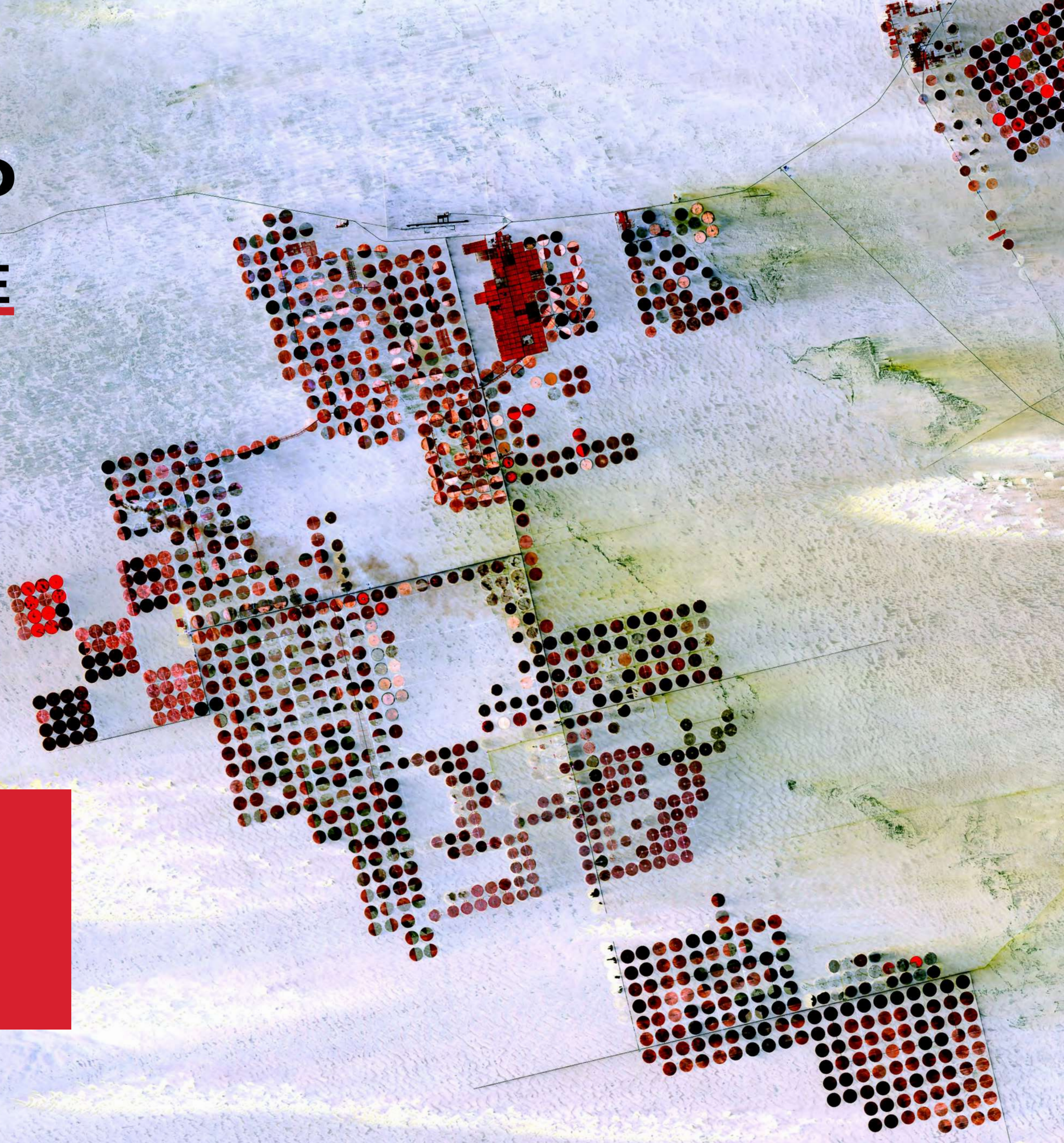# DATA
POWERED
POSITIVE
DEVIANCE

DPPD Handbook

A step-by-step guide for development practitioners to apply the Data Powered Positive Deviance method

Version 1.0 | November 2021

Acknowledgments

Contents

Introduction

Technical Terms

Positive Deviance
An approach that seeks to identify outperformers to understand and replicate their strategies and practices within a community

Positive deviants
Individuals or groups who achieve significantly better outcomes than their peers, despite having similar challenges and resources

Data Powered Positive Deviance
A method that combines non-traditional digital data and traditional data to identify and characterize outperformers in development-related challenges

Non-traditional data
Data that is digitally captured (e.g. mobile phone records and financial data), mediated (e.g. social media and online data), or observed (e.g. satellite imagery)

Traditional data
Data that is captured manually such as official statistics, observation data, surveys, and interviews

The Method

Positive Deviance (PD) is based on the observation that in every community or organization, there are a few individuals who achieve significantly better outcomes than their peers, despite having similar challenges and resources. These individuals are referred to as positive deviants, and adopting their solutions is what is referred to as the PD approach[1].

The method described in this Handbook follows the same logic as the PD approach but uses pre-existing, non-traditional data sources instead of — or in conjunction with — traditional data sources. Non-traditional data in this context broadly refers to data that is digitally captured (e.g. mobile phone records and financial data), mediated (e.g. social media and online data), or observed (e.g. satellite imagery). The integration of such data to complement traditional data sources generally used in PD is what we refer to as Data Powered Positive Deviance[2] (DPPD).

The digital data opportunity

Recent developments in the availability of digital data provide an opportunity to look for positive deviants[3] in new ways and in unprecedented geographical and on temporal scales. A number of studies[4] have described the challenges related to the application of the PD approach in development. Given these challenges, there are obvious opportunities for innovation in PD and our particular interest here is in the innovative opportunities offered by non-traditional data, following the increasing "datafication" of development and the growing availability of big datasets in a variety of development sectors[5]. DPPD builds on this and expands our ability to extract value from non-traditional digital data while providing a systematic process for leveraging local know-how and the collective wisdom of communities.

Data Powered Positive Deviance

The DPPD method described in this Handbook emerged from a process of research and testing and follows the same stages as the PD approach. The difference is that DPPD integrates pre-existing, non-traditional data across the five stages, requiring a series of new and specific methods and practices that are not required in the PD approach. The first stage is also somewhat different because it not only defines the problem, but it also checks if it is suitable and feasible to use the DPPD method for the proposed project.

Table 1 lists the five stages of the DPPD method. This Handbook dedicates a section to each stage.

---

[1] Positive Deviance: *https://en.wikipedia.org/wiki/Positive_deviance*
[2] Data Powered Positive Deviance: Combining traditional and non-traditional data to identify and characterise development-related outperformed: *https://www.gdi.manchester.ac.uk/research/publications/di/di-wp91*
[3] Positive Deviance, big data, and development: A systematic literature review: *https://onlinelibrary.wiley.com/doi/abs/10.1002/isd2.12063*

[4] The Positive Deviance approach: Challenges and opportunities for the future: *https://pubmed.ncbi.nlm.nih.gov/12503241*; Present promise, future potential: Positive deviance and complementary theory: *http://www.laurelfelt.org/wp-content/uploads/2011/06/PositiveDeviance-CodyMayer.LaurelFelt.Quals_.May2011.pdf*;
[5] Big data for development: a review of promises and challenges: *https://onlinelibrary.wiley.com/doi/full/10.1111/dpr.12142*

Stage 1

Assess problem-method fit	Define the problem and check if DPPD is a suitable and feasible method by assessing the required data and capabilities and by ensuring that potential outcomes are desirable for the target group.

Stage 2

Determine positive deviants	Divide the population of study into homogeneous groups and measure the performance of the observed units to identify potential positive deviants. Conclude this stage with a preliminary validation of identified potential positive deviants.

Stage 3

Discover underlying factors	Conduct field research on the performance of both positive deviants and non-positive deviants. Collect and analyze data to identify predictors that distinguish both groups. Uncover positively-deviant behaviors, practices, and other factors that explain the outperformance of positive deviants.

Stage 4

Design and implement interventions	Assess the potential of the identified practices to be replicated and scaled. Based on the insights generated, design and implement community interventions to scale those practices. Further insights on contextual factors influencing behavior and outcomes within communities can be considered for future policy interventions.

Stage 5

Monitor and evaluate	Monitor and evaluate the effectiveness and suitability of the community and/or policy interventions.

Table 1:
DPPD stages

Methodological differences between PD and DPPD

Both PD and DPPD start with the same premise: in every population, there are individuals or communities who, despite facing similar challenges and limitations, achieve better results than their peers. Both acknowledge the value of locally-sourced solutions, the agency of people, and the need to value the relative advantage that a positive deviant might have over others. However, they somewhat differ methodologically, as highlighted in Table 2.
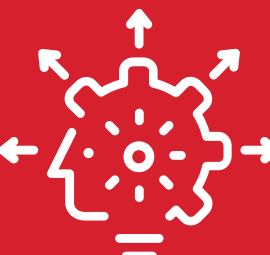
	Positive Deviance	**Data Powered Positive Deviance**
Data	Positive deviants and their practices are identified using primary data collected through direct observation, like interviews and surveys.	Positive deviants are identified using existing digital data sources, as well as more traditional secondary data to control for potential confounds. Their practices are then identified using primary data collected during the field research phase.
Coverage	Positive deviants are generally individuals identified within very specific areas that are largely homogeneous (e.g. an individual village).	Positive deviants can be individuals, communities, administrative units, or geographic areas that can be identified in large spatial areas and on temporal scales.
Capabilities	Requires local knowledge, domain expertise, basic statistical analysis, appreciative inquiry, and ethnography skills.	Requires local knowledge, domain expertise, and a combination of domain-specific data knowledge, advanced statistical analysis, appreciative inquiry, and ethnography skills.
Costs	The initial cost of positive deviant identification is directly proportional to the sample size, as it relies mainly on primary data collection.	The initial cost of positive deviant identification can be low because it relies on readily available data. However, there can be significant additional costs associated with data access and analysis.

Table 2:
Differences between Positive Deviance and Data Powered Positive Deviance

The Initiative

In April 2020[6], a network of partners launched the DPPD initiative. The Initiative was set up to design, test, and document a method combining the use of novel, digital data sources with one of the most prominent asset-based approaches to development practice — Positive Deviance.

The DPPD network

The DPPD network was initiated by the GIZ Data Lab, the UN Global Pulse Lab Jakarta, the UNDP Accelerator Labs Network, and the University of Manchester's Centre for Digital Development. It is an action-learning network that explores the use of new digital data sources to systematically identify and understand positive outliers in various domains, including deforestation, agriculture, and gender-based violence. The goal is to build a shared understanding of the opportunities and limitations of the DPPD method, to develop a set of tools and techniques to apply the method, and to test the method across locations, sectors, and data types and sources.

While the DPPD method may seem niche and very technical, it has the potential to help development practice shift from top-down identification and tackling of development challenges to focusing on communities' inherent assets and capabilities as the ultimate starting point in the search for solutions[7]. It provides a path to move away from external solutions to the diffusion of local practices and strategies that take into account contextual variables, making them more likely to stick[8] and less vulnerable to social rejection.

Positive outliers matter

---

[6] Launching the Data Powered Positive Deviance initiative: *https://dppd.medium.com/launching-the-data-powered-positive-deviance-initiative-6872cd9208f9*

[7] Fall in love with the solution, not the problem: *https://www.nesta.org.uk/blog/fall-in-love-with-the-solution-not-the-problem*

[8] What are the new skills we need in development: *https://acclabs.medium.com/what-are-the-new-skills-we-need-in-development-ad4dde64d131*

The Pilots

To test, refine, and document the DPPD method, the network initiated a number of pilots covering a variety of domains, countries, data sources and types, and scales (see Figure 1 and Table 3). A core group was responsible for developing and revising conceptual frameworks, tools, and techniques to gradually refine the DPPD method as the pilots progressed. The implementation of the method was led by in-country practitioner teams that brought together domain and technical specialists who worked closely with the core team.

Districts in Germany that are better able to contain the spread of COVID-19 infections

Publication outperformance among Egyptian researchers

Safer public spaces for women in Mexico City

Cattle farmers in Ecuador who deforest less than their peers

Farmers who achieve higher than expected cereal crop productivity in Niger

Communities in Somalia that preserve their rangelands despite frequent droughts

Farmers who achieve higher than expected rice productivity in Indonesia

Figure1:
Map of DPPD projects

	Positive deviants	How to tell if they are positive deviants?	Data used	Partners
Ecuador	Cattle farmers in the Ecuadorian Amazon who operate in areas of potential forest clearance for farming without deforesting themselves	Cattle-raising farms with deforestation rates that are significantly lower than expected for three consecutive years, while controlling for the size of the farm, the land use, soil adaptability, and cattle density	Remote sensing data, vaccination data, cadastral data, official statistics, land use data, observation data, and semi-structured interviews	UNDP Ecuador Accelerator Lab, GIZ Ecuador, UNDP PRO Amazonia, GIZ Data Lab, and University of Manchester
Egypt	Researchers in Egypt who have high citation metrics scores	Information system researchers in public universities in Egypt who achieved significantly higher-than-average scores in one or more of six citation metrics	Citation data from Google Scholar, research publications on Scopus, university websites, interviews, and surveys	University of Manchester
Germany	Districts in Germany that were better able to contain the spread of COVID-19 infections	Districts with lower than expected weekly reproduction factors, considering two timeframes, while controlling for weather, mobility, ruralness, and socioeconomic status	Public data on infections per district, official administrative statistics, CDR-derived mobility data, weather report data, and focus group discussions	GIZ Data Lab, Teralytics
Indonesia	Rice-farming villages in Indonesia that have higher rice productivity than expected	Villages with higher productivity than expected, as measured by enhanced vegetation index scores, while controlling for their climatic, socioeconomic, and demographic conditions	Remote sensing data, official statistics, administrative boundary data, and crop masks	UN Global Pulse Lab Jakarta, GIZ Data Lab, and University of Manchester
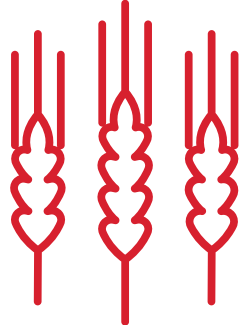
	Positive deviants	How to tell if they are positive deviants?	Data used	Partners
Mexico	Public spaces in Mexico City where women are safe(r)	Areas where gender-based crimes and crimes with female victims are lower than expected given their population density, demographics, and socioeconomic status	Mexico City open data portal, 911 calls, administrative data, official statistics, observation data, cartography, and semi-structured interviews	UNDP Mexico Accelerator Lab, GIZ Mexico Project on 2030 Agenda, GIZ Data Lab, and University of Manchester
Niger	Cereal-growing communities in Niger that achieve healthier rainfed yields despite conflict and climatic challenges	Communities with a higher soil-adjusted vegetation index than expected, while controlling for soil, precipitation, and land use	Remote sensing data, administrative boundary data, land use data, observation data, semi-structured interviews	UNDP Niger Accelerator Lab, GIZ PromAP Niger, GIZ Data Lab, and University of Manchester
Somalia	Pastoral communities in Northern Somalia that are better able to preserve the health of their rangelands	Communities in the same land capability class (i.e. having similar biophysical conditions) that have high vegetation productivity, calculated using the soil-adjusted vegetation index	Remote sensing data, settlement location data, observation data, focus group discussions, and semi-structured interviews	UNDP Somalia Accelerator Lab, GIZ Project on Sustainable Land Management Somalia, GIZ Data Lab, and University of Manchester

Table 3:
DPPD pilot projects

The choice of locations and development challenges addressed by the pilot projects[9] was made to ensure diversity in terms of non-traditional data types used, such as remote sensing data, mapping, and cadastral geographic data, as well as both proprietary and open data sources. These were further complemented by a variety of traditional data sources, including official statistics, administrative data, surveys, and interviews.

The units of analysis were selected to cover different aggregation levels, from individuals and communities up to administrative and geographical units in urban areas or villages. The diversity of domains, countries, data, and scales was important for testing the DPPD method and to gauge the breadth of its applicability.

[9] Find out more about the pilots at *https://www.datapoweredpd.org/projects*

The Handbook

This Handbook is a practical guide for the application of the DPPD method. It aims to help development practitioners identify and scale uncommon but successful practices in different sectors by mixing analytical insights from traditional and non-traditional data sources. This should help amplify innovative, locally sourced, and evidence-informed solutions to development challenges.

The step-by-step guidance contains an initial assessment of the suitability, feasibility, and desirability of the method given a particular project scope all the way to the more technical details of how to identify potential positive deviants in large data sets and subsequently validating these findings through fieldwork. It builds on the documented experience of pilot teams who have worked on six different challenges across seven countries.

This Handbook targets all development practitioners. The language and terms used are therefore not specific to any of the institutions involved in the DPPD initiative.

How to use the handbook

The Handbook is divided into five sequenced stages, as shown in Figure 2.

Stage 1 helps define the problem and validate if it is suitable and viable to use the DPPD method. Stage 2 provides guidance on identifying positive deviants within available datasets. Stage 3 details how to uncover factors underlying outperformance. Stages 4 and 5 cover the different aspects of designing and implementing DPPD interventions, as well as means to monitor and evaluate those interventions.

You will find detailed guidance with practical examples in Stages 1 through 3. Most of the pilot teams have worked through these stages and their experiences and learnings were documented for the Handbook. The descriptions of Stages 4 and 5 in the corresponding sections is different from the first three. While we provide initial thoughts and suggestions on how to go about Stages 4 and 5, they do not have the depth and practical learnings from the pilots.

The pilots have yet to reach these final stages of the DPPD method where they design interventions to scale and amplify the positively-deviant practices identified and monitor their implementation.

The Handbook will be useful for data for development experts interested in exploring a new method that looks at identifying and scaling solutions rather than understanding and describing problems using digital data. PD practitioners will find value in DPPD as a way to explore and use new sources of data and expand the scope of their work.

Finally, domain experts will see value in the DPPD method to find uncommon practices to inform and complement their current portfolio of development interventions. It is important to emphasize that the DPPD method is best suited for teams, as it requires a range of skills unlikely to be accessible to one single individual.

The DPPD handbook is written by practitioners for practitioners. It is meant to be brief, practical, and easily applicable. For those interested in diving deeper into theory and the underlying conceptual frameworks, the Handbook contains links and references to relevant resources. Each section of the Handbook contains descriptions of the steps to follow, practical examples from the pilots, as well as templates to fill in when designing and implementing a DPPD project.
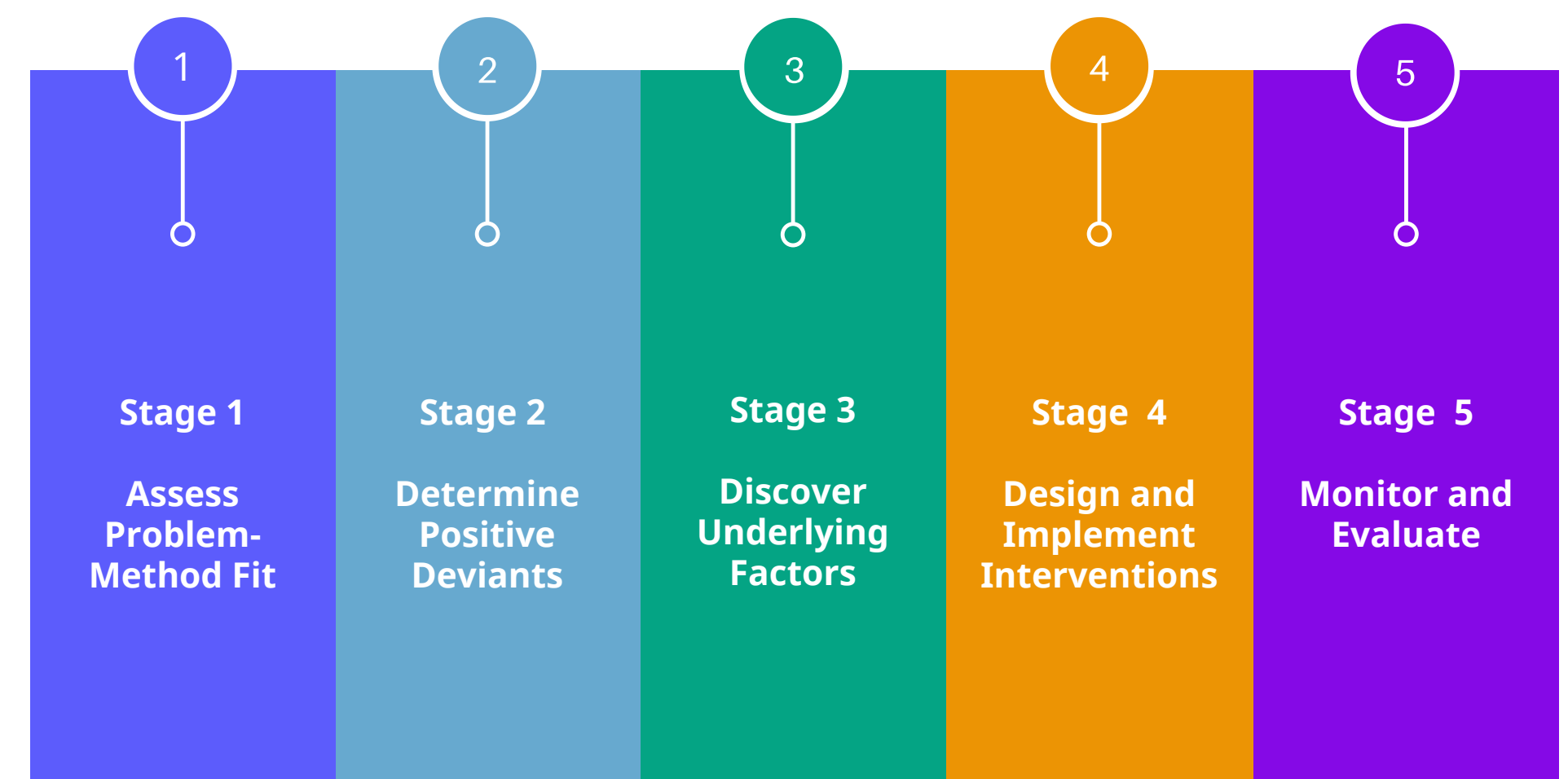
Figure 2:
DPPD stages
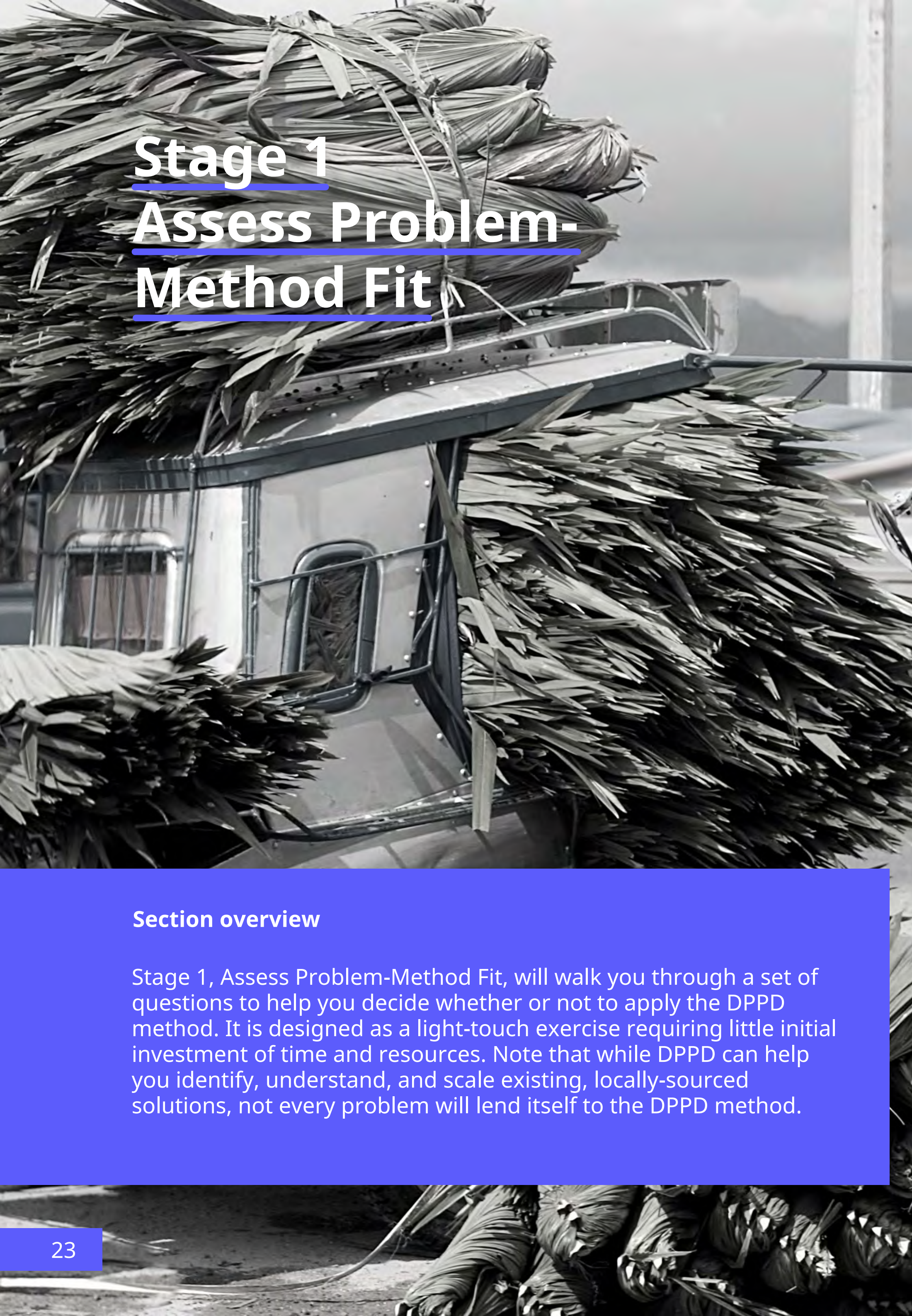
Stage 1
Assess Problem-Method Fit

Section overview

Stage 1, Assess Problem-Method Fit, will walk you through a set of questions to help you decide whether or not to apply the DPPD method. It is designed as a light-touch exercise requiring little initial investment of time and resources. Note that while DPPD can help you identify, understand, and scale existing, locally-sourced solutions, not every problem will lend itself to the DPPD method.

Technical Terms

Unit of analysis	The major entity within the target group that is being analyzed and the level at which you expect to find positively-deviant solutions
Target group	The group that will be included in the search for positive deviants and their uncommon practices and strategies
Desired outcome	The state you want the target group to be in after the intervention
Outcome data	The data that can help directly or indirectly measure the performance of the target group, given the desired outcome
Behavioral data	Data that is collected during the fieldwork to understand the uncommon practices and strategies of positive deviants
Contextual data	Data that allows to control for factors that are likely to impact performance but are not related to practices or behavior
Data readiness	The number of conditions regarding data availability, accessibility, and adequacy that need to be met for the DPPD method to work
Data availability	The extent to which the data is collected or generated in the first place
Data accessibility	The extent to which you can access and use the data
Data adequacy	The extent to which the data fits the scope of the project, including the know-how and resources required to access and use it

Problem Definition

The first step in this stage is to properly describe the problem and its root causes. It is also important to define the target group of individuals, communities, or geographic units that are impacted by the problem, as well as the units that will be analyzed in that target group. Finally, it is crucial to consider the desired outcome, which is what should be achieved at the end of the intervention. Try to develop or validate each of these components with people affected by the problem using Tool 1.1 at the end of this section.

Problem

Briefly describe the essence of the problem. This does not need to be long and detailed but should give you an initial idea of where to look for positive deviants.

Root causes

Describing the likely root causes of the problem will help you better understand the problem, who would be considered positive deviants, and what factors need to be controlled for when identifying them. For example, if the root cause is drought, you need to identify positive deviants from within groups that experience similar rainfall conditions and that have similar access to water resources.

In the Somalia rangelands project, the problem statement was as follows: population growth, increased settlements, pressure on natural resources, and recurring droughts all led to the severe degradation of rangelands. This led to losses in livestock numbers while threatening the sustainability of pastoral livelihoods.

Target group

The target group is the group that will be included in the search for positive deviants and their uncommon practices and strategies. It is sometimes also referred to as the study population or the study sample when it comes to the more analytical steps later on.

The target group can be individuals, communities, or administrative or geographic units that are negatively impacted by the problem, or, in some cases, that cause the problem.

Keep in mind that the target group needs to be big enough to provide a sufficiently large sample size (see Stage 3 for details), and it needs to be in a location that is accessible to undertake the qualitative fieldwork in Stage 3 of the DPPD method.

In the Ecuador deforestation project, the target group was the farmers at the frontier of the Amazon forest, as they were the ones harming the forest the most. In the Somalia rangeland project, the West Golis region was the focus — it has the majority of the pastoralist communities and sufficiently large sample size with over 200 settlements across four administrative regions mostly accessible for fieldwork.

Unit of analysis

The unit of analysis refers to the level at which you expect to find positively-deviant solutions. They are the major entity within the target group that is being analyzed. This could be individuals such as farmers, groups such as clusters of farms, or even geographic units like villages. While you don't need to settle on the unit of analysis at the beginning, it is important to reflect on it when determining the scope, as it will impact the data sources you will be using later on.

In the Somalia rangelands project, units of analysis were villages with a 5 km buffer around their centers to represent the average movement radius of livestock.

Desired outcome

The desired outcome is the state you want your target group to reach after the intervention.

In the Somalia rangelands project, the desired outcome was healthier rangelands that enable pastoralists to maintain and enhance their livelihoods.

It is important to involve stakeholders early on, ideally those negatively impacted by the problem or with considerable stakes involved in identifying and scaling practices. Understanding their perspectives and securing their buy-in at the very beginning helps develop a much stronger understanding of the problem, its perceived causes, the community's challenges and constraints, and existing common practices and normative behaviors. It also makes the adoption of positively-deviant practices much more likely through increased ownership.

The team in Somalia met with government officials, researchers, and members of agro-pastoral communities to develop a more nuanced understanding of the problem, the main drivers of rangeland degradation, and what could be considered a positively-deviant community.

Suitability

The second step is to assess the suitability of the DPPD method. You should ask yourself a simple question before moving forward: "Should I be applying the DPPD method?" To help you answer this question, use Tool 1.2 and consider these two secondary questions[1]:

- What is the nature of the development challenge I am trying to address?
- What is the likelihood that positive deviants exist?

---

[1] There are other criteria to take into account such as the buy-in and commitment of local leadership, but we highlight these two here as they are specific to the Positive Deviance approach and the DPPD method.

The nature of the problem

There is no single and exact way of answering the first of the two questions. However, as a rule of thumb, consider this: If the problem requires a technical solution, say building a road, constructing a dam, or introducing a new IT system, the DPPD method will likely not be suitable. The positive change you are aiming for is likely not related to a change in practice or behavior.

If addressing the development challenge, however, (also) depends on a change of practices or behavior, the DPPD method will likely make a difference. It will allow you to identify and scale practices that can shift the status quo. This applies to challenges in areas where the behaviors, practices, and coping strategies of individuals or communities are key determinants of success, such as health, nutrition, agriculture, or education.

The likelihood that positive deviants exist

The more likely the existence of positive deviants in the problem space you are engaging in, the greater the chances are that you can identify them for the particular challenge you are tackling. While you will not know for sure at the beginning, you can at least get a sense by searching for signs of positive deviance and grassroots solutions in other communities living in comparable contexts and facing similar challenges. This should help you determine the chances of finding positive deviants.

Note that this should not require significant time or resources. Talk to people in your network who are familiar with the specific sector you are focusing on or review academic and grey literature. You can also browse existing collections of grassroots innovations like the Honey Bee Network's database[2] which contains more than one million ideas, innovations, and traditional knowledge practices.
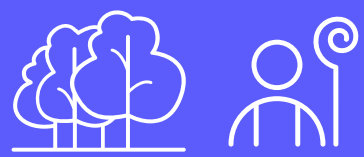
In the Ecuador cattle-farming project, the team learned from agriculture experts that certain farmers adopt more sustainable cattle ranching practices and deforest less than others. In Somalia, the team learned through interviews with an officer from the Ministry of Environment and Rural Development that a positively-deviant community was protecting its trees from being cut down or burned for charcoal production.

The suitability matrix

The PD suitability matrix (Figure 1) shows four quadrants that can be used to judge how suitable Positive Deviance (PD) is, including the DPPD method, for a particular problem. Problems for which PD is best suited generally lie in the top right quadrant, where positive deviants are likely to exist and where scaling their practices is likely to contribute to addressing the challenge. Problems in the bottom left quadrant, where positive deviants are unlikely to exist and where scaling practices will likely have a limited impact on the challenge, are not suitable for PD.

The greater the role of behavioral change in addressing the problem, the more suited Positive Deviance is

The nature of the problem

Positive deviants are unlikely to exist, but identifying and scaling their practices could have a significant impact on the problem.

Positive deviants are likely to exist, and identifying and scaling their practices could have a significant impact on the problem.

Positive deviants are unlikely to exist, and identifying and scaling their practices could have limited impact on the problem.

Positive deviants are likely to exist, but identifying and scaling their practices could have limited impact on the problem.

The likelihood that positive deviants exist

The more positive deviants are likely or known to exist, the more suited Positive Deviance is

Figure 1:
PD suitability matrix

---

[2] Check the Honey Bee Network database for examples of grassroots innovations:
http://www.honeybee.org/seeking_solution.php

Feasibility

The third step in the Problem-Method Fit Stage is to assess the feasibility of applying the DPPD method. While many factors are involved, the two most important ones are data and capabilities.

Data

The DPPD method leverages various data types from different sources. Three broad categories of data are needed for the method to work: outcome data, contextual data, and behavioral data.

Contextual data	Behavioral data	Outcome data
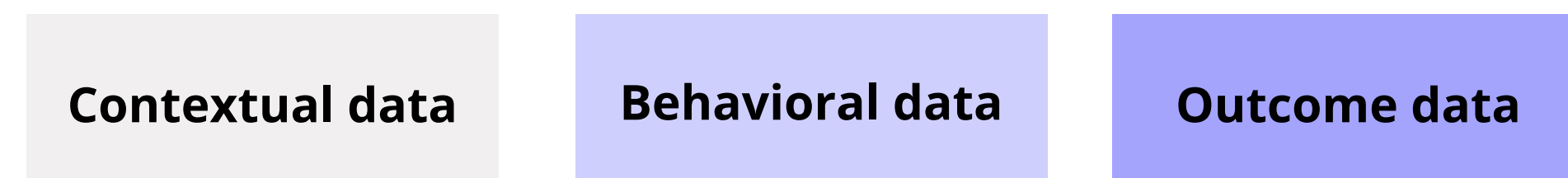
Outcome data

We refer to outcome data as data that can help directly or indirectly measure the performance of the target group, given the desired outcome — for example, crop yields for sustainable agriculture or food insecurity, crime rates for public safety, or test scores for school performance. The outcome data helps identify individuals, groups, or, more generally, units within the target group that outperform their peers. The DPPD method leverages the potential of readily available non-traditional data, like earth observation data, online or mobile data, citizen-generated data, or sensor data, to capture these outcomes.
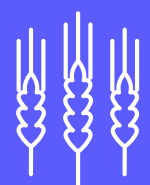
In the agricultural project in Niger, villages that produced better crop yields than others were identified using remote sensing data, from which the team derived a vegetation index that indicated the health of sorghum and pearl millet crops during the rainy season.

Contextual data

Contextual data allows controlling for factors that are likely to impact performance but are not related to practices or behavior.

A large portion of those factors should emerge from determining the root causes when defining the problem. Contextual data helps put the outcome measure in perspective, enabling the identification of positive deviants in relation to their context — a relative rather than absolute approach.

Contextual data	Behavioral data	Outcome data

Identification of positive deviants (see Stage 2)

In the agricultural project in Niger, positively-deviant villages were not identified simply based on their absolute vegetation health values, but rather using relevant contextual factors that would affect the yield of sorghum and pearl millet, such as precipitation, climate, and soil. This prevented us from falsely identifying a village as a positive deviant, simply because it had more rainfall than others.
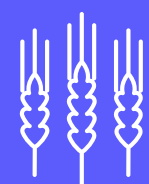
Behavioral data

Behavioral data is data that is collected during the fieldwork at Stage 3. It fills the gap between the outcome data that shows certain units are outperforming and the contextual data that only in part explains why they are. What we refer to as behavioral data is essential information on the uncommon practices and strategies of identified positive deviants.

Contextual data	Behavioral data	Outcome data
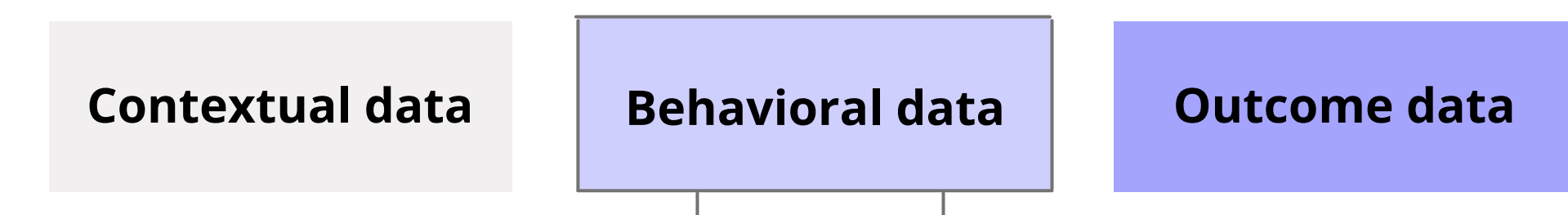
Understand underlying factors (see Stage 3)
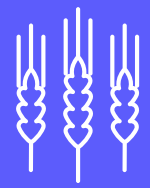
In the Niger agricultural project, the team used a semi-structured survey to uncover uncommon practices and strategies of farmers, such as farming tools, techniques, and inputs in the positively-deviant villages that were identified in the quantitative data analysis (Stage 2).

Data readiness: availability, accessibility, and adequacy

A number of conditions regarding data availability, accessibility, and adequacy need to be met for DPPD to work. The questions presented in Figure 2 and Tool 1.3 at the end of this section can help you assess whether your data fulfills these conditions.

Availability

First, it is important to ensure there is a non-traditional data set that can be used to capture the outcomes of the observed population (outcome data). You will also need data to contextualize the findings to make sure positive deviants are identified based on relative performance rather than absolute performance (contextual data).

Search for suitable data and contact potential data providers, such as government offices, telecom operators, research centers, and social media data aggregators that might know of or hold this data. The data needs to be granular enough for the envisaged unit of analysis and provide the required spatial coverage of the envisaged study population and the temporal coverage for the type of performance you have in mind.

In the Ecuador cattle-farming project, the team tried to identify cattle farmers who operate in areas of potential forest clearance without deforesting themselves. Sufficient temporal coverage, across several years was needed in the outcome data — a farm could be a positive deviant one year but not in the following years, as it may start to deforest heavily, turning it into a "negative deviant."

Public vs non-public data
Data partnerships

Accessibility

Do you have access to and are you allowed to use the data?

Availability

Does the relevant data exist?

Adequacy

Is the data adequate for your intended use?

Data that captures outcomes
Data on contextual realities

Geographic coverage
Temporal coverage
Spatial resolution
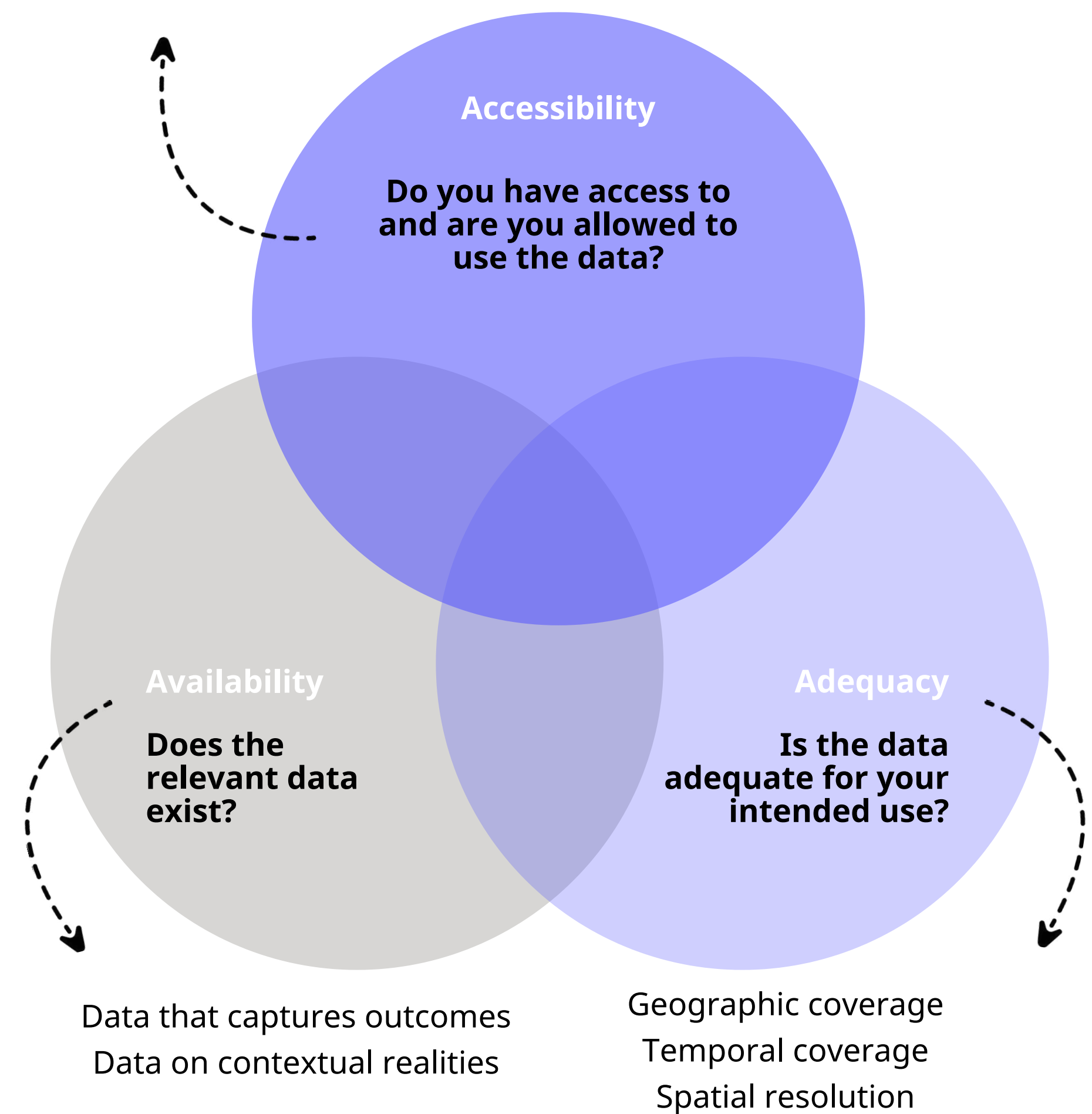
Figure 2:
Data readiness assessment

Accessibility

Check if you have access to the data sets in question and if you are allowed to use them. If you need to use non-public data, having data access agreements in place with data providers is a real asset[3]. It can take significant time to negotiate the conditions of data access.

The cattle-farming project in Ecuador grew out of an existing partnership between the UNDP ProAmazonia project and the Ecuadorian Ministry of Environment. This made it possible to access cattle vaccination data from the Ministry of Agriculture, training datasets for land cover analysis through the Ministry of Environment, and cadastral data as well as farm boundary data from the national agricultural survey through ProAmazonia.

Adequacy

The data you are able to access and use should fit the scope of the project. That means the know-how needed for the analysis should be attainable and the choices of both data and skills should consider the project's time and budget limitations.

Achieving data adequacy can take several iterations between framing the problem, scoping the project, and mapping the necessary and available data until you find a fit. However, make sure you do not compromise the initial purpose of the project. We suggest you start with a flexible framing of the development challenge you want to address. This will allow you to explore different data sources to capture the performance of the target group.

In the Somalia rangelands project, the team pivoted from trying to find drought-resilient pastoral communities, using their livestock numbers as outcome data, to identifying them based on their ability to sustain healthy rangelands, a necessary condition for them to maintain their livestock. The livestock numbers were very difficult to reliably derive from available data, as this required extensive and complex analytical skills. On the other hand, the rangeland health data was accessible and could more easily be analyzed.

---

[3] Sharing is caring. Four key requirements for sustainable private data sharing and use for public good: https://www.vodafone-institut.de/wp-content/uploads/2019/11/DPA_VFI-Sharing-1.pdf
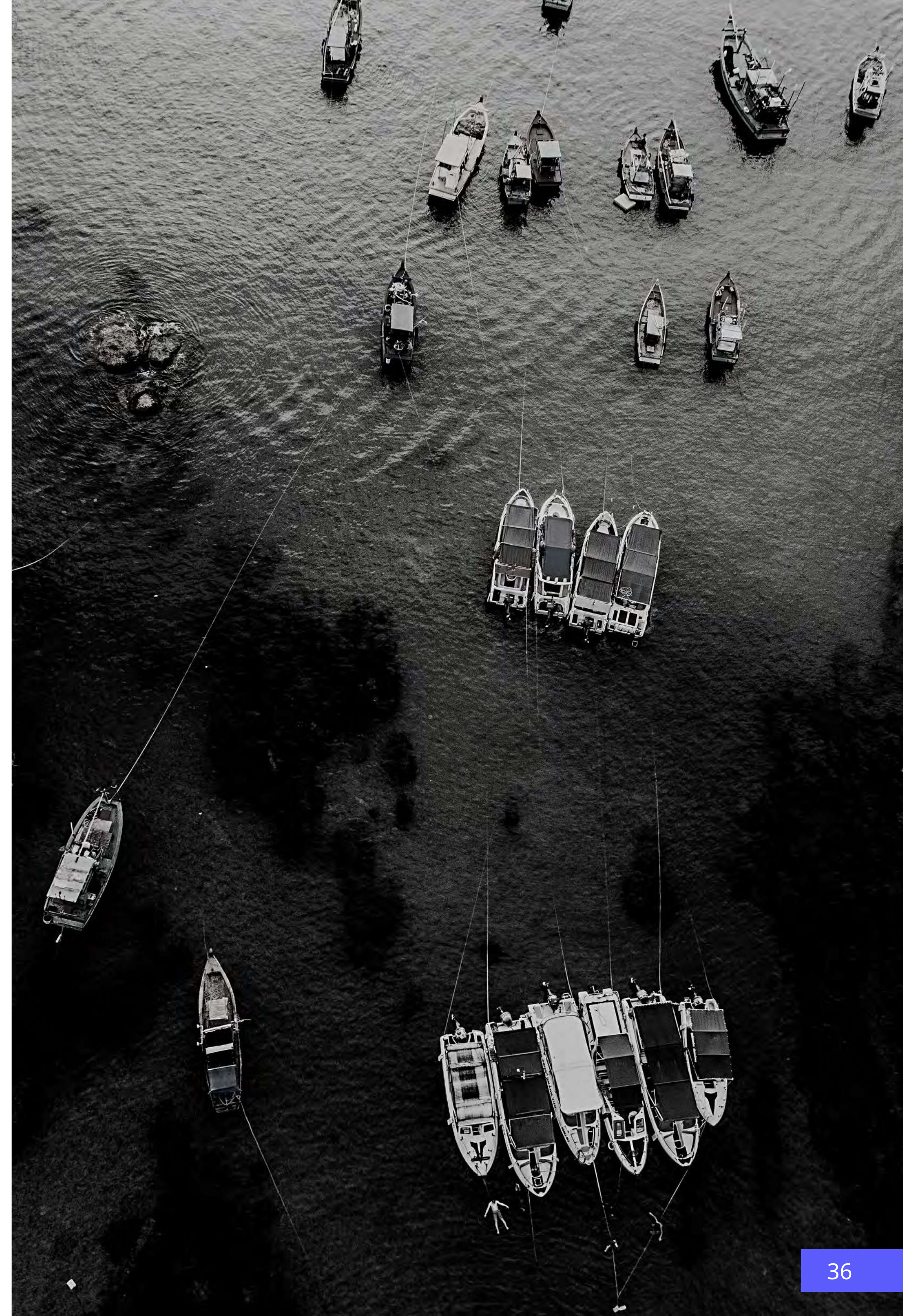
Capabilities

The capabilities needed to apply the DPPD method typically include local knowledge, domain expertise, and a range of computational and analytics skills. The questions presented in Figure 3 and Tool 1.4 at the end of this section can help you assess whether your data fulfills these conditions.

Local knowledge

The performance and underlying practices of potential positive deviants need to be understood within the specific context they exist in. This requires a multidimensional understanding of socioeconomic, cultural, structural, and other factors that might affect the outcome and are very much linked to the local conditions.

Domain expertise

Identifying positive deviants and controlling for potential confounding variables requires the ability to select domain-relevant performance indicators and contextual variables that can only be understood through deep domain expertise.

Data skills

The DPPD method includes the analysis of both quantitative and qualitative data that requires respective skill sets. They range from analyzing large digital data sets — which requires advanced computer science skills — to collecting ethnographic data and observations about positive deviant practices.

Try to have a data analyst on your team who is experienced in working with non-traditional data from the particular sector or domain you are addressing and who has knowledge of the different data sources that can be used. You will also need someone who has experience in conducting mixed-methods studies to collect and analyze the field data. As such a combination of skills is rare for a single person, you might need to build a multidisciplinary team.

Community resources, challenges, ongoing interventions and normative behaviors

Local Knowledge

Do you have sufficient knowledge about the local conditions?

Data Skills

Do you have the required data skills?

Domain Knowledge

Do you have sufficient know- ledge about the respective domain?

Both quantitative and qualitative data skills

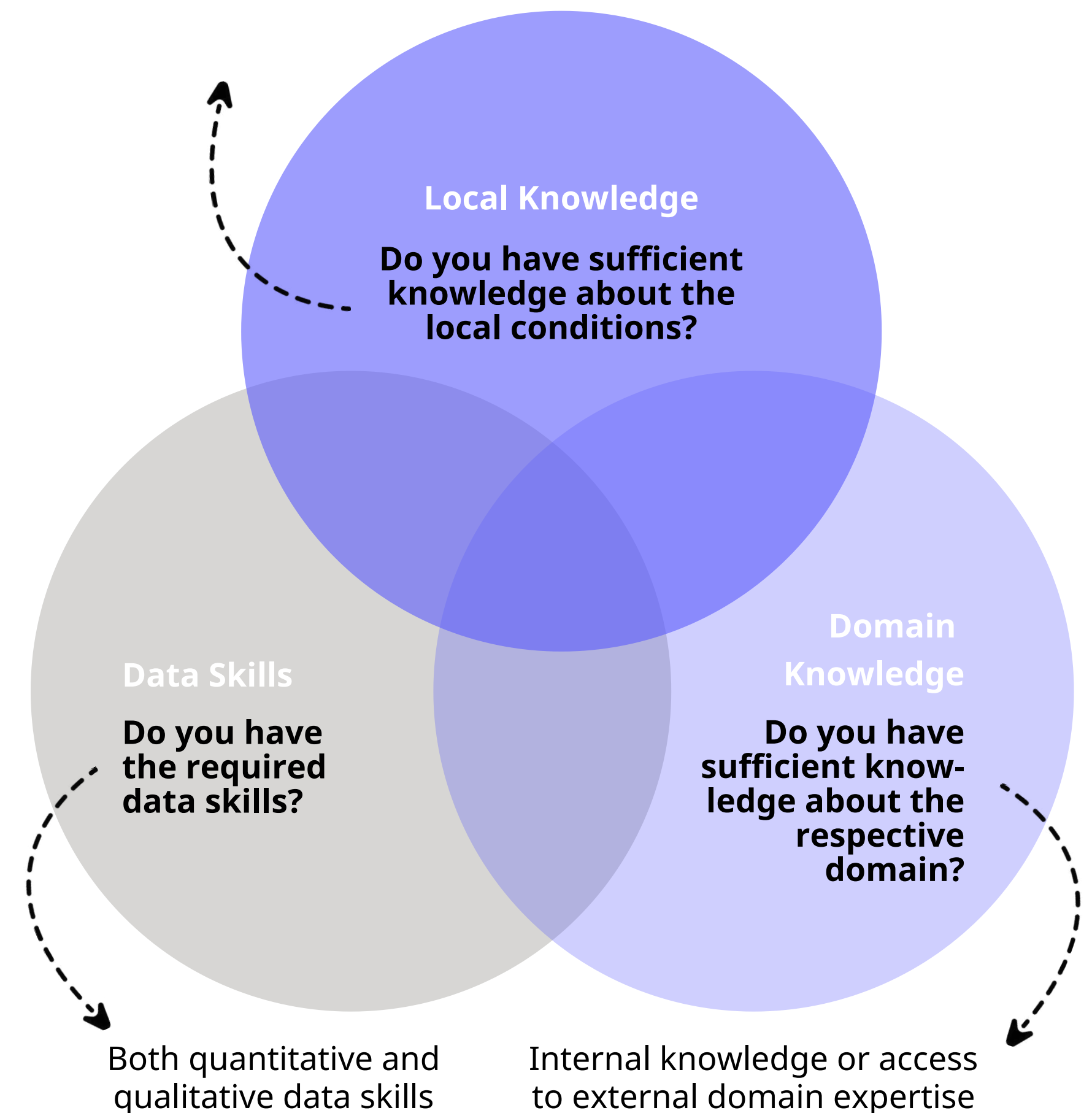Internal knowledge or access to external domain expertise

Figure 3:
Capabilities assessment

Desirability

The fourth step is to assess the risks and benefits (or the desirability) of applying the DPPD method using Figure 4 and Tool 1.5 The PD approach entails acknowledging the value of locally-sourced solutions, the agency of the people you are working with, and respecting the right for positive deviants to decide whether they want to share their practices or not. Moreover, DPPD is about using data responsibly to minimize the risks related to potentially sensitive data, especially when that data is personally identifiable.

As in every development intervention, if you think the outcome might disadvantage or even harm the positive deviants, their communities, or the wider systems they are embedded in, it is important to adjust the project or to ultimately not move forward with it. Just because the DPPD method is suitable and feasible does not mean it should be applied.

Unintended consequences

This section will guide you through the desirability assessment by helping you identify potential unintended consequences of a DPPD intervention, as well as potential risks related to data privacy and protection.

Positive deviants

When applying the DPPD method, you can generally assume that positive deviants are not aware of their innovativeness and/or of the success of their uncommon practices and strategies. In some cases, however, they may be aware of this and may choose deliberately not to share their strategies with others in their community.

For example, positive deviants might fear losing their competitive advantage over others or depleting a resource they alone are aware of, if it were to be shared with others, making their solution unsustainable. It is important to assess whether it is desirable for positive deviants to share their practices with others. Note that this is less likely an issue where cultural norms dictate against competitive strategies, such as child malnutrition or health care, but it might be problematic in areas where people more overtly compete with one another.

Key questions to ask here include:

- Is it generally safe to assume that it will be desirable to scale the practice in question?
- Are we endangering the competitive advantage of positive deviants by sharing their practices and strategies with others?
- Might we risk harming a positive deviant by revealing their identity?

Community

The individual positive deviant(s) might be supportive of the idea of sharing their successful practices. However, you should also ask what the likelihood is of these individuals or communities, who you intend to be at the receiving end of accepting, adopting, and embracing such practices.

Another important question to ask is to what extent a successful practice will really lead to positive effects in the long run. For instance, identifying and scaling the practice of a positively-deviant pastoralist might lead to an increase in pastoral activities and livestock numbers. However, over time this might result in a shortage of grazing land, eventually leading to the obsolescence of the positive practice. In other words, it is important to consider the sustainability of PD practices.

Key questions to ask here include:

- In the short term, are non-positive deviants likely to accept and adopt the PD practice? Or might there be issues linked to limited capabilities or cultural factors that will hinder such adoption?
- In the longer term, will the adoption of the practices lead to foreseeable negative effects?

Systems

While it may be difficult to go beyond the needs and aspirations of an immediate community, it is also worth reflecting on potential negative impacts on the environment, natural ecosystems, governmental and social institutions, and other wider political, social, and economic systems, sub-systems, and institutions.
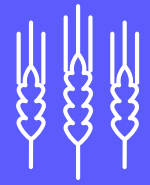
In the Niger agricultural project, the team found that some farmers left millet stalks on their land between agricultural cycles, while others sold it to livestock farming communities. Keeping it on the land protects the soil from wind erosion and helps restore the organic matter, which contributed to higher yields in the next cycle. This gave a competitive advantage in crop production to the prior group of farmers. However, simply scaling this practice to all farms might lead to a shortage of fodder for livestock and pastoral communities. Hence, this might lead to negative consequences for communities in the wider system.

Key questions to ask here include:

- What could be undesirable (or even harmful) outcomes for the system(s) when scaling PD practices?
- Is there a risk that wide adoption of a practice will lead to negative consequences, beyond the positive deviants and the community/ target group?

Positive Deviants	Have an advantage over others
Communities (Target Group)	Mostly non-positive deviants without a relative advantage
(Social) Systems	Systems, sub-systems, institutions that positive deviants are embedded in or connected with
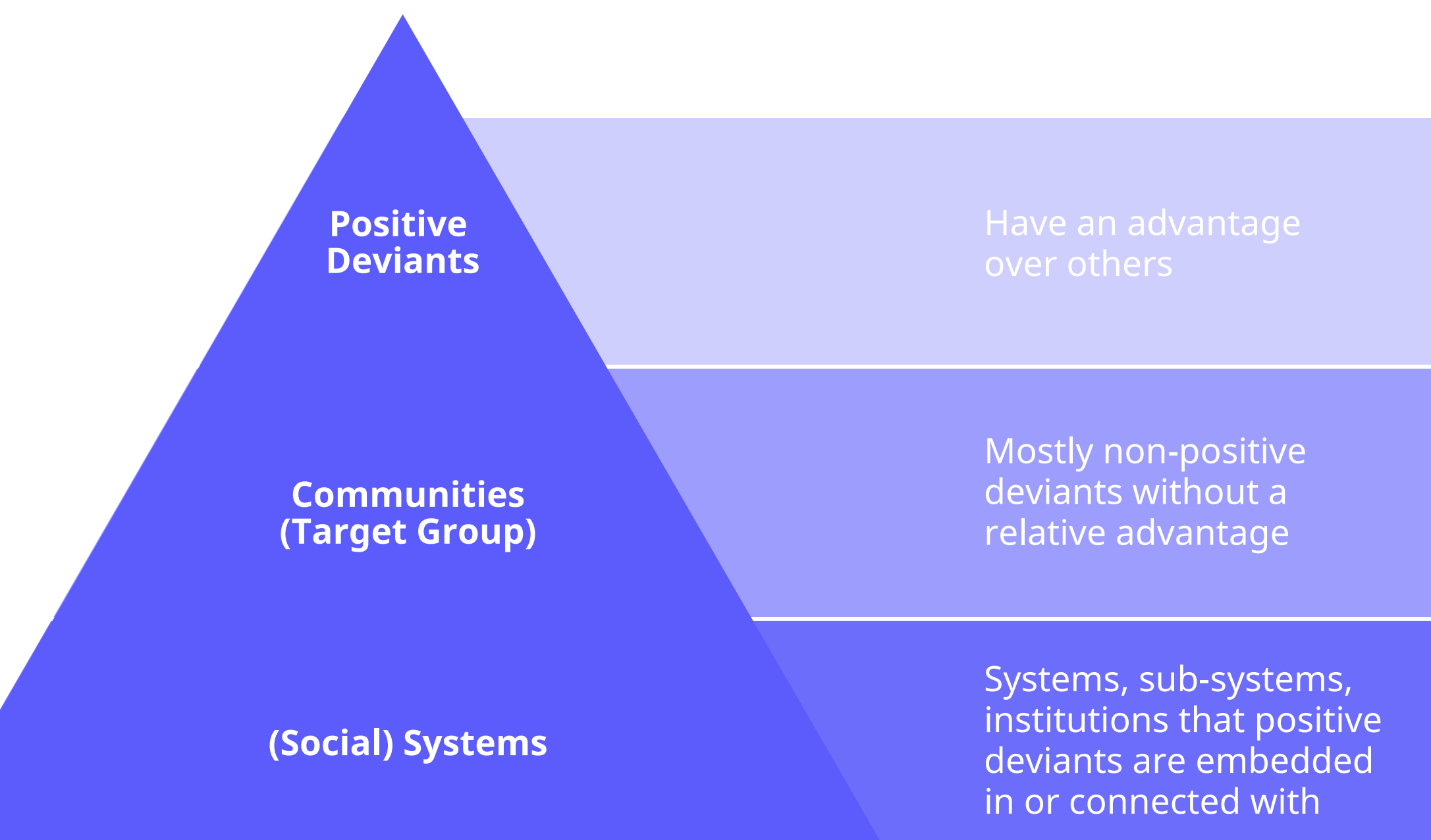
Data privacy

If you believe the DPPD method is suitable given the scope of your project, and that you will have access to the necessary data and skills, it is still important to think about the protection of privacy of individuals and communities involved.

Key questions to ask here include:

- Is the data source to be used of sensitive nature? (e.g. does it contain personally identifiable information?)
- Are safeguards in place for secure data access and processing?
- Has consent been given (directly or indirectly) by the data subjects to use their data?
- To whom can the identity of positive deviants and non-positive deviants be revealed, if at all?

Given the next stages involve the identification of positive deviants, such questions must be thought through at this point. Below is a list of helpful resources on the responsible use of data. The specific assessment to be performed will depend on your project scope and the data envisaged for the analysis.

- GIZ Responsible Data Guidelines[4]
- UN Global Pulse Privacy Assessment Tool[5]
- Responsible Data Handbook[6]
- Ten simple rules for responsible big data research[7]

Figure 4:
DPPD desirability pyramid

---

[4] https://d-nb.info/1175764698/34
[5] https://www.unglobalpulse.org/policy/risk-assessment
[6] https://responsibledata.io/resources/handbook
[7] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5373508
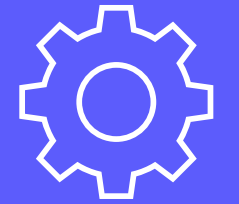
Tool 1.1
Problem Definition

The Problem Definition tool will help you describe the problem and its root causes, define the target group and the units that will be analyzed in that target group, and the desired outcome of a potential intervention. Try to develop or at least validate each of these components with people affected by or with expert knowledge of the problem.

What is the problem? 1.

Examples:

Deforestation, unsafe public spaces, rangeland degradation

Why does the problem exist? 2.

Examples:

Population growth, increased settlements, recurring droughts, pressure on pastures

Who is impacted? 3.

Examples:

Pastoral communities, rain-dependent farmers, farmers at the frontier of the Ecuadorian Amazon

What is the unit of analysis? 4.

Examples:

Individuals such as farmers, groups such as pastoral communities, or geographic units like urban blocks

What does a positive outcome look like? 5.

Example:

Healthier rangelands that enable pastoralists to maintain and enhance their livelihoods or safer public spaces for women

Tool 1.2
Suitability Assessment

Use the tool below to describe what or who the positive deviants might be, given your particular project scope. This will help you determine the data sources you will need. Remember that in DPPD positive deviants can be individuals, communities, or even administrative or geographic units. Use the boxes below to develop an initial description of potential positive deviant(s).

General Tips:

- In some cases, there might be an agreed definition of who is considered an outperformer. Check the academic literature!
- Engage with people who might be aware of individuals or communities who are adopting uncommon practices and strategies.
- Talk to people working on the problem including sector experts, government officials, or NGO staff.
- If you have access to data on the performance of potential positive deviants and easy access to data analysis skills, run a preliminary analysis to understand the data distribution and to see if you can spot any immediate outliers.

Who am I?   1.

An individual, a group, or a geographical unit? Add other attributes or demographics you can think of (e.g. age, gender, socioeconomic status).

Where am I?   2.

What is my geographic location?

What outcomes can I achieve better than my peers?   3.

Example: I didn't deforest for three years in a row.

Am I known (or likely) to exist?   4.

See the guidance in the bullet points for ways to find out if positive deviants might exist.
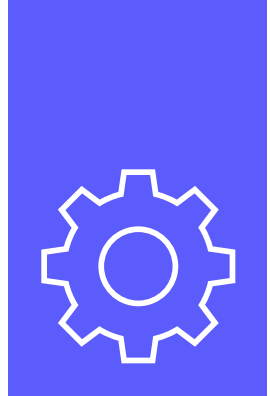
Tool 1.3
Data Mapping

Use the tool to identify potential data sources for each of the three categories. Start with the outcome column then move to the context. Leave the behavioral data category for later, after you have performed initial quantitative data analyses. Use the factors row as an intermediate step to help you brainstorm potential data sources.

	Context	**Behavior**	**Outcome**
Factors	Contextual factors that are beyond the control of individuals and that are hard to change and transfer	Behavioral factors that are within the control of individuals and can be changed and transferred	A measure that can capture the perfomance of the target group to identify the units that were able to achieve better outcomes
Examples	Biophysical, socio-economic, demographic, infrastructure factors	Attributes, practices, and strategies of individuals and groups	Exam scores (pass or fail), crop yields, and vegetation health
Data	Data to control for contextual variables	Data to uncover the uncommon practices and strategies of positive deviants	Data to directly or indirectly measure outcomes
Examples	Both traditional and non-traditional data sources (e.g. official statistics, administrative and environmental data)	Mainly traditional data sources (e.g. surveys and interviews). Very limited use of non-traditional data	Non-traditional data sources that are readily available and accessible (e.g. earth observation data, online data, mobile data, citizen-generated data, sensor data)

Tool 1.4
Capabilities Assessment

It can be helpful to think about how these different skills will be required for the different types of data introduced in the data sub-section: outcome, context, and behavior. The tool provides a process to map existing skills and skill gaps along these three dimensions.

	Context	**Behavior**	**Outcome**
Domain knowledge Do you have sufficient knowledge about the respective domain?	Example: The ability to propose domain-specific contextual variables	Example: The ability to select the most suitable conceptual/theoretical frameworks that could inform the design of the data collection tools	Example: The ability to choose potential proxies and variables that are highly correlated with the outcome measure
Local knowledge Do you have sufficient knowledge about the local conditions?	Example: Knowledge about socioeconomic, cultural, structural and other factors that might affect outcomes	Example: Knowledge about the normative behaviors and practices of positive deviants, existing interventions, current resources, challenges or constraints.	Example: Knowledge about local factors that could influence the selection of the outcome measure
Data skills Do you have the required data skills?	Example: The ability to map, identify and analyze relevant contextual data and dividing/clustering the population into peer groups	Example: The ability to collect and analyze primary data that can uncover predictors of outperformance	Example: The ability to select relevant digital outcome indicators and predicting this outcome measure using contextual factors as inputs

Tool 1.5
Desirability Assessment

Using the tool below, ask yourself whether it is desirable for a positive deviant, should you identify one, to share their practices with others. If so, would the community be able and willing to adopt the practice and benefit from it? Might there be negative consequences on the (social) systems that the positive deviant and the community are embedded in or connected with?

	Positive Deviants	**Communities (Target Group)**	**(Social) Systems**
	A positive deviant has an advantage over others given their knowledge of and application of PD practices. Sharing their practice with others could endanger their competitive advantage.	In the short term, non-positive deviants might not accept, adopt, or embrace the PD practice. In the long term, adoption of the practice could lead to overall negative effects.	The wide adoption of a common practice could have negative consequences beyond the positive deviants and the community (e.g. on governmental and social institutions, the natural environment or political, social or economic systems).
Potential negative consequences	A competitive advantage is more likely to exist in areas where people, groups, or units compete with one another, e.g. a market with a limited number of sellers, and less likely for non-competitive issues (e.g. malnutrition, new-born health or climate resilience).	Factors that hinder acceptance and adoption of practices could be cultural factors or limited capabilities. Long-term negative effects could be overuse of natural resources like overgrazing.	Scaling a particular practice could, for example, result in increased air, water or noise pollution.

Stage 2
Determine Positive Deviants

Technical Terms

Homologues
Units having similar structural factors that are beyond the level of the individual and are generally common to a group

Structural variables
Contextual factors that are above the level of the individual and the social network; includes factors that vary as a function of geographic unit

Dependent variable
The outcome variable being tested and measured and is dependent on the independent variables

Independent variable
The variable that is the predictor of a change in or effect on the dependent variable

Proxy
An indirect measure of the desired outcome which is strongly correlated to that outcome

Ground truth data
Data collected by direct observation as opposed to data provided by inference (i.e. proxy)

Outliers
In statistics, an outlier is an observation that differs significantly from other observations

Cut-off point
The limit used to distinguish positive deviants from non-positive devaints

Univariate/ multivariate outlier detection
In univariate outlier detection techniques, outliers are identified from a single dimension or measure while in multivariate techniques outliers are identified across multiple dimensions

Section overview

This section outlines the different steps of Stage 2, Determine Positive Deviants, which starts with performance measurment, followed by the homogeneous grouping and positive deviant identification, and finally the preliminary validation of the identified positive deviants.

Performance Measurement

After defining the problem and ensuring the applicability of the Data Powered Positive Deviance (DPPD) method comes the stage of looking for the positive deviants. The first step of this stage is to establish a strategy for measuring the performance of the units of analysis in your target group. Essentially, those who will perform high on this measure will be your positive deviants. Take the following requirements into account:

Digital component

The DPPD method derives the performance measure from non-traditional, mostly digital data sources. These include mobile transaction data, remote sensing data, social media data, citizen-generated data, among others. Figure 5 provides an overview of useful digital data sources for the seventeen Sustainable Development Goals (SDGs). You can use the digital data alone or in combination with other, more traditional data sources, like census data or ground measurements.

Figure 5:
Examples of digital data sources that can be used to help achieve the SDGs[1]

1 NO POVERTY
Spending patterns on mobile phone services can provide proxy indicators of income levels

2 ZERO HUNGER
Crowdsourcing or tracking of food prices listed online can help monitor food security in near real-time

3 GOOD HEALTH AND WELL-BEING
Mapping the movement of mobile phone users can help predict the spread of infectious diseases

4 QUALITY EDUCATION
Citizen reporting can reveal reasons for student drop-out rates

5 GENDER EQUALITY
Analysis of financial transactions can reveal the spending patterns and different impacts of economic shocks on men and women

6 CLEAN WATER AND SANITATION
Sensors connected to water pumps can track access to clean water

7 AFFORDABLE AND CLEAN ENERGY
Smart metering allows utility companies to increase or restrict the flow of electricity, gas or water to reduce waste and ensure adequate supply at peak periods

8 DECENT WORK AND ECONOMIC GROWTH
Patterns in global postal traffic can provide indicators such as economic growth, remittances, trade and GDP

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
Data from GPS devices can be used for traffic control and to improve public transport

10 REDUCED INEQUALITY
Speech-to-text analytics on local radio content can reveal discrimination concerns and support policy response

11 SUSTAINABLE CITIES AND COMMUNITIES
Satellite remote sensing can track encroachment on public land or spaces such as parks and forests

12 RESPONSIBLE CONSUMPTION AND PRODUCTION
Online search patterns or e-commerce transactions can reveal the pace of transition to energy efficient products

13 CLIMATE ACTION
Combining satellite imagery, crowd-sourced witness accounts and open data can help track deforestation

14 LIFE BELOW WATER
Maritime vessel tracking data can reveal illegal, unregulated and unreported fishing activities

15 LIFE ON LAND
Social media monitoring can support disaster management with real-time information on victim location, effects and strength of forest fires or haze

16 PEACE, JUSTICE AND STRONG INSTITUTIONS
Sentiment analysis of social media can reveal public opinion on effective governance, public service delivery or human rights

17 PARTNERSHIPS FOR THE GOALS
Partnerships to enable the combining of statistics, mobile and internet data can provide a better and real-time understanding of today's hyper-connected world

[1] Big data for sustainable development: *https://www.un.org/en/global-issues/big-data-for-sustainable-development*

Proxy validation

Much of the non-traditional data that can be used for performance measurement is collected for other purposes. This means any insight into your topic of interest will be indirect — you will seldom get a direct measurement of the performance of your units of analysis. We use the term proxy to define this level of indirection between what is observed in data collected for a different purpose and the actual measure of the phenomenon of interest. Beware that it is crucial to validate proxy measures before using them widely. This usually requires a strong correlation with ground truth data.

In the Indonesia agricultural project, the Enhanced Vegetation Index (EVI)[2], a remotely sensed vegetation index, was used as a proxy for crop yields. A number of studies[3] have previously validated the use of such remote sensing-derived vegetation indices as good proxies for yields, so the team did not need to do the validation themselves. However, had the proxy measure not been previously validated, we would have to use statistical methods to compare the direct and proxy measures[4].

Desirable outcome

The performance measure should reliably capture the desired outcome that is defined or validated by domain experts and, ideally, the target group. As already mentioned in Stage 1, it is crucial to involve the community in the framing of the problem, as they are directly confronted with the problem and likely know what desired outcome they want. The domain experts' main role is then to help you identify and validate the data sources you can use to capture this outcome. The desired outcome, for example, could be farmers achieving healthier vegetation or livestock farms with low deforestation rates.

In the Niger agricultural project, the team used the Soil Adjusted Vegetation Index (SAVI)[5], rather than the EVI used in Indonesia, as an indicator of vegetation health and crop productivity, following the advice of a local domain expert. She told them that SAVI was more suitable for semi-arid areas like Niger, given that it incorporated a soil brightness correction factor.

Multiple performance measures

Depending on how you define your desired outcome, you may have to deal with one or multiple performance measures.

In the Mexico safe public spaces project, the team used three performance measures related to crime severity. The first labelled "Severity 1" was concerned with crimes that threatened life and aimed to damage the physical and sexual integrity of women. The second, "Severity 2," was concerned with physical and psycho-emotional violence that did not threaten life. The third, "Severity 3," was concerned with incidents where physical violence did not occur, but which had a strong psycho-emotional impact.

These categories were used to test whether it is possible to find different positive deviants within and/or across multiple measures. This enabled them to identify a set of positive deviants for "Severity 1," the most severe crimes such as femicide and rape, and another set of positive deviants for "Severity 3," the less severe crimes such as robbery to passerby without violence. This way, it might be possible to identify distinctive factors that are conducive to more or less severe types of crimes when they conduct fieldwork.

---

[2] A remote sensing measure of the state of plant health based on how the plant reflects light at certain frequencies. It is used to forecast agricultural productivity and monitor crop health.
[3] A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation : *https://www.sciencedirect.com/science/article/pii/S016819231400149X? casa_token=ILYSkzsPYGQAAAAA:VluIYoaaELW7eavpHW7fuXjNHmoepKfsYRlVXmUeCweBGpJI6f1_AA6_oVNxQKsUo _ oe_BDvebLI;* Forecasting crop yield using remotely sensed vegetation indices and crop phenology *metrics: https://www.sciencedirect.com/science/article/pii/S0168192313000129? casa_token=f7ktduIZkQMAAAAA:TrWZFvcHLbYD-wJ1DmvvppAPg7Etpd9D-H0Bcxo4FwieBqvUK- 6D_UcKllxNHVIWMVsIuajnBWk*

[4] Statistical considerations in a systematic review of proxy measures of clinical behaviour: *https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-5-20*
[5] A vegetation index similar to EVI and NDVI but it attempts to minimize soil brightness influences using a soil-brightness correction factor. This is often used in arid regions where vegetative cover is low

Dealing with multiple performance measures

Statistical techniques can be used to summarize multiple performance measures into a single index or fewer measures[6]. The simplest way is to develop a weighted average[7] applied directly to the standardized performance measures, forming a single weighted average across all types of measures, but this would require you to determine the weight (importance) of each measure.

Another more complex technique is principal component analysis[8] which replaces the original set of measures with a smaller number of uncorrelated measures that account for most of the information in the original set. The new set of measures are referred to as components. You can decide to use the first component or the first two or more components based on the amount of variance they were able to explain.

Homogeneous Grouping

The second step of Stage 2 is concerned with dividing the target group into homogeneous groups (or clusters) that share a common set of structural factors (e.g. climate, socioeconomic conditions, or geographic proximity). You want to identify Positive Deviance (PD) related to particular practices and behaviors that can be transferred, not to in-group "structural" attributes. In other words, we want to find the units of analysis that perform better in similar circumstances.

To do the homogeneous grouping, you need to determine a set of variables that define the different contexts of your target groups. Check Tool 2.1 in the tools section which can help you map out the different variables that will be used in this stage.

In the Indonesia agricultural project, biophysical factors such as rain and temperature were used to do the homogeneous grouping. The team identified these variables by following the results of both theoretical and empirical research, as well as common sense, that showed they had a significant impact on the outcome measure.

Why is clustering important?

A number of studies[9] have demonstrated how clustering a population into homogeneous groups can improve predictions in the later steps of this stage, where you will try to identify and understand positive deviants within each cluster. This is because you will not need to build models that explain the natural variation between clusters. You will only have to focus on within-cluster variations, which generally increase the model's performance in prediction. Additionally, findings can be extrapolated with more confidence because more detailed and localized information can be extracted from homogenous groups because they share similar conditions.

Moving from this set of structural variables to homogeneous groups requires some form of clustering. This can be done manually, based on professional experience and intuition, or it can be done through unsupervised machine learning techniques such as cluster analysis[10]. Cluster analysis is a way of grouping, or clustering, units in relation to certain variables. In your case, these will be the structural variables. The prime objective of cluster analysis is to identify clusters that are internally homogeneous and heterogeneous between them. In other words, the point is to minimize the variance of structural factors within homogeneous groups and to maximize the variance between the groups.

---

[6] Multidimensional poverty measurement and analysis: Chapter 3 – Overview of methods for multidimensional poverty assessment: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2564782
[7] How to calculate weighted average in 3 steps (with example): https://www.indeed.com/career-advice/career-development/how-to-calculate-weighted-average
[8] Principal component analysis: https://en.wikipedia.org/wiki/Principal_component_analysis

[9] The Utility of Clustering in Prediction Tasks: https://arxiv.org/abs/1509.06163; Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions: https://web.cs.wpi.edu/~nth/pubs_and_grants/papers/2011/AIED/Clustering%20Students%20to%20Generate%20an%20Ensemble.pdf
[10] Data Science for Business and Decision Making, Chapter 11: Cluster Analysis: https://link.springer.com/chapter/10.1007/978-3-642-21869-9_49

Figure 6 is an illustrative example of clustering observations according to two variables: income and age. There are four main clusters. Each cluster represents individuals that fall into similar age and income groups, each individual being closer to others in the same group than to others in other groups. You might also be able to find predetermined homogeneous groups or clusters that you can reuse for your DPPD project.

In the Niger agricultural project, the team found a map[12] developed by the Adapt'Action Facility that divided Niger into agro-ecological zones with similar biophysical, ecological, and climatic conditions (Figure 7). The team decided to use these zones instead of creating their own since they took into account valuable local and contextual information they could not get access to.

Figure 6:
Illustrating the variability within the clusters and between them[11]
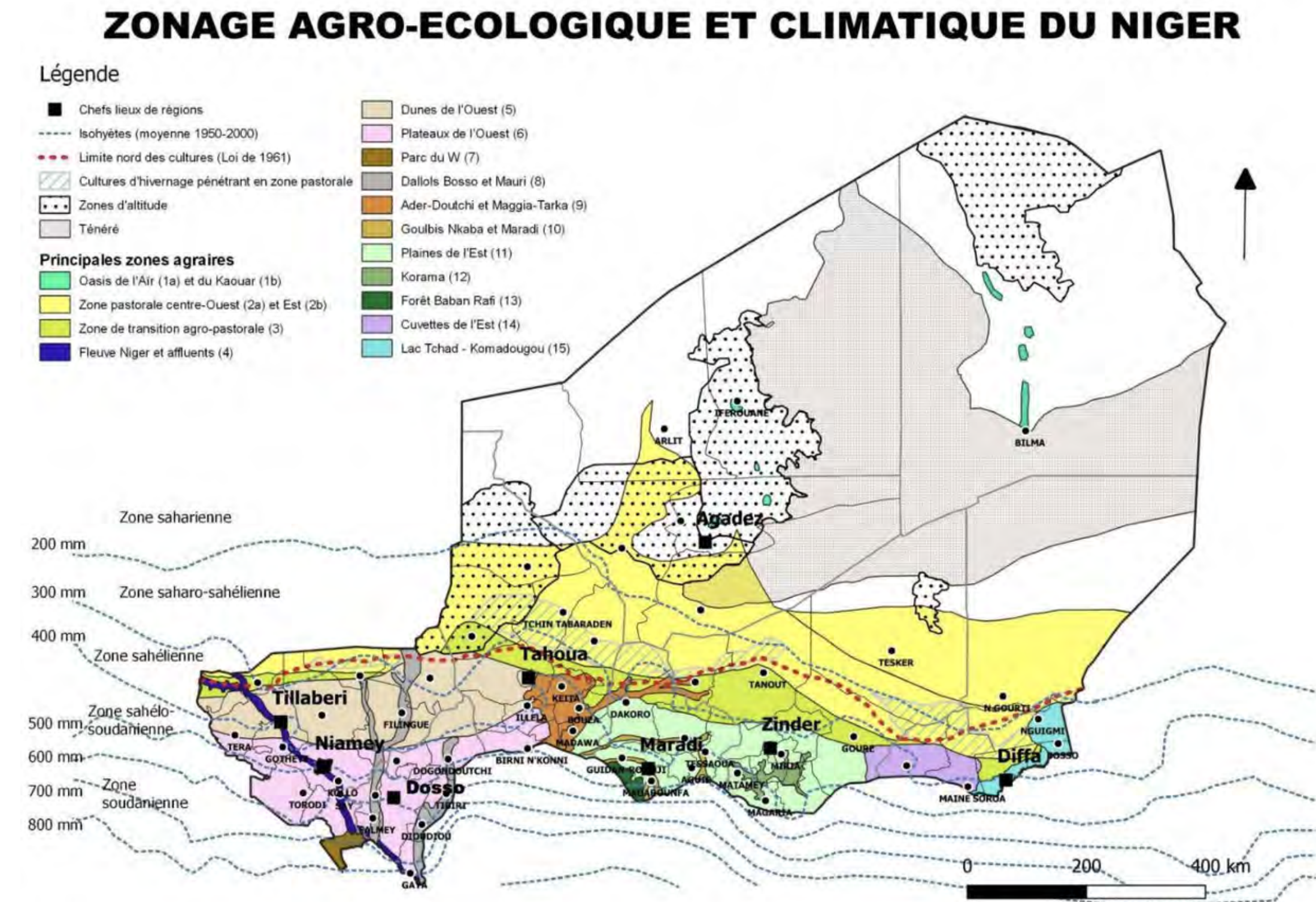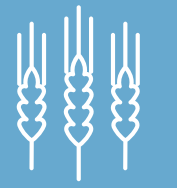
Figure 7:
Niger agroecological zones

---

[11] Data science for business and decision making, chapter 11: Cluster analysis: https://www.sciencedirect.com/science/article/pii/B9780128112168000112#ab0010

[12] Identification and evaluation of farming options climate-smart priorities for adaptation to climate change in Niger: https://bibliosud.omekas.mind-andgo.net/files/original/048098db5a4623fe2932755ca73a5d59e7aeea78.pdf

Additional methods for clustering

The COVID-19 pilot in Germany used yet another strategy to ensure fair comparisons between homologues. Instead of sorting every district into a static set of clusters, each district was assigned its own group of peer districts, based on the euclidean distance in the geometric space defined by relevant structural factors. The districts' raw deviances were then standardized relative to each individual district's peer group, by subtracting the peer group's mean and dividing by its standard deviation.
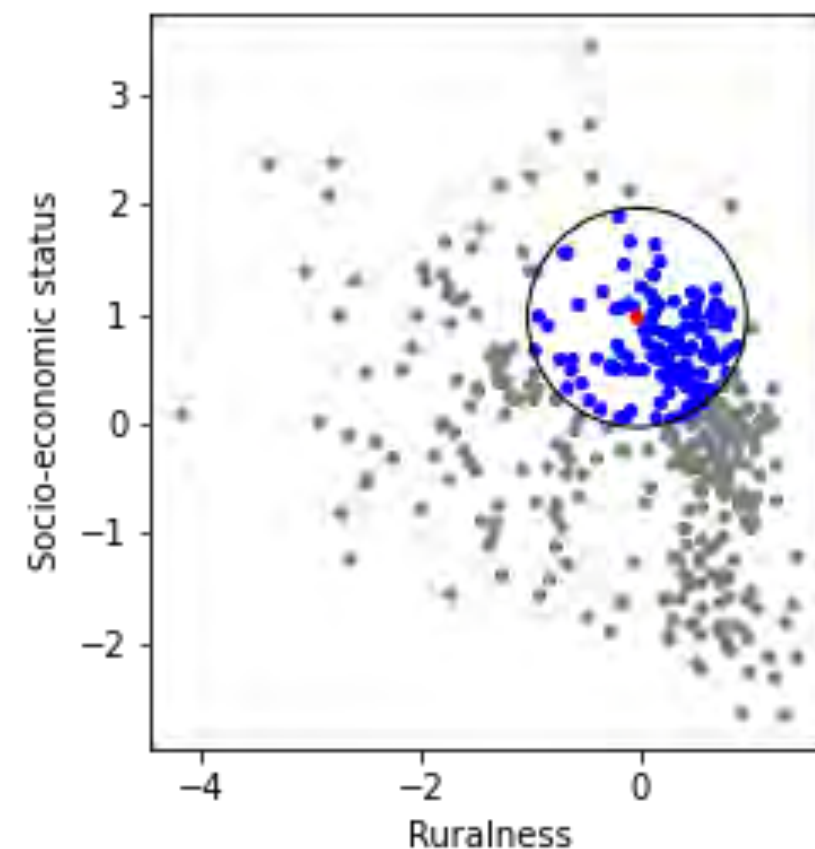
These means and standard deviations were calculated with the focal district included. (red: focal district, blue: included in peer group, grey: not included).

Positive Deviant Identification

The third step of this stage consists of identifying outliers, your positive deviants, within each homogeneous group. This requires defining the techniques and cut-off points that distinguish positive deviants from non-positive deviants. Depending on your performance measure, which we refer to here as the dependent variable, there can be several ways to do this.

Univariate analysis

Only one variable, the performance measure, is used to identify positive deviants. This variable can be categorical (i.e. pass/fail, win/lose, or healthy/sick) or continuous. If it is categorical, then positive deviants are simply those who succeed when a majority of others fail. If the performance measure is continuous, then the way to identify positive deviants depends on the distribution of performance within homogeneous groups. If performance can be assumed to follow a normal distribution, then positive deviants are those who lie at the extreme high end of the distribution. In this case, a typical cut-off point can be two standard deviations from the mean. If no assumption of normality can be made about the distribution of performance, different types of extreme value analysis can be used. These generally deal with extreme deviations from the median.

Proximity-based models can also be used. These define outliers as observations that are far from dense clusters in the data. For example, clustering methods like K Means, segment the data into clusters, and positive deviants can be defined as observations that lie far away from important clusters[13]. Similarly, the Interquartile Range (IQR) method segments the data into quartiles, and positive deviants can be defined as observations that lie beyond 1.5*IQR[14].

In the Egypt research publication case study, positive deviants in each citation metric were identified using the IQR. Researchers who had citation metrics scores that are beyond 1.5*IQR were considered positive deviants.

Multivariate analysis

Multiple variables are needed to confidently identify positive deviants. In addition to the performance measure, these variables usually account for contextual variations among units of analysis that belong to the same homogeneous group — they need to be controlled for (i.e. to reduce their effect). Here you will typically model the performance measure (dependent variable) as a function of the structural factors (independent variable).

---

[13] How to detect outliers using parametric and non-parametric methods: *https://clevertap.com/blog/how-to-detect-outliers-using-parametric-and-non-parametric-methods-part-ii/?utm_source=ref_article_outlier1*
[14] What is the interquartile range rule?: *https://www.thoughtco.com/what-is-the-interquartile-range-rule-3126244*

If the performance measure is categorical, then positive deviants can be identified using probabilistic models like a logistic regression in which they are "false negatives," observations that, based on the independent variables, are expected to fail but, in fact, succeeded.

If the performance measure is continuous, then positive deviants can be identified through other types of regression analysis. The models obtained can be used to predict the performance of the units of observation, and this prediction can be compared to the actual, observed performance. The "error" that distinguishes predicted from observed can then be run through an extreme-value analysis to identify positive deviants.

In the Ecuador cattle-farming project, the team developed a model that predicted farm deforestation rates as a function of farm cattle density, size, soil adaptability, and type of land use. Positively-deviant farms were then identified using the residual values, the difference between predicted and observed deforestation rates. In the Germany COVID-19 project, the team predicted weekly reproduction factors using prior infection spread, mobility, and weather. These weekly values were then compared to the weekly observed reproduction rates, yielding weekly deviances. After peer-group standardization based on the groups they belong to, positively-deviant districts were identified as the top three districts with the highest average deviance across one of two study timeframes.

A posteriori expectation

The cutoff for PD is already established or known and a posteriori expectation of an outcome exists. Here, positive deviants exceed this with some level of magnitude. For instance, according to the International Union for Conservation of Nature, species under threat are defined as some level of decline for three generations or over 30 years. A positive deviant, in this case, would be a population of a species whose size is increasing or stable for three generations or more when the species as a whole is declining rapidly.

Exceptional responders

Exceptional responders are units that respond better to interventions than what was expected at the onset. An example would be an intervention to protect forests. Forest cover could be measured both inside and outside a protected area. The difference between the inside and outside can be used to generate an average expected effect of protection and positive deviants would be the protected areas significantly exceeding the expected effect.

This can be done using the difference in differences method where positive deviants would be the units having the largest difference in differences[15]. Understanding positive deviants as exceptional responders can be particularly interesting when it comes to exploring why certain communities responded better to an intervention than others and which underlying factors contributed to this "better response." This could then be used to inform the design of future interventions.

Exceptional responders in tiger conservation

This study[16] evaluated the effectiveness of tiger reserves within India using the PD approach. The opinions of three tiger experts were used to generate a list of seven tiger reserves classified as successful (or exceptional responders) and five reserves as failures. Expert opinion was also used to identify any key individuals that garnered widespread support for tiger conservation at any of the identified reserves. They found two uncommon insights:

- Tiger conservation success is achievable even within a densely populated human landscape where a high percentage of the population is involved in agriculture
- The presence of "conservation champions" can dramatically affect the performance of individual reserves and have positive outcomes for tiger conservation

---

[15] An intuitive introduction to difference-in-differences: *https://www.youtube.com/watch?v=J7q2H8aB8bQ*

[16] Comparative evaluation of tiger reserves in India: *https://link.springer.com/article/10.1007/s10531-013-0554-9*

Positive Deviant Validation

The last step of this stage consists of initial validation of the potential positive deviants you identified. While field research is necessary for full confirmation of their deviance, there are intermediate ways to validate whether observed deviance in the data is simply due to random noise or whether it is based on actual signals of outperformance. Here are several ways to do so.

Time series analysis

Evaluate the performance of positive deviants over time and check whether they consistently outperform their peers. The timeframe will vary according to the focus of the project, and selecting it can require expert opinion.

In the Indonesia agricultural project[17], the team conducted a time-series analysis to see whether the performance of rice farming villages was independent of climatic patterns. They developed a model to predict the EVI at the village level as a function of precipitation and temperature in 2013, where they used historic climate and EVI data from 2000 to 2012 to train the model. The observed performance of positively-deviant villages was significantly higher than the observed performance of non-positively-deviant villages. This implied that outlier villages had likely adopted specific approaches and practices that others had not, and had established production systems that delinked climatic patterns and productivity. This provided an initial validation of them being potential positive deviants.

Local knowledge

Reach out to community leaders, government officials, local domain experts, and development practitioners who are engaged in activities, projects, or services related to the targeted areas before doing the field research. Sharing with them the initial set of potential positive deviants can lead to possible explanations for outperformance that may have been overlooked, and could bias the results without reflecting the actual practices you are interested in. Explanations such as the existence of development interventions in positively-deviant areas (i.e. external support) or having better access to resources, like irrigation points.

Triangulating different data sources

Check whether possible predictors of PD can be identified through official statistical data, like a census or survey, or other sources of data, such as high-resolution imagery, covering the targeted areas. Although the PD approach is mainly concerned with identifying grassroots practices, which are seldom captured through such data, some higher-level management practices can be distinguished. If their association with positively-deviant units is in accordance with existing literature and local knowledge, this can count as means of validation in itself.

---

[17] Identifying potential positive deviants across rice-producing areas in Indonesia: https://www.unglobalpulse.org/document/identifying-potential-positive-deviants-pds-across-rice-producing-areas-in-indonesia-an-application-of-big-data-analytics-and-approaches/

In the Somalia rangeland project, the use of the remote sensing datasets was complemented by the use of open-source, high-resolution imagery available from Google Earth for pre-fieldwork visual inspection. The imagery was used to rule out false-positive deviants whose vegetation scores were inflated by interventions (e.g. government reserves), and to look for early signs of pastoral and agro-pastoral activities, visible soil and conservation techniques, and other rangeland management practices. Through this remote inspection, the team identified patterns indicating the existence of soil and water conservation techniques at a number of potential positively-deviant communities. Figure 8 presents some of those interventions.

Figure 8:
Examples of soil and water conservation techniques. On the left, there is a shrub barrier in the frontline with soil erosion to limit its crawling. On the right, half-moon techniques to reduce water run-off.

Similarly, in Niger, where the team was looking for positive deviants cultivating purely rain-fed crops, they visually inspected all potential positive deviants. They excluded those that were close to river basins, as these are a source for irrigation and thus crops were not purely rain-fed.

Different techniques for outlier identification

Experiment with various methods for outlier detection (e.g. regression vs. neural networks, univariate vs. multivariate) to see if there are outliers that persistently appear across the different methods.

In the Ecuador cattle-farming project, the team modelled deforestation rates using yearly predictors and inter-annual variations in predictors. When doing the fieldwork, priority was given to positive deviants identified across multiple approaches.
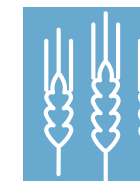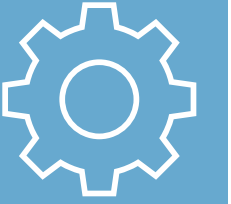
Tool 2.1:
Variable Mapping

This tool will help you in mapping out all the possible control and independent variables in order to select the most suitable variables for the homogeneous grouping and the positive deviant identification steps of Stage 2 and the positive deviants' characterization step of Stage 3.

The X-axis shows to what extent a variable is behavioral and could be transferable, at the level of the individual or community. The further to the right, the more effective it is to recommend this practice, strategy or behavior to other members belonging to the same homogeneous group.

The Y-axis shows to what extent the independent variable is structural or supra-individual (i.e. above the level of the individual). The more supra-individual (structural) and less transferable the variable is, the more suitable for homogeneous grouping it becomes; the less structural and more transferable a variable is, the more suitable it becomes for characterizing positive deviants and understanding their underlying behaviors.

Using the guidance provided below you will be able to categorize and map the study variables into four main types:

- **Homogeneous grouping:** Factors beyond the control of individuals that can influence the outcomes of the target group, and at the same time they cannot be changed e.g. rain.

- **PD identification:** Individual-level factors that can influence the outcomes of the target group, and at the same time they cannot be changed. Those variables should be controlled for when identifying positive deviants within each homogeneous group or cluster e.g. socio-economic status.

- **PD characterisation** (Supra-individual): Factors beyond the control of individuals that can influence the outcomes of the target group but can be transferred to peer communities e.g. policies.

- **PD characterisation** (Individual): Individual-level factors that can influence the outcomes of the target group and can be changed and transferred e.g. soil and water conservation techniques.

How structural is is this variable?

Supra-individual

Homogeneous Grouping

Examples: rain, temperature, elevation, soil, land cover, community assets such as irrigation

PD Characterization

Examples: policies, successful development interventions, governance schemes

Individual

PD Identification

Examples: race, gender, socio-economic status, age, individual assets

PD Characterization

Examples: soil and water conservation techniques, sustainable cattle ranching practices, attitudes towards conservation, income diversification, saving groups

Non-transferable   Transferable

How transferable is this variable?

Stage 3
Discover Underlying
Factors

Section overview

This section outlines the different steps needed to discover positively-deviant underlying factors. It follows the Determine Positive Deviants stage which resulted in a list of potential positive deviant units that will be targeted for the Positive Deviance inquiry. The inquiry starts with fieldwork planning, followed by data collection, and ends with the data analysis step.

Technical Terms

Positive Deviance inquiry
Finding uncommon but successful strategies and behaviors that can be shared and acted upon by others in the community

Deductive reasoning[1]
Deductive reasoning, or deduction, starts out with a general statement, or hypothesis, and examines the possibilities to reach a specific, logical conclusion[1]

Inductive reasoning[1]
Inductive reasoning makes broad generalizations from specific observations: make many observations, discern a pattern, make a generalization, and infer an explanation or a theory

Conceptual framework
It defines the relevant variables for your study and maps out how they might relate to each other

Study design
The set of methods and procedures used to collect and analyze data on variables specified in a particular study

Qualitative research
A research approach that relies on data obtained by the researcher from first-hand observation, interviews, focus group discussions and participant-observation

Quantitative research
A research approach that focuses on quantifying the collection and analysis of data. It is used to find patterns and averages, make predictions, test causal relationships, and generalize results to wider populations

Mixed methods
The type of research which combines elements of qualitative and quantitative research approaches

Community-based participatory approaches
An approach to research that involves collective, reflective and systematic inquiry in which researchers and community stakeholders engage as equal partners in all steps of the research process

---

[1] Deductive reasoning vs. inductive reasoning: *https://www.livescience.com/21569-deduction-vs-induction.html*

Fieldwork Planning

By now, you should have identified a list of potential positive deviants. In the first step of Stage 3, you will prepare all the instruments necessary for conducting the fieldwork for the Positive Deviance (PD) inquiry.

The goal of the fieldwork is twofold:

- To validate if the potential positive deviants are indeed true positive deviants
- To uncover the underlying factors responsible for their deviance

You should expect to consult with a variety of stakeholders, all of whom should have something to do with your performance measure. Consider a stakeholder anyone who has an indirect or direct relationship with your units of analysis and could be part of the problem or the solution. Start by asking yourself who these stakeholders are. Get familiar with their social and cultural environment, and keep them in mind when approaching them. For further guidance check Tool 3.1 in the tools section.

In the Ecuador cattle-farming pilot, the unit of analysis was cattle-raising farms. The field inquiry targeted farm owners, workers, families, farmer associations, and community leaders.

Conceptual framework

Before developing your data collection tools, it is important to read any background literature and understand the conceptual frameworks that others may have developed and used when investigating similar topics. Are there any dominant theories or models that are used? Are there any approaches or variables that have not yet been taken but which you or any of the stakeholders think could add something new? Can you combine elements of different approaches which could help you better understand your positive deviants? Look for frameworks that are based on a thorough understanding of the studied topic and are aligned with existing literature.

We recommend discussing your conceptual framework with relevant stakeholders and the colleagues who were involved in the previous quantitative stage to make sure you did not miss anything, such as the control variables that were used in positive deviant identification that might require validation.

Study design

After mapping out the different stakeholders, ask yourself how you are going to collect the information you need to validate your positive deviants and uncover the underlying factors of their deviance. There are generally two approaches: a qualitative approach (e.g. interviews) and a quantitative approach (e.g. surveys). But you can also use a mix of both (i.e. mixed-methods). The data collection section presents the methods you can use for each of the approaches. We generally encourage a mixed-methods approach. It enables you to first gain highly specific and valuable insights from a small, information-rich sample of positive deviants. These will help you generate hypotheses about the individual, cultural, social, and structural factors that are likely to influence deviance, which you will validate through more structured investigations with a larger sample including both positive deviants and non-positive deviants.

Figure 9 proposes a mixed-methods study design. Note, however, that in cases where you only have a small number of potential positive deviants, or where you only have limited resources to conduct large-scale surveys, you may prefer an entirely qualitative approach. Conversely, when doing retrospective studies using secondary data sources, or when it is difficult to have face-to-face engagements with stakeholders, you may prefer an entirely quantitative approach.

In the Egypt research publication case study, positively-deviant researchers were first interviewed to generate hypotheses about the basis for their performance. Quantitative data was then collected from both positive deviants and non-positive deviants to validate those hypotheses, and to identify significant differences between both groups[2].

---

[2] Publication outperformance among global South researchers: An analysis of individual-level and publication-level predictors of positive deviance: *https://link.springer.com/article/10.1007/s11192-021-04128-1*

Step 1:
Gather data from positive deviants

Step 2:
Look for patterns

Step 3:
Hypothesis generation

Inductive Reasoning

A social phenomenon is observed. In our case, this social phenomenon will be PD and we are going to focus on observing and collecting data from positive deviants and relevant stakeholders using qualitative approaches.

After collecting the data and observations, step back and get a bird's eye view of the data in search of patterns.

Infer tentative hypothesis and potential PD explanators from this data which will be validated through deductive reasoning.

Use the hypothesis generated from the inductive reasoning together with established theories from prior research to develop the quantitative data collection tools.

Collect data from positive deviants and non-positive deviants to compare both groups. You can also use digital data sources to generate and test additional predictors of performance.

Analyse the collected data to identify significant predictors of positively-deviant performance.

Deductive Reasoning

Step 4:
Hypothesis

Step 5:
Collect data from (non)-positive deviants

Step 6:
Test hypothesis

Note that the first step of Figure 9 could include non-positive deviants to establish an understanding of the population's normative and common behavior before interviewing positive deviants. This could make it easier to identify uncommon practices and strategies among positive deviants.

Figure 9:
DPPD study design

Sampling

When doing qualitative and quantitative data collection, it is important to properly determine your sample size. Unfortunately, there is no golden rule for this. For the qualitative investigation, it is a good idea to start with 20-30[3] units and to grow your sample size if necessary until you reach data saturation[4]. For the quantitative data collection, you can target a sample size of around 100 units for each major sub-group (i.e. non-positive deviants) in the population, and between 20 to 50 units for each minor sub-group (i.e positive deviants)[5].

Sampling tools

There are tools[6] that suggest the ideal sample size based on the confidence level, confidence interval, and the population size. However, such tools should be used with caution, as they enable inferences on entire populations, not subgroups within the population. That said, for comparing the prevalence of certain practices between two groups, positive and non-positive deviants, there are also case-specific sample size tables that can be used, as shown in this guide[7].

It is also important to ensure diversity in the selection of the sample units — make sure your sample covers the different attributes of the units that are being analyzed.

---

[3] Sample size for qualitative research: *https://www.emerald.com/insight/content/doi/10.1108/QMR-06-2016-0053/full/html#loginreload*
[4] Data saturation is the point at which no new information or themes are observed in the data from the completion of additional interviews or cases
[5] The sampling issue in quantitative research: *https://files.eric.ed.gov/fulltext/EJ919871.pdf*
[6] Sample size calculator: *https://www.surveysystem.com/sscalc.htm*
[7] Sample size guide: *https://www.researchgate.net/profile/Ankit-Gaur-9/post/What-is-the-Sample-Size-above-which-the-data-is-assumed-to-have-normality/attachment/5b2b370bb53d2f892896fbcf/AS%3A639830115495943%401529558795646/download/samplesizecalculation.pdf*

In the Ecuador pilot, a convenience sampling was carried out that achieved diversity in the farms' animal load, pasture (%), forest (% and ha), poverty level (index), access to roads, and proximity to the main city. The team used a qualitative study design in which the key informants were farm owners or those in charge of the farm.

Things to consider

Concept communication

What message do you want to bring along? Why would your target group be willing to achieve the desired outcome? How would they benefit from applying the positively-deviant practices? Think of a narrative that will encourage communities to engage.

Incentive model

When conducting a quantitative investigation, you might face the problem of getting enough responses. Think of ways to encourage respondents to fill out your survey without having a negative impact on the data quality. For example, consider getting approbation through community leaders. Convincing people who are influential in the community to participate in the study can incentivize other community members to participate.

Gatekeeper

Somewhat related to the previous point, are there individuals or entities who could act as shortcuts to interviewees and respondents?

Ethical considerations

The PD inquiry is based on a deep respect for the community, its members, and its culture and focuses on interactive engagement and the capacity to let the community lead. Make sure you take that into account when you reach out to communities. Who will collect the data and how? Will the data expose stakeholders to potential harm? How might you mitigate the risks? For further guidance check Tool 3.2 in the tools section.

Resources

What resources do you have (financial, time, and personnel) for reaching out to communities? And for collecting and analyzing the data you will gather? Reflect on how this might affect your choice of study design and research methods.

Data Collection

In the second step of Stage 3, you will collect the data needed to identify factors that favor PD. While the conventional PD approach focuses mainly on individual-level factors, the Data Powered Positive Deviance (DPPD) method enables you to take a more systemic lens that also considers structural-level factors like infrastructure, public policies, or social dynamics. This should help you gain a more comprehensive understanding of the complex forces at play behind positively-deviant practices or solutions.

It is important to design your data collection instruments in such a way as to capture this mix of both individual and structural factors. There are several methods you can use, and the choice of participants to include in each method depends on your study design. If you follow a mixed-methods approach, as recommended in the previous step, you should include positive deviants in the initial qualitative data collection. In the following quantitative data collection, you should include the larger target group (having both positive and non-positive deviants). If you follow either an entirely qualitative or an entirely quantitative approach, you should include both positive and non-positive deviants simultaneously. Some of the qualitative and quantitative methods that can be used for the data collection are outlined below.

Qualitative methods

The main benefit of qualitative methods is that the information you gather is richer, and therefore it generally provides deeper insights into the phenomenon you are investigating. It might also help you understand the interaction between relevant factors. Examples of common qualitative methods include:

- Individual interviews: Interview your positive deviants to identify any uncommon attributes, attitudes, and/or practices that might distinguish them from non-positive deviants
- Focus group discussions: Have a collective discussion with a group of positive and non-positive deviants
- Observations: Observe positive deviants' practices closely to understand what they do and how they do it

In the Ecuador cattle-farming project, 19 farmers — 9 positive deviants and 10 non-positive deviants — were interviewed at their farms. The interview schedules contained closed-ended questions and observational checklists to capture contextual, demographic, and socioeconomic variables, in addition to variables derived from the conceptual framework. Open-ended questions were also asked to uncover uncommon strategies, attitudes, and practices that positive deviants may have by comparing them with those of non-positive deviants.

Community-based participatory research

Community-based participatory research is also commonly used in the conventional PD approach, as it recognizes the importance of including community stakeholders in the different steps of the PD inquiry. Practice shows this involvement mobilizes people, creates buy-in, increases knowledge, and changes attitudes. The researcher or practitioner acts as a facilitator who creates a space for the community to elicit and share knowledge from and for action. Examples of common participatory tools include:

Discovery and action dialogues

This type of dialogue is central to the PD approach which focuses on the solution rather than the problem[8]. The aim of this method is to ensure that in the presence of a facilitator, people in the group, unit, or community elicit by themselves the positively-deviant practices. The dialogues also create favourable conditions for stimulating participants' creativity in spaces where they can feel safe to invent new and more effective practices. Resistance to change diminishes as participants are given the freedom to choose the practices they will try out and the problems they will tackle. Further, social proof increases the likelihood that solutions will be adopted. The discovery and action dialogues (DADs) make it possible to achieve frontline ownership of solutions. It acts as a tool that does both, discovering positively-deviant practices and mobilizing communities to take action.

Employing DADs in the PD approach

This study[9] used PD in the control of healthcare-associated infections (HAIs) due to methicillin-resistant staphylococcus aureus (MRSA) in a Colombian hospital with the aim of reducing HAI rates through a cultural change in processes. They used DADs in the presence of a facilitator, a group of around eight positive deviants and other non-positively deviant participants to discuss the identified problem and how barriers may be overcome to find a solution. The group made proposals and commitments and identified specific actions to be taken forward as interventions. The most feasible and low-cost economic proposals were selected for implementation.

---

[8] Check this guide on how to do DADs: *https://www.liberatingstructures.com/10-discovery-action-dialogue/*

[9] Using Positive Deviance in the prevention and control of MRSA infections in a Colombian hospital: a time-series analysis: *https://www.emerald.com/insight/content/doi/10.1108/QMR-06-2016-0053/full/html#loginreload*

Participatory sketching

Participatory sketching is a method of collective drawing employed to obtain enriched narratives. Participants are divided into subgroups that jointly draw a sketch that describes what they envision as good practice or an ideal model in a physical space. It is mainly used when visual aids are required to identify positively-deviant practices. Sketches are then shared and discussed with the entire group. This study[10] used PD to inform the design of an anti-triatomine prototype house by identifying knowledge, attitudes and practices used by families that have remained triatomine-free. In this study, participants were divided into subgroups to jointly draw a sketch describing what they envision as a healthy environment at home in intra and peridomestic areas. Sketches were then presented to the entire group.

Photo elicitation

Photo elicitation is a method used in visual anthropology that introduces pictures to elicit comments. Pictures taken during the interviews with positive deviants are presented to focus groups and participants are asked to reflect on the captured practices and solutions.

Community mapping

Community mapping describes a process where the community gets together to map their own assets, values, beliefs, spatial units of interest, or any other self-selected variable[11]. This method is about mapping "by the community for the community, using relatively informal processes. Community mapping has been used to identify positively-deviant households, the location of a community's resources and infrastructures, or where the most vulnerable families in a community live. The method can also be used to better understand and identify spatial factors contributing to the observed deviance.

Data-driven participatory approaches

These include methods that engage community members in interpreting the data that were collected about them (in Stage 2) to encourage dialogue and debate around the challenges they face and possible means to address them. This shifts community members from a passive role of producers of data to an active role of investigators[12].

General tips for qualitative methods

- During the PD inquiry stage, focus on what positive deviants do and also on how and why they do it.
- Interview potential positive deviants as well as others around them, anyone who might have an influence on the deviance (community members, family members, community leaders).
- Usually, interviews and focus group discussions are conducted in parallel. The themes uncovered in the focus groups can then help orient the interviews.
- Focus groups and participatory activities should include between 6 and 10 people, as smaller groups may limit the amount of information collected, and more may make it difficult for all participants to participate and interact. Bigger groups will also make it harder for the interviewer/facilitator to make sense of the information gathered.
- Focus groups are better suited when you have limited resources (time, manpower, money). Greater insights might be developed from the group dynamics, and the phenomena being researched requires a collective discussion in order to understand the circumstances, behaviors, or opinions.

---

[10] Positive deviance study to inform a Chagas disease control program in southern Ecuador:
https://www.scielo.br/j/mioc/a/qMBfF39f568Wm8gyB3YJmbH/abstract/?lang=en
[11] Check this guide on how to do community mapping:
https://web.archive.org/web/20160415103218/http://www.positivedeviance.org/about_pd/Mappingweb.pdf

[12] A few examples of such methods are provided in this resource: *https://medium.com/data-empowerment/three-examples-of-data-empowerment-5f3e964ffbdc*

Quantitative methods

The main benefit of quantitative methods is that the data you collect can be analyzed statistically. Such methods are generally used to test hypotheses, so it is important to identify a suitable conceptual framework that incorporates the hypothesis generated from the qualitative inquiry, in addition to established theories in the literature that might explain the investigated phenomenon. You will use the framework as the basis for questionnaire development. Examples of common quantitative methods include:

- Surveys: Collect structured data from both positive and non-positive deviants to identify statistically significant differences between both groups.
- Observation checklists: Lists of things that the observer will look at when observing positive and non-positive deviants.

The nature of the DPPD method, which leverages digital data in the PD approach, provides an opportunity to use such data, not just for positive deviant identification but also for understanding their underlying behaviors and practices. While this is not possible in most cases, it is still important to ask, "Are there digital traces that can shed light on positively-deviant behaviors and practices?"

In the Egypt research performance case[13], which looks into outperforming positively-deviant researchers in a resource-constrained country, the publications of those researchers were analyzed to identify paper extrinsic features (e.g. a number of pages), intrinsic features (e.g. topics covered), and publication outlets (e.g. where do they publish their research) that could shed light on their publication strategies and tactics.

General tips for quantitative methods

- Before you build tools from scratch, always look for existing questionnaires, checklists, or tests that can be adapted to your specific investigation.
- Tools should always begin with an introduction outlining the purpose of the data, collection efforts, the ethical norms that govern the use of the tool and the data collected, and how the data will be used.
- Before you use the tool, make sure to follow these steps:
  a) Translate the tool into any local languages required
  b) Train researchers and test the tool
  c) Revise and finalize your tool based on the results of your testing
  d) Plan for implementation and data management
- Keep it short and simple.

---

[13] Publication outperformance among global South researchers: An analysis of individual-level and publication-level predictors of positive deviance: *https://link.springer.com/article/10.1007/s11192-021-04128-1*

Data Analysis

The main goal of this step is to identify significant attributes of positive deviants. The data analysis techniques you use will depend on your study design and whether you have used a qualitative approach, a quantitative approach, or a mixed-methods approach. Specific analytical techniques for each approach are outlined below.

Qualitative data analysis

Thematic analysis

This type of analysis starts with a verbatim transcription of all recorded interviews and focus group materials. The transcriptions are then analyzed to identify themes that capture recurring patterns in attributes, attitudes, practices, and strategies of positive deviants[14]. You can also bring a quantitative lens to this kind of qualitative data through open and closed coding processes. Through these, you either use an existing codification of topics and themes or create your own before coding your transcriptions and counting the number of times different themes appear across the two groups of positive deviants and non-positive deviants.

In a mixed-methods approach, the themes identified are used to develop a survey instrument that is grounded in the views of the participants. Its main objective is to validate the qualitative findings (uncommon factors that contribute to positive deviance) using a large, representative sample of the population.

In the Ecuador cattle farming project, the team conducted a thematic analysis of interviews using the Dedoose software[15] to identify differences between positively-deviant and non-positively-deviant farms.

The central themes that emerged included the typology of the farm, farmer characteristics, farm characteristics, enablers and barriers of deforestation, attitudes towards conservation, practices and innovations.

Positive deviance accessibility sieve

In this case, findings from interviews with positive deviants are presented to a large group of non-positive deviants, where potential uncommon practices are presented one by one to the entire group. Participants are then asked to point out those that are known to them by a show of hands. This method is mainly used to uncover the distinguishing practices and behaviors of positive deviants. Photo elicitation can also be used here.

Accessibility sieve in action

This study[16] used PD to improve student clinical performance. They employed the PD accessibility sieve to identify behaviors that were truly uncommon but replicable. Here, individual in-depth interviews were first conducted with 20 positively-deviant students who were asked what they thought were the reasons for their good performance. The transcribed responses were then analyzed to list potential behavioral predictors of their outperformance.

In a large group session with the entire batch of students, these behavioral predictors were displayed one by one to the students, and they were asked to point out behaviors that were common to many other students by a show of hands. A list of behaviors was compiled, eliminating those behaviors that were voted to be common for the batch; behaviors of seven students were found to be truly novel.

---

[14] Check this resource on how to do a thematic analysis: *https://www.interaction-design.org/literature/article/how-to-do-a-thematic-analysis-of-user-interviews*
[15] Dedoose software: *https://www.dedoose.com/*

[16] Change in action: using positive deviance to improve student clinical performance: *https://link.springer.com/content/pdf/10.1007/s10459-011-9301-8.pdf*

Quantitative data analysis

Descriptive statistics

These are used as a first step to quickly compare results from different groups: positive deviants, non-positive deviants, average performers, or negative deviants. This provides a basic overview of the data (e.g. mean, standard deviation, min and max).

Inferential statistical tests

They are used to identify significant differences between the means of the analyzed groups for each variable using statistical tests like student t-test, Mann Whitney, and Fisher exact test when comparing two groups and tests like ANOVA, Kruskal–Wallis and Chi-square when comparing three or more groups (e.g. PDs, average performers, and NDs)[17].

---

[17] Check this guide for further guidance on which test to use:
https://www.scielo.br/j/mioc/a/qMBfF39f568Wm8gyB3YJmbH/abstract/?lang=en

Regression analysis

This analysis is used to examine the relationship between the captured performance measure (dependent variable) and the independent variables. Regression analysis enables you to estimate the effect of each independent variable on the dependent variable while holding all the other independent variables constant. In this stage, the independent variables will include both the structural variables used in the positive deviant identification step and the sociodemographic and behavioral variables captured in the data collection step. For each homogeneous group, you will have a separate model to identify significant predictors of performance that are relevant to the respective group.

Cluster-then-predict

You might consider regressing the performance measure using different subsets of the independent variables. For example, modeling the performance measure as a function of structural variables that are beyond the level of individuals, modeling separately behavioral predictors of performance, then creating a full model containing both subsets.

This way you might be able to understand the effect of each subset of variables separately which could guide the design of the interventions. If structural variables have more influence, then you might consider a systematic policy-focused intervention. If the behavioral variables stand out then you should consider a community-focused intervention. If both have significant effects, then you should consider an intervention combining both.

Tool 3.1:
Fieldwork Design Canvas

This tool will help you cover the different aspects of fieldwork planning, starting from the characterization of the identified positive deviants and mapping the relevant stakeholders to determining the suitable data collections tools and sampling strategy.

Positive deviants' profiles 1.

What are the characteristics of the identified potential positive deviants? Where are they located? Which homogeneous groups or clusters do they belong to?

Stakeholder mapping 2.

Who should be involved in the PD inquiry? Who will be the key informants? Think of anyone who is part of the problem and possibly part of the solution. Or anyone who has a direct or indirect relationship with your unit of analysis.

Conceptual framework 3.

Are there any dominant theories, frameworks, models that are used to explain the behavior/outcomes of the target group? What are those models or frameworks? Are there additional variables that you can include to explain performance?

Study design 4.

What methods will you use to collect and analyse your data? Are you going to follow a qualitative approach, a quantitative approach or a mixed-methods approach?

Data collection tools 5.

How are you going to collect data from the study subjects? Are you going to use interviews, surveys, or both? What are the key variables that you will collect in your field work?

Sampling 6.

How many positive deviants and how many non-positive deviants will your study sample include ? How will you achieve diversity in the sample and what will be your inclusion and exclusion criteria?
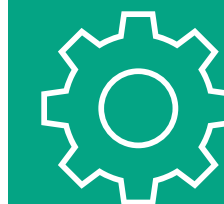
Tool 3.2:
Responsible Data Collection

In this tool, each guiding principle and its question will help you ensure responsible and ethical data collection when you go to the field for data collection[17]. Note that the list below is not exhaustive and should serve as starting point in developing your own context-specific guiding principles and/or specific questions to address them.

Principle	Do no harm	Diversity	Consent	Respect	Participation
Question	What could go wrong, who would be negatively impacted, and how can we minimize and mitigate these risks? ...	Have you selected your sample in a way that there is diversity and representation? ...	Have the participants been informed in a non-technical way about the study, their rights to privacy, and confidentiality and given their consent? ...	Are facilitators and researchers trained in interpersonal skills and intercultural awareness to adequately engage with the participants? ...	Have you engaged community members in activities that enables them to view, understand, and interpret data collected about them? ...

[17] Adapted from a tool developed by Paulina Jimenez - DPPD Ecuador cattle-farming project

Stage 4
Design and Implement Interventions

Technical Terms

Community interventions

Following the same logic as the conventional Positive Deviance approach, these interventions are designed to enable people to share, learn, and practice the identified positively-deviant strategies and practices

Policy interventions

These interventions make use of the insights generated from the contextual data analysis to inform interventions beyond the replication of individual practices

Scaling

Scaling in the DPPD context can take on different forms. Community interventions will mostly seek to scale-out positively-deviant practices to reach greater numbers of people. Policy interventions would rely on scaling up to impact policy and laws or scaling local conditions to improve infrastructure

Collective intelligence

The process of a group of people working together, often with the help of technology, to mobilize and share information, ideas, and insights

Data empowerment

The process where people, on their own or with the help of intermediaries, take control of their data to promote their and their society's wellbeing

Section overview

This section describes how the practices identified and documented in the previous stages can be scaled either through community interventions or through policy interventions.

Scaling of solutions can be done in at least two different ways. The first is what we refer to as community interventions that focus on activities that enable members of the community to apply the behaviors and strategies of positive deviants. The second seeks to use the contextual insights generated to design what we broadly refer to as policy interventions. The process of designing and implementing interventions as part of the application of the Data Powered Positive Deviance (DPPD) method is yet to be fully explored and tested in the pilots, but initial learnings point to potential opportunities for both types of interventions. It's important to point out that DPPD interventions should follow established principles like "Leave No One Behind" and that every organization will have its own frameworks and protocols of how to design and implement interventions.

Community Interventions

Designing and implementing DPPD community interventions follows the logic of the conventional Positive Deviance (PD) approach described in the Positive Deviance Handbook[1]. It includes a number of activities to enable people to share, learn, and practice the behaviors and strategies identified in Stage 3. Reflecting on our DPPD pilot projects, this could include farmers sharing their agricultural practices on cultivating rainfed cereal crops in the South of Niger or cattle breeders in Ecuador helping their peers reduce their deforestation imprint.

Things to consider

It is crucial for positive deviants and the community to be at the center of, and closely involved, in every step of the process. Participatory design offers a range of practical methods and tools[2] to enable the active participation of those who have developed and own a given solution (the positive deviant), those who stand to benefit from adopting a positively-deviant practice (the non-positive deviant or members of the community), as well as those who might have influence over the design and implementation process.

While the list of suitable methods and tools is long and ranges from co-design workshops and living labs to focus groups and usability testing, there is no single right way of going about the design and implementation of activities. Here are some suggestions on how to enable active and meaningful participation throughout the process:

- Further refine and contextualize the understanding of positively-deviant practices together with one or several positive deviants.
- Jointly develop a plan of action to support members of the community in adopting the practices. Do not design interventions that simply tell people about them. The point is not to transfer knowledge but to encourage a change in practices. Try instead to design your interventions in a way that will enable those who are not positive deviants to access and act upon the practices and strategies of positive deviants (the discovery and action dialogues presented in Stage 3 could be a useful technique).

> *"It's easier to act your way into a new way of thinking, than to think your way into a new way of acting".*
>
> Millard Fuller

- Collectively generate activities to enable the adoption of positively-deviant practices. Encourage and enable the people who discovered the deviations to spread the word in their groups or homologues. Note that the scaling of positively-deviant practices and strategies should be done for each homogeneous group separately (clusters created in Stage 2). In other words, practices identified within a certain cluster should be amplified and scaled only within this cluster because this is where they are relevant.

---

[1] The Positive Deviance handbook contains a useful list of very practical tips for facilitators:
https://static1.squarespace.com/static/5a1eeb26fe54ef288246a688/t/5a6eca16c83025f9bac2eeff/1517210135326/FINALguide10072010.pdf

[2] See the participatory design of evidence-based online youth mental health promotion, intervention and treatment handbook for a good overview of resources:
https://www.westernsydney.edu.au/__data/assets/pdf_file/0005/476330/Young_and_Well_CRC_IM_PD_Guide.pdf

While the process mainly relies on the qualitative insights generated about positively-deviant practices from the inquiry stage, you might also want to consider sharing (some of) the quantitative data collected, generated, and analyzed in the previous stages for people to generate their own insights and draw their own conclusions. Doing so might require training or targeted guidance by someone with the necessary technical skills, but it could allow you to go from having a one-way conversation informing people about the conclusions drawn from the data to a collective process of analysis, discovery, and discussion. The Data Zetu Shareback Sessions[3] along with other data empowerment-based interventions[4] could be drawn upon in the design of such activities.

Policy Interventions

The DPPD method employs a wide range of data, both traditional and non-traditional as well as quantitative and qualitative. Using what we referred to in Stage 1 as behavioral data, we can capture the factors within the control of individuals. But we can also learn about external factors that are beyond the control of individuals by using the insights generated from contextual data. The DPPD method might therefore not only allow us to identify positively-deviant practices of individuals or groups, but also provide a better understanding of factors that are beyond individual practices, but which enable, facilitate, or support a positively-deviant practice. Understanding these structural or supra-individual factors and their interactions with behavioral factors can inform the design of nuanced interventions and thereby increase their effectiveness and contextual fit.

In the Somalia rangelands pilot, the team knew that managing rangelands sustainably is very complex and cannot be successful if community interventions that focus on individual practices and policy-level interventions are not combined.

Policy interventions should set the "rules of the game" and clarify what the right use of the land should be through a general land-use plan. On the other hand, how the rules are implemented on the ground, such as which areas are closed, who gets access at what time and under what conditions, and individual-level practices such as soil and water conservation techniques, can only be amplified through community interventions that engage all stakeholders on the ground.

In addition, as the notion of PD in the DPPD method not only goes beyond behavior but also beyond individuals or (small) groups to include communities and administrative or geographic units, adapting human practices should not be the only focus of DPPD interventions.

If positive deviants are public spaces, as in the Mexico safe public spaces pilot, it will be more appropriate to focus on how structural attributes like infrastructure or governance mechanisms affect their "performance." Finding positively-deviant public spaces in Mexico City that have lower crime rates than others means there is a focus on identifying and understanding non-behavioral factors such as lighting, the existence of alarm buttons or safe pathways, or the distribution of informal commerce as determinants that enable positive deviance to emerge.

Things to consider

As is the case with community interventions, there is no single right way of designing and implementing policy interventions as part of the DPPD process. When designing DPPD policy interventions the following three questions can serve as a starting point:

---

[3] Data Zetu Shareback Sessions: *https://medium.com/data-empowerment/three-examples-of-data-empowerment-5f3e964ffbdc*
[4] Data empowerment-based interventions: *https://medium.com/data-empowerment/how-to-do-data-empowerment-7a5686e5aa93*

1. What policy areas do you want to address? These can be at different levels (from local to national) and address different domains and sector-specific policies.
2. Who should you engage with to develop and implement the policies? As mentioned in Stage 1, see if you can involve such policy-level stakeholders from the start of the project.
3. How will you go about feeding the project findings into policies? Activities can range from meetings with policymakers to present the project findings and suggested policy changes to working with NGOs to formulate and advocate for changes in policy.

Policy interventions will benefit from different ways of scaling, such as scaling up and scaling initial conditions (see Figure 10). Community interventions will likely scale-out by replicating positively-deviant practices. Given the complexity of most of the challenges DPPD will be applied to, and the range of behavioral, contextual, and other factors that need to be addressed for an intervention to be successful, it might make sense to employ systems thinking, collective intelligence, and experimentation for conceptual guidance to address multiple leverage points.

There is a growing number of practical resources from these fields applied to address development challenges that could be used for guidance, for example:

- The Portfolios of Options Green Paper[5] of Chôra Foundation provides comprehensive guidance on how to design for and enable system transformation.
- The UNDP Accelerator Labs team have put together a set of Collective Intelligence resources[6] for organizations to understand complex problems, make better decisions and find new solutions.

A. Scale Up

Impacting laws and policy

Changing institutions at the level of policy, rules and laws

B. Scale Out

Impacting greater numbers

Replication and dissemination, increasing number of people or communities impacted

C. Scale Deep

Impacting cultural roots

Changing relationships, cultural values, and beliefs, 'hearts and minds'

D. Scree Scaling

Impacting norms & expectations

Legitimizing a multitude of different small, and more relevant solutions

E. Scaling Initial Conditions

Impacting infrastructure

Changing access to capital, data, talent, knowledge, and networks
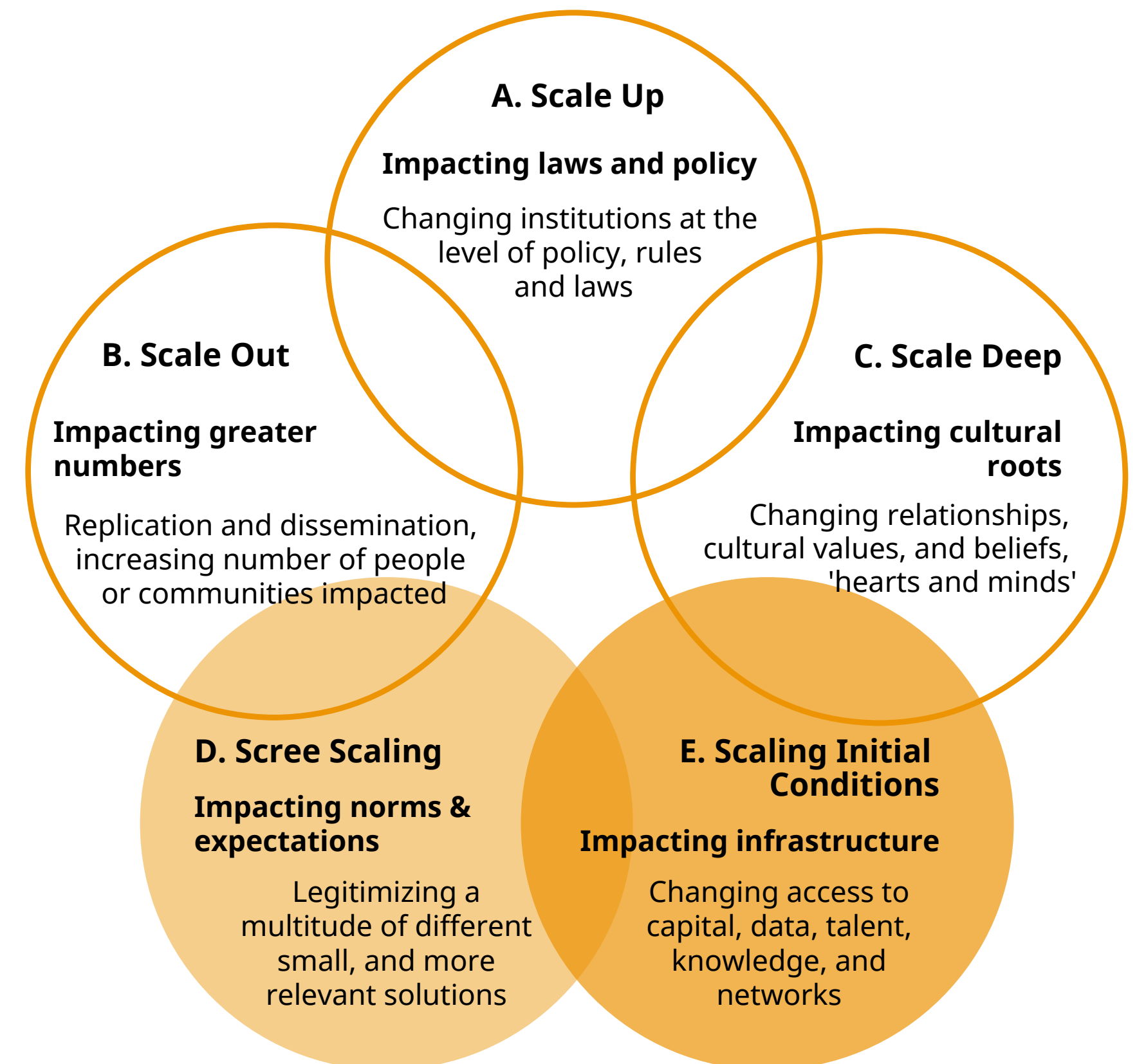
Figure 10: Scaling can mean different things and different types of scaling can be effective under different conditions. Community interventions will normally seek to scale out positively-deviant practices to reach greater numbers of people. Policy interventions would rely on scaling up to impact policy and laws and scaling local conditions to improve infrastructure[7].

Stage 5
Monitor and Evaluate

Section overview

This section provides an overview of how to measure, monitor, and evaluate the effectiveness of interventions that aim to scale and amplify the uncommon practices, strategies, and other factors underlying the performance of positive deviants.

Technical Terms

Intervention group — The group that was exposed to a DPPD intervention that aims at scaling and amplifying positively-deviant practices and strategies

Control group — The comparison group having similar properties to the intervention group but was not exposed to the DPPD intervention

Randomized control trials — An experimental form of impact evaluation in which the population receiving the intervention is chosen at random from the eligible population, and a control group is also chosen at random from the same eligible population

Quasi experiments — A type of experiment that attempts to establish a cause-and-effect relationship but the main difference with a randomized control trial is that the groups are not randomly assigned

The focus of Stage 5 is to assess and evaluate the impact of the Data Powered Positive Deviance (DPPD) interventions, both policy and community interventions. It is important to undertake this step to make sure data is being used to make interventions as effective as possible and to be able to report on results at the end of the project. It is also important to note that, usually, there are specific monitoring and evaluation (M&E) guidelines for each organization and the suggestions provided in this stage are for a general audience. There are three main elements related to the M&E of DPPD interventions:

1. Measurement: What data will be collected to track the outcome measure?
2. Analysis: How will the monitoring data be analyzed?
3. Dissemination: How will the results of the data collection be disseminated across the different stakeholders?

Measurement

The measurement can be done using the same digital outcome measure that was used to identify the positive deviants in Stage 2. However, here in Stage 5, this measure will be used to see if the intervention group experienced a significant enhancement in performance (i.e. better vegetation, lower deforestation, or fewer crime reports). Such as the remote sensing derived index in the Somalia, Niger, and Indonesia pilots, the crime reports and calls in Mexico, and the deforestation rates in Ecuador.

You should then decide whether traditional data (e.g. surveys or focus groups) is also needed to track the progress and the impact of the intervention. For example, alongside a survey, will you also need to do a focus group discussion with the intervention population to capture the challenges, unexpected dynamics, and benefits the group faced in adopting the solutions? You might also be interested in identifying positively-deviant practices and strategies that had the largest contribution to the positive outcomes or the highest prevalence among the intervention group.

Once you determine how the data will be collected, it is necessary to decide how often it will be collected and how long after the intervention to undertake the measurement. Too soon and it will not fully have had an effect. Too late and exogenous factors might be impacting the community. Will you need to capture the outcomes of the intervention group continuously, every six months, or once a year? And for how long? This should depend on the intervention timeline (check Tool 5.1 in the tools section for further guidance).

Analysis

The main focus of your analysis should be measuring the change caused by scaling the positively-deviant practices and strategies. Other changes that are not triggered or influenced by the intervention might exist. They can have an effect on the outcome measure, but you should consider these as secondary in your evaluation. To isolate the effect of the intervention from the non-intervention related effects, the changes in the outcomes of the intervention group should ideally be compared with changes in the outcomes of the control group, which wasn't exposed to the intervention (only in the case where a control group is ethically justifiable). Alternatively, if you use the same digital data as in Stage 2, you might consider looking at the same group before and after the intervention. This should give you a closer estimate of the intervention-caused effects or differences.

The most prominent ways to do this comparison are randomized controlled trials (RCTs) or quasi-experiments. In RCTs, both the intervention and control group units are randomly selected from the same sample of non-positive deviants within the same homogeneous group. Both groups are considered similar at the beginning, and only the intervention sets them apart. The outcomes of the intervention group are then compared with the outcomes of the control group to check if there is a significant difference in the performance of both groups. In quasi-experiments, the intervention and control groups are not randomly selected, and the units are assigned to those groups based on their choice or on convenience[1].

---

[1] Experimental vs quasi-experimental design: which to choose?: *https://quantifyinghealth.com /experimental-vs-quasi-experimental-desig*n

It is also possible to approximate effects by before-versus-after designs without control groups if confounding effects are negligible, in a way similar to how it was done in this study[2]. However, not all evaluators use such methods in assessing the impact of the interventions. Some prefer to use more accessible approaches that depend on the common sense of the evaluator and on the theory of change of an intervention (e.g. outcome mapping[3], contribution analysis[4], or building on stakeholders' capacities in grasping and valuing causal relations (e.g. Utilization-Focused Evaluation[5])).

Dissemination

The last step of the M&E process is to identify how and to whom data will be disseminated. In order to do so, ask yourself the following questions:

1. How and to whom will you communicate the M&E results internally (within the organization you work at)?
2. How and to whom will you communicate the M&E results externally (the different project stakeholders and donors)?
3. How will this data be used to design more effective interventions both internally and externally?

For instance, a team may want to review data on a monthly basis to make programmatic decisions and develop future work plans, while meetings with the donor to review data and program progress might occur quarterly or annually. Dissemination of printed or digital materials might occur at more frequent intervals.

[2] Positive deviance: a program for sustained improvement in hand hygiene compliance:
https://www.sciencedirect.com/science/article/pii/S0196655310007388?casa_token=_jA2VzImyhkAAAAA:
Bs8otGDRoMEBe84IqmrQ-w15WwBiQ-VUNKsZUBGBKbLHzrZ07RfiYj0rHx-9hUXSOPoGTSwWmw
[3] Outcome Mapping: a Realistic Alternative for Planning, Monitoring and Evaluation: https://odi.org/en/
publications/outcome-mapping-a-realistic-alternative-for-planning-monitoring-and-evaluation/
[4] Adressing Attribution through Contribution Analysis: Using Performance Measures Sensibly:
https://www.betterevaluation.org/en/plan/approach/utilization_focused_evaluation
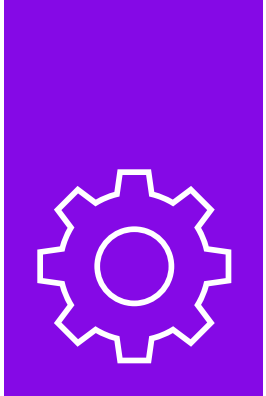[5] Utilization-Focused Evaluation: https://www.betterevaluation.org/en/plan/approach/utilization_focused_
evaluation

General tips when disseminating results[6]

- Honouring and amplifying the success stories by storytelling
- Enabling the community members to tell their stories and coach others within their networks
- Creating a "living university" for other communities to discover how the DPPD method could help them solve the same problem and reflect on what worked best

[6] Basic field guide to the positive deviance approach: https://static1.squarespace.com/static/
5a1eeb26fe54ef288246a688/t/5a6eca16c83025f9bac2eeff/1517210135326/FINALguide10072010.pdf

Tool 5.1
Measurement Tool

This tool presents an example of the different indicators that could be captured in a positive deviance intervention, along with their corresponding data sources and frequency of capture.

Performance indicator	Data source	Frequency
Village vegetation health	Remote sensing vegetation	Every six months over a period of 18 months
Percentage of promoted practices that were adopted by the intervention group	Surveys	Annually
Number and percent of facilitators who were trained to scale the positively-deviant practices	Project team	Annually
Number and percent of farmers in the intervention group who adopted the positively-deviant practices	Interviews with the farmers in the intervention group	Every six months over a period of 18 months

Postscript

The Data Powered Positive Deviance (DPPD) Handbook provides a portable, step-by-step guide to applying the DPPD method — a mixed-methods approach that relies on a combination of traditional and non-traditional data for identifying grassroots solutions to complex development problems.

The Handbook covers extensively the first three stages of the method: assessing the problem-method fit, determining positive deviants, and discovering positively-deviant practices and strategies. The two final stages that are concerned with designing and implementing interventions and monitoring and evaluating their effects are less elaborate because the pilot work conducted at the time of writing the Handbook has not finalized these stages. Further testing is likely needed.

Another application of the DPPD method that is not described in much detail here is the identification of exceptional responders to development interventions. At the time of writing the Handbook, the relevance of this application is being investigated in India, by looking for farmers who have adopted an early sown wheat innovation that aims to reduce the impact of climate change on crop yields. Further applications or derivatives of the DPPD method might include reusing and integrating the outputs of Stage 2 (e.g. homogeneous groups, the digital performance indicator) for other purposes.

To conclude, DPPD is a novel method that has the potential to lead development practitioners to identify and nurture local know-how and community-level collective intelligence using increasingly available digital datasets. That said, the method is still in its operational infancy. We hope that this Handbook will help practitioners pick it up, and refine it in new and enhanced ways.

Contact:

Catherine Vogel (GIZ Data Lab):
catherine.vogel@giz.de

Jeremy Boy (UNDP Accelerator Labs):
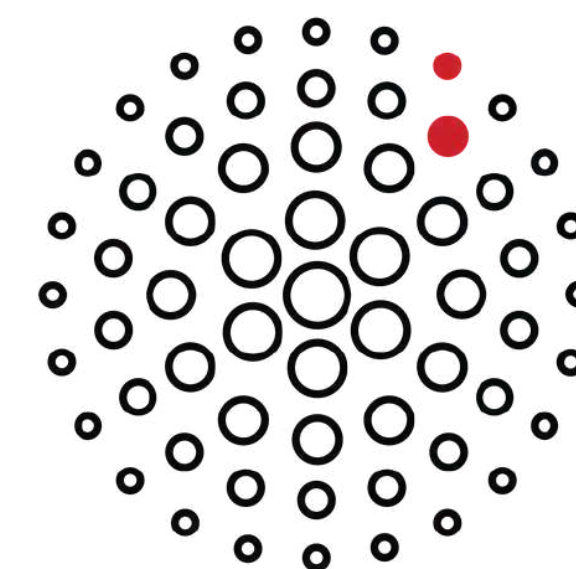jeremy.boy@undp.org

Basma Albanna (University of Manchester):
basma.albanna@manchester.ac.uk

Website:
https://www.datapoweredpd.org/
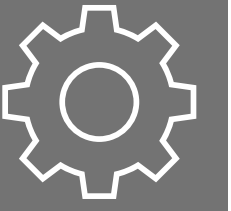
Photo credits:
Title Page: Geometric shapes lie across the emptiness of the Sahara Desert in southern Egypt, Sahara Desert 2020, Photograpger: USGS
Page 5: Shepperd, Photographer: Gueli54 - Canva
Page 8: Rice fields and farming, Photographer: ArtRachen01 - Canva
Page 10: Women working in rice field, Photographer: Langdu - Canva
Page 23: Overloaded motorcycle sidecar, Photographer: Onfilm - Canva
Page 36: People in red and white boats on water, Photographer: Alesia Kozek - Canva
Page 54: Lives of Burmese women, Myanmar 2019, Photographer: Sippakorn Yamkasikorn
Page 67: From a naming brainstorm at a talk, United States 2017, Photographer: Patrick Perkins
Page 73: Shadow player, Photographer: Jan Mohammad Shaikh - Canva
Page 75: A Myanmar mom carry her kid, Myanmar 2019, Photographer: Sippakorn Yamkasikorn
Page 82: Cattle grazing, Norway 2021, Photographer: Daniel J. Schwarz
Page 89: Person writing, Photographer: Rattanakun - Canva
Page 93: Web developer, Photographer: Oatawa - Canva
Page 100: People gardening, Slovenia 2017, Photographer: Daniel Funes Fuentes
Page 107: A real-photography of boats on river surrounded by desert, Photographer: Pok Rie - Canva
Page 111: Mexican carrot worker, Edinburg, Texas. 1939. Photographer Lee Russell
Page 115: Rainforest surrounded by fog, Photographer: David Riaño Cortés

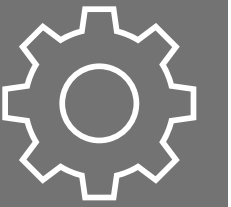**DATA POWERED POSITIVE DEVIANCE**

Appendix

Tool 1.1
Problem Definition

The Problem Definition tool will help you describe the problem and its root causes, define the target group and the units that will be analyzed in that target group as well as the desired outcome of a potential intervention. Try to develop or at least validate each of these components with people affected by or with expert knowledge of the problem.

What is the problem?	1.	Why does the problem exist?	2.	Who is impacted?	3.

What is the unit of analysis?	4.	What does a positive outcome look like?	5.

Tool 1.2
Suitability Assessment

Use the tool below to describe what or who the positive deviants might be, given your particular project scope. This will help you determine the data sources you will need. Remember that in DPPD positive deviants can be individuals, communities, or even administrative or geographic units. Use the boxes below to develop an initial description of potential positive deviant(s).

General Tips:

- In some cases, there might be an agreed definition of who is considered an outperformer. Check the academic literature!
- Engage with people who might be aware of individuals or communities who are adopting uncommon practices and strategies.
- Talk to people working on the problem including sector experts, government officials, or NGO staff.
- If you have access to data on the performance of potential positive deviants and easy access to data analysis skills, run a preliminary analysis to understand the data distribution and to see if you can spot any immediate outliers.
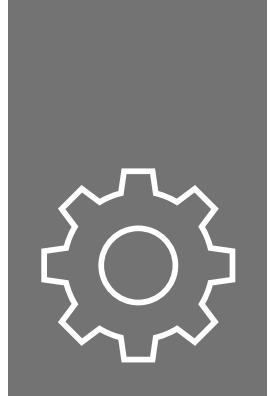
| **Who am I?** | 1. |

| **Where am I?** | 2. |

| **What outcomes can I achieve better than my peers?** | 3. |

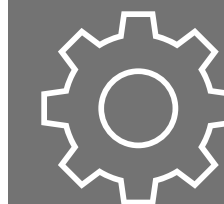| **Am I known (or likely) to exist?** | 4. |

Tool 1.3
Data Mapping

Use the tool to identify potential data sources for each of the three categories. Start with the outcome column then move to the context. Leave the behavioral data category for later, after you have performed initial quantitative data analyses. Use the factors row as an intermediate step to help you brainstorm potential data sources.

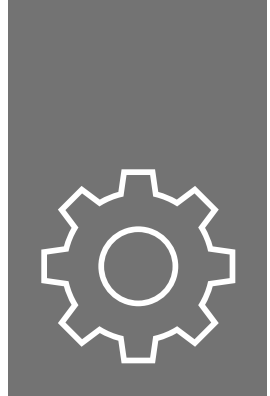|	**Context**	**Behaviour**	**Outcome**
Factors	Contextual factors that are beyond the control of individuals and that are hard to change and transfer	Behavioural factors that are within the control of individuals and can be changed and transferred	A measure that can capture the perfomance of the target group to identify the units that were able to achieve better outcomes
	....	....	....
	....	....	....
Data	Data to control for contextual variables	Data to uncover the uncommon practices and strategies of positive deviants	Data to directly or indirectly measure outcomes
	....	....	....
	....	....	....

Tool 1.4
Capabilities Assessment

It can be helpful to think about how these different skills will be required for the different types of data introduced in the data sub-section: outcome, context, and behavior. The tool provides a process to map existing skills and skill gaps along these three dimensions.
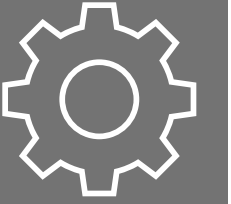
	Context	Behaviour	Outcome
Domain knowledge Do you have sufficient knowledge about the respective domain?	....	....	....
Local knowledge Do you have sufficient knowledge about the local conditions?	....	....	....
Data skills Do you have the required data skills?	....	....	....

Tool 1.5
Desirability Assessment

Using the tool below, ask yourself whether it is desirable for a positive deviant, should you identify one, to share their practices with others. If so, would the community be able and willing to adopt the practice and benefit from it? Might there be negative consequences on the (social) systems that the positive deviant and the community are embedded in or connected with?

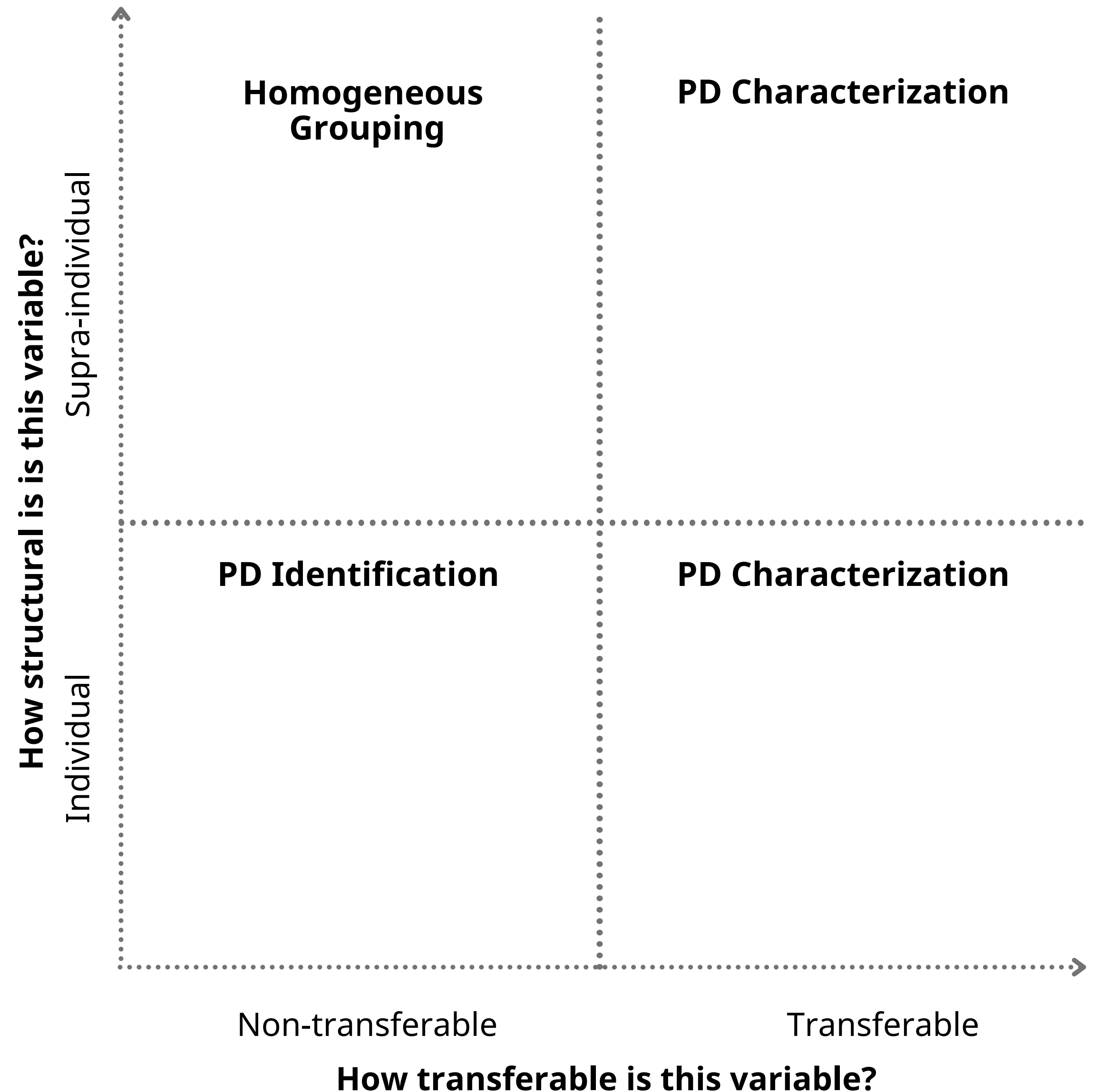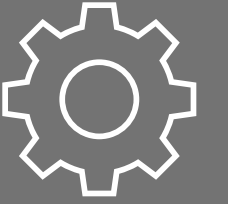|	**Positive Deviants**	**Communities (Target Group)**	**(Social) Systems**
	A positive deviant has an advantage over others given their knowledge of and application of positively-deviant practices. Sharing their practice with others could endanger their competitive advantage.	In the short term, non-positive deviants might not accept, adopt or embrace the positively-deviant practice. In the long term, adoption of the practice could lead to overall negative effects.	The wide adoption of a common practice could have negative consequences beyond the positive deviants and the community, e.g. on governmental and social institutions, the natural environment or political, social or economic systems.
Potential negative consequences			

Tool 2.1:
Variable Mapping

This tool will help you in mapping out all the possible control and independent variables in order to select the most suitable variables for the homogeneous grouping and the positive deviant identification steps of Stage 2 and the positive deviants' characterization step of Stage 3.

Using the guidance provided below you will be able to categorize and map the study variables into four main types:
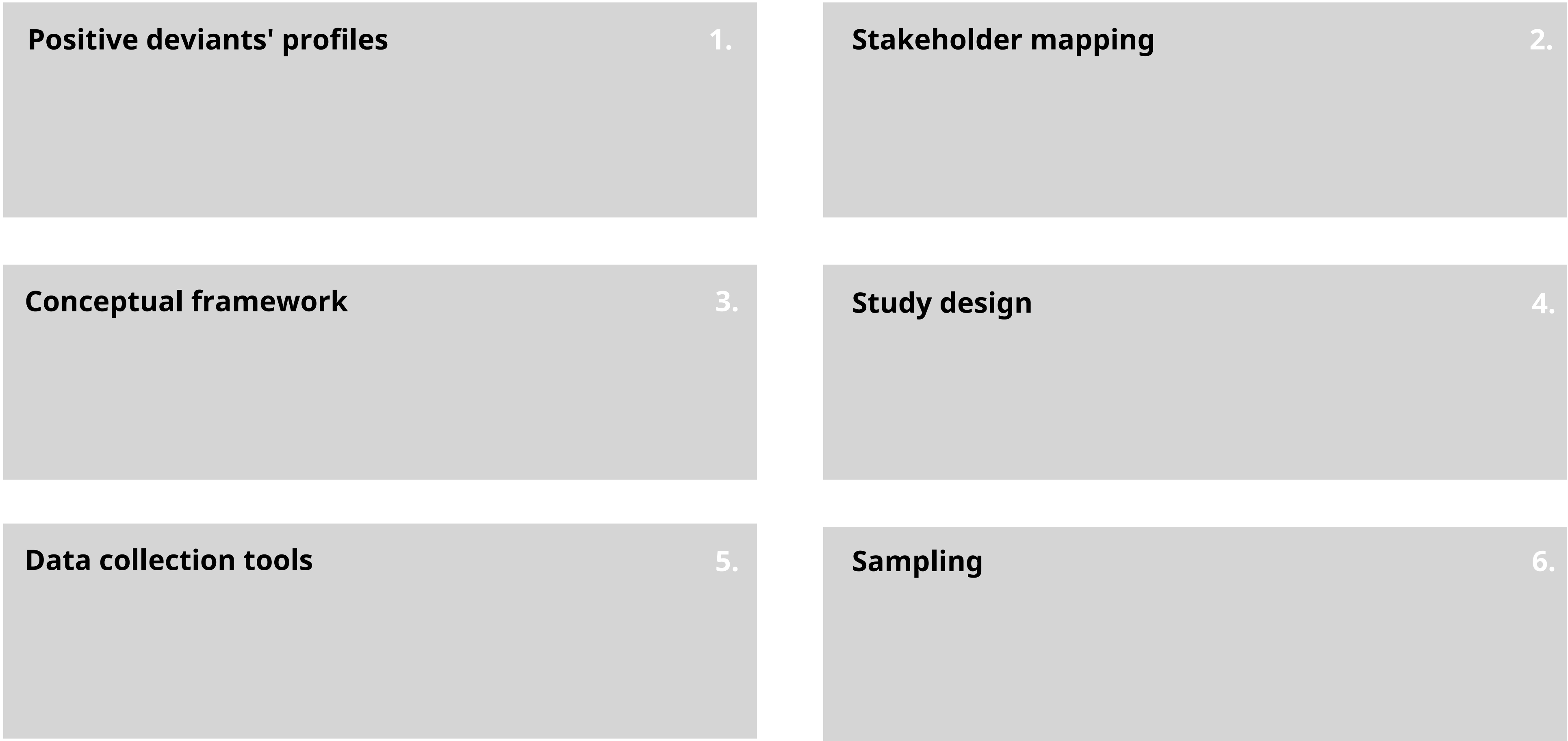
- Homogeneous grouping: Factors beyond the control of individuals that can influence the outcomes of the target group, and at the same time they cannot be changed e.g. rain.

- PD identification: Individual-level factors that can influence the outcomes of the target group, and at the same time they cannot be changed. Those variables should be controlled for when identifying positive deviants within each homogeneous group or cluster e.g. socioeconomic status.

- PD characterization (Supra-individual): Factors beyond the control of individuals that can influence the outcomes of the target group but can be transferred to peer communities e.g. policies.

- PD characterization (Individual): Individual-level factors that can influence the outcomes of the target group and can be changed and transferred e.g. soil and water conservation techniques.
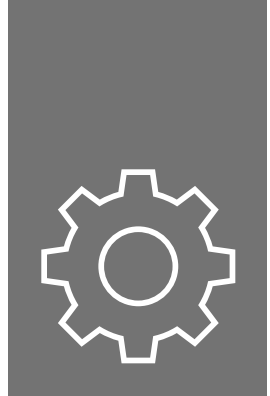
How structural is is this variable?

Supra-individual

Individual

Homogeneous Grouping

PD Characterization

PD Identification

PD Characterization

Non-transferable

Transferable

How transferable is this variable?

Tool 3.1:
Fieldwork Design Canvas

This tool will help you cover the different aspects of fieldwork planning, starting from the characterisation of the identified positive deviants and mapping the relevant stakeholders to determining the suitable data collections tools and sampling strategy.
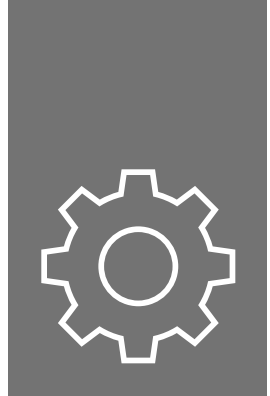
Positive deviants' profiles 1.	**Stakeholder mapping** 2.
Conceptual framework 3.	**Study design** 4.
Data collection tools 5.	**Sampling** 6.

Tool 3.2:
Responsible Data Collection

In this tool, each guiding principle and its question will help you ensure responsible and ethical data collection when you go to the field for data collection. Note that the list below is not exhaustive and should serve as starting point in developing your own context-specific guiding principles and/or specific questions to address them.

Principle	Do no harm	Diversity	Consent	Respect	Participation
Question	What could go wrong, who would be negatively impacted, and how can we minimize and mitigate these risks? ...	Have you selected your sample in a way that there is diversity and representation? ...	Have the participants been informed in a non-technical way about the study, their rights to privacy, and confidentiality and given their consent? ...	Are facilitators and researchers trained in interpersonal skills and intercultural awareness to adequately engage with the participants? ...	Have you engaged community members in activities that enables them to view, understand, and interpret data collected about them? ...

Adapted from a tool developed by Paulina Jimenez - DPPD Ecuador cattle-farming project

Tool 5.1
Measurement Tool

This tool presents an example of the different indicators that could be captured in a positive deviance intervention, along with their corresponding data sources and frequency of capture.

Performance indicator	Data source	Frequency