, AND EVALUATION OF ALL PROGRAMMING.** The scale of the challenges we face as a global community, and the ambitious goals we have set, mean we must strive to maximize the impact of our efforts at every turn. As set out in this report, data offers a powerful tool to identify, target, and impact those most in need around the world.

B. **SHARING DATA OUTPUTS AS PUBLIC GOODS.** Ensuring new and existing data repositories are open (e.g., publicly accessible, machine-readable, non-proprietary) and that data is consistently integrated into the activities of all ecosystem participants will help to build a culture of data sharing and data-driven decision making. In turn, more impactful and efficient data-driven activities will lead to greater recognition of the potential of data, and greater innovation in the application of data tools for sustainable impact.

C. **CONTRIBUTING TO FILL PERSISTENT KNOWLEDGE GAPS IN THIS AGENDA.** We must harness our going experience to strengthen the evidence base around the powerful role of data in improving human outcomes. This research agenda should include key topics such as increasing our communal understanding of the scale and implications of data's indirect benefits; analyzing the political economy of country investments in data and data systems; and building on the work in this report to develop robust analyses of the diverse returns on investments in data ecosystems.

**Taken together, these efforts will integrate support for data ecosystems into a wide array of stakeholder programming. The resulting virtuous cycle will unleash data's potential to enhance the impact and efficiency of humanitarian, development, and domestic spending, accelerating headway on shared goals and setting a course for lasting progress.**

# ANNEX 1

# Methodology: Return on investment analysis

## Introduction

**We conducted this analysis to produce the sector's first quantitative estimates regarding the economic benefits of investing in data ecosystems.** Data ecosystems are inputs that drive better program design and implementation. This indirect link to change makes it inherently challenging to quantify their impact—a difficulty that helps explain why efforts to estimate the financial impact of data ecosystems have so far been limited. More than anything, this gap highlights the need for new analyses to serve as a starting point for sector-wide discussion and refinement.

**The economic and social impact estimates presented in this report rely on original modeling that builds on the best available data and evidence.** The model uses existing case examples of discrete data ecosystem investments with direct causal link to impact. This approach produces a conservative estimate of the benefits of data ecosystem investments, as more integrated investments over time will see reduced costs with consistent or increasing benefits. However, it also offers the most robust approach to measuring the direct impacts of data ecosystem investments.

## Overview of approach

**The analysis takes a bottom-up approach to estimating costs and returns by building estimates for individual cases that are then aggregated into a global average.** The model follows a three-step approach:

1. **Analysis of individual cases –** estimate case-specific costs and associated social and economic benefits of each sample investment
2. **Segmentation by thematic groups –** group cases according to 1) their relevant economic sector and 2) income category, according to the countries in which the investments were located
3. **Aggregation of thematic groups into global weighted average –** estimate a global average based on sector relevance and country representation within each income group.

## PART 1: Analysis of individual cases

**As the basis of the model, we selected discrete case studies of data investments with a strong causal link to impact.** Selected cases needed to have all the necessary data regarding the costs and direct impacts, or to provide sufficient information to take reasonable assumptions to fill information gaps. For example, sample investments that had not yet conducted impact evaluations (due, e.g., to too-recent implementation) or that led to *policy* changes but not yet measurable impact were omitted. While we recognize that confounding factors may still be present within the case examples, this provides the clearest look at direct benefits of improved decision making in this context.

**The filtering process yielded 12 case studies of data ecosystem investments that were included in the analysis.** The case study selection includes examples from four sectors (agriculture, education, health, and government) across three country income groups (low income, middle income, and high income). Note that countries were assigned to income groups based on World Bank classifications; the middle-income category was grouped to include both lower- and upper-middle income countries.

*TABLE 1: LIST OF CASE STUDIES INCLUDED IN THE ROI MODEL*

| Country | Name | Description | ROI | Sector | Income Category |
|---|---|---|---|---|---|
|  Uganda | **Deployment of U-Report to fight banana bacterial wilt** | The Ugandan government used U-Report, UNICEF's community polling project, to fight the spread of the banana bacterial wilt. U-Report helped spread awareness of the disease, mobilized a network of nearly 330,000 volunteers across the country, and provided vital information to the government about the disease and its pattern of spreading. | 67.0 | Agriculture and food security | Low income |
|  Guatemala | **Strengthening statistical capacities to tackle school dropout in Guatemalar** | Since 2018, Guatemala's Ministry of Education has strengthened its Education Registry System; trained 38 Ministry of Education staff members on how to collect, manage, analyze, and use high-quality statistical information; and prevented 850 at-risk students from dropping out by implementing an early warning system. | 20.2 | Education | Middle income |
|  United Kingdom | **Impact of school performance statistics in the UK** | A policy change in 2001 resulted in the provision of school statistics in England but not in Wales. Burgess et al (2013) found a significant and sizeable negative effect on pupil progress in Wales compared to England following the policy change. | 15.0 | Education | High income |
|  South Sudan | **Introduction of the Early Warning Alert & Response System** | In 2013, WHO helped introduce the Early Warning Alert & Response System (EWARS) for conflict-affected regions through the distribution of EWARS kits to relevant health workers, leading to lower fatality rates. | 22.1 | Health | Low income |
|  Bangladesh | **Introduction of the Maternal and Perinatal Death Review system** | In 2010, the Bangladeshi district of Thakurgaon introduced the Maternal and Perinatal Death Review (MPDR), which collected data on maternal and neonatal deaths to inform remedial action plans at the community and facility level. Thanks to this initiative, maternal and neonatal deaths and stillbirths decreased between 2010 and 2014. | 14.2 | Health | Middle income |
|  Uganda | **Introduction of mTRAC** | In 2011, the Ugandan government introduced an SMS-based health reporting program called mTRAC. mTRAC has cost-effectively enhanced the quality and exchange of health data and strengthened the capacity of practitioners to use this data to improve health outcomes. This program resulted in fewer malaria-related deaths. | 39.2 | Health | Low income |

| Country | Name | Description | ROI | Sector | Income Category |
|---|---|---|---|---|---|
|  Philippines | Creating a unique identification number | In 2018, President Duterte signed a law that will provide a unique identification number for every Filipino. Currently, there are 33 different identification cards issued by a variety of government agencies. The new system will collect a person's common reference number and basic information (e.g., biometrics, passport number). The reform is expected to create significant cost savings for the Philippines Statistics Authority. | 12.2 | Government | Middle income |
|  Tanzania | Geographic information system for tax collection | The WB and DANIDA invested in the introduction of a geographic information system (GIS) platform to support local governments' tax reporting, revenue collection, operations and maintenance, urban planning, permitting, and land management systems. One year after the system was introduced, the eight participating cities significantly increased their revenue collection. | 23.6 | Government | Middle income |
|  Ukraine | Integrating a public e-procurement platform | In 2015, Ukraine piloted the ProZorro electronic system, an e-procurement platform developed with the support of donors, businesses, and volunteers. It is a public-private partnership consisting of over 100 members from civil society, the private sector, and government institutions. The initiative aims to make all public procurement documents publicly available through a dedicated online e-auction platform. The reform has led to significant cost savings for the government. | 72.8 | Government | Middle income |
|  New Zeland | Valuing the impact of a national census | New Zealand has conducted its census since 1851, providing a vital source of data about the size and demographics of the country's population. In 2014, a valuation of the census was performed. Even though the scope was limited to a select number of quantifiable applications, the study found that over the next 25 years, the census would have significant returns to the national economy. | 7.4 | Government | High income |
|  USA | Valuing the impact of Landsat | The Landsat program consists of a series of Earth-observing satellite missions managed by NASA and the U.S. Geological Survey. Landsat has produced substantial public cost savings as well as world economic benefits. | 30.9 | Government | High income |
|  USA | Risk terrain modelling to improve policing | The US Atlantic City Police Department introduced Risk Terrain Modelling (RTM) to optimize resources for predicting and preventing crime. RTM is an analytical technique combining crime data and environmental risk factors to identify high-risk areas. This data informed new approaches that led to reductions in homicides, shooting injuries, and robberies in the first year alone. | 34.2 | Government | High income |

**For each case study, we first estimated the relevant social and institutional benefits.** The benefits vary by case. For example, case studies related to the health sector focused on estimating the number of lives saved, disability-adjusted life-years (DALYs) averted, or infections prevented. In the education sector, benefits came in the form of increased years of education or number of pupils enrolled. In most cases, these figures were taken from existing literature pertaining to the given case example. Where this data was not provided directly, we built on available information regarding, for example, observed changes in local disease incidence in order to project estimates of resultant DALYs averted in the region relative to baseline trends.

**Next, we estimated economic benefits based on the social benefits created.** Although some cases already provided information on the economic benefits created, most required additional calculations. For case studies in the health sector, the process involved estimating the additional working years generated and the expected wages received during this period. For case studies in the education sector, the calculations estimated the added lifetime income derived from the additional years of quality schooling per child. In either case, future wages were calculated based on historical trends and converted to US dollars.[71] Once we forecasted future wages, we estimated their present value using the discount rate suggested by Haacker et al. (2020) for the defined income group at the start of the data ecosystem investment.[72]

**Meanwhile, we typically derived the costs of each program from the source material.** Some of the selected case studies provided detailed costs broken down by fixed and variable expenses. However, in other cases, costs were provided only for the first few years. In those cases, we linearly inferred the remaining costs based on the observed rate of change. In other cases, when only upfront costs were provided, we estimated variable costs based on comparable programs. Lastly, one case did not provide the total program cost but provided cost and impact figures at a per-unit level. Accordingly, all calculations were made at the unit level (with no impact on the final ROI estimate).

**Next, we calculated the overall economic benefits of a given case study as the present value of observed new value and/or reduced costs.** It should be noted that this estimate reflects only the observed financial benefits of a given investment. See Equation 1 below for an illustrative sample calculation of the economic benefits of a case study in the health sector (via increased lifetime earnings as a result of DALYs averted). While the figure is linked to social benefits (e.g., lives saved leading to prolonged opportunity for productive work), this does not reflect a general valuation of the saved lives themselves.

***EQUATION 1: ECONOMIC BENEFITS, FOR ILLUSTRATIVE HEALTH-RELATED CASE EXAMPLE***

$$\textit{Economic benefits} = \mathbf{n} \sum_{i=1}^{Y} w\,(1+\chi)^{i}\,\delta^{i}$$

Where **n** = *number of people impacted,* **i** = *year of working live saved,* **Y** = *total years of live working saved,* **W** = *annual wage (baseline),* **x** = *annual wage growth rate,* **δ** = *time-discount factor*

**Finally, we calculated the overall return on investment (ROI) of the given case as the ratio of benefits to costs.** This gives an indicative sense of the monetary value created by each dollar invested in the given case study.[73]

***EQUATION 2: RETURN ON INVESTMENT***

$$ROI = \frac{(\textit{Estimated economic benefits}) - (\textit{Total costs})}{\textit{Total costs}}$$

71. Once the expected wages exceeded a threshold defining the World Bank's country income groups, we assumed that wages would continue growing at the same rate as the new income group's gross national income (GNI) per capita historical trend. For example, Uganda was a low-income country in 2010 when the mTRAC system (see *Spotlight 2*) was implemented. We assumed women's future annual wages would grow at Uganda's historical GNI per capita growth rate until 2023 (i.e., 5.45%). In 2023, women's predicted wages would exceed the lower-bound threshold defining lower-middle income countries. We assumed that women's wages would hence continue growing at the average GNI per capita growth rate exhibited historically by lower-middle income countries (i.e., 5.0%). It follows that, in 2051, the predicted annual wage for women would again exceed the lower-bound threshold defining upper-middle income status. Accordingly, we set the applied wage growth to the average GNI per capita growth exhibited historically by this income group (i.e., 6.2%).
72. Haacker et al., On discount rates for economic evaluations in global health, 2020.
73. We note that as with any ROI analysis, this should not be taken to suggest that all similar investments will show linearly proportional returns at any investment size. Rather, this gives a directional sense of the efficiency of the assessed investment in creating economic impact, and may require some minimum investment size to generate such returns.

## PART 2: Thematic groupings

**We segmented cases by sector and country income category to account for any potential variation in the typical costs or scope of benefits derived from data ecosystems under diverse contexts.** The segmentation reflects the assumption that the cases present within each grouping have a representative ROI. The sectors covered by the case studies include health, education, agriculture, and government administration. They were chosen based on the public-facing nature of the implemented programs, enabling a clearer distinction of the impact pathway (acknowledging also a potential bias toward those with available data). ROI figures may differ within each sector based on the observed outputs from the respective investments. For example, saving working lives through better health programs may have an economic return different from improving crop production through agricultural programs. In addition, we grouped countries by income level (low, middle, and high)[74] given the strong correlation between data ecosystem development and economic development.[75]

**This process created a matrix of 12 thematic groupings.** Due to the limited availability of suitable case studies, not all sector-income intersections had available datapoints. We linearly inferred the estimation of the average ROIs of missing groupings based on the available data in other groupings. For example, the ROI for health investments in high-income countries was based on the ROI linear trend for health investments between low- and middle-income countries.[76]

## PART 3: Aggregation and global estimate

**After estimating the case study ROIs and segmenting them by grouping, we estimated the global ROI through a three-step process:**

1. Averaging within thematic groupings (i.e., within each sector - income intersection)
2. Averaging across sectors, within each country income grouping (e.g., arriving at a "low-income country" ROI across component sectors)
3. Averaging income groupings to calculate the global estimate

**First, the thematic groupings aggregation took the unweighted mean of the available data points.** However, we applied a discount factor to certain case studies based on the "disruptiveness" of the investment. Through this process, we recognize that impact can be more easily estimated in data ecosystem investments that are markedly different from the previous status quo. However, disruptive investments are, in the long run, likely scarcer in more developed data ecosystems where most investments build on existing infrastructure and may provide more incremental returns. To account for this bias, we applied a discount factor to "de-weight" case studies classified as "disruptive" in those contexts where similar impact is less likely in the future. This was smallest (5%) on "disruptive" cases identified in LICs, and largest (15%) for those in HICs.

**Second, we averaged the thematic groupings within each income grouping by weighting according to their relative share of current government spending.** The indicators we used were:[77]

1. Government health expenditure as a share of GDP
2. Government expenditure on education as a share of GDP
3. Agriculture, forestry, and fishing value added as a share of GDP
4. Government expenditure as a share of GDP

---

**74.** Income groups follow the World Bank 2022 categorization.
**75.** Chen, Mapping data governance legal frameworks around the world, 2021.
**76.** This approach provides only a rough approximation of the expected benefit of data ecosystem investments across income groups and relies on an assumption that the returns of data investments will diminish as country income grows. While this approach would benefit from further study, we also note that a sensitivity analysis showed the final aggregate ROI estimate to be only weakly affected by changes in the assumed values for "empty" matrix cells, and therefore this assumption has minimal impact on the final results in either direction.
**77.** All indicator values were extracted from The World Bank Open Data.

This weighting approach accounts for the fact that some sectors are more impactful on national economies and/or may receive a greater share of public spending, thereby enhancing the scale and potential impact of associated data ecosystem investments.

**Finally, we averaged the income-group figures with weights according to the number of countries in the respective income category.** This aggregation approach provides more relative weight to the ROI values where these investments are more likely to be seen.

## Limitations and assumptions

**Given the data limitations highlighted above, we note that these analyses come with inherent uncertainties and constraints**. These limitations, along with our assumptions, include the following:

- **The bottom-up, case-based approach is limited to the findings of the investments with available data.** This may introduce bias toward certain kinds of investments, funders, or other factors that contribute to certain investments generating a strong evidence base. This also naturally biases the selection toward investments that were successful. Nonetheless, this risk can be mitigated in future investments through adherence to the integrated data ecosystems principles outlined elsewhere in this report.

- **The case studies focused on examples with discrete data investments.** Consequently, the model's output is a conservative ROI estimate, as it includes higher one-off costs that may not be seen in future investments in more integrated systems. The incremental costs of future data-based decision making can be expected to decrease, further increasing the observed ROI.

- **Due to data gaps, we have at times needed to make assumptions to interpolate missing figures from the data we do have.** Wherever possible we have erred toward making conservative assumptions.

- **The analyses aim to estimate the incremental costs and impacts beyond a counterfactual situation.** There may, nevertheless, be other factors that contributed to the observed changes in outcomes that are not accounted for, or that would have altered historic trends in the counterfactual scenario.

# ANNEX 2

# Methodology: Percentage pledge target

## Overview

Chapter 4 of this report calls on donors and governments to allocate a percentage of their respective annual spending to data ecosystems,[78] through or alongside CRAF'd and GDF. The calculations behind these pledge targets are outlined below.

## Equations and subcomponents

### *EQUATION 3: PERCENTAGE OF PUBLIC SPENDING IN IDA-ELIGIBLE COUNTRIES REQUIRED TO REALIZE CTGAP[79]*

$$\textit{Pledge for IDA eligible governments} = \frac{(N)(1-p)}{G}$$

Where **N** = *total funding per annum required in IDA-eligible countries to meet the needs of integrated data ecosystems under the Cape Town Global Action Plan for Sustainable Development Data;* **p** = *proportion of IDA-eligible funding needs expected to be covered by donors;* **G** = *total value of government spending by IDA-eligible governments per annum*

### N = TOTAL FUNDING PER ANNUM REQUIRED IN IDA-ELIGIBLE COUNTRIES TO MEET THE NEEDS OF INTEGRATED DATA ECOSYSTEMS UNDER THE CAPE TOWN GLOBAL ACTION PLAN FOR SUSTAINABLE DEVELOPMENT DATA (CTGAP)

**Value:** USD 2.4 billion per annum to 2030 (2021 USD)

**Source:** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019

**Approach:**

- Calleja and Rogerson estimate the total funding needed to meet the objectives of CTGAP using planned expenditure cost lines of LIC and MIC National Statistical Development Strategies. Averaging the annual cost of each activity needed to realize CTGAP, they build a bottom-up estimate of the total cost of achieving CTGAP in each IDA-eligible country and multiply this by the number of IDA-only and IDA-blend states.
- These cost estimates equated to a total per country of roughly USD 30 million. This number was then multiplied by the number of IDA-eligible countries to provide an estimate of total needs (roughly USD 2.2 billion).
- This investment case updates these numbers to account for inflation and changes in the number of IDA-eligible countries, which fell from 75 to 74.
- For further detail on the original methodology used and its underlying assumptions, please refer to Calleja and Rogerson (2019).

**Notes:**

- Needs may well be higher than the USD 2.4 billion outlined above. Firstly, data ecosystems have been subject to a sizable funding gap totalling multiple billions per annum since Calleja and Rogerson's estimate was first produced. Secondly, as the data landscape widens and the demands for increasingly granular and timely data increase, the investments required to create effective integrated data ecosystems are likely to grow.

78. Note that, while crucial and impactful, investments in data (e.g., through new sector-specific sub-systems, cutting-edge analytical software, or visualization tools) are distinct from investments in the foundational pillars (e.g., institutions, governance frameworks, or infrastructure and economic policies) and enablers (e.g., demand, funding, human capital, incentives, and trust) of data ecosystems and, therefore, do not count towards the systems-focused percentage pledge. All figures in this annex reflect needs and spending for data ecosystems specifically.
79. CTGAP provides a roadmap outlining the activities needed to modernize statistical systems and expand statistical capacity to meet future needs, including the SDGs. For more information, please refer to UNDES, Cape Town Global Action Plan for Sustainable Development Data, 2017

### p = PROPORTION OF FUNDING NEEDS EXPECTED TO BE COVERED BY DONORS

**Value:** 50%

**Source:** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019 & Open Data Watch, The State of Development Data, 2016

**Approach:**

- In line with Calleja and Rogerson (2019) and Open Data Watch (2016), we assume domestic governments will cover 50% of the costs of IDA-eligible countries

### G = TOTAL VALUE OF GOVERNMENT SPENDING IN IDA-ELIGIBLE COUNTRIES PER ANNUM

**Value:** USD 262 billion (2021 USD)

**Source:** World Bank, General government final consumption expenditure, 2020

### *EQUATION 4: PERCENTAGE OF PUBLIC SPENDING IN IBRD-ONLY COUNTRIES REQUIRED TO REALIZE CTGAP*

$$\textit{Pledge for IBRD only governments} = \frac{(N)(1-p)}{G}$$

Where **N** = *total funding per annum required in IBRD-only countries to meet the needs of integrated data ecosystems under the Cape Town Global Action Plan for Sustainable Development Data;* **p** = *proportion of funding needs of IBRD-only countries that are expected to be covered by donors;* **G** = *total value of government spending by IBRD-only governments per annum*

### N = TOTAL FUNDING PER ANNUM REQUIRED IN IBRD-ONLY COUNTRIES TO MEET THE NEEDS OF INTEGRATED DATA ECOSYSTEMS UNDER THE CAPE TOWN GLOBAL ACTION PLAN FOR SUSTAINABLE DEVELOPMENT DATA (CTGAP)

**Value:** USD 3.8 billion per annum to 2030 (2021 USD)

**Source:** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019; Open Data Watch, The State of Development Data, 2016; World Bank, Population figures, 2021

**Approach:**

- We assume that the absolute cost of implementing statistical plans and building statistical capacity is higher in IBRD countries. In line with Calleja and Rogerson (2019) and Open Data Watch (2016), we scale the cost of the 74 IDA-eligible countries to IBRD-only countries by taking the geometric average of the ratio of population sizes (4.90 billion/1.75 billion, or 2.80) and the number of countries (70/74 or 0.95), to produce a multiplier of 1.63. This gives a total cost of USD 3.4 billion per annum.

- This investment case updates these numbers to account for inflation.

### p = PROPORTION OF FUNDING NEEDS EXPECTED TO BE COVERED BY DONORS

**Value: 95%**

**Source:** Calleja and Rogerson, Financing challenges for developing statistical systems, 2019 & Open Data Watch, The State of Development Data, 2016

**Approach:**

- In line with Calleja and Rogerson (2019) and Open Data Watch (2016) we assume domestic governments will cover 95% of the costs of IBRD-only countries, with the balance provided through ODA or philanthropic support

### G = TOTAL VALUE OF GOVERNMENT SPENDING IN IBRD-ONLY COUNTRIES PER ANNUM

**Value:** USD 5,082 billion (2021 USD)

**Source:** World Bank, General government final consumption expenditure, 2020

***EQUATION 5: PERCENTAGE OF DONOR SPENDING REQUIRED TO REALIZE CTGAP IN IDA-ELIGIBLE AND IBRD-ONLY COUNTRIES***

$$\textit{Donor pledge} = \frac{A - B}{O + P}$$

Where **A** = *total of value of IDA-eligible AND IBRD-only countries' needs under the Cape Town Global Action Plan for Sustainable Development Data;* **B** = *value of IDA-eligible AND IBRD-only countries' needs that are met by governments;* **O** = *total of funding per annum;* ***P*** = *total value of philanthropic foundation funding per annum*

### A = SUM OF IDA-ELIGIBLE AND IBRD-ONLY COUNTRIES' NEEDS UNDER THE CAPE TOWN GLOBAL ACTION PLAN FOR SUSTAINABLE DEVELOPMENT DATA

**Value:** USD 6.2 billion (2021 USD)

**Source:** See above

### B = SUM OF IDA-ELIGIBLE AND IBRD-ONLY COUNTRIES' NEEDS THAT ARE MET BY GOVERNMENTS

**Value:** USD 4.8 billion (2021 USD)

**Source:** See above

### O = TOTAL VALUE OF ODA FUNDING PER ANNUM

**Value:** USD 185 billion (2021 USD)

**Source:** OECD, ODA Levels in 2021, 2020

### P = TOTAL VALUE OF PHILANTHROPIC FOUNDATION FUNDING PER ANNUM

**Value:** USD 11 billion (average value of funding per annum between 2016-2019, updated to 2021 USD)

**Source:** OECD, Private Philanthropy for Development, 2019