# Getting the Most Out of SDG Data Investments:

## *A Living Manual for Increasing Value by Focusing on Decision Needs and Portfolio Function*

**PROJECT LEAD:** MARC LEVY, CIESIN AT COLUMBIA UNIVERSITY

MARCH 2017

TRENDS
Thematic Research Network on Data and Statistics

# Getting the Most out of SDG Data Investments

A Living Manual for Increasing Value by Focusing on Decision Needs and Portfolio Design

## Contents

## 1. Motivation for this Guide

The world is embracing the most ambitious set of international goals and targets ever. The 17 Sustainable Development Goals (SDGs) and the associated 169 policy targets articulate for the first time a comprehensive set of goals and targets that span not only the entire range of global ambitions identified in the UN Charter, but also the more specific formulations that have emerged since 1945 and new issues that were not even imagined when the Charter was signed.

This far-reaching policy agenda, without historical analog, emerges at a time when the task of providing for the public welfare has become more difficult, owing to new risks, such as climate change and the growing complexity stemming from heightened linkages across multiple natural and social systems. A number of systemic transitions underway in geopolitics, demographics, economics and technology, for example, all have the potential to generate reverberating shock waves as they unfold and interact.

*This Guide is a tool for those seeking to design a cost-effective and fit-for-purpose data and information system for the Sustainable Development Goals. It is aimed at key decision-makers seeking to harness the full power of data to achieve the SDGs.*

The SDGs require a new approach to data and monitoring. The number of issues is greater, the complexity is higher, and measurement technology is changing more rapidly than when the dominant measurement systems were first designed. The gap between what the international community can measure and what it needs to measure was dramatically increased with the adoption of the SDGs.

Creating a measurement collection portfolio that is fit for purpose across multiple sectors and places, and that is responsive to cross-sector and cross-location challenges, requires effective planning.

By making wise data investment choices governments can shrink the measurement gap and enable the effective decisions and actions that will make the vision behind the SDGs a reality.

*Moving forward from historical data paradigms*

For most of human history the ambitions that nations took on expanded slowly and were reasonably well matched by data capabilities. The nation states that emerged in the 18th and 19th centuries were largely able to measure what mattered to them -- they could conduct population censuses; they could carry out territorial surveys; they could monitor what passed through their ports; and they could quantify their military power.

This relationship began to change as pressure for social reform at the end of 19th century and beginning of the 20th century led governments to take on additional ambitions at a more rapid pace. And these expansions were often ahead of corresponding improvements in data capabilities. For example, although governments began adopting Keynesian economic policies in the early 1930s, it was almost a decade until rudimentary national income accounts first became available.

The gap between what governments aspired to manage and what they were able to measure grew enormously at the end of the 20th century and into the 21st. This growing gap was driven in part by the embrace of the global environmental policy agenda, which has notoriously lacked a data infrastructure that functions as well as that supporting economic and social measures. For example, over the 1990s governments accepted responsibility for managing problems such as biodiversity loss, water scarcity, land degradation, and marine fishery decline that lacked any robust measurement capabilities. But the gap also grew as a result of the deepening of the traditional social and economic goals, in particular their reformulation in the 2015 UN SDG resolution as applying equally to men and women and universally to each country, region and social group. Even the comparatively robust measurement processes that support social and economic goals do not currently match this radically universal ambition.

There are many reasons to believe that the gap between policy ambition and measurement supply has now peaked. Technological advances in a wide range of applications have combined to dramatically reduce data collection costs compared to a generation ago. The innovation cycle has shrunk remarkably in time and expanded remarkably in participation. Large communities are able to design, prototype, test and improve data approaches far more quickly.

Over the course of the 2030 development agenda, therefore, it is more likely that success at developing appropriate data systems will be constrained more by imagination and organizational skill than by measurement costs and technologies. A transistor may cost a thousand times less today than it did twenty years ago, but it is not a thousand times easier for a government to mobilize a functioning information system. For most of what matters for the SDGs, measurement is system-constrained, not technology-constrained.

A stark conclusion follows from the above: If we treat the ability of governments and other actors to structure effective information systems as a limited resource, it is vital that we find effective, efficient mechanisms for closely matching data supply efforts with concrete decision-making needs.

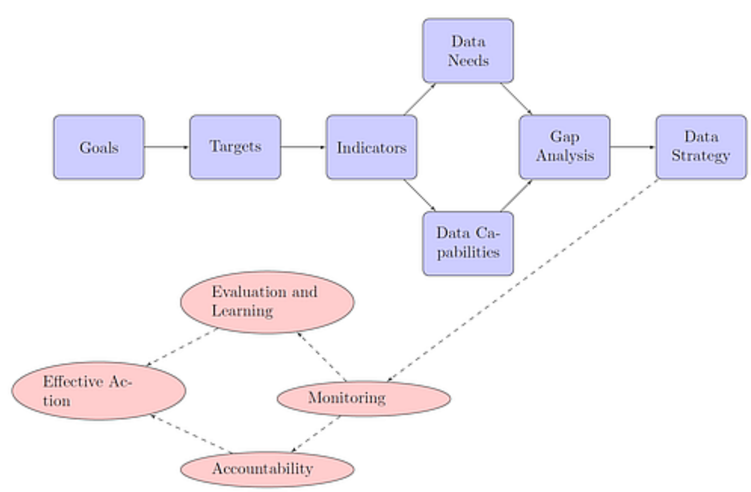


*Figure 1: Data Strategy from a Monitoring Perspective*

*Looking ahead to new models and better designs for data systems*

There is a great deal of high-quality guidance on how to design data systems that conform to international standards and best practices and that are capable of monitoring progress toward the SDGs in a comprehensive manner. Bodies such as the UN Statistical Commission, the Interagency Expert Group on SDG Indicators, and Paris 21 are active producers of such guidance. A government or other actor seeking insights into how to develop data systems for effective SDG monitoring will find consistent, comprehensive, and useful advice. And there is no doubt that sound monitoring systems are a vital component of the SDG data ecosystem.

When it comes to moving beyond monitoring, however, toward designing data systems from the perspective of driving effective action, the picture is far murkier. Relevant knowledge tends to be organized in specific sector and technology communities. There is virtually nothing that provides comprehensive guidance to those seeking to develop the most effective action-oriented data systems. The guidance available is not consistent, even within specific technology or sector communities. In contrast to processes that generate guidance on monitoring data systems, there are few institutional mechanisms that drive consensus in the data for action realm. Although there is considerable high-quality advice organized around specific entry points such as satellite data, mobile phone data, household surveys, electronic sensors, and so on, it remains quite challenging to understand how to choose from among the many instruments available in a way that provides the greatest enhancement to decision-making and implementation.

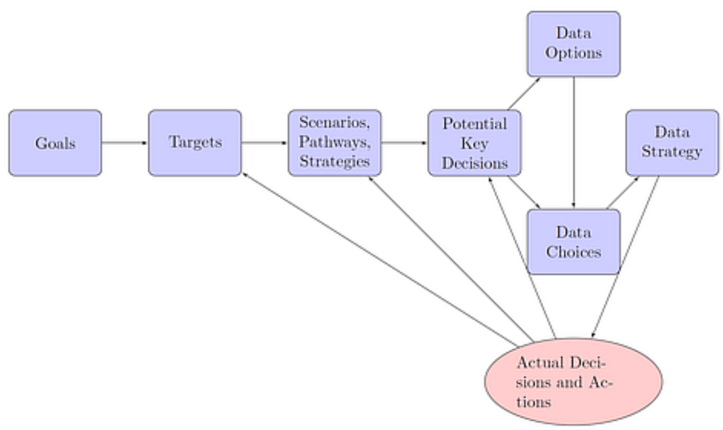

*Figure 2: Data Strategy from a Decision-Support Perspective*

This guide is meant to provide an initial platform for meeting the unmet demand described above. It focuses on two crucial aspects of the designing for action question: how best to tailor data systems to decision-maker needs, and how best to combine data technologies to create the most value. Together, these comprise the most important considerations for designing effective action-oriented data systems. They are complementary to the valuable guidance regarding data for monitoring. They are not meant to replace such guidance, but rather to augment it and increase its impact. Although it is possible that a country that vigorously pursues the ideas presented here, regarding data systems to support action, may choose to modify some standard approaches to monitoring, it is not necessary that countries do so.

# 2. Making Data Investments Relevant to Decision-Making

Early in the SDG process the need for a fresh approach to development data was recognized. A 2014 report commissioned by the UN Secretary General demonstrated that achieving the SDGs would require concerted focus on improving data quality; on disaggregating data with respect to gender, social groups, and regions; and on filling critical data gaps, such as those concerning environmental conditions (IAEG 2014). Achieving minimal data needs for monitoring and reporting purposes is within reach of all countries, though far from a trivial matter. For the 77 lowest-income countries to develop such capabilities will require an investment of $1 billion per year (SDSN 2015).

Countries that create data systems that meet monitoring needs will have the best opportunity for reporting on SDG progress in globally-sanctioned manner. On the other hand, by considering data investment choices from the additional perspective of enabling effective decision-making, countries can increase the likelihood that they will choose and implement the most effective actions for achieving the SDGs. Data that enable good reporting play an *indirect* role in promoting progress toward achievhing the SDGs, through their impact on accountability, evaluation, and learning processes. Data that enable good decision-making play a *direct* role by shaping immediate actions.

When data are primarily playing a descriptive role, they can be collected in a way that pays most attention to a small number of criteria. Typically, data systems designed to support national development processes strive to achieve high levels of accuracy; to have official, authoritative status; and to be consistent and comparable. As one moves up the analytics ramp in Figure 4, however, other considerations often weigh heavily. If investment and policy decisions are to be made quickly, for example, then how fast the data can be generated becomes important. If many of the most important decisions concern location (such as where to invest in power grids or schools or clinics), then the spatial precision of the data may matter as much as the accuracy at the national level. If the data are meant to assist in decision-making by being used in a specific set of algorithms or decision-support models, then it becomes important how easy it is to incorporate the data into such tools. If the data are meant to drive decisions in part by identifying interventions that are most effective, then it becomes highly important that the data are well-suited at tracking changes over time.

## 2.1. Practical steps toward matching decision needs to data solutions

The economist Doug Hubbard (2014) has developed methods for matching decision needs to data systematically. The steps are simple to grasp in the abstract, though often challenging in complex settings. Hubbard observes that it is not easy at first, but it is possible to get good at it; the data that people think is important often turn out not to be important; and focusing on the value of data helps prioritize choices. Further, he finds that data that are good for reporting are not always good for decision making and that what people think is most important is often not the most important. Therefore it is worth going to the trouble of being as systematic as practical in cataloging decision needs and matching such needs to data choices.

The implications of focusing on decision-making needs when it comes to data systems for sustainable development have been elaborated (O'Connell et al 2013, Shepherd et al 2015, Rosenstock et al 2017). They have also been applied in a number of practical settings. This section summarizes the insights gained from these exercises.

### 2.1.1. Clarify the decisions that matter

It is common for governments and other stakeholders to put significant effort into convening dialogues about how the global goals will be pursued in their locale; it is also common to convene dialogues to formulate strategies for measuring relevant aspects of target and goal achievement. For actors seeking to make their data investments as effective as possible, just as much effort should be deployed to create lists of key decisions whose outcomes will determine whether the country succeeds or fails at meeting the goals.

What choices will be made over the coming decade that will determine whether or not your country achieves the SDGs? Focus on the most influential investment and policy decisions. Identify the decisions most likely to be improved with better data.

Such an exercise can be challenging for stakeholders not accustomed to thinking in this way. It can help to provide examples and narratives showing how the decisions matter for SDG achievement and how focusing on such decisions can help identify priority data needs effectively.

Table 1: Illustrative Decisions by SDG

| SDG | Illustrative Decisions |
|---|---|
| 1) No Poverty | How to sequence investments in schools, clinics, and light manufacturing zones. |
| 2) Zero Hunger | Where to intensify agricultural production.<br>Whether to subsidize fertilizers for poor farmers. |
| 3) Good Health and Well-being | Where to locate new hospitals and clinics.<br>Whether to charge for community health worker visits. |
| 4) Quality Education | Where to locate new schools.<br>Where to concentrate teacher training efforts. |
| 5) Gender Equality | Where gender-separate restrooms would have the biggest impact. |
| 6) Clean Water and Sanitation | How much sewage treatment capacity to create in each city. |
| 7) Affordable and Clean Energy | What energy source to choose for remote regions. |
| 8) Decent Work and Economic Growth | What skill training program to develop. |
| 9) Industry, Innovation and Infrastructure | Whether to invest in tech hubs. |
| 10) Reduced Inequalities | Whether to adopt minimum income guarantees. |
| 11) Sustainable Cities and Communities | What new transportation infrastructure to build in large cities. |

| | |
|---|---|
| 12) Responsible Consumption and Production | What new products should a company develop to meet consumption needs sustainably?<br>Should there be a tax on waste? |
| 13) Climate Action | What economic instruments to use in order to accelerate adoption of low-carbon technologies. |
| 14) Life Below Water | How to lower negative impacts of coastal sewage effluent. |
| 15) Life on Land | Where to locate new protected areas. Which protected areas to receive additional support. |
| 16) Peace, Justice and Strong Institutions | What resettlement, rehabilitation and job training programs to use to reduce risk of conflict resumption |
| 17) Partnerships for the Goals | In which policy areas to encourage more active public-private partnerships. |

At the end of such an exercise, a country will have a list of crucial future decisions across the most important investment and policy areas, relevant to the stakeholders that will be making these decisions. This exercise is scale-neutral. It can be carried out at the national, regional or local level; or at the firm, facility or group level.

Sometimes a country may want to understand a particular SDG or target but lack the most elementary descriptive data about how important an issue it is, which social groups and regions are most affected, and what the trends are. In such cases it might be tempting to revert to basic monitoring approaches to data investments instead of the decision-centered approach described here. In most cases that is not a necessary or advisable approach. Data for monitoring purposes are necessary and worthwhile, but they are not the only source of descriptive data on severity and distribution. For example, in 2012 Niger developed a strategic response to drought risk in partnership with the World Bank . The strategy consisted of rank-ordered prioritization of issues and selection of most appropriate policy responses (World Bank 2013). Identifying priority issues and responses is an act of decision-making. While the strategy development process made use of much official monitoring data, it also made use of considerable data outside this realm, including qualitative expert judgments and back-of-the envelope boundary estimates. For example, a decision had to be made on how much effort to devote to vaccinating cattle against major livestock diseases, which become more threatening during times of drought. The prevalence data were considered unreliable, though there was enough information to conclude that the problem was serious. Instead of investing in more accurate prevalence data collection, those involved in developing the strategy instead looked at cost estimates for the vaccination program and compared that to ballpark scenarios of livestock death. They quickly and easily concluded that the low cost of the vaccination program would be justified even under very conservative estimates of potential loss. Therefore a decision could be even though the monitoring data were far below par.

This exercise can interact with national planning in two ways. Where there is a national SDG plan in place that has strong support from all relevant stakeholders, the exercise can focus on decisions necessary to have the plan succeed. Where there is not yet a national SDG plan in place, or key elements of the plan are contested and under consideration for modification, the exercise can be focused on formulating or modifying the plan.

### 2.1.2. Clarify the consequences of getting it right, compared to getting it wrong

To know how much should be invested in data systems it is necessary to know their value. Within a decision-support context, the value of an information system lies in the ability to make the right decision instead of the wrong one. Therefore it is necessary to characterize in some way what would happen if the wrong decision were made. For example, if the decision concerns where to locate the most intensive efforts to increase agricultural production, then making the wrong decision could result in wasted financial resources, damage to local terrestrial ecosystems and downstream water resources, and possibly financial hardships for participating farmers. The worse the consequences of making the wrong decision, in comparison with the right one, the greater the investment in data systems can be justified.

For example, in some locations identifying school teachers who need targeted training is an important pathway toward improved educational outcomes. Therefore whether or not the right decision is made regarding which teachers need training has high consequences. An exploratory study conducted in five schools in Ghana used real-time data and corrective teacher-feedback as a mechanism to improve children's reading skills. The aim of the study was to collect data to understand the components and processes that improve learning. The treatment also included teacher training, discussions of student-learning outcomes, and materials development, as well as increase pedagogical support for lesson planning. The percentage of learners able to read as result of this instructional treatment almost doubled to 34% from 18% (Iyengar 2016).

### 2.1.3. Identify the data that will most efficiently enable you to make the right choice

Even if a high level of data investment could be justified based on the consequences of making the wrong decision, that doesn't mean that the high benefits justify high expenses. To the contrary, the logic of the decision-oriented approach to data systems is to focus on the most efficient approach to providing the information that will enable decision makers to avoid mistakes.

A clear finding that emerges from the experience with systematic application of the decision-oriented approach to data systems is that there is no single standard for judging data that works across contexts. One must be clear about what characteristics matter for the specific purposes.

For example, in Mexico a major effort to encourage sustainable agricultural intensification undertaken by the government in collaboration with the International Maize and Wheat Improvement Center (CIMMYT) was vigorously pursuing a wide range of advanced data technologies. Project leaders came to discover that a weak link in the initiative was the proliferation of multiple incompatible crop calendars. Because the crop calendars played a strong role in the operation of credit markets, subsidy programs, input markets, and advisory services, the lack of compatibility was a major obstacle to the integrated approach that was a cornerstone of the initiative. Under the project's auspices seven organizations with competing crop calendars came together and agreed on a common calendar (Govaerts 2016). In this case the most important criterion was that the new crop calendar be treated as authoritative by the relevant parties.

| Table 2: Examples of multiple criteria relevant for evaluating data alternatives<br>Plausibly most relevant criteria by type of decision need | | | | |
|---|---|---|---|---|
| | Nature of Decision Needs | | | |
| Criteria | Meet legal reporting obligations | Make judgments with high, irreversible consequences | Make quick decisions in response to needs, opportunities | Make decisions under high uncertainty, requiring ongoing evaluation to get right |
| Accuracy | | x | x | |
| Conformity with international standards | x | | | |
| Official Status | x | | | |
| Spatial Precision | | X | | x |
| Time to Generate | | | x | |
| Frequency of Update | | | | x |

Table 2 illustrates that the criteria used to judge suitability of alternative data will vary by the nature of the decision in question. The table is not meant to identify the only way to prioritize criteria, but only to illustrate the point that how important certain characteristics of data are will vary according to the different purposes to which data could be put. A country choosing how to estimate gross national income, for example, will probably place very high emphasis on conformity with international standards, and this would be appropriate. However, if there is a decision to be made with high stakes and potentially irreversible consequences (for example, what building code will prevent catastrophic damage from plausible earthquakes), the accuracy of the data probably matter more than whether they conform to any standard. If there is a lag in getting more accurate methods into international standards, a rational government will go with the accurate over the compliant approach for this class of decision. If a decision requires quick information, then the other criteria are of no value if the data are provided too slowly to act on. Therefore time to generate will matter a great deal, whereas such a factor may not matter for other decisions. And finally, if the nature of the decision has to do with making judgments on the effectiveness of interventions so as to permit learning and improvement, then frequency of update will matter a great deal. Where there is high uncertainty and good outcomes require progressive improvement, the best baseline data will be of little use in the absence of suitably frequent updates.

Soon after the Department of Transport of the Metropolitan Assembly of Accra, Ghana was created in 2015, it contemplated policy changes regarding the city's public transit. Public transit in Accra is dominated by small privately-operated jitney services. Although the

Department had jurisdiction over the jitney services, it had little systematic information about routes, fares, vehicle quality, or ridership (Saddier et al, 2016). A reporting-oriented approach to public transit data would have focused on estimating SDG target 11.2.1, proportion of population with convenient access to public transport. The Department focused in 2016 primarily on collecting data to support decisions it had to make regarding what kinds of new regulations were necessary. In this context what mattered most was getting accurate data about ridership, fares, and routes, with high spatial precision. In partnership with Concordia University and the French Development Agency the Department was able to collect detailed information on 315 individual jitney routes, using a smartphone-based collection effort integrated with free cloud storage services and GIS software. By taking the time to articulate decision needs, the Department was able to translate a modestly-funded data exercise ($10,000) into highly valuable information (Saddier et al, 2016). For example, by comparing the routes traced by the smartphones with routes that had been registered by jitney operators with Accra authorities, over 150 “ghost routes” were discovered. These ghost routes were claimed preemptively by operators to keep out competition, but not being services. This discovery enabled the Department to plan regulatory changes. The exercise also provided the first ever estimate of public transit demand in Accra, including identification areas of unmet demand.

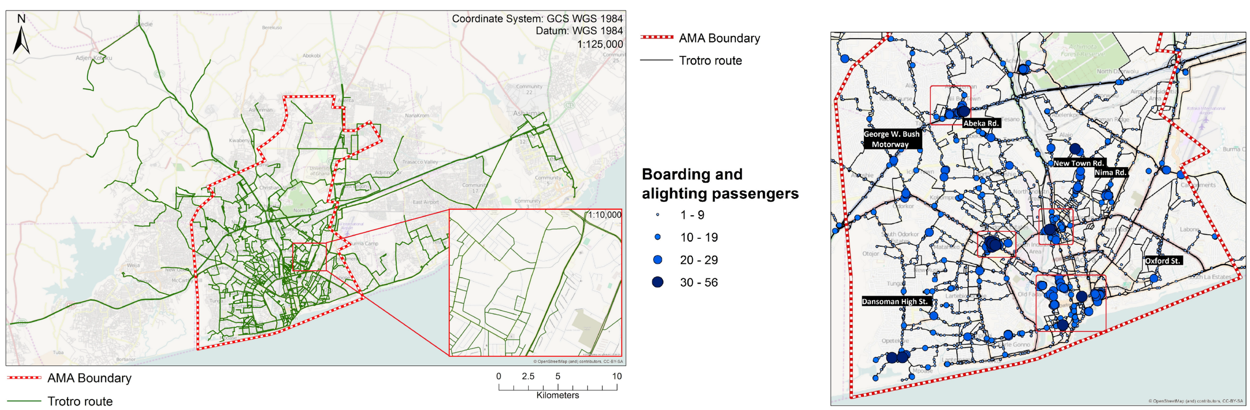


*Figure 3: Map of jitney routes (left) and passenger distribution (right) in Accra Metropolitan Area (Saddier et al, 2016).*

## 2.2. It is worth the effort

Many of the most successful corporations have converged on a framework for information management that seeks to move beyond data as a descriptive tool toward data as a vital component of planning and prescription. A popular diagram prepared by the consulting firm Gartner captures the new thinking:

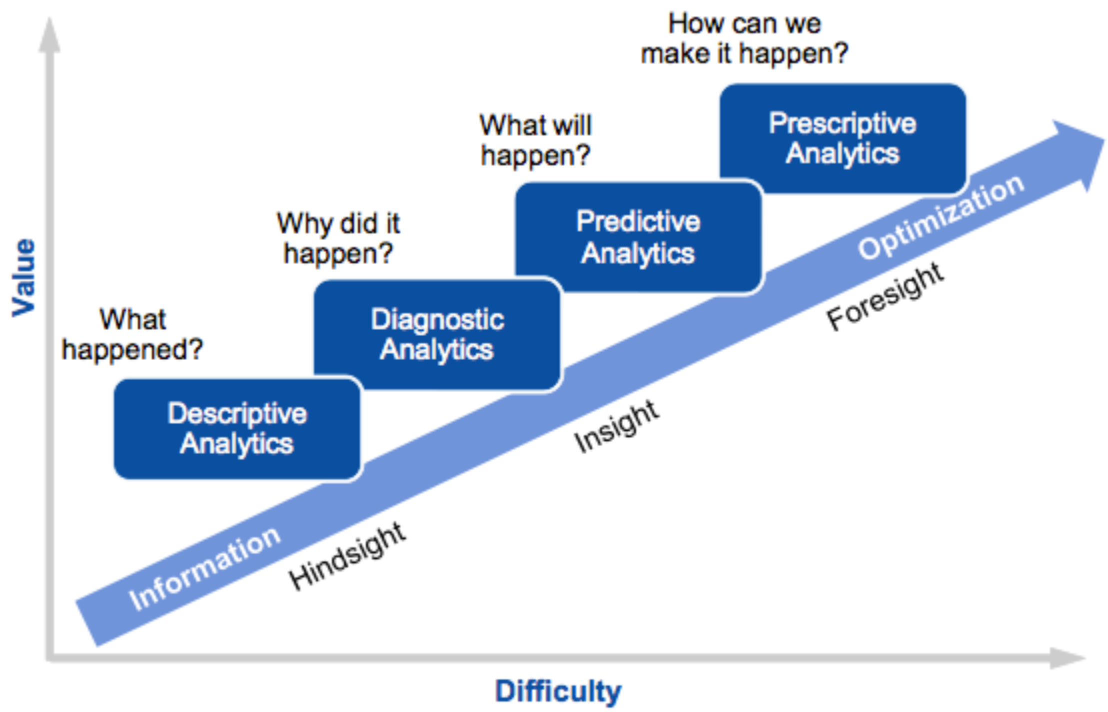


*Figure 4: Gartner Analytic Ascendancy Model. Gartner (2012)*

When a government seeks to make data choices based on priority decision needs it is moving up the value ramp in Figure 4.

### 2.3. An example from Kenya

Researchers at the World Agroforesty Center (ICRAF) in Nairobi have been working to incorporate some of Hubbard's ideas into landscape-scale sustainable development planning. They used a formal modeling exercise to help analyze a proposed water pipeline in Northern Kenya.

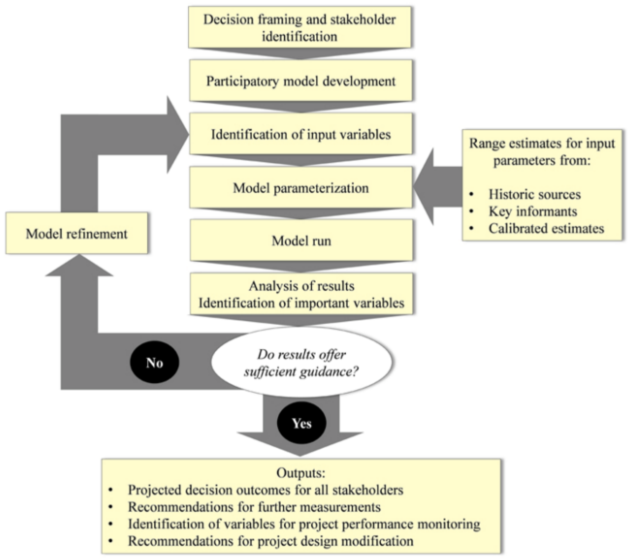


*Figure 5: Analytic Model use in N. Kenya Water Pipeline Project*

Many would expect that big investment decisions in developing countries are informed by more scientific methods than what is normally used, however most traditional scientific approaches are inadequate for such situations and limited reliable data on important issues can make fact-based decision-making a challenge. When the data required for making decisions do not exist, then it is important to make data fit the needs.

To help clarify investment decision uncertainties, anticipated project outcomes, costs/benefits and project risks under conditions of data scarcity, business analysts regularly use tools and techniques in the business context that can now be considered in the development context to make data fit for purpose.

ICRAF teams have been exploring and adapting such approaches amid imperfect information. By using participatory processes to create models that display the impacts of decisions and Applied Information Economics that offer guidelines for analysis, teams devised a decision model for a water supply intervention.

The decision analysis approach was applied in the case of the proposed Wajir water supply pipeline project in Northern Kenya, where water was to be extracted from a major aquifer near Habaswein, about 100 kilometers away, and piped to the growing city of Wajir.

This project was plagued by several risks. The information derived from decision analysis indicated that the hydrological risks of salt-water intrusion were significant, however not in comparison to the risk of political interference due to inadequate benefit sharing. The analysis showed that the greatest uncertainties were about how to value decreasing infant mortality and the reduction in water-borne disease incidence, as well as how to financially sustain the water supply systems operations. The data and analysis from the models enabled the team to modify project design plans and mitigate risks to ensure the project's chances of success.

## 2.4. Ecological Sequestration Trust in Accra

An even more ambitious modeling exercise is being carried in Accra, Ghana. The Ecological Sequestration Trust and their partners are developing and demonstrating the world's first open-source, integrated human-ecology-economics systems platform that enables resilient, disaster risk sensitive planning, policy-making, investment and procurement for city-regions globally.

Resilience.io is designed as a computer-based platform that provides an integrated systems view of a city-region. It will be an analysis and decision-support tool for collaboration and resilience decision-making. The resilience.io platform combines computer representations of resource flows, human and business activities and infrastructure systems. The platform contains a growing library of process models of typical human, industrial and ecological systems, the relevant ones of which are used in a local instance to create a tailored integrated systems model for a city-region.

In Ghana, the team is part of the Future Cities Africa programme, funded by the UK Department for International Development, and has implemented the first prototype of resilience.io. It focuses on the Water and Sanitation systems of the Greater Accra Metropolitan Area (GAMA). GAMA has a population of 4 million and less than 55% of residents enjoy access to piped potable water.

The platform uses an Agent Based Model, which enables 'bottom-up' generation of demands across socio-economic characteristics of individuals and groups (in this case for Water and Sanitation requirements) using a synthetic population developed based on census data. This demand is then used to drive a Resources - Technology Network model which aims to 'solve' demand requirements with user input scenarios of technology and policy interventions. In this way users can quickly run the model and understand interventions, which bring more favorable outcomes economically, as well as environmentally and socially.

Process Blocks include comprehensive descriptions of processes (such as a water treatment plant or anaerobic digestor) and define the input and output values of labor, energy, materials and wastes. The user can place these processes onto a geo-spatial environment in order to build up a model of the functioning city-region that is material and energy balanced.

A GAMA Technical Group has been established to facilitate data collection and to own the resilience.io prototype in Ghana. It acts as an interface between the detailed modeling and data-driven environment and decision makers in business and government. This group is expected to transition into a collaboratory 'Collaborative Laboratory' which drives integrated systems thinking to identify pathways to achieving city and national ambitions and of course the Global Goals.

Following a series of interactive workshops in June 2016 (where groups would interrogate the model and early findings) over 71% of participants thought to a high/very high degree this approach is suitable to inform policy and investment decision-making. And after a subsequent training and installation session 10/10 users would like to use resilience.io as part of their roles and within their institutions.

Decisions that are currently being tested in the model in Ghana include 1) consideration of a centralized vs. decentralized approach to potable water supply to meet SDG 6 and universal access, 2) investment decision investigating the benefits of producing biogas from wastewater treatment in terms of economics and CO2 and energy provision.

Trained users include the University of Ghana Centre of Remote Sensing and GIS, The National Planning and Development Commission, The Ministry of Local Government and Rural Development, Ghana Water Company and Zoomlion in the private sector, and the district assemblies (local government).

Next steps are to:

- Establish an investment facility, which enables users of the platform to gather evidence and build a case to secure new investment in projects that help to meet the global goals.
- Add in the other sectors such as energy, transport and agriculture
- Scale throughout Ghana
- Establish further demonstrator city-regions globally

The modeling facility, by identifying very specific opportunities to bring city planner needs in line with investor opportunities, makes the data collected for the platform of maximum value to investment and planning decision-making. Considering the information value chain summarized in Figure 4, the architects of Resilience.IO have started from the predictive analytics that they think will drive the most effective actions in support of achieving the SDGs. In particular, they focused on decision frameworks focused on approving large-scale investments in development interventions, on the part of both the private and public sectors.

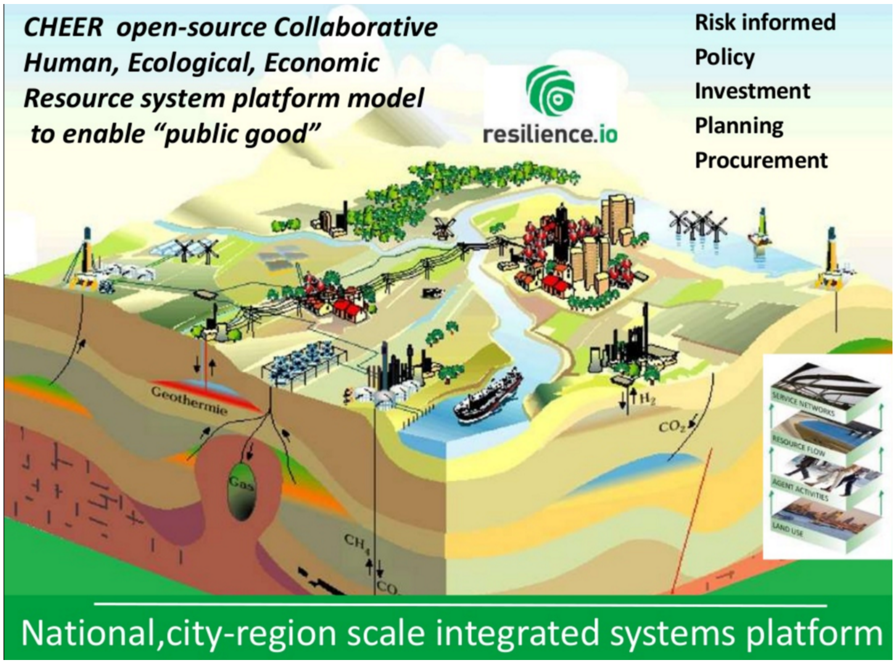


*Figure 6: The Resilience.IO platform utilizes multiple information streams to improve decisions regarding SDG-related urban investment. Data needs are prioritized according to how well they serve this purpose.*

## 2.5. Implications of Decision-oriented Data Design are Profound

A good example of how designing a data system with specific decision-support needs in mind shapes the criteria that are applied can be found in the Nigeria MDGs Information System (NMIS). Begun in 2010, NMIS was designed to support a very specific set of decisions: what are the priority needs at the Local Governmental Authority (LGA) level; which proposed projects submitted by LGAs to the Conditional Grant Scheme were most appropriate for their circumstances; which ongoing projects required higher levels of oversight and review; what broader infrastructure investments would have the biggest impact on enabling LGAs to achieve the MDGs. With these criteria in mind, an information system was developed that placed a high premium on capturing data at the facility level (school, clinic, water points, etc.) This was important because information about these facilities was crucial to making decisions about what interventions would pay off the best, and would enable very easy evaluation of progress. Because it was important to capture as many of the existing facilities as possible in a rapid manner, it was deemed impractical to rely on extant administrative data or to seek methods for enhancing administrative data collection. Instead, a new data protocol was developed, an input tool was designed for handheld Android-based smart phones, and enumerators were trained to collect all the relevant data. In this case, being complete at the facility level and having attributes that are directly relevant to planning and evaluation were so important that the value of official data collected through time-tested mechanisms less, in comparison.

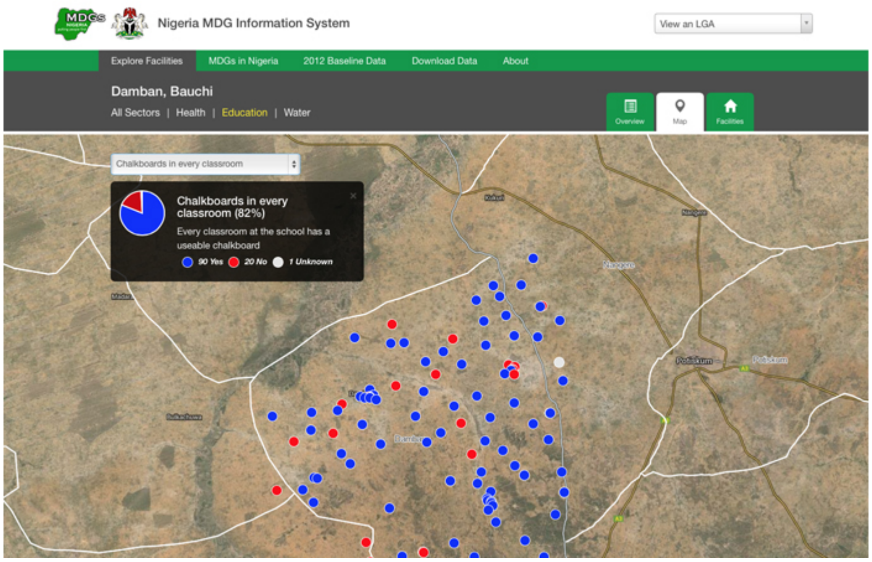


*Figure 7: NMIS data can be easily used to generate reports, maps and other output as needed by decision-makers. Here an education map is displaying schools according to whether or not they have chalkboards in each classroom, for Damban, Bauchi (http://qsel.colum)*

### 2.6. Conclusion to this section

The fact that clear, comprehensive guidance on how to align SDG data system design with decision needs is lacking should not deter governments from trying. Although optimizing data for decision support across all decisions relevant to all SDGs is not realistic, doing better than the status quo is a realistic objective, for every government. The examples above demonstrated that small investments in new data systems can often quickly generate data that are highly useful to very important decisions.

## 3. The Portfolio Approach to Investing in Data Systems

Before the data revolution, governance data systems tended to be designed in a way that applied the single best approach to measurement for a given data need.  When data collection was comparatively expensive and technology change comparatively slow, such an approach served well.  The methodology for counting population through a national census has remained robust and powerful for most of the last millennium.  Today, however, change is fast and costs are falling, though unevenly. Moreover, much more is expected from data systems – instead of serving a single governance function they more often now support many policy communities. Therefore data systems that generate the most value for the investment are designed in a manner that aims at a package of data collection strategies that produces the best value-for-cost combination.  And when such a strategy is pursued it is possible to reap an additional benefit. In much the same way that modern portfolio theory has shown how investors can improve risk-adjusted returns by combining distinctive asset categories, architects of data systems are discovering that they get better results by combining multiple approaches to data collection.

The implications are profound. Those charged with designing data systems in the past had to know one thing well; today they must be comfortable with a broad array of technology and institutional choices. In the past there was a premium on consistency and stability; today the imperative is to adapt to change fast enough to take advantage of innovation but without triggering unhelpful disruption and distrust. As a result the task of designing a fit-for-purpose data system is increasingly a task that only a purposive community can take on, because no single individual or organization can plausibly have the right information.

*The increased emphasis on data systems as portfolios of measurement technologies is driven by multiple, converging forces:*

*Technology is changing fast, putting cost-benefit relationships in constant flux*

*Policy processes are more integrated, so data systems have to be responsive to multiple communities*

*Portfolios are more efficient than single-technology systems*

In the private sector, most leading firms have responded to the rapidly changing data technology landscape by concentrating responsibility for strategic planning with respect to information systems in the position of a Chief Information Officer (CIO), as opposed to the head of Information Technology department. Within firms, CIOs are able to map organizational needs to the broad landscape of information technology in order to design data and analytic systems that add value. Where such strategies thrive they can go even further, as in the case of the innovations in information technology and informatics developed by Jack Levis at UPS which have been so successful that the core work is baked into the entire DNA of the firm.

Few countries or international organizations yet have Chief Information Officers, but all need to start building the kinds of capabilities associated with them. Successful data decisions today require carefully calibrating a portfolio of measurement solutions to meet decision-making needs.

## 3.1. Illustrations

A few examples can help make the portfolio approach to data idea more clear.

### 3.1.1. Satellite Data

Satellite data used to be employed by specialists working in narrow policy domains. The data were costly, they were difficult to work with, and they were not easy to blend with other forms of data. In the development context most applications concerned land use issues such as agriculture, forestry and mining.

A major transformation took place in 2008 when the US government started making Landsat data available at no charge. In the framework outlined above, this constituted a drastic fall in costs. Overnight it became cost-effective to combine satellite data with other data technologies. It also helped that the data integration and analysis technologies had also become more powerful and less costly. The impact on improved measurement was enormous. In a single year, for example, an estimated $2.2 billion in value is generated by use of the free Landsat data (National Research Council, 2014).

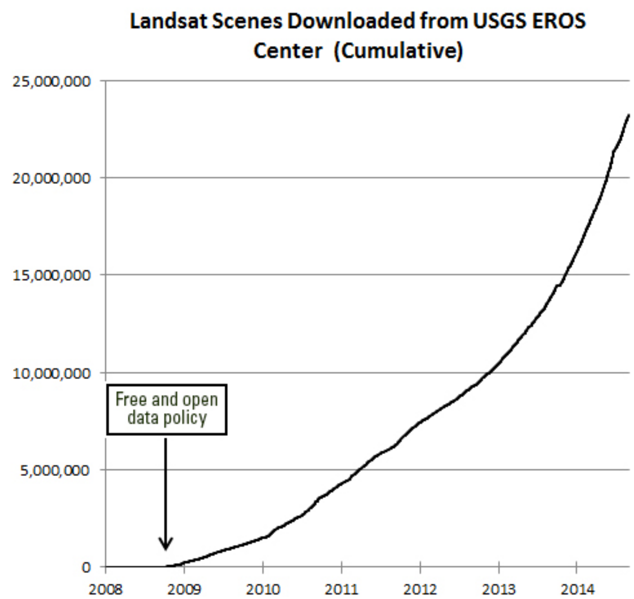


*Figure 8: When Landsat data became available at no cost the data's value could be exploited by a much larger community (National Research Council 2014)*

The European Space Agency also now makes its satellite data available at no cost, after a similar exercise found the benefits of open access far exceeding costs.

As a result, measurement efforts that used to be done with a single technology are now routinely blending in free satellite data. Instead of measuring urban extents through census records and other time-consuming data sources that are infrequently updated, several countries now track their urban areas with Landsat or other satellite data. In fact, it was the widespread adoption of such practices that made it possible for the UN Interagency Expect Advisory Group to recommend that target 11.3.1 (concerning making cities and human settlements inclusive, safe and resilient) the ratio of land consumption rate to population growth rate at comparable scale. This indicator is an excellent example of the power of combining measurement technologies. It combines census data with satellite data to calculate something in a comparable way that would have been overwhelmingly expensive just ten years ago.

The measurement of air quality has also been revolutionized by the practice of combining data on pollutants from ground stations with data from satellites. Ground stations are expensive and sparsely distributed among low-income countries. Satellite-based estimates of air pollution have been available since the late 1990s. The satellite data are less accurate than the ground station data, but cover virtually the entire land surface of the Earth and are far less expensive (the cost derives from the processing effort). By combining the two it is possible to generate air pollution exposure estimates that are more valuable than either alone. Adding a third data source, georeferenced census data, permits the calculation of an even more valuable measure, the number of people exposed to dangerous thresholds of air pollution. Robust methodologies have been published, via an expert process initiated by the World Health Organization (Shaddick et al 2016).

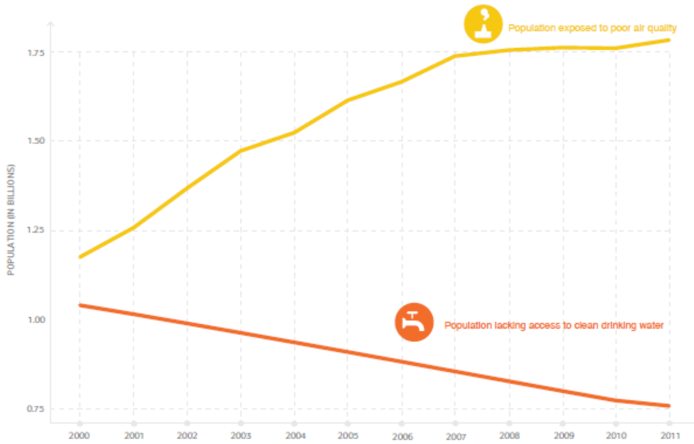


*Figure 9: The Millennium Development Goals included a target for improving access to improved drinking water, but did not have targets regarding air quality. The drinking water situation improved significantly but the air quality situation got much worse. Source: 2014 Environmental Performance Index.*

The lack of a robust metric for exposure to air pollution was one of the reasons the Millennium Development Goals (MDGs) lacked a target regarding air pollution. A comparison of drinking water and air pollution data, as shown in **Error! Reference source not found.**, shows how much worse the air pollution got. The international community did not even know how much worse the air pollution problem was until satellite-based estimates began to be published in 2010 (de Sherbinin et al 2014). One reason that there is an SDG target on substantially reducing the number of deaths and illnesses from air pollution is that the integration of these three data sources – satellites, ground stations, and censuses – makes it possible to reliably measure exposure.

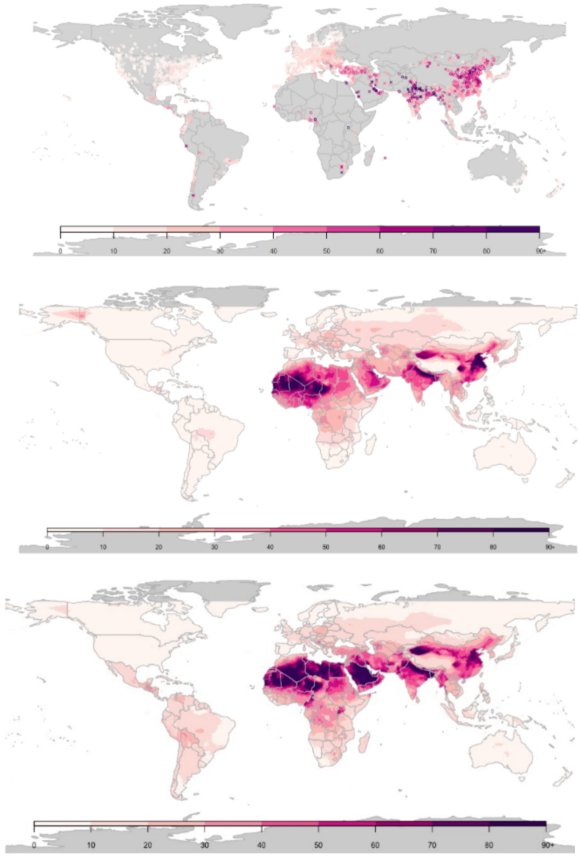

*Figure 10: By combining ground station air pollution data (top) with satellite-based estimates (middle) it is possible to generate superior estimates (below) at*

The impacts of freely available satellite on data system design provides a good example of another important attribute of the portfolio approach. A technology disruption regarding data can generate cascades of additional disruptions as multiple actors in the system adjust. For example, Google responded to the decision to make Landsat available for free by loading the entire archive into its servers and developing algorithms that made the data easily usable through an online tool, Google Earth Engine, along with additional US and European satellite data. That has meant that only are the data free, but access and use is within practical reach of far more users. That move in turn has accelerated the incorporation of satellite data in government metrics relevant to sustainable development, in diverse policy areas including flood risk management, wildfire control, nature conservation, and humanitarian disaster response.

### 3.1.2. High-speed spectroscopy labs

The big impact of freely available satellite data on data system design is a case of producing a transformative public good. But similar positive disruptions can emerge from technologies that do not lend themselves to public good provision. One example is the application of high-speed spectroscopy to agricultural soil measurement. Using a spectrometer housed at the World Agroforestry Center (ICRAF), a growing number of governments, NGOs, and commercial actors are collecting soil samples and getting highly detailed and accurate measurements from the spectrometer. Traditional methods for measuring chemical soil qualities required expensive, time-consuming and labor-intensive lab procedures. A spectrometer can determine the chemical makeup of the soil very quickly by directing light of a specific set of wavelengths at a sample and measuring the nature of the light reflected back.

Advances in spectroscopy technology have been driven by the pharmaceutical industry and other fields where precise measurements of chemical qualities is extremely important. The costs are now low enough that it is feasible to use the technology for agricultural applications in developing countries.

The emergence of this technology in the agricultural realm has had the effect of dramatically changing the incentives to collect and measure soil samples. The old technology cost about $100 per sample and took several days. With the new technology the cost is only $1 and hundreds of samples can be measured in day. The drop in price (and increase in accuracy and number of parameters measured) has led to an increase in the number of samples being collected and measured. Ethiopia collected and measured over 18,000 samples, example (http://www.ata.gov.et/highlighted-deliverables/ethiosis/).

And the cascading impacts continue from there. The new technology makes it attractive to collect and measure more soil samples, and that in turn makes it possible to create "digital soil maps" characterizing soil attributes over an entire country (Sanchez et al 2009). That greatly increases the value of carrying out trials testing the impact of fertilizer and other inputs on crop yields, because the trials can be designed more scientifically based on the soil maps and because the results of the trials can be used more powerfully, because there is good information on the distribution of the relevant conditions.

A spectrometer cannot be operated as a public good, because its use has limits. The number of Landsat users, by contrast, could go from thousands to millions to tens of millions without creating any scarcity. Landsat and other satellite data made available by the US and European space agencies has an additional advantage. The benefits to the producer of the data are large enough to justify the entire investment. But there are physical limits to how many samples a spectrometer can process. That makes it potentially more of a "club good," in which a group of users come together to share in the benefits.

### 3.1.3. High-resolution population mapping in northern Nigeria

When a Bill and Melinda Gates Foundation initiative to eradicate polio began working in northern Nigeria, it quickly became apparent that existing data on settlement and population distribution was not appropriate suited to the decisions that had to be made. There was inconsistency, incompleteness and inaccuracy regarding the location and names of settlements. The population counts in each settlement were not up to date and in some cases highly inaccurate. As a result it was impossible to design and implement an effective vaccination campaign.

No single approach to measurement would have been capable of meeting decision support needs. Field surveys were too slow and prone to error. High-resolution imagery could detect structures but not generate estimates of populations. By combining these two sources of information, along with other data such as road networks and land cover maps, it was possible to generate highly accurate estimates of population distribution in less than three months. These new estimates were instrumental in the effort to deliver vaccines where and when they were needed.

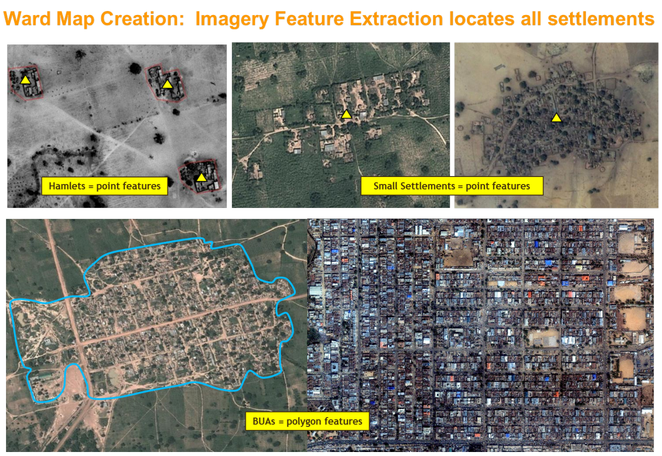


*Figure 11: High-Resolution Imagery Helped Create More Accurate Population Estimates in Northern Nigeria*

## 3.2. Getting Started with the Portfolio Approach

A government seeking to design a data collection portfolio in support of the SDGs must focus on several interdependent questions. The exercise is more complex than comparing existing data capabilities against international best practices, but the additional effort is made worthwhile by the increased value that results.

There are formal methods for identifying appropriate data portfolios, but they have not been used in cases as complex as the SDGs and therefore are not yet appropriate for high-level data design. However, the core elements of the methods can be approximated through consultative processes as outlined below.

| Table 3: Determining a Useful Data Portfolio | |
|---|---|
| Step | Description |
| 1 | Assemble highest-priority decision needs, as described in Section 2. Use the SDGs as the foundation. Utilize consultative dialogues with experts, national and local government officials, private sector representatives, and civil society. |
| 2 | Inventory the data that can be counted on from existing sources. Utilize desk research and consultations with local and international data experts. |
| 3 | Catalog data technologies in use elsewhere for similar decision needs. For the decisions that are the most important, what data sources are in use worldwide, and which seem most effective? Utilize consultations with international development and data experts. |
| 4 | For the most important decision needs, evaluate fitness for use for the relevant data sources identified in steps 2 and 3. To the extent possible, use common benchmarks such as cost, accuracy, timeliness, spatial precision, authority, and update frequency. If quantitative estimates are not possible, |

| | |
|---|---|
| | use expert judgment to assign an estimate or a qualitative score. Bring together participants in step 1 with local and international data experts. |
| 5 | Using the results from step 4, create first-order matches between available data sources and decision needs. Which data sources are most promising for meeting decision needs? Use the same participants as in step 4. |
| 6 | Begin an iterative process to move from single-data-source evaluations to portfolio-based evaluations. For decision needs where data availability is heavily cost-constrained, explore options for bringing costs down through blended approaches. Typically such options will emerge when there is a costly approach that is inadequate because of spatial coverage or update frequency and a less expensive approach with good spatial coverage and update frequency, but is lower on other criteria such as accuracy or authority. Blending satellite and ground station air pollution data is an example.<br><br>If the data options are constrained by some other parameter, e.g. accuracy, explore options for mounting a selective new measurement campaign optimized to boost that parameter in the overall portfolio. Such options are likely to emerge when the value of information is very high (because of its impact on priority decisions) but none of the available data sources generate adequate information, even in combination. An example is the use of Open Street Map campaigns to collect spatial data relevant to hazard vulnerability in Jakarta. |
| 7 | Test the recommendations that emerge from step 6 by generating sample data products and reviewing their value with the stakeholders from step 1. |
| 8 | Where specific approaches emerge that appear to provide high value relative to cost, begin producing relevant data products. Having the first products generated in an open manner, with opportunities for review by experts and users, and with appropriate validation exercises, is important for achieving quality and trust. |

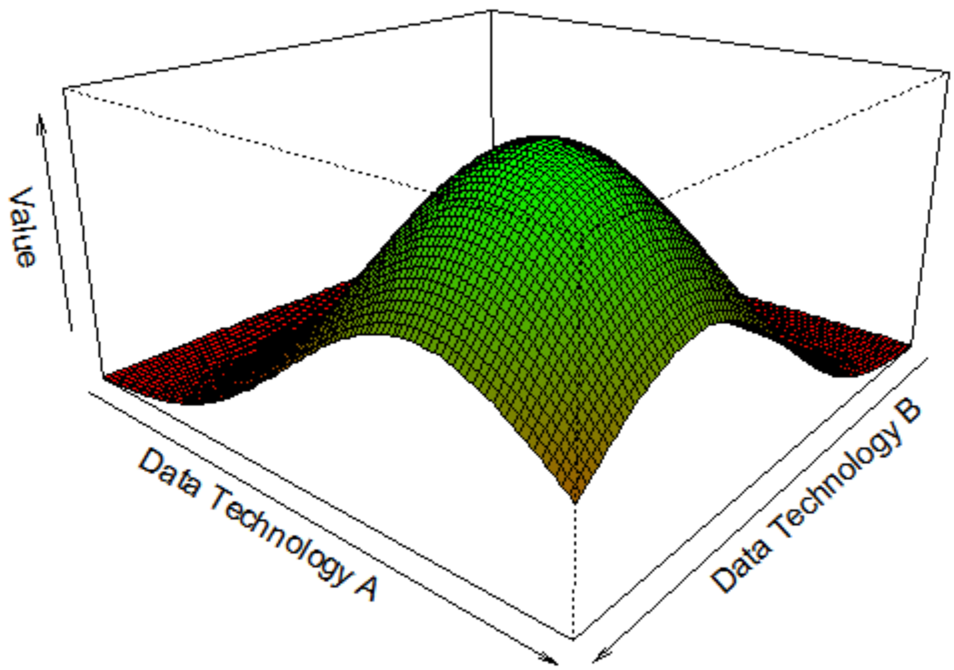


*Figure 12: Hypothetical Value of Data Portfolio as a Function of Contribution from Two Data Technologies*

Figure: By combining data technologies into a portfolio, it is possible to increase the value to decision makers beyond the sum of the value of the individual technologies.

### 3.3. Putting all the pieces together in a complex problem: Landscape Surveillance

Much of the world faces threats to sustainable land management. Pressures in the form of agricultural intensification, deforestation, water extractions, and settlement growth are increasing. Although goals are in place to limit the damage from these pressures so that human development needs can be met sustainably, these goals are seldom accompanied by clear metrics to permit even the most basic planning and evaluation.

Simply scaling up traditional approaches to measuring such phenomena is not a viable option. Such approaches are slow and expensive, and often do not generate the most relevant measures for shaping decisions.

An alternative approach has arisen that blends information from multiple sources, generating useful information that can rapidly be put to effective use. The approach can be summarized as follows:

- Define the measurement targets in terms of relevant parameters including time, space and uncertainty.
- Develop a model for estimating parameters at appropriate temporal and spatial resolutions and with appropriate levels of uncertainty.
- Develop a sampling frame that is optimized for the measurement targets and the estimation model, relying on available information.
- Use no-cost or very-low-cost data to populate the estimation model with foundational information.
- Collect additional data using the most appropriate data-collection technology, using the estimation model and sampling frame to guide deployment to the optimal level.
- Regularly and rapidly validate the estimations to identify opportunities for improvement.
- Although such methods are not yet the norm in official statistics, they are based on statistical methodologies that have been in active development for over a decade, and are in routine operational use in much of the private sector.

For example, the Africa Soil Information Service (AfSIS) has used this approach to generate detailed maps of soil characteristics in a number of sub-Saharan African countries, along with Africa-wide maps on select metrics (Hengl et al 2015). Freely available satellite data, in combination with existing maps of soil conditions and crop patterns, was used to generate robust sample frames for field campaigns in which soil samples and observations were collected. The introduction of satellite data at the sample frame stage made the field campaigns more efficient than traditional methods. The samples were processed in a spectroscopy facility at the World Agroforestry Center (Viscarra Rossel et al 2016). Using the spectrometer reduced the cost of sample processing from 50-100 fold as compared to earlier methods. The spectroscopy results were combined with additional satellite data and incorporated into a statistical model that generated detailed spatial estimates of soil properties. Because accurate information about land use was helpful in improving the quality of the estimates, a low-cost crowd-sourcing platform was created that enabled thousands of points to be classified, utilizing free high-resolution imagery available through Google Earth and volunteer participants.

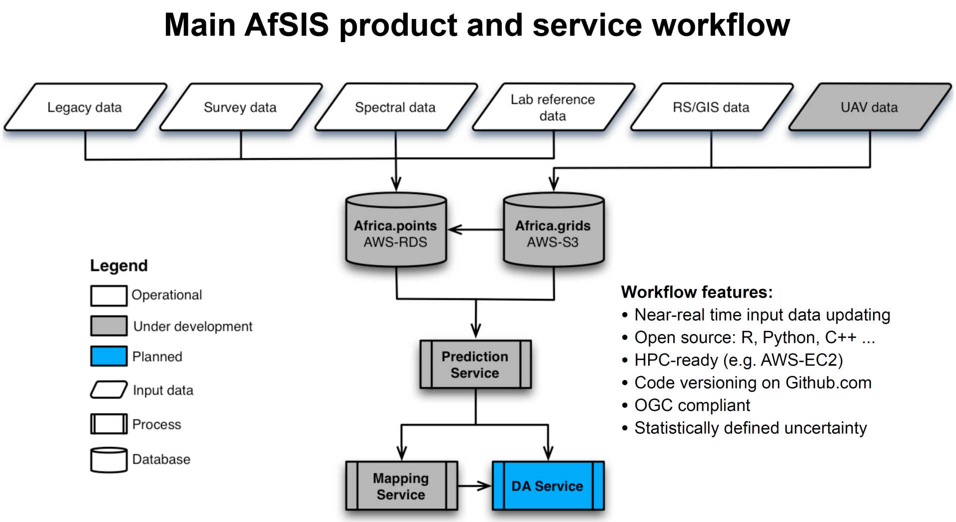


*Figure 13Multipe Data Inputs Are Utilized by AfSIS to Generate Maps of Soil Properties*

These models can be put to work generating maps and tables in support of specific decisions. A major example concerns the question of what fertilizer blends are best suited to different locations. Prior to creating the Ethiopia Soil Information Service (EthioSIS), in partnership with AfSIS, Ethiopia had a single fertilizer blend for use throughout the country. Detailed soil maps produced by EthioSIS permitted specification of additional blends, supporting sustainable intensification of agricultural output (Wösten et al 2015).

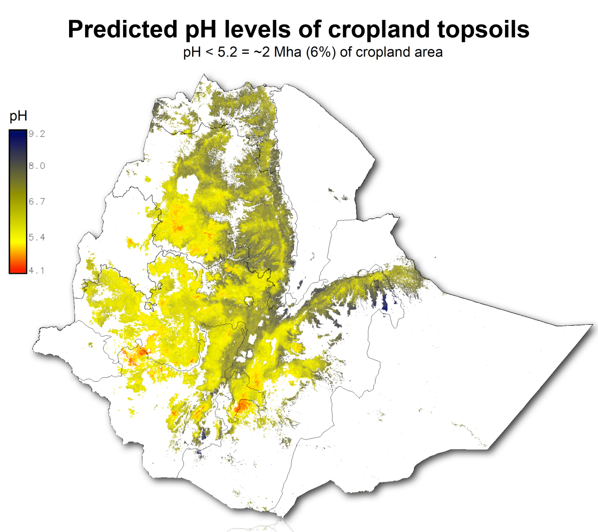


*Figure 14: Detailed Maps of Soil Properties can be Generated by Statistical Models that Incorporate Multiple Data Inputs*

## 4. Fostering Sustained Innovation

Continuity and reliability have always been important virtues when governments have designed and implemented data and information systems. Where continuity and reliability are of the highest importance, change is often slow. National censuses have tended to change slowly in most countries, for example. The UN Statistics Division, for example, found that for their 2010-round national censuses 83% of 138 surveyed countries were using a “traditional” methodology (UNSD nd). This methodology was first elaborated in the late 18th century and its core elements remain unchanged.

Increasingly, however, as decision-maker needs and data technologies change, sometimes quite rapidly, governments find that effective innovation is often as important as continuity and reliability, sometimes more so.

Here we summarize some guidelines for supporting sustained innovation for SDG data and information systems.

- Legislation and mandates for data systems that specify the types of methods and technologies to be used are more restrictive than those that refer to what is needed from the data system. For example, many OECD countries have shifted from censuses as a means to count and characterize their populations to use of registries and administrative data, because such alternatives cost less and provide more information. The United States, has been unable to make such as shift, in large part because its constitution requires a census.
- Develop capabilities to explore and experiment with emerging technologies. Some countries have started creating innovation nodes within their statistical apparatus, and given them mandates to test promising uses of new data approaches. Colombia’s National Administrative Department of Statistics, for example, created a “Smart Data” unit to support innovation and learning.
- Take user needs seriously, and evaluate the worth of your data systems from the perspective of how well they meet user’s decision-support needs. Rather than judging data systems by well they meet their target specifications, also look at how well they perform from the decision makers’ point of view. During the 2015-2016 El Niño that severely altered weather patterns East Africa and other regions, Kenya undertook an effort to quickly assemble useful data for guiding effective responses to the cycles of floods and droughts that were expected. What they found was that the data systems that had emerged around climate risks in the years since the last major El Niño in 1998 were doing a good job at delivering useful climate information, but that they were neglecting vital information about the toolkit for responding. What the Kenyan government most wanted was data on what interventions have been attempted during previous El Niño events, and how much they cost and what level of benefits did they produce (Thigo, 2015). Such knowledge about the fit of a data system to user needs only emerges when there is a practice of routinely asking the question.

### 4.1. Principles for lasting success

Principles for Designing and Sustaining Effective Data and Information Systems

Feedback within a sector and across sectors.

To help you get started, these are eight important principles to consider when designing effective information systems:

Think about how data will be used

Seek to have your data drive the future, not monitor the past.

Think about making timely and impactful decisions.

Do things rapidly and deliver results.

Think about how to make data as powerful as possible

Figure out what is needed to maximize the power of your data in practice.

Data are almost always more powerful in combination, so design the combination instead of leaving it to chance.

Find a sustainable path to continuous innovation

Figure out how to create new programs and mechanisms to support ongoing improvement.

Design for learning, not being perfect or following specs

What matters is how data are used and what people learn from engaging, not only what kind of product it is

*Better to focus our measurement on things that matter

## 5. References

Clapp, A.; DauSchmidt, N.; Millar, M.; Hubbard, D.; Shepherd, K. A, 2013. “Survey and analysis of the data requirements for stakeholders in African agriculture,” World Agroforestry Centre (ICRAF), Nairobi, Kenya.

de Sherbinin, Alex, Marc A. Levy, Erica Zell, Stephanie Weber, and Malanding Jaiteh (2014). “Using Satellite Data to Develop Environmental Indicators, " *Environmental Research Letters:* 9.

Gartner, 2012. “Predicts 2013: Information Innovation,” https://www.gartner.com/doc/2278715/predicts--information-innovation.

Govaerts, Bram, 2016. Personal Communication (25 August).

Hengl, T., G.B.M. Heuvelink, B. Kempen, J.G.B. Leenaars, M.G. Walsh, K.D. Shepherd, A. Sila, R.A. MacMillan, J. Mendes de Jesus, L. Tamene and J.E. Tondoh (2015), Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS ONE* 10(6): e0125814.

Hubbard, Douglas W., 2014. *How to measure anything: Finding the value of intangibles in business*. John Wiley & Sons.

IAEG (Independent Expert Advisory Group on a Data Revolution for Sustainable Development), 2014. "A world that counts: mobilizing the data revolution for sustainable development."

Iyengar, Radhika, et al. 2016. "Using Real-time Data and Corrective Teacher-feedback as a Mechanism to Improve Children’s Reading Skills: An Exploratory Study in the Millennium Village Site of Bonsaaso, Ghana." J*ournal of Education and Training Studies* 4.9: 82-99.

Luedeling, Eike, et al. "Fresh groundwater for Wajir—ex-ante assessment of uncertain benefits for multiple stakeholders in a water supply project in Northern Kenya." *Frontiers in Environmental Science* 3 (2015): 16.

National Research Council, 2014. *Landsat and Beyond: Sustaining and Enhancing the Nation's Land Imaging Program.* National Academies Press.

O’Connell, Deborah; Braid, Andrew’ Raison, John; Hatfield-Dodds, Stephen; Wiedmann, Thomas; Cowie, Annette; Littleboy, Anna; Clark, Megan, 2013. “Navigating sustainability: measurement, evaluation and action.” CSIRO, Australia.

Rosenstock, Todd S., Christine Lamanna, Sabrina Chesterman, James Hammond, Suneetha Kadiyala, Eike Luedeling, Keith Shepherd, Brian DeRenzi, and Mark T. van Wijk. "When less is more: innovations for tracking progress toward global targets." Current Opinion in Environmental Sustainability 26 (2017): 54-61.

Saddier, Simon; Zachary Patterson; Alex Johnson; and Megan Chan. 2016. "Mapping the Jitney Network with Smartphones in Accra, Ghana: The AccraMobile Experiment." *Transportation Research Record: Journal of the Transportation Research Board* 2581 (2016): 113-122.

Sanchez, Pedro A., et al. 2009. "Digital soil map of the world." *Science* 325.5941, 680-681.

SDSN (Sustainable Development Solutions Network) 2015. “Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.”

Shaddick, Gavin et al (2016), “Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution,” arXiv:1609.00141v2

Shaman, Jeffrey, and Alicia Karspeck. "Forecasting seasonal outbreaks of influenza." Proceedings of the National Academy of Sciences 109.50 (2012): 20425-20430.

Shepherd, Keith et al, 2015. "Development goals should enable decision-making." Nature 523, no. 7559: 152.

Thigo, Philip, 2015, personal communication (October 2).

Viscarra Rossel, R.A., Behrens T., Ben-Dor E., Brown D.J., Demattê J.A.M., Shepherd K.D., et al, 2016. A global spectral library to characterize the world's soil. *Earth-Science Reviews* 155; 198-230.

World Bank, 2013, “Agricultural Sector Risk Assessment in Niger: Moving from Crisis Response to Long-Term Risk Management,” Report 74322-NE.

Wösten, H, Leenaars JGB, Eyasu E, Keestra S, Mol G, Zaal A, Wallinga J, and Jansen B, 2015. Soil map providing basic information for crop and site specific water and fertility recommendations in Ethiopia, paper prepared for presentation at Wageningen Soil Conference 2015.