POLICY BRIEF SERIES | MAY 2021, PB NO. 5

PULTE INSTITUTE
FOR GLOBAL DEVELOPMENT

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

MICHAEL COOPER AND PAUL PERRIN

UNIVERSITY OF NOTRE DAME | KEOUGH SCHOOL OF GLOBAL AFFAIRS

OVERVIEW

This brief, and larger report, was developed to assist International Development (referred to here as "Development") decision-making regarding when and how to best utilize data science methods. It is meant to assist all levels of decision-making - from the field to the board room and across the full spectrum of Development actors, from Donors to community-level NGOs. This general applicability stems from a research focus on general principles and processes that can be adapted as needed. The final recommendations include a general decision-making process that outlines how data science methods can integrate private sector best practices with legacy Monitoring, Evaluation, and Learning (MEL) systems. It also presents next steps for operationalizing this process. Feedback from an advisory group consisting of donors, service providers, and thought leaders then establishes the foundations for a future-focused research agenda.

DATA-DRIVEN DECISION-MAKING AND THE FUTURE OF DEVELOPMENT

Data is only as valuable as its use; to invest in data without accounting for its end use introduces risk that the investment in data will have suboptimal returns and raise barriers to social outcomes. With varying levels of digital transformation within societies, new types of data are becoming available to guide decision-making within the Development industry. This new data has created both the demand and an opportunity for increased uptake of data science methods and processes by Development policymakers, practitioners, and other stakeholders in a wide variety of applications and decision-making settings.

WORKING WITH "NEW" DEVELOPMENT DATA FOR DECISION-MAKING

There are a wide variety of data science uses in Development, from ad hoc responses to inform high-level decision-making to empowering local communities to sustainably identify and solve their own "data problems." As data science uses different types of data, gathered in different ways and often meant for different purposes than traditional Development data sources, Development cultures and practices will need to adapt to this new data and learn how to use it most effectively.

In consideration of this need and after initial consultations with donors, thought leaders, and practitioners, two research questions were developed:

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

To answer these questions, an iterative research process was utilized. Literature review findings were shared and discussed, then used to inform additional literature reviews and Key Informant Interviews (KII) with public and private sector data scientists, thought leaders, policymakers, and field practitioners. The research was carried out in two phases. First, initial results and recommendations (immediately below) were reviewed by the Advisory Board and then additional discussions were held. Second, these discussions highlighted next steps (labeled "Conclusions and Next Steps" below) to build on the initial results and drive further research or learning agendas.

RESEARCH RESULTS

Results and recommendations stemming from the literature reviews and KIIs are organized by their relevant research question.

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?

1. An inclusive and iterative process for problem solving.

Data should solve problems. If it does not, then the resources expended on the data have little to no return. To this end, problem-solving should focus on barriers and facilitators for behavior changes given the necessity of behavior change in achieving social outcomes. Given that the most valuable data comes from those closest to the problems, there is a need to be more inclusive of beneficiaries in iterative problem-solving given their proximity to the problems at hand and their perspectives on barriers to the required behavior changes. Research participants noted both a need to develop profiles for those closest to possible "data problems" and a need to use behavioral diagnostic tools to help unpack data problems to more easily correlate with data science methods.

However, such an approach requires high levels of experimentation and risk-taking that are often contrary to the requirements and culture of Development. A realistic approach would thus entail higher tolerance for error to 1) produce data that speaks to the ground level complexity and can be used to 2) identify problems in a highly participatory and iterative manner.

2. An Understanding of the Incentives for Decision Makers.

A better understanding of decision-making processes and influences was a consistent theme across KIIs and desk review material. This theme begs the question, "How would data science applications and outputs be done differently if there were a perfect understanding of decision-making influences?" Development experience in this area has mostly focused on knowledge transfer through data visualizations and other communications, perhaps partly due to resource constraints and the often-false assumption that "understanding" leads to use. Meanwhile, as highlighted 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.

3. Treatment of Data as an Asset with Decision-making as the Defining Mechanism for Returns on Data Investments.

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 or impact. 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 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, complex, and effective it is. To develop and manage this value chain 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 alongside traditional MEL data practices.

QUESTION #2: WHAT LESSONS FROM THE PRIVATE SECTOR CAN HELP INFORM BETTER PRACTICES FOR USING DATA SCIENCE PRODUCTS IN PUBLIC SECTOR DEVELOPMENT?

1. Private Sector Use of Business Problems

In the private sector, data is for problem-solving. Entire protocol structures are established to maximize this potential. This starts with the definition of a Business Problem, which includes two considerations: 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. Ideally, between the scope of the two values, any decision could be clearly understood and communicated.

2. The Role of Value

Data value chains unpack 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 differs 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 even less, if any, focus on the value created. Introducing the use of a data value chain model could serve as 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.

3. Solving Business Problems with Data

Data Action Plans are often used by private sector entities to explain how 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.

The role of Business Translators is critical to ensuring that nothing is "lost in translation" through the data value chain. This includes ensuring that all relevant stakeholders are not only involved, but are involved according to optimal protocols.

4. Public versus Private Sector Decision-Making

While neither sector is homogenous and there is overlap, private sector data science primarily focuses on identifying correlations and patterns with far less priority placed on explaining causation. As Development primarily uses hypotheses to drive decision-making for things like program design, a culture around theories of change has developed. However, a sizable portion of data science applications in Development have not been used to explicitly test hypotheses, but instead use exploratory analytics to extract value from data. There is a need for guidance to inform decision-making when using exploratory analysis and testing hypotheses (like a theory of change) using data science methods.

RECOMMENDATIONS

The primary recommendation is a general decision-making process that is adaptable depending on the needs of users, whether it be a field-level NGO or a global donor. The process was developed around two primary requirements, informed by the research findings:

- Providing measures for the potential value of possible data investments to guide decision-making and summative value measures for completed data investments for longer-term learning.
- Creating mechanisms of coordination and oversight for integrating new data science methods and applications into legacy MEL systems to allow for both new data science methods and legacy MEL methods to be optimally utilized.

These requirements are met by adapting private sector tools to Development decision-making needs. This includes a **Data Problem Tool** which examines needs across three dimensions: **1) Expected Value, 2) Business Value, and 3) Decision-Making Requirements,** with guiding questions for each. Modeled after private sector use of Data Translators, the process is managed by a Data Facilitator who practices oversight and helps determine the design and management of a Data Action Plan. The **Data Action Plan** provides the operational solution to the Data Problem and provides the consistency needed for longer-term learning around a **Data Learning Agenda**.

The table below gives an overview of this process with next steps for operationalization.

Table: Development Decision-Making Process with a Data Facilitator				
Phase	**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

FEEDBACK AND A RESEARCH AGENDA

Upon review of the results and recommendations, discussions with advisors provided insight into a future research agenda that builds on this report.

A Data Science Operational Framework for Development

Data science operational frameworks are meant to curate and standardize skill-sets definitions, terms, and concepts to provide the commonality needed for larger-scale applications and learning. Advisors noted a need to build on the findings of this research to create a fully designed operational framework around the proposed decision-making framework that donors can use to identify and procure necessary skill-sets as needed.

A Need for Better Contextualization

Advisors noted the need for more contextualized guidance that is specific to the wide variety of Development data requirements. Possible dimensions for organizing this guidance could include:

- Types of Decisions (as defined by complexity/level of uncertainty or function)
- Types of Development Actors (as defined by capacity or function)
- Types of Problems (defined by behavior change mechanism and/or sectors)

Experimentation, Learning, and Accountability

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 faster than the guidance to inform it or the requirements that allow it.

Whose Problem is Being Solved?

Advisors noted what was framed as "beneficiary centric" versus "beneficiary driven" Development. Some considered the former to be more in line with "current top-down donor dictated data science applications," while the latter is collaborative in soliciting beneficiary level data on what problems should be addressed and the barriers to the behavior changes needed to address them. 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 (i.e. the beneficiaries). Advisors noted that if data is to be treated as an asset with an expected return on the investment 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.

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.

Visit pulte.nd.edu/cooper-sources for a full list of sources used to write this policy brief.

POLICY BRIEF TEAM

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.

Recommended citation: Cooper, Michael and Paul Perrin. Maximizing Returns on Data Science Investments: The Evolution of Data-Driven Decision-Making in Development. Notre Dame, IN: Keough School of Global Affairs; Pulte Institute for Global Development, 2021. https://doi.org/10.7274/r0-t0xq-4p69

Pulte Institute for Global Development
3150 Jenkins Nanovic Halls, Notre Dame, IN 46556
(574) 631-2940 • globaldevelopment@nd.edu
pulte.nd.edu