



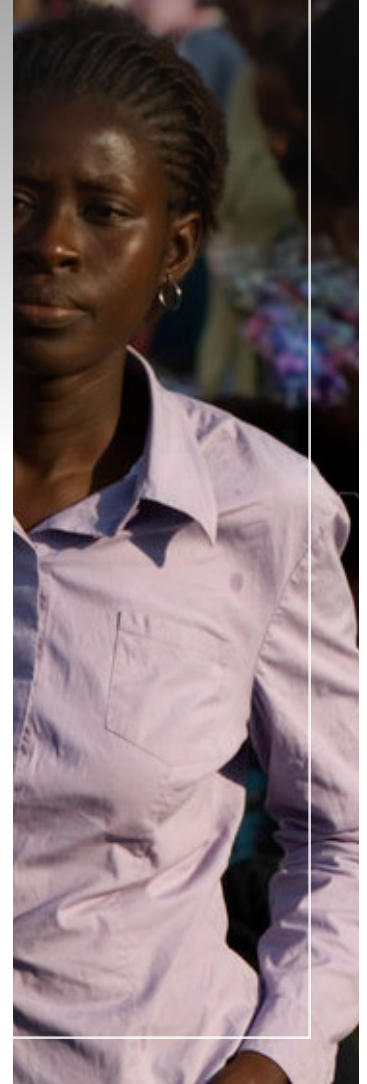
Global
Partnership
for Sustainable
Development Data

TRÉNDS

Thematic Research Network
on Data and Statistics

Value of Data Case Studies: Synthesis Report

Written by Jessica Espey
and Hayden Dahmm,
SDSN TRéND





1 Project Summary

Since negotiations to shape and define the Sustainable Development Goals (SDGs) began in 2013, the importance of data for improving sustainable development outcomes has played a starring role. There have been calls for additional investments into the data sector. Yet the international community has so far little considered the incentives of funders and investors. Framing these discussions with funders and investors in mind and using some of their common language—e.g. “return on investments” or ROI—will be critical to engage and sustain a broad investment community for these data needs. Equally important is careful documentation of data impact examples that can be used to make a compelling investment pitch.

How can these returns be contextualized for these audiences? In the data sector, returns might take the form of positive economic impacts—for example, better data on agricultural productivity enabling an increase in agricultural yields and exports, or more systematic national data collection encouraging higher rates of lending or borrowing from international financial institutions. But the returns might also be social or environmental—for example, improvements in civil registration and vital statistics (CRVS) systems enabling better tracking of children’s educational access and attendance, or better climate projections enabling the development of climate resilient infrastructure saving future costs due to damage and loss.

Over the course of six months, SDSN TReNDS and the Global Partnership for Sustainable Development Data (GPSDD) reported on the value of investments made into data applications and systems, including the impacts of capacity development and data use. We systematically reviewed and then shortlisted compelling data impact stories that clearly explain the level and type of investment going into a specific type of data system, and the resulting outcomes. (Methodological learnings from the research process are laid out in Annex 1). To the extent possible, the case studies attempt to quantify the impact and suggest the magnitude of ROI for funders and investors. Yet the case studies also detail the technical solutions, the process of implementation, and the challenges overcome. They span a range of data collection methods and technologies, from SMS-based health administrative data collection to satellite imagery, from census to local risk modelling (see Box 1). They also consider institutional investments to improve data quality, such as the reform of the Philippines statistical system to create one centralized Philippine Statistics Authority (PSA). Together, the case studies provide clear evidence of the benefits of investing in a comprehensive national data system not only rooted in traditional data collection approaches, such as census and CRVS, but also capitalizing on innovative tools and methods, such as mobile technologies and Earth observation.



Box 1 | The Value of Data Case Studies

An SMS-based system called **mTRAC**, implemented in Uganda, has supported significant improvements in the country's health system—including halving of response time to disease outbreaks and reducing medication stock-outs, the latter of which contributed to a reduction in malaria-related deaths.

NASA's and the U.S. Geological Survey's **Landsat** program—satellites that provide imagery known as Earth observation data—is enabling discoveries and interventions across the science and health sectors, and provided an estimated worldwide economic benefit as high as US\$2.19 billion per year as of 2011.

BudgIT, a civil society organization making budget data in Nigeria more accessible to citizens through machine-readable PDFs and complementary online/offline campaigns, is empowering citizens to partake in the federal budget process, and is helping to minimize waste and corruption.

International nonprofit **BRAC** is ensuring mothers and infants in the slums of Bangladesh are not left behind through a data-informed intervention combining social mapping, local censuses, and real-time data sharing. BRAC estimates that from 2008 to 2017, 1,087 maternal deaths were averted out of the 2,476 deaths that would have been expected based on national statistics.

Police in Atlantic City, New Jersey are developing new approaches to their patrolling, community engagement, and other activities through **risk modeling** based on crime and other data, resulting in reductions in homicides and shooting injuries (26 percent) and robberies (37 percent) in just the first year of implementation.

In 2013, the Philippines merged multiple data producing agencies into a single institution: **The Philippine Statistics Authority**. The creation of

the PSA has improved timeliness of national and regional accounts; opened up national statistical data, including microdata; innovated the way the Philippines conducts household survey and censuses; and is deriving value from a new national identification system.

According to a 2014 study, the **New Zealand census** returns to the national economy NZ\$5 for every NZ\$1 invested. The census's contributions to other areas, such as inclusion and empowerment of the Māori, are documented in this case study.

Household surveys are a powerful analytical tool that can shed light on how households interact with services and how interventions affect their wellbeing. This case study evaluates the return on investment from the **Living Standards Measurement Survey (LSMS)**—for example, helping to improve the targeting of programs associated with the Nicaraguan Emergency Social Investment Fund.

Civil registration and vital statistics are the backbone of effective national service delivery. CRVS data is also key to monitoring 12 of the 17 Sustainable Development Goals and 67 of the 230 SDG indicators. This case study shows the immense value that can be derived from CRVS investment for governments and for society at large.

With two-thirds of the world's population facing water scarcity at some point during the year, increasing the reliability of water access is essential to sustainable development. The sensor-driven **Smart Handpump** project showcases one data technology that is revolutionizing the way water services can be delivered.

Available at: sdstrends.org/valueofdata and data4sdgs.org/valueofdata



2 | Cross-Case Study Learning

Although the cases compiled relate to diverse methodologies or stages in the data value chain, there are some common lessons that emerge from comparative assessment. Among these:

- » It is possible to deduce a numeric return on investment for many different data approaches.
- » Most forms of socioeconomic data can have a positive social impact when systematically used to inform policy.
- » Innovation requires both high-level leadership and local government buy-in if it is to be sustainable.

a. Strong economic returns

The cases demonstrate a wide range of economic benefits from investing in data systems, including more efficient and/or targeted government services and wider business opportunities. Although we had hoped to develop an aggregate figure for the cost savings or ROI for data, the nature of the economic evidence is so varied that this proved infeasible. Nevertheless, these cases underscore that the economic effects of more frequent, higher-quality data can be profound. For instance, according to one study, the value of the Landsat program in 2011 was equivalent to US\$2.19 billion per year. Landsat also highlights the benefits of recurrent long-term investments, producing estimated annual cost savings in the United States ranging from US\$350 million to US\$436 million for federal and state governments, non-governmental organizations, and the private sector. Likewise, a valuation report found that the New Zealand census returns NZ\$5 to the national economy for every NZ\$1 invested. The Philippines has invested into a new ID system and expects to see resulting taxation efficiency savings of 2 percent of GDP over the next five years, equivalent to US\$6 billion. Meanwhile, the BudgIT project in Nigeria has exposed corruption scandals, such as a 41 million Naira (approximately US\$110,000) investment that claimed to be funding a non-existent youth center in Kebbi State.

However, identifying robust economic evidence for data can be challenging as it requires investment in subsequent academic analysis. In the case of Landsat, the New Zealand census, and the Philippines ID program, cost-benefit analyses were used to help make the case to government as to the value of the investment, to influence how much public money would be allocated to the projects. For all data programs, however small, investing in subsequent cost-benefit analysis can be a compelling way to ensure sustained investment in the program and encourage its expansion and replication.



b. Clear social benefits of more accurate data

Building better, more detailed, and more timely data systems is crucial to understand the situation of the poorest and most vulnerable members of society. The World Bank's LSMS program, for example, enables us to understand individual experiences of poverty and how policy interventions affect livelihoods over time. A census can be key to ensuring representation and services for marginalized members of society, as with the indigenous Māori community of New Zealand. Additionally, CRVS systems provide insights into the trends and demographics of health issues, gender inequities, marriage patterns, and more—in turn helping us to understand and expand services, such as in refugee communities in Ethiopia.

In addition to improving our understanding, some forms of data, such as CRVS, are in and of themselves empowering. For example, building robust CRVS and ID systems can give individuals proof of identification, which in turn can facilitate their access to services, enabling them to vote, get a passport, or get a driver's license.



c. Poor quality administrative data necessitating survey-based approaches

Across the cases, there was a recurrent theme relating to the poor quality of cross-governmental administrative data at both national and local levels. For example, in the absence of robust local establishment surveys in Atlantic City, police officers had to manually categorize some 1,000 business establishments before a location-based crime risk model could be developed. Similarly, BRAC community health workers had to follow an involved process of mapping Bangladeshi slums in order to reliably identify pregnant women. Although cumbersome, these exercises generated previously unavailable information essential to decision-making processes, thereby helping to fill gaps in the administrative record.



d. Scaling innovation and ensuring its sustainability requires political leadership

Scaling up innovative data solutions often requires high-level championing or political leadership. A fundamental overhaul of the national statistical system in the Philippines was initiated by a mandate from the president, and was supported by two sitting ministers actively participating on the PSA's board. Another example is Landsat, which was created thanks to the leadership of NASA and then received ongoing support from the U.S. Congress.

Additionally, local government support can help ensure sustained innovation and data quality. For example, risk terrain modeling in Atlantic City is managed entirely by local officials, while in the Philippines the government is actively training up a new cadre of local statisticians who can work with the central government to improve the quality and granularity of local statistics, including by using new technologies and innovative approaches.

e. Data collection should be iterative, where possible harnessing new technologies, while being cognizant of their limitations

A common lesson from these case studies is that methods for collecting and using data are not necessarily static. With the Smart Handpump project, waterpoint data transmitters were originally the primary source of information about pump maintenance. Yet, as communities have grown to trust the service and more complex piped schemes have been implemented, immediate community feedback has become an important complement to automated data collection. This data has also proven to be valuable for ensuring the accountability of maintenance providers. Uganda's health system historically collected health data on paper forms, but the mTRAC program now allows this same data to be submitted by SMS, and health reporting and responses to health issues have been improved.

A dynamic and iterative approach can realize new benefits and efficiencies but complications can also arise. For example, New Zealand adopted a digital-first approach during its 2018 Census, and critics claim that this was responsible for an "undercount," with minority communities expected to be disproportionately underrepresented.

Innovation will be essential to meet ever-increasing demands for data and to achieve the SDGs. However, if we are serious about our commitment to leave no one behind, it is important to be cognizant of ever-widening technological inequalities and how these might bias data collection efforts moving forward.



3 Conclusions and Recommendations

Data have been recognized as an engine of the modern economy and as paramount to achieving the SDGs. Investments into data systems can support a range of solutions, from addressing crime to improving water infrastructure. Communicating the value of these investments will be essential to realizing the full potential of data-driven solutions. Increased investment from funders and investors alike will be necessary to discover innovative data solutions. This will also ensure the continuing success of more traditional data products, such as censuses and civil registration and vital statistics systems, that are foundational to a robust national data system. Specific recommendations for data funders and investors, based upon this research, include:

Invest across all parts of the data value chain, from production to analysis, from curation to communication. This limited selection of case studies suggests that funding data communication, as well as systematic evaluations of data impact, will be crucial to encourage sustained, long-term funding and support for robust national data systems.

When funding data projects that are not initiated by government, make local and government engagement a core condition so that there is a clear plan to take the innovation to scale and ensure sustainable implementation after the non-governmental actor steps back.

Encourage innovation and support data collection approaches that employ new technologies and methods. But this should not occur at the expense of investing in traditional official statistical methods, such as undertaking censuses and related post-enumeration surveys, building robust CRVS systems, and supporting regular household surveys. These are the building blocks upon which innovation can take place, and against which we can benchmark results.

Support iteration when testing and scaling up new data collection approaches or other related methodologies. Many of these cases provide examples of how solutions can be improved over time, so donors should not be afraid of failure when encouraging innovation.



Annex 1: Methodological Learning

Attempting to find compelling case studies that document the value of data and returns on data investments resulted in a number of important methodological lessons. First, it is necessary to set a scope or bound for the types of data and data collection approaches under study, as with the advent of big data and the so-called “data revolution,” there is a nearly overwhelming array of methodologies outside of official statistics. Second, it was important to set clear criteria for what was and was not sufficiently rigorous evidence to constitute a compelling case study.

a. Defining a data typology

To make a compelling case for investments in national data systems, it is important that there are compelling stories of data impact for many, if not all, of the types of data governments collect. In 2015, SDSN and partners identified a typology of data types required to monitor the SDGs, in the report “Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.” The types of data identified in this typology are as follows:

- » Census
- » CRVS
- » Household surveys
- » Administrative data
- » Economic statistics
- » Establishment surveys
- » Agricultural surveys
- » Environmental statistics
- » Geospatial data



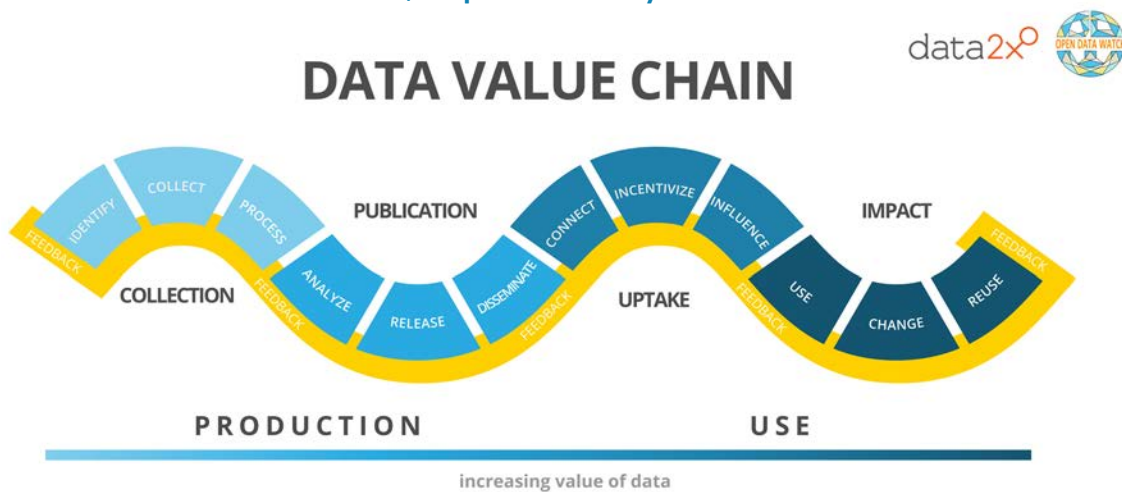


There are, however, various new methods and means for collecting data on similar sustainable development outcomes that are increasingly important both for governments, but also for citizens and other non-governmental actors. These include, but are not limited to, the use of telecommunications data, big data, and citizen-generated data. A review of academic and “grey” literature in data science and on the data revolution for sustainable development suggested that there was no comprehensive catalogue or listing of the diverse data collection methodologies in regular use around the world. We therefore

defined one, expanding on the typology in the “Data for Development” report to also include telecommunications data, big data, sensor- and ICT-based collection approaches, and citizen-generated data.

However, simply employing this method falls short of a comprehensive picture of the data investment landscape, as identifying cases solely by the data collection methodology risks leaving out investments in other stages in the Data Processing Cycle. This cycle, also dubbed the Data Value Chain (see Figure 1) includes data collection, curation, publication, analysis, uptake, and use or impact.

Figure 1: The Data Value Chain, as presented by ODW and Data2X



Reprinted with permission of Open Data Watch

As a result, we worked to identify case studies across the value chain. The majority of identifiable cases were focused on collection and analysis—such as mTRAC in Uganda, Landsat, the New Zealand census, the LSMS program and investments in CRVS systems—while the PSA and BudgIT case studies focus on the latter part of the value chain (uptake, storage, and access). In general, it was challenging to find compelling cases documenting uptake, e.g. how certain data had been used in a specific policy scenario with resultant impacts. Further research in this regard—tracing the use of specific data in policy and programming decision and documenting resultant impact—could make an important contribution to the extant literature and debate on how best to prioritize data investments.



b. Identifying cases across that typology with a robust evidence base and secondary literature

At the outset of this project, we assumed there would be a wealth of cases to choose from based on initial discussions with contacts and basic online searches. In more considered research, however, we encountered difficulties. A number of suggested cases had little to no project or government documentation—or if they did it was written by the project leads themselves and therefore was highly subjective. To ensure that our case studies would be grounded in strong evidence, we set three criteria to help us decide whether a case should be considered:

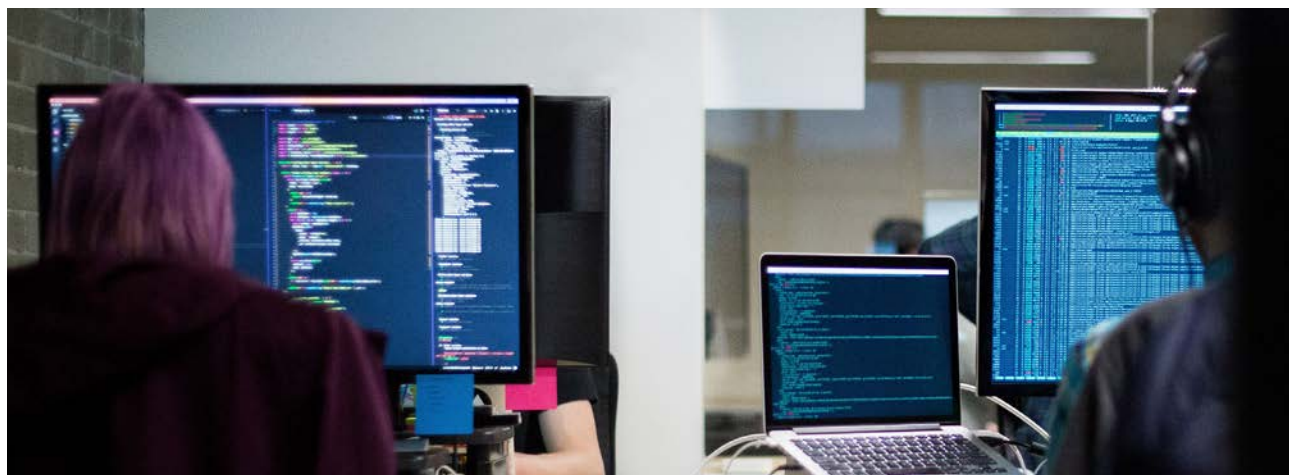
First, has sufficient time lapsed to be able to document and carefully assess the impact of the investment?

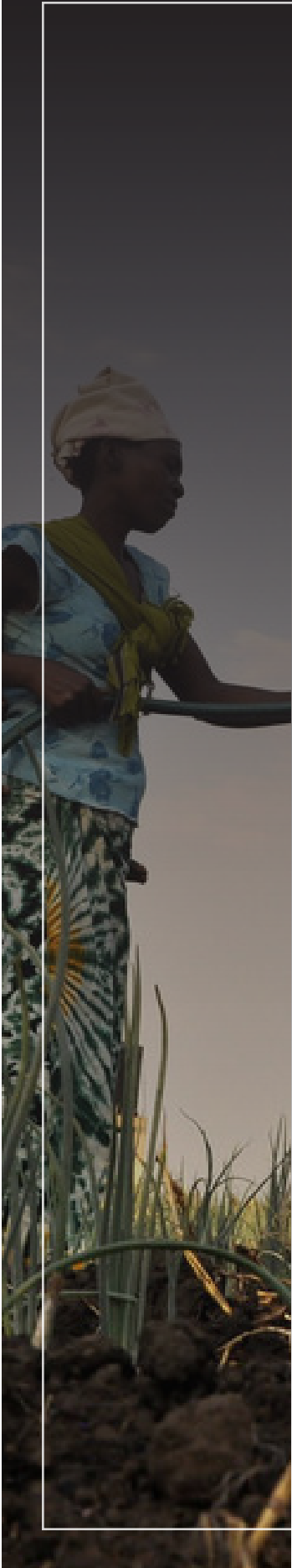
Second, has the impact been rigorously assessed and findings published (ideally within academic or peer-reviewed literature)?

And third, is the impact of the data investment quantifiable, economically or in terms of social benefit?

These criteria limited the number of cases we were able to draw upon, but we considered it an important trade-off to ensure the cases were robust. In the end, all but one of the cases (BudgIT) have been cited in academic or peer-reviewed studies. In the case of BudgIT, an exception was made, as it received considerable funding from the Bill & Melinda Gates Foundation. As such, there were comprehensive impact reports, as well as BudgIT's annual reports and audits, available online.

The result is a concise set of 10 case studies that use robust secondary literature and key informant interviews to make a compelling case as to the value of investing in a comprehensive national data system, that makes best use of new technologies and alternative, non-traditional monitoring approaches.





Global
Partnership
for Sustainable
Development Data



TRENDS

Thematic Research Network
on Data and Statistics

