PULTE INSTITUTE
FOR GLOBAL DEVELOPMENT

Maximizing Returns on Data Science Investments: The Evolution of Data-Driven Decision-Making in Development

Cooper, M.

Perrin, P.

UNIVERSITY OF NOTRE DAME | KEOUGH SCHOOL OF GLOBAL AFFAIRS

PULTE INSTITUTE
FOR GLOBAL DEVELOPMENT

Maximizing Returns on Data Science Investments: The Evolution of Data-Driven Decision-Making in Development

Cooper, M.
Perrin, P.

Report prepared for the Pulte Institute for Global Development,
part of the Keough School of Global Affairs, University of Notre Dame

3150 Jenkins Nanovic Halls
Notre Dame, Indiana 46556
(574) 631-2940
Website: pulte.nd.edu
Email: globaldevelopment@nd.edu

TABLE OF CONTENTS

INTRODUCTION

ACRONYMS

CSO: Civil Society Organization

DCLI: Data Collaboratives for Local Impact

DSC: Data Steering Committee

KII: Key Informant Interview

MCC: Millennium Challenge Corporation

M&E: Monitoring and Evaluation

MEL: Monitoring, Evaluation, and Learning

PEPFAR: President's Emergency Plan for Aids Response

FOREWORD AND ACKNOWLEDGMENTS

This research is not only due to the funding of the Pulte Institute; it is the result of individuals who donated their time to improve its value for the community of practice. These contributions are often assumed in our field but they should not be underestimated. The design (especially Phase 2) of this research asked a lot from key informants, advisors, and others who were gracious in making themselves available. Any value that comes from this research can be attributed to the collaborative action of this group; however, we specifically want to highlight the contributions of Agnieszka Rawa, Mitch Blazer, Brian Fogerty, Michael Bamberger, Chris Watson, Alberto Lizzi, Ric Shreves, Rivandra Royono, Kara Morgan, and Valentine Ghandi. A very special thank you to Dr. Paul Perrin whose leadership was crucial, as well as Heather Asiala, Cory Hankins, and Edward Jurkovic who had the daunting task of making the work readable.

ABOUT THE AUTHORS

Michael Cooper has worked with the U.S. State Department, USAID, Millennium Challenge Corporation and others in designing, measuring, and learning about the effectiveness of development interventions with a focus on decision-making. His latest work on future trends includes data science, decision science and blockchain technology for social impact and has been featured by the American Evaluation Association, European Evaluation Society, and various other publications. He currently lives in Colorado where he manages Emergence, a research and advisory consultancy firm. He can be contacted at emergence.cooper@gmail.com.

Dr. Paul Perrin leads the Pulte Institute's Evidence and Learning Division and is an international health, humanitarian, and development research and practice professional with over a decade of work experience in academic, government, and non-governmental settings. To date, his work has largely focused on supporting the conceptualization, design, implementation, analysis, and reporting on monitoring and evaluation and implementation science activities in resource-challenged environments. His latest work centers on optimizing the integration of new principles, protocols, and methods from emerging technologies into the supply of evidence for Development learning. He can be contacted at pperrin@nd.edu.

---

*The views expressed in this paper are the authors' alone and do not reflect the opinion of the Pulte Institute, Notre Dame at large or any person or organization mentioned in this paper. All content of this paper is the sole responsibility of the authors.

INTRODUCTION

Data science is becoming more of a tool in extracting value from data across industries. Within international development data science, big data, and other related terms have largely been relegated to buzz words with expanding, but limited, real-world applications. As the fields of big data and advanced analytics (amongst others) converge under the auspices of data science, new ways of extracting value from data are emerging at a pace that requires a need for strategic thinking that can guide tailored operational guidance and tools.

International Development (referred to as "Development" in this report) has been an industry that has prioritized data-driven decision-making to emphasize "learning" for some time (World Bank, 2019). Many consider the rise of data science as the next evolutionary (some would say "revolutionary") cycle of data-driven decision-making.

In response, the Pulte Institute for Global Development[1] – part of the University of Notre Dame's Keough School of Global Affairs – partnered with Michael Cooper of Emergence[2] to fund a perspectives project that would propose recommendations for how the Development industry could see better returns on their investments in data science.

Improving Development decision-making has been a priority for decades as different campaigns[3] have attempted to provide the evidence-driven guidance that improves the next generation of Development decision-making practice. With the advent of the digital economy and the corresponding explosion in the supply of data, the next generation is here and the need for guidance is increasing.

The need for this research rests on two primary assumptions that lead to a Development problem:

- **Assumption #1:** Contexts in which the Development community primarily operate will see an increase in the use of digital technologies that result in more data for use in decision-making, while at the same time there is insufficient understanding of data science capacities to unlock the value in this data.
- **Assumption #2:** The current trends in the use of data science methods and techniques in Development will increase and diversify across decision-making levels, sectors, and types of data used.
- **Development Problem: Current decision-making practices within Development do not always make optimal use of data science to create both value from data and returns on data investments.**

Assumption #1 and #2 go hand in hand and have evidence supporting them[4]. Evidence around the saturation of smartphones amongst poorer populations, generation of big data sets by governments, and the use of data science-related approaches to analyze and use this data is fairly well known. Hence this research focuses on filling the need left by the problem, namely the gaps in current thinking around data-driven decision-making.

RESEARCH QUESTIONS

These research questions were chosen after some initial discussions with advisors from the Development industry (outlined below) and private sector data science practitioners. This feedback resulted in two questions to answer:

1. What lessons from previous and current experience with data-driven decision-making can help inform better practices for using data science products in **public sector** Development?
2. What lessons from the **private sector** can help inform better practices for using data science products in **public sector** Development?

The implicit argument presented is that the lessons learned from current data science applications in both the public and

[1]https://pulte.nd.edu/

[2]Emergence.consultant.com

[3]Examples of such campaigns include the World Bank's "Data-Driven Development" and the Learning Lab's "What is Adaptive Management?" to name just a couple.

[4]Evidence of assumptions 1 and 2 can be found in The World Bank's World Development Report 2016: Digital Dividends and the International Telecommunication Union's Measuring Digital Development Facts and Figures 2020 report.

However, it should be noted upfront that while this paper uses phrases like Development, Private Sector, and general statements about their mandates and perspectives, this is not to assume homogeneity within these groups and their functions, much less their respective data science applications. A higher level of abstraction was needed for this project given the abstract nature of the research questions decided upon after input from advisors.

While limited in its ability to provide specific operational guidance on particular applications, this level of abstraction was able to highlight general lessons that are directly applicable to any use case. Hence the real value of this research will come in how its general results are modified to fit individual use cases, something far beyond the scope of this current project.

This research seeks to focus on how data science can better inform decision-making at all levels. As will be seen, the evidence that informs this research was gathered from data science efforts aimed at national, sub-national, and community-level decision-making amongst government, Civil Society Organizations (CSOs), development donors, etc. **Hence the overall recommendations are meant to be broad to enable actors to modify them to their specific use; whether they are working with donor decision-making at the global policy level or CSO adaptive management decision-making at the field level.**

Term Definitions

Terms such as decision-making and data science are nebulous with shifting and contextualized definitions; therefore, any final definition of data science, big data, and other associated terms is beyond the scope of this research. For the purposes of this research, the authors have created the definitions below:

- **Data Science:** Data science combines multiple fields including statistics, scientific methods, and data analysis to extract value from data that cannot be analyzed using "traditional" methods due to the size and complexity of the dataset.
- **Big Data:** Large data sets that cannot be analyzed using traditional methods, require additional computational power to analyze and are sourced digital ecosystems.
- **Data Ecosystem:** The technology infrastructure, social dynamics, data analytics, and decision-making used to capture, analyze, and use data for specific ends.
- **Decision-Making:** The process of identifying a need for action that must stem from a decision that needs a calculated way of gathering data to inform it. This research defines "decision" to mean any selection between options by an individual or group authorized to make that specific decision.
- **Stakeholder:** An individual or group that has an incentive or influence on the data problem in some way. Stakeholders can be beneficiaries with contextualized knowledge of the operating environment; hence their influence is in their knowledge of the problem sets.
- **Data Science Application:** The use of data science methodologies and/or use of big data to extract value from data.
- **Data Science Team:** The entire team of computer scientists, coders, statisticians/mathematicians, etc. that help identify data problems, propose data science solutions, and implement data science applications (model selection, analysis, communication, etc.) in order to produce data science products.
- **Data Science Products:** The products (outputs) that result from a data science application. This includes data interpretations, communications, visualizations, and data sets.
- **Data User:** Any person or entity that uses data at some point in a data lifecycle. This can be policy leaders who use the data for decision-making, programmers who use it for their analysis, community leaders who use it for planning, etc. The needs of data users are as varied as the dimensions of the data ecosystem, meaning the more complex the ecosystem the more varied the needs of data users. Data user needs typically drive capacity development efforts.
- **Development:** The goals, strategies, activities, and actors within the international development industry.

RESEARCH DESIGN OUTLINE

While largely a perspectives piece, this research is meant to be formative and exploratory. A more open-ended design was intended given that the application of data science in Development is still in its infancy stage. In addition, the wide variety of data science applications covered in the research did not lend themselves to standardized questions for surveys, etc.

This research was comprised of two phases. Phase One consisted of Key Informant Interviews (KIIs)[5], which were semi-structured and exploratory in order to identify areas of success, problems, and lessons learned. KII respondents were purposely selected based on their experience in applying data science methods in private, public, and Development sectors as a policymaker, thought leader, implementer, or donor. This maximized the representativeness of use cases as much as possible, acknowledging that responses would be "a mile wide and an inch deep." In addition, an extensive literature review was conducted on private and public sector data science applications. This review had more of a focus on decision-making in addition to literature on the uptake of decision science practices in the private sector.

Phase One resulted in preliminary findings and recommendations. Phase Two consisted of feedback on Phase One findings as applied to the real-world experience of an advisory board consisting of representatives from:

- MCC/PEPFAR (Data Collaboratives for Local Impact)
- Premise Data
- The Development Café
- United Nations Development Programme: Global Pulse Jakarta
- Mercy Corp
- Pulte Institute for Global Development at the University of Notre Dame
- Lucy Family Center for Data and Society at the University of Notre Dame
- Michael Bamberger - Rockefeller Foundation Consultant
- Kara Morgan - Ohio State University

RESEARCH LIMITATIONS

This research cannot be considered fully representative of data science use cases across sectors, methodologies, and geographic areas in the Development and private sectors. Given the parameters (especially time) for this perspectives research, we sought to maximize the value of the research by 1) selecting KIIs and case studies from different points of view (implementers, thought leaders, donors, etc.) and 2) introducing Phase Two of the research to allow for the initial research results to be "white-board" tested against real-world problems.

OVERVIEW

This paper was organized by an initial review of the research findings for each research question, in order to build the understanding required for the recommendations and conclusions. The recommendations take the findings of the research questions and integrate them into a general decision-making process that Development actors can modify per their capacity and priorities.

The conclusions section is a summary of the feedback received by advisors. This feedback includes perceived gaps in the perspectives and research, needs for additional research, and points of agreement. This feedback is perhaps the most important piece of this project since it is the only opportunity of any original thinking to be tested against real-world problems. The feedback is also extremely helpful in highlighting next steps in moving beyond the thinking captured here.

To help the reader, key points that are directly related to the final results and recommendations are in bold to help distinguish them.

[5] A full list of reference materials and KIIs are included in Annex 1.

FINDINGS

Research findings (from KIIs and literature review) are organized by research question.

Research Question #1:

What lessons from previous and current experience with "data-driven decision-making" can help inform better practices for using data science products in development?

The Development industry has emphasized the importance of using traditional data as evidence in decision-making for decades[6]. Development data is primarily used for accountability purposes but with increased importance on Monitoring & Evaluation (M&E), learning for performance improvement has grown in priority. With the emergence of the digital revolution and the exponential growth in the supply of data, data is increasingly seen as an investment in decision-making across a wide variety of data uses. There has been outstanding research done with Development leaders, practitioners, and thought leaders to provide a landscape "state of the field" that confirms the wide variety of thinking and use cases related to Development use of data science[7].

To answer Research Question #1 we used two use cases, a variety of KII's with subject matter experts, and a literature review to better understand how the Development industry is adapting to this explosion in data supply.

The use cases were selected due to their alternative mandates. The Data Collaboratives for Local Impact (DCLI) program seeks to empower communities to define and solve their own social problems with data solutions, hence there is a large focus on building capacity within data ecosystems. Meanwhile, UN Global Pulse is primarily responsive to demands from government and other stakeholders to assist in identifying solutions and value from emerging big data sets. While these are just two approaches amongst many, they were selected because they are representative of the ends of a functional spectrum. One spectrum end is primarily characterized by responsiveness to data problems as they emerge from stakeholders (Ministries, CSO's, etc.) to the other end that is solely focused on empowering and building capacity within local systems (collaborations of government, community, and private actors) to sustainably identify and solve their own problems with data.

Data Science Use Cases in Development: Empowering Systems and Responsive Solutions

In 2015, the United States President's Emergency Plan for AIDS Relief (PEPFAR) and the Millennium Challenge Corporation (MCC) entered a partnership to create the Data Collaboratives for Local Impact (DCLI) program. DCLI seeks to improve the capacity of communities and organizations to use data to solve problems relating to HIV/AIDS, global health, gender equality, and economic growth.

DCLI focuses on creating an enabling environment where various government and community actors can identify data solutions for jointly identified problems through collaborative planning. MCC and PEPFAR bring expertise in designing data solutions for decision making at the country, sub-national, and organizational level. Through its partnerships, DCLI takes an innovative approach to developing data ecosystems through community engagement that focuses on building data literacy, awareness, and linkages within the local ecosystem. The aim of this approach is to build sustainable capacity to better articulate the data problems (i.e. demand) and match them with a data solution (ie. supply). This is not to say that all data solutions start with an articulated problem within DCLI, as new supplies of data provide fuel for collaboration around potential value within the data as well.

These partnerships are common amongst various applications covered by this research[8]. For example, UN Global Pulse Jakarta uses similar approaches to achieve **the level of inclusivity that is needed for effective problem-solving**. But a key takeaway seems to be that "data problems" (i.e. the data requirements for specific decision-making) also **require inclusive**

[6]Data Driven Development, World Bank Group. 2018.: https://www.worldbank.org/en/topic/digitaldevelopment/publication/data-driven-development

[7]MERL TECH's State of The Field Series provides an excellent landscape on this topic

[8]Similar partnerships include the Tanzania Data Lab, Data Zetu, or funding mechanisms like the DCLI Innovation Challenge.

processes to identify and define them in the first place.

As part of an independent evaluation of DCLI (Development Gateway, 2020), three principles of success stood out amongst stakeholders:

- **Adopting a community-led planning process**
- **Providing high-quality training coupled with support**
- **Creating demand for data through individualized problem solving**

Maturing the data ecosystem is not just a DCLI priority, it is the primary priority. There is additional evidence that DCLI has been able to "reach" actors beyond its immediate activities as communities integrate some of the DCLI functions, such as using convening authority to facilitate problem-solving through community dialogue.

An additional finding of the DCLI evaluation is that **increasing data use is not merely a question of technology used or data science methods, it is also dependent on an understanding of the incentives of the actors involved**. DCLI works at multiple levels of decision-making, the relevant decision-making "actor" could be a community leader, regional government official, or Public Utility Director.

Per the experience of DCLI, understanding the incentives of decision-making requires an approach that can be applied across a wide variance of applications. This point was paramount in KII's with DCLI leadership from MCC in Washington D.C. and is a consistent point highlighted elsewhere by Key Informants and literature review evidence. **Inclusivity of problem-solving processes and the need for understanding the incentives of decision-making are two consistent best practices.**

However, DCLI offers a unique view on future data science applications in Development. Given the DCLI focus on maturing data ecosystems, there is the inevitable question of moving from initial "prototypes" to data ecosystems that can sustainably identify and solve their own data problems. As DCLI shifts to longer-term thinking, **the limitations of "enabling environments" within data ecosystems are becoming apparent. As one DCLI leader stated in an MCC KII on October 15, 2020, "at some point there has to be a demand, in terms of use, for the supply of the data. Otherwise, what are we doing?"**.

Thinking about this tension between the explosion in the supply of data and the demand for it was a consistent theme throughout this research. The importance DCLI places on data governance could be seen as a way to establish ground rules for managing the emergence of all the new data (i.e. Big Data). The supply of data is ready-made as many DCLI activities are meant to help data ecosystem actors discover what data is already there and how it can be used. But there is a caution within DCLI to have data planning be purely "supply-led" where the availability of new data sets is the only thing that spurs collaborative action on how the data can be used. **Having a process to 1) define problems, and 2) define the data requirements that decision-making needs to solve the problem (i.e. demand) has been a key feature of DCLI success** in managing this tension and is a need moving forward if this success is to be sustained.

The UN Global Pulse Lab Jakarta carries out applied research along three primary tracks: 1) implementing specific data science applications (data mining government data sets, analyzing ICT supplied data for humanitarian planning, etc.), 2) facilitating relationships between different actors in digital ecosystems (government, civil society, academia), and 3) technical assistance on methodologies, guidance on data ethics, security, etc.

Like DCLI, Global Pulse works in all dimensions of data ecosystems, including smart infrastructure, building data literacy, policy development, etc. Global Pulse builds different types of capacity within these ecosystems depending on current priorities and frequently uses tailored teams that engage with data users to help determine those priorities.

Because DCLI and Global Pulse actively map to engage the ecosystem, they are not only able to gain more granular data on problem drivers but also on what data products would prompt action from specific stakeholders. This systems approach mitigates a risk outlined by Development Gateway on Development leadership use of data:

"**Absent a visible demand for development data from citizens and officials, there is little incentive for those funding or producing data and statistics to change the status quo.** Additionally, when existing development data is not seen as sufficiently timely, relevant, and credible for their purposes, it is unlikely that there will be an uptick in demand by citizens and officials for this information……The wider the gap between those producing and using development data, the greater the risk of a mismatch between supply and demand." (Masaki & Cluster, 2017)

Per Global Pulse, creating an understanding of the data amongst stakeholders was a critical requirement for cultivating demand and eventual use in decision-making. Priority is given to engagement on the communication (particularly the visualization) of data, which has possibly seen a disproportionate share of attention within Development as the supply of data in evaluations and data science applications becomes more readily available and varied (USAID, 2015). This emphasis on communication of data results is partly built on **the assumption that understanding the data will lead to its use in decision-making**. This is similar to the "Field of Dreams" theory of change (i.e. "if you build it they come") in that all it takes is producing an output (like a data set) and it will be utilized (Laverty & Littel, 2020).

Such an assumption partly differs from the private sector experience (covered under research question #2) where understanding data starts with first understanding the problem, as data that speaks to the problem is more understandable and valuable than data that does not.

This is not to say that the private sector is purely demand-driven when it comes to data investments. Private sector data science applications are often exploratory, where data is mined to simply look for unknown correlations that could provide value. **Managing the tension between these supply-led exploratory exercises and demand-driven problem solving is a space in need of an evidence base to inform effective policy and practice.** In other words, how does one keep from being data supply-led while still being effective in extracting value in existing supplied data that could be explored for value?

Moving forward, Global Pulse finds itself in a similar space as DCLI, looking to shift from "prototypes" to systems that can sustain their own data lifecycles. Currently, many of Global Pulse's prototype applications have expiration dates on them, usually defined as the period of performance on a services contract. Over time Global Pulse has identified a need to better understand the requirements for sustaining and improving the prototype (whether it is a specific data science product or a technical assistance curriculum) after closeout (Laverty & Littel, 2020). These requirements include promoting data literacy, as well as hardware and protocols for sustainability, but focus more on the incentives involved in the decision-making processes that stimulate demand for data science products.

These incentives go beyond the influence of data science products and include political, financial, cultural, and organizational aspects. **In short, those incentives that often go un-stated (or at best under-stated) need to be articulated so that they can be accounted for in the design of data solutions.** As will be seen, in the private sector these incentives are formally captured as part of the process to decide whether the data investment will have positive returns. However, the private sector has a formalized structure to decide whether to make a data investment built by protocols and guiding principles that are much different than what the Development sector needs to create. This is not only due to the difference between public and private sector decision-making, but also in the experience of the Development sector in generating and using data in the past and the culture, policies, and practices that have developed as a result.

Working with "New" Development Data

Development actors (who are used to working with "traditional" data from M&E activities, administrative data, etc.) have entire organizational cultures and practices set up to make decisions using this traditional data. **As data science products make use of many different types of data, gathered in different ways and often meant for different purposes, these organizational cultures and practices will need to adapt to this new data and learn how to use it most effectively.** In order to better understand what these adaptations need to be, we could first understand the attributes of this new influx of data.

A recent Rockefeller Foundation report outlines primary constraints to the integration of data science into traditional Development data lifecycles, specifically MEL activities. The barriers include:

"i) evaluators and data scientists traditionally work in different contexts; ii) the two professions tend to use different methodological frameworks and analytical tools; iii) there are weak institutional linkages between data science and evaluation offices, even when both offices are in the same organization; and iv) training for big data and training for evaluation often do not cover the methodological approaches and philosophies of the other." (York and Bamberger, 2020)

The authors further note that there is a risk of data science methods, skill sets and philosophies becoming overly prominent within traditional Development data processes, which could result in less effective decision-making that does not understand the data that is being spoon-fed rather than debated. **Hence there is a need for an integration function, a person or team that understands the skill sets, functions, limitations, and value of new data science and traditional Development data practices.**

In spite of the potential costs, York and Bamberger note the benefits that data science can contribute to development evaluation, including the ability to strengthen counterfactuals, evaluate complexity, identify unintended outcomes, lower data collection/management costs, integrate new sources of data that is more granular and reaches previously "unreachable" groups, etc. However, realizing these benefits requires **navigating the tradeoffs associated with integrating data science into legacy MEL systems**, which in turn requires the skill sets to map, calculate, and communicate these tradeoffs to stakeholders. Table 1 below gives an overview of some of these trade-offs.

What these tradeoffs communicate is not only **a need for these new quality control protocols that can assess the impact of these differences on the problem at hand, but also the need for a new data literacy amongst those who work closest with the data whether they are a government, private sector, or community stakeholder**.

Developing an Evidence Base

When considering the data science evidence base on what works, there are currently two categories: 1) "what works" in terms of applying data science methods (with an emphasis on analysis) to produce data science outputs, and 2) evidence for outcomes of interest (agriculture, finance, etc.) where data science methods were used to produce the evidence[9]. As data science applications are in their infancy within Development, and because these applications are primarily being used for traditional Development outcomes in hybrid approaches (where data science supplements traditional M&E), there is much more evidence on #2 than #1.

However, frameworks have been developed and piloted that aim to build each evidence base simultaneously. In a Devex 2018 opinion piece, Chris Watson from Premise Data and Alberto Lizzi of the United Nations Development Programme made the argument that

TABLE 1: BIG DATA ATTRIBUTES AND TRADEOFFS (ADAPTED FROM YORK AND BAMBERGER)
Basic Attributes
• Timely, often real-time production of data • Always on and non-reactive (i.e. remote sensors, smart infrastructure, smartphones) • Ability for multiple data points around one behavior of interest
Advantages for Development MEL
• There usually is at least some data for the entire population of interest • More granular data that can be more longitudinal • Highly quantifiable data that are easy to perform statistical and economic calculations on • New sources, quantity, and frequency of data are offering new insights on behavior change
Barriers for Development MEL
• Data is often collected for other purposes (indirect) and can be proprietary to specific actors (inaccessible) • Specific measures that the data is collected on could shift over time (inconsistency) • Data can be incomplete (dirty) and have errors

[9]Resources such as MERLtech.org and the International Initiative for Impact Evaluation support evidence for these two categories.

inefficiencies in decision-making, stemming from the lack of granular level data that directly measure the dynamics of the problem at hand, are a critical barrier moving forward. In order to generate this granular level **data that speaks to the ground level complexity** of Development problems, they argue for more decentralized decision-making that allows for field-level operators to have more flexibility to **iteratively identify problems in a highly participatory and adaptive manner**. Many of the approaches cited in their argument advocate for more participatory methods that are participatory precisely because they seek more granular level data[10]. These approaches might demonstrate a trend in Development towards its own "**customer-centric decision-making**" approach.

While Development does not have customers in the sense the private sector does, all Development investments are meant to improve the lives of beneficiaries who either benefit from the behavior change of others or from their own behavior change. For the purposes of using data to effectively inform decision-making, the "customer" could be the actor(s) who you want to use your good or service. In the private sector it could be an insurance policy or a can of soda. In Development, it could be a policy reform or a bar of soap. These customers, or more appropriately their mindsets (World Development Report, 2015), are what is most important to decision-making if the Development outcomes are of primary importance. Evidence shows this is not often the case (Development Gateway, 2018).

In short, **"customers" are those whose behavior you want to change**. In the private sector, the targeted behavior can be putting a product in a shopping cart where the public sector might target the uptake of an agricultural farming technique. Hence the next generation of Development decision-making that uses data science products **must find ways to be more customer-centric in order to generate the granular level data around the problems of interest**.

Current problem-solving in Development primarily focuses on the "right design" problems during initial design/planning then management problems around the "right implementation". During the initial phases of the program lifecycle, decision-making focuses on what the right approach should be; however, evidence suggests that once design decisions are made, decision-making turns to accountability regarding the fidelity of implementation. This means that decision-making is less focused on learning optimal paths to outcomes and more on ensuring that what was designed to be done was done according to the original plan.

This shift from data for formative learning to data for accountability is often not conducive to managing the level of complexity involved in Development interventions (Lizzi, 2019). It also does not lend itself to effective use of data science products which require more collaborative learning-centered problem solving throughout the entire data lifecycle. Research from the University of Oslo lays out this tension between accountability, learning and risk by stating:

"The current situation of "big aid data" does not remedy the widespread experience that we know too little and learn too little. This problem is only deepened by intensified calls for transparency, accountability, audit, and control, which, while serving critical democratic functions, are currently operationalized in ways that do not necessarily harmonize well with the ambition to learn. One learns not least by making mistakes, and one must expect aid work to involve making many mistakes. **A realistic approach would thus entail high tolerance for error. In reality, the expectations to aid are much stricter than this.**" (Reinertsen, Bjorkdahl and McNeill, 2017)

At present, some would say "learning" (i.e. the result of good decision-making) is crowded out by accountability, a ratio that would need to evolve if optimal returns are desired on data investments.

Research Question #1 Summary of Findings

Development Gateway perhaps stated it best in Designing Data Strategies: A Playbook for Action by positing "we have found there is no one path to an "ideal" internal data ecosystem" but that evidence-driven principles can still be tailored for individual applications. In that line of thinking, the findings from Research Question #1 are summarized in a series of principles that will later be integrated with private sector best practices into a general decision-making process with tools for Development.

[10]Another example that supports this argument can be found in the United Nations University Working Paper It's All About MeE

1. An inclusive and iterative process for problem-solving.

The rise in digital economies has also increased the complexity of Development interventions and amplified the necessity of having granular data at more frequent intervals. This data must be collaboratively managed, interpreted, and utilized. **More often than not, this problem-solving focuses on barriers and facilitators for behavior changes, thus the best data comes from those closest to the problems around the targeted behavior change. Often these are those from whom the behavior change is required.**

2. An understanding of the incentives for decision-makers.

A better understanding of decision-making processes and influences was a consistent theme across Key Informants and desk review material, **but it begs the question of how data science applications and outputs would be done differently if there were a perfect understanding of decision-making influences**. Development experience in this area has mostly focused on data visualizations and communications, perhaps partly out of resource constraints and the assumption that "understanding" leads to use. Meanwhile, as will be seen in the next research question, **private sector experience has much to offer on how to 1) develop an understanding of decision-making incentives and 2) create data science products that account for these incentives**. As noted later, there is an increase in the use of tools like Political Economy Analysis to better understand the demand for data and the possible results of its use.

3. Treatment of data as an asset with decision-making as the defining mechanism for returns on data investments.

Data as an asset has two primary dimensions: 1) as a tool to help create value, and 2) as an investment requiring a positive return. Each data investment can have management planning and measures established to track the performance of the data through its value chain. **All data has a value chain, it is a question of how simple or complex it is.** While the explosion in the supply of data and the advances in computational technology make data cheaper and faster to get, significant investments must be made in the human capital required to manage it. Hence "data assets" include tracking the return on investment (ROI) of the human capital as well.

The use of data in decision-making is not sufficient enough to warrant a positive return on the data investment. The decision must result in an action that improves performance; therefore, the use of the data results in an improvement in actual outcomes (i.e. efficiencies in process, innovations in strategy, etc.) (Spradlin, 2012). Measuring the return on the data investment requires measuring the "value-add" for the outcome which resulted from the decision-making informed by the data.

Research Question #2:

What lessons from the private sector can help inform better practices for using data science products in public sector development?

The general differences between private and public sector decision-making could be critical to Development integrating private sector best practices when it comes to using data science. However, it should be noted that the private sector does not take a unified or homogeneous approach to data science and decision-making. The results for this research question focus on general themes and practices that often have different names but very similar functions across private sector usage. For example, while this paper uses the phrase "business translator" it may not be that every private sector application has a "business translator" but they have a similar function.

Relevant Differences between Public and Private Decision-Making

Dr. Kara Morgan (2019), Research Scientist with Ohio State University Department of Food Science and Technology, ensures that her public policy students often hear the phrase "Businesses make things and governments make decisions." For our purposes here, "businesses make things" correlates to the private sector specialization in making very specific products, including services, or baskets of products for a specific customer base. As a result, their decision-making usually focuses on solving simpler problems and is profit-driven.

For example, the private sector is concerned with answering the question "Who is most likely to be a repeat customer?" while a Development problem could be "what is the best way to incentivize agricultural savings plans for the ultra-poor?" While not universally applicable, this generalization is a helpful way to frame a larger discussion about how Development decision-making can adapt to new types of data.

Development problems tend to be more complex, meaning they entail more uncertainty and risk. Conceptually, this could necessitate much higher levels of experimentation, through decentralized and flexible decision-making. Many in the public sector have limited awareness of decision-making processes as an area for research and improvement. It is an area in which people generally do not seek out support because they feel qualified based on their experience in another domain. Public sector decision-making tends to be risk-averse, in part due to the inherent complexity of most public sector decision-making[11]. This complexity can be due to:

- The number of stakeholders that have an interest in the decision (not only those within the public institution but those who are "served" by it)
- The number of variables and assumptions (i.e. risk/uncertainty) involved in the cause and effect relationships
- Conflicting evidence and viewpoints on the best solution amongst alternatives
- The detachment between those who fund the public goods/services and those who are intended to benefit from them

In addition, public agencies are accountable to a wider variety of stakeholders (e.g. administrators, elected officials, citizens). Measuring the value that analytics generates for Development is more challenging than in the business sector as Development success metrics are more varied and complex (McKinsey, 2014). Higher levels of complexity require a structured but flexible decision-making process in order to maintain a level of organization (which produces a degree of predictability/stability). Complexity also requires the ability to update the process as new best practices emerge. This begs the question regarding the relationship between a "good process" and a "good decision".

A rising sentiment amongst decision scientists is that "good" decisions are a result of a sound process that views data as an investment that necessitates a return in the form of improved performance (improved performance being the result of "better" decisions)[12]. **Because public sector decision-making often deals with higher levels of uncertainty, decision-making cannot be solely judged by its final outcome, but instead, by the process it uses to reach that outcome. This means that decision-making has its own "theory of change" type model, process, and targeted behavior change, etc., that can all be measured – a necessity if a public sector actor wants to measure the performance of its decision-making.**

Differences in Desired Behaviors

Generally, public sector problem solving requires more of a focus on behavior change than does the private sector (Eppel, 2017; Head, 2010). This does not mean that public sector decision-making (especially for intervention design) uses behavior change models effectively, if at all, just that there is a necessity to do so. Given social outcomes are the result of behavior change, one would be hard-pressed to find a Development intervention that does not require some type of behavior change to achieve the targeted outcome. **Hence the attributes of desired behavior change could be the defining characteristics of the data problem to be solved with data science methods.**

Private sector data science primarily focuses on identifying correlations and patterns where far less priority is placed on the path of causation behind the correlation or pattern (Forbes, 2013). This is because **the private sector business models are more concerned with understanding how customers behave and not necessarily why they behave that way.**

Because Development primarily uses hypothesis to drive the decision-making around what interventions should be funded, how they should be designed, implemented and measured, an entire culture around theories of change has developed. However, a sizeable portion of data science applications in Development have not tested hypotheses but instead are using exploratory analytics to extract value from data sets (think back to the DCLI and Global Pulse case studies already mentioned). **There is no guidance on deciding between using exploratory analysis, despite the exponential increases in data supply and the use of data science to test hypotheses like a theory of change.**

[11] KII conducted on November 16, 2020. See Annex 2 for full list of KII's.

[12] An example of this can be found in the 2013 PNAS article Bridging the gap between science and decision making.

The Role of Value

The private sector adopts a Value Chain approach that encourages the treatment data as an asset by charting and measuring its value add to decision-making. Some in the Development sector have experience working with value chain models as part of programmed interventions in various sectors, notably in agriculture. However "lifecycle" models are much more familiar to Development practitioners, as they are used to organize program management and other functions.

Data value chains un-pack the transformation of a data problem into a solution through phases, the functions of those phases, and the value created by each phase/function. This is different from a standard public sector lifecycle model that primarily outlines the phases and steps (tasks) to be completed with less of a focus on function and less, if any, focus on the value created. **Introducing the use of a data value chain model could be an initial step in improved Development decision-making by providing more detailed feedback on the "value-add" of its data, as opposed to simply knowing that data tasks were completed.**

As the research shows, the use of a value-chain model creates a possible dilemma for Development decision-making, namely "value for whom?", meaning that "value" must be defined in order to assess the performance of the data. In the private sector, "value" is largely defined as value to the customer[13]. But as noted above and further discussed below, the Development sector does not have "customers" per se, although there are normally contractual relationships to consider. **Hence there could be the tendency to define "value" from the perspective of the Development leaders and practitioners as opposed to intended beneficiaries.** As Development seeks to use data more effectively, it will have to ensure that "value" is defined, measured, and communicated in a "customer" centric way that leads to social impact.

Managing the Data Value Chain

Below are the general functions, tools, skill sets, and people involved in developing and managing an effective data value chain in the private sector. They are introduced here and subsequently integrated into a general decision-making process for Development actors to modify per their own priorities.

Value Chains: can take place within a lifecycle model; however, the critical difference between the two is that a lifecycle is task-oriented, while in a value chain those tasks are designed and measured to add value to the "product" as it passes through the phases. Steps that do not add demonstrable value to the chain are subsequently re-examined.

Private Sector Use of Business Problems: For the private sector, data is for problem-solving, and entire governance structures are established to maximize this potential. **This starts with the definition of a Business Problem that is composed of two parts: Expected Value and Business Value.** Expected Value is an evaluative framework used to assign a monetary value to the expected results of decision-making. Business Values are the variables that influence decision-making outside of the Expected Value, with the idea being that between the two values any decision could be clearly understood and communicated (Provost and Fawcett, 2013, p.32).

Expected Value: is quantified through scenario planning, which means that the inputs for the scenario planning model must come from the "completeness" of the decision-making requirements. "Completeness" is used to refer to standard variables used as inputs for the scenario planning. These variables come from the attributes of the data needed for decision-making. Highly structured templates to "unpack" decision-making are used to not only ensure standardized quality control – which in turn is used for institutional learning – but also for the development of evidence bases around standardized constructs (p. 44). However, the competitive nature of the process demands flexibility in teams being able to explore additional inputs into the model or even challenge components of the model itself if necessary.

Due to the "completeness" of the decision-making requirements, scenario planning models are able to run various cost-benefit calculations that include sensitivity analyses that focus on assumption testing. **These values are critical and a primary driver of discussion, as the various offices that provide data for the calculations have a stake in the decision and thus are motivated to understand the model and its limitations. Ultimately it is these values that the ROI will be calculated against.**

[13]Customer preferences and desired behaviors are usually the basis of the business problem driving the value chain. Articles in Analytics Vidhya and Deloitte Insights further outline this idea.

The use of scenario planning, cost-benefit analysis, sensitivity analysis, and assumptions testing are methods already in use (to various degrees and in a variety of ways) within Development decision-making[14]. Hence adapting them to the uses outlined here could encourage more immediate uptake.

Business Values often need to be acquired from other sources. Quantifiable, data-driven modeling may help estimate components of Expected Value, while the rest come from external domain knowledge (p. 46). This knowledge can include organizational strategic priorities (ownership preferences, for example) and organizational culture values, among others. In the private sector, this external domain knowledge is usually limited as Expected Value calculations are much more informative given the importance of financials. **However, for public sector decision-making, especially Development decision-making, calculating Business Value can be a much more difficult activity.**

These calculations focus on expected results. In order to find a reasonable solution to the "data problem," different variations of a Data Action Plan are used (Leong, 2016). These Data Action Plans include a restatement of the problem analysis, and then assesses various data sources and models of analysis against the requirements of the Expected and Business Value calculations. This Data Action Plan is often presented back to stakeholders for review to ensure that it is accurate (sometimes using standardized scorecards).

The Data Action Plan explains how the data will be used for value creation by breaking the data problem into specific data science tasks and sub-tasks. While a data team usually leads the development of the Data Action Plan, it is a highly participatory effort where stakeholders are consulted to validate and inform the design. In collaboration with these stakeholders, data scientists decompose a business problem into subtasks like model selection and then further into coding activities, all the way through to data visualization and communication. Some of these subtasks are unique to the particular business problem, but others are common data science tasks (Provost and Fawcett, 2013, p. 48).

Data Translators:

"The biggest barriers companies face in extracting value from data and analytics are organizational; many struggle to incorporate data-driven insights into day-to-day business processes. Another challenge is attracting and retaining the right talent—not only data scientists **but business translators who combine data savvy with industry and functional expertise.**" (p. 52)

These "business translators" are critical to ensuring that nothing is "lost in translation" through the data value chain. This includes not only ensuring that all relevant stakeholders are involved, but are involved according to optimal protocols. To do this, **these "business translators" not only need data science and decision science skillsets but also should be familiar with the organizational processes, priorities, and sector expertise if they are to manage the data value chain effectively.** It is the "business translator" who facilitates the problem solving associated with creating additional value with the data.

The role of the translator is critical as they are the primary, often only, person to see the entire ecosystem and value chain. Other actors only see parts depending on their function as outlined in Table 2. Decision science has made significant gains in the private sector as data science is used at larger scales. Decision Scientists bring unique skill sets that improve the ability to act on data science results, but even with their inclusion, there is still a need for the translator to manage the process end to end.

Research Question #2 Summary of Findings

Private sector use of data science largely stems from decision-making requirements to better understand what their customers prefer given a variety of options. Hence data is treated as an investment in an asset that requires a return in improved performance that results from the informed decision-making. This entire process is managed via a data value chain that uses tools like Business Value and Expected Value to better understand the dimensions of the Data Problem. The Data Action Plan is a primary tool used to operationalize the collaborative efforts to solve the data problem in a process facilitated by Data Translators who ensure that requirements are correctly communicated and integrated into the eventual data solution.

[14]The Millennium Challenge Corporation outlines Economic Rates of Return online at https://www.mcc.gov/our-impact/err.

Table 2: Development Decision-Making Process with a Data Facilitator

Phases	Decision-Making Using Data Science	Data Facilitator within Overall Process	Data Facilitator with MEL/ Data Science Team	Next Steps
1	Identify Decision- Making data requirements and relevant stakeholders, data sources, processes, and timelines for defining the Data Problem	Develop a Data **Problem Tool** with dimensions **(Expected Value, Business Value, Attributes of Decision-Making Requirements)** and guiding questions for each. Facilitate development o **Data Problem Action Plan**	Communicate possible requirements and changes to team work plans with initial assessments of possible skill sets, level of effort, and timeline requirements	Pilot, test, and adapt Data Problem Tool Dimensions Create templates for **Data Action Plan**
2	Assess the completeness of the Data Problem Action Plan against Decision-Making Requirements	Facilitate approval for Action Plan before its implementation	Work with data engineers to identify the structure and attributes of relevant data sets	Create **Data Steering Committee** (DSC) with approval authority
3	Complete the dimensions of the Data Problem (Expected Value, Business Value, etc.)	Provide technical oversight and facilitate collaborations to complete Data Problem Tools	Use Data Problem Tools to identify personnel to perform assessments that identify analysis options	Specialist tool kits for completing Data Problem Tool Dimensions (sector-specific guiding questions)
4	Assess the adequacy of various data analysis options	Use Data Problem Tool results to identify data analysis model options with trade-offs for decision-making	Perform and communicate assessments of various design and analysis options with MEL team, Data Analysts, etc.	Tool kits for assessing and communicating trade-offs of various models to DSC
5	Perform analysis using selected model	Ensure that decision making requirements are correctly correlated to design (data sources, model of analysis, etc.)	Perform analysis with necessary quality controls, identify previously unknown risks and limitations	Tool kits that ensure decision-making requirements are correctly translated into Data Problems, Analysis Models and Coding to ensure that final products reflect what was approved by DSC
6	Communicate products to stakeholders	Facilitate communication and feedback from stakeholders on products through various mediums depending on need/capacity	Work with Data Science/ MEL team to develop sets of visuals and data packages depending on stakeholders	Typology of communication packages (visuals, talking points, etc.) for different types of stakeholders
7	Assess the adequacy of data products on targeted decision-making (have conditions changed, etc.)	Lead discussion on the adequacy of data products against decision-making requirements and capture feedback	Identify any additional data sets or analysis that could meet outstanding requirements	Templates that assess the data products per the individual dimensions of the Data Problem Tool
8	Make the targeted decision based on data products and other factors (captured in the Business Value dimension of the Data Problem Tool)	After the decision is made, finalize measures for the resulting action from the decision to determine the return on investment	Communicate results of decision-making back to teams and perform necessary after-action review	Training materials to help stakeholders understand how measuring the results of decision making is done to measure the return on the data investment
9	Develop a **Data Learning Agenda** that tracks positive deviation, risks, common errors, etc., in treating data as an asset	Lead a **Working Group** that has technical oversight, quality control and convening authority for maintaining the Data Learning Agenda	Coordinate input and assistance from teams into Data Learning Agenda discussions	**Data Learning Agenda** Template Working Group Scope of Work

It is these functions, skill sets, and processes that could be integrated into Development decision-making processes as outlined below.

OVERALL RECOMMENDATIONS

The overall recommendations focus on how the results of the two research questions can be best operationalized in Development decision-making. Hence the recommendations initially focus on a general decision-making process that integrates tools and functions like Data Action Plans before speaking to the staffing and skillsets needed to manage the process.

Integrating Data Facilitation into Development Data and Decision-Making Processes

Many Development organizations use Decision-Making processes similar to the steps outlined below for deciding what interventions are needed, how they will be designed and, eventually, implemented. Titles of the phases and steps differ as do the protocols across different Development actors. Some are general with more open-ended thresholds for decision-making[15] while others are more prescriptive with quantitative threshold definitions. Most can generally be described using these broad General Public-Sector Decision-Making Steps (Morgan, 2019):

Table 2 unpacks a process based on the general public sector process outlined above but captures the key recommendations of this research: 1.) the role of a Data Facilitator to facilitate the process and 2.) the use of new tools like Data Problems, Data Action Plans, and Decision-Making ROI.

This process, through the expertise of the Data Facilitator, has the flexibility to integrate other value-adding tools and methods like value chain modeling (Fayyad and Hamutcu, 2020), organizational learning, and others mentioned in this research. This process is not meant to be overly-prescriptive but could provide the basic structure that Development actors can build on and modify per their priorities.

Table 2 takes decision-making process phases and maps each one to a "Data Facilitator" function within the overall process as the various tools are utilized, and then as a coordinator with the data science team. The last column outlines possible next steps to further develop the tools and improve the process.

A Data Science Operational Framework for Development

Data science operational frameworks are meant to curate and standardize skillsets, definitions, terms, and concepts to provide the commonality needed for larger-scale applications. Organizations have unique ways of defining roles in data science and associated skills and knowledge which has resulted in a confusing landscape for all involved (Fayyad and Hamutcu, 2020). The confusion stems not only from "which" skills an analytics or data science professional must possess, but "what" level is required for a professional to fulfill a particular function. There are various attempts at creating data science knowledge maps that can be very useful in creating one specific to the needs of the Development industry. Namely, one that can be used to inform the early piloting needed to begin building the evidence base for what works.

Table 3: Data Science Roles (adapted from Fayyad and Hamutcu)

While the phrase "Business" appears in the above table several times, no role in this table is specific to private sector data science applications. In fact, given the research results in this paper, it is these "Business" roles that could add the most value in Development data science applications. As alluded to previously, these "Business" roles are deemed the most critical in private sector applications because these roles "fit" the data to the specifics of the problem, a function often neglected in Development.

Just as the level of complexity in the problems at hand vary, so too will be the composition of the responding data science teams. As will be outlined in the last section, data problems can be deconstructed and their parts correlated to types of data functionality needed to solve them. These functions are then mapped to skill sets, level of skill, level of effort, and other characteristics that can drive the composition of the data science team. But the first step is building a data science knowledge map specific to Development.

---

[15] Examples include USAID and MCC.

As noted previously, Bamberger and York (2019) have not only highlighted the value-add of data science to Development but are careful to note that Development will not solely be dependent on data science to inform its decision-making moving forward. From a technical perspective, this is partly due to the fact that some problems either do not require a data science produced solution. They either require a hybrid data science/MEL produced solution, or the prerequisites for performing a data science application are not present.

Just as evaluability assessments have seen an increase in use in helping to determine whether and what MEL design best fits the problem at hand (i.e. need for an evaluation, monitoring, etc.), Development will need similar tools and processes for data science. By understanding the data science design that has the highest potential to produce the required solution, practitioners can then unpack that design into the required skillsets. Otherwise, data science resources could be mismatched against the identified data problems.

Feedback and Research Agenda

Advisors, listed on pg. 17, provided feedback on these recommendations and the evidence/argument used to reach them. This feedback helps ensure that these recommendations could at least conceptually meet real-world problems faced by current practitioners. The two overall conclusions of the feedback were that this research did not go far enough, that in order to provide real value, the final results should be further elaborated upon to provide more specific guidance and tools to build an evidence base on "what works".

A Need for Better Contextualization

Development requirements for data science solutions have high levels of variance due to the global context Development works in. This breadth creates the need for specialized evidence to inform tools and other guidance for those operating in specific contexts. However, there is a question of how best to generate this evidence given the lack of standards to generate it and a common framework to assess it. Given the results of this research, several advisors had suggestions for how this could be done using 1) Types of Decisions, 2) Types of Development Actors, and 3) Types of Problems as the primary categories of a standardized framework to build an evidence base.

Types of Decisions (as defined by complexity/level of uncertainty or function): Some decisions can be classified for primarily accountability purposes, while others could be more formative. Likewise, some decisions have higher levels of uncertainty stemming from the varying types of complexity involved. Categorizing decisions in this way could not only help ensure that decision-making is prioritized, but that because it is the priority, it could be the starting point for determining data requirements. By unpacking decision-making along the dimensions of complexity and function, specific data science tools and guidance could be better tailored to fit specific types of decisions.

Types of Development Actors (as defined by capacity or function): Government ministries have different needs and capacity versus a local CSO which requires modifying the scope of Data Action Plans or how data is collected for calculating Business Value, etc.

Types of Problems (defined by behavior change mechanism and/or sectors): Given Development outcomes are dependent on behavior change, correlating data solutions to specific types of behavior change problems could go a long way towards ensuring data investments have a positive return through improved performance.

Experimentation, Learning, and Accountability

Advisors highlighted the need for better "sandboxes" where experimentation (from conceptual to real-world piloting) takes place. This need for experimentation was consistently stated and emphasized as a priority barrier to improved performance. This need is tied to the tension between accountability and learning. Several advisors noted that their data science use cases are predominantly used for reporting/accountability as opposed to experimentation/learning. Where experimentation does happen, it is often (not always) conceptual and done for marketing and communication purposes. There was near consensus amongst advisors that there is an urgent need for Donors to address the tension between accountability and learning, given that the need for experimentation is growing.

However, advisors noted that an operational framework is needed in order to really experiment and find out what works. This envisioned operational framework would answer critical questions around what skill sets are needed, and when, to identify and solve data problems. This operational framework could integrate emerging data science functions and skillsets with traditional MEL ones already operating within much of the Development organizational infrastructure. Such a framework could include guidance on identifying necessary skillsets, timing, coordination, planning, and even procurement tools.

Lastly, advisors noted that intervention design and management itself will need to be done in a way that facilitates experimentation. Interventions will need to be capable of reacting and adjusting to the more real-time data that is possible given different data science tools and approaches that are used. Given the "pace" at which data science is able to generate data results, it could be that it could facilitate the long-awaited Adaptive Management progress many in Development have been advocating for.

Whose Problem is Being Solved?

There was a fascinating discussion amongst numerous advisors on what was framed as "beneficiary centric" versus "beneficiary driven" where some considered the former (as written in the paper) to be more in line with "current top-down donor dictated data science applications".

Advisors noted that if data is to be treated as an asset with an expected return on the investment (as measured in improved performance) and data is primarily meant to solve a "data problem", then the origin of the problem must be clearly articulated in order for the ROI to be calculated. The origin of the data problem is important since, as outlined above, the most valuable data comes from those who understand the root causes of the problem the best. For example, DCLI seeks to empower local communities to sustainably identify and solve their own data problems. The level of transparency and collaboration involved help to ensure that there is a common understanding of the origin of the problem. Any audit on the data investment could clearly identify the origin of the demand that drove the investment through documentation of collaboration and due diligence activities on the data problem.

Advisors noted that if funders do not begin to equip their national and community partners with the tool to leverage their own data assets, then they will not be able to make the most of data investments. With PEPFAR, for example, an absence of a mature data ecosystem means that the sustainability of their investments is at risk because they will never be able to hand over management of epidemic control to the country.

However, not all data science applications are focused on community empowerment or are equally transparent. Often data science applications are ad hoc and in response to more open-ended requests from organizational, donor, or government leadership with vague criteria and little to no indication of intended use. Often this type of application uses data mining methods to find value in pre-existing data sets. Such a scenario proves impossible for calculating reliable ROI on any data investment, but the costs can often be negligible for quick response activities.

Advisors noted that a "slippery slope" culture can quickly develop where there is no data strategy for data as an asset. Instead, ad hoc requests for data science applications become commonly accepted and Standard Operating Procedure, making any measurement of its effectiveness incredibly difficult.

Ethics

Lastly, and perhaps most importantly, is the role of ethics. Advisors noted that while there is general guidance available, predominantly in the form of principles, there is little guidance on how to operationalize these principles in specific contexts. Further, there is a gap in due diligence requirements for ethical standards in these data processes, often leaving operators with greater flexibility to do what seems "right" to them.

ANNEX 1: REFERENCES

Akiko, R. (2019, May 29). Kyela District Council's data road map: Harnessing local data for sustainable development. Medium. https://medium.com/data-zetu/kyela-district-councils-data-road-map-harnessing-local-data-for-sustainable-development-bb9c2f77af75

Chen, J. (2017, January 30). Solve Business Problems with Data Science: A generalized framework to data-driven consulting projects. Medium. https://medium.com/@jameschen_78678/solve-business-problems-with-data-science-155534b1995d

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_3

Deshmukh, O. (2019, August 22). How can you Convert a Business Problem into a Data Problem? A Successful Data Science Leader's Guide. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/data-science-leader-how-effectively-transition-business-requirements-data-driven-solutions/

Development Gateway. (2020). Delivering Data Where it Counts. https://www.developmentgateway.org/sites/default/files/2019-11/DeliveringData_Final.pdf

Eppel, E.A., & Rhodes, M.L. (2018). Complexity theory and public management: a 'becoming' field. Public Management Review, 20(7), 949-959. https://doi.org/10.1080/1471.20903717.1364414

Fayyad, U., & Hamutcu, H. (2020). Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards. Harvard Data Science Review, 2(2). https://doi.org/10.1162/99608f92.1a99e67a

Head, B. (2010). How can the public sector resolve complex issues? Strategies for steering, administering and coping. Asia-Pacific Journal of Business Administration, 2(1), 8-16. https://doi.org/10.1108/17574321011028954

International Telecommunication Union. (2020). Measuring Digital Development: Facts and Figures. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf

Kirby, P., Powell, J., Kilroy, A., & Orton-Vipond, S. (2020). Designing Data Strategies: A Playbook for Action. Development Gateway. https://developmentgateway.org/wp-content/uploads/2020/10/DataStrategy_Final-1.pdf

Laverty, M., & Littel, C. (2020, January 16). Entrepreneurship. OpenStax. https://openstax.org/books/entrepreneurship/pages/11-1-avoiding-the-field-of-dreams-approach

Leong, L. (2019, July 19). Turning a Data Vision into a Data Action Plan. Towards Data Science. https://towardsdatascience.com/turning-a-data-vision-into-a-data-action-plan-820976c22a4f?gi=83535b4a890d

Lizzi, A. (2019, March 22). Making International development Programmes More Adaptive. Medium. https://medium.com/@albertolizzi/data-driven-adaptive-management-for-international-development-b1eaa7cea460

Masaki, T., Custer, S., Eskenazi, A., Stern, A. & Latourell, R. (2017). Decoding Data Use: How do leaders source data and use it to accelerate development? AidData. https://www.researchgate.net/publication/338686253_Decoding_Data_Use_How_do_leaders_source_data_and_use_it_to_accelerate_development

McDowell, T., & Mallon, D. (2020). Getting decision rights right: How effective organizational decision-making can help boost performance. Deloitte Insights. https://www2.deloitte.com/content/dam/insights/us/articles/6360_getting-decision-rights-right/DI_Getting-decision-rights-right.pdf

McKinsey Global Institute. (2016). The Age of Analytics: Competing in a Data Driven World. https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.pdf

Morgan, K. (2019, October 7). Decision-making in the public sector: What can government learn from decision sciences?

Public Policy. https://doi.org/10.1287/orms.2019.05.11

Powell, J., Orton-Vipond, S., Bhatia, V., & Kilroy, A. (2018). Decision-Making and Data Use Landscaping: Better Data, Better Decisions. Development Gateway. https://developmentgateway.org/publication_landing/decision-making-and-data-use-landscaping-better-data-better-decisions/

Press, G. (2013, May 28). A Very Short History of Data Science. Forbes. https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=1c6e14de55cf

Provost, F. & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media.

Pritchet, L., Salimah Samji, Jeffrey Hammer. It's All About MeE: Using Structured Experiential Learning ('e') to Crawl the Design Space. CID Working Paper No. 249 December 2012 https://bsc.cid.harvard.edu/files/bsc/files/249_pritchett_samji_hammer_mee.pdf

Reinertsen, H., Bjørkdahl, K., & McNeill, D. (2017). Confronting the Contradiction: An Exploration into the Dual Purpose of Accountability and Learning in Aid Evaluation. University of Oslo.

Spradlin, D. (September 2012). Are You Solving the Right Problem? Harvard Business Review. https://hbr.org/2012/09/are-you-solving-the-right-problem

USAID Bureau for Policy, Planning and Learning. (January 2018). Discussion Note: Adaptive Management. https://pdf.usaid.gov/pdf_docs/PBAAJ032.pdf

USAID Bureau for Policy, Planning and Learning. (2015, May 1). Data Visualization. https://usaidlearninglab.org/library/data-visualization

Verma, D. (2014). Research and Analysis of Various Decision-Making Models in an Organization. Journal of Business and Management, 16 (2), 171-175. DOI:10.9790/487X-1621171175

Warren, C. (2019, March 5). Don't Do Data Science, Solve Business Problems. Towards Data Science. https://towardsdatascience.com/dont-do-data-science-solve-business-problems-6b70c4ee0083

Watson, C. & Lizzi, A. (2018, September 27). Opinion: Data-driven adaptive management for international development. Devex. https://www.devex.com/news/sponsored/opinion-data-driven-adaptive-management-for-international-development-92994

Winterfeldt, D. (2013, August 20). Bridging the gap between science and decision making. Professional Network of American Sciences.

World Bank. (2015). World Development Report 2015: Mind, Society, and Behavior. Washington, DC: World Bank. doi: 10.1596/978-1-4648-0342-0

World Bank. (2016). World Development Report 2016: Digital Dividends. Washington, DC: World Bank. doi:10.1596/978-1-4648-0671-1

World Bank. (2018, October 30). Information and Communications for Development 2018: Data-Driven Development. Washington, DC: World Bank. doi:10.1596/978-1-4648-1325-2.

York, P., & Bamberger, M. (2020). Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology. The Rockefeller Foundation. https://www.rockefellerfoundation.org/wp-content/uploads/Measuring-results-and-impact-in-the-age-of-big-data-by-York-and-Bamberger-March-2020.pdf

ANNEX 2: KEY INFORMANT INTERVIEWS

Michael Bamberger - Rockefeller Foundation/Consultant (9/22/20)

Lungi Okoko - USAID Senior Technical Advisor (9/25/20)

Rivandra Royono - Director Un Global Pulse Jakarta (10/8/20)

Alberto Lizzi - Dir at UNDP (10/15/20)

Kara Morgan - Prof at Ohio State University (11/16/20)

Agniezka Rawa - Managing Director at Millennium Challenge Corporation (11/5/20)

Chris Watson - Director at Premise Data (10/23/20)

PULTE INSTITUTE
FOR GLOBAL DEVELOPMENT

pulte.nd.edu

3150 Jenkins Nanovic Halls
Notre Dame, Indiana 46556 USA
(574) 631-2940

globaldevelopment@nd.edu