Learning from the Edges

Lessons Learned from Applying the Data Powered Positive Deviance Method to Identify Grassroots Solutions Using Digital Data

By Andreas Pawelke,
Andreas Glücker,
Basma Albanna,
and Jeremy Boy.

With thanks to
Richard Heeks,
Catherine Vogel and
Gina Lucarelli for their
feedback, as well as to
Amy Elizabeth Bennett
for her editing.

Contents

Executive Summary

This report presents six learnings from four pilot projects conducted by the Data Powered Positive Deviance (DPPD) initiative, a global collaboration between the GIZ Data Lab, the UNDP Accelerator Labs Network, the University of Manchester Center for Digital Development, and UN Global Pulse Lab Jakarta. The pilots are run in Ecuador, Mexico, Niger and Somalia to learn about grassroots solutions to development challenges that range from the interaction between livestock farming and deforestation to gender-based violence and insecurity in dense urban environments. The learnings relate to the early stages of the DPPD method, originally proposed by Albanna & Heeks (2018), and focus mainly on the access to, and use of digital data. They are summarized as follows:

1. Remain flexible in the face of data unavailability
2. Leverage existing partnerships for data access
3. Map and fill know-how gaps early
4. Scale with caution
5. Look at deviance over time
6. Look beyond individual or community practices and behavior

The report is written for development practitioners, data analysts, domain experts, and more generally anyone interested in using new data sources and technologies to uncover successful local practices to development challenges.

Data Powered Positive Deviance

The Data Powered Positive Deviance initiative is established on the concept of Positive Deviance [1], which assumes that in every community there are individuals or groups who develop unconventional practices that help them deal better with challenges than their peers. The goal of the Positive Deviance approach is to identify these people, to understand what makes them and their practices different and successful, and to mobilize the rest of their communities to emulate those practices.

Its origins date back to the 1990s, and the work of Sternin and colleagues [2] on child malnutrition in Northern Vietnam. Their idea was simple: flip the logic of traditional development efforts. Rather than spend time, effort and money on bringing resources into communities from the outside, Sternin and colleagues decided to seek out and scale solutions to malnutrition that might exist on the inside. The team found families who were able to keep their children properly nourished, while others around them could not. They learned that caregivers in these families would add small shellfish and sweet-potato greens to their children's meals. They would also feed their children more often than the customary two meals a day. While these practices were accessible to everyone in the community, they were deemed inappropriate and thus only a few "positive deviants" employed them. By supporting other families to adopt such practices, Sternin and colleagues claim to have impacted the lives of two million people by the turn of the century.

Recent developments in the availability of digital data offer an unprecedented opportunity to look for positive deviants across large geographical areas and analyze their performance over time. The integration of such novel data sources to complement traditional data sources used in the Positive Deviance approach is what we refer to as the Data Powered Positive Deviance (DPPD) method. This method can be regarded as a new tool for development professionals and Positive Deviance practitioners alike to identify what works and why, by mixing analytical insights from traditional and non-traditional data.

DPPD is based on a paper by Albanna and Heeks [3]. The authors suggest that the DPPD method* should help identify positive deviants in new and more generalized ways because of the geographic, demographic and temporal scales afforded by digital data sources. It should also help expand the common focus of previous Positive Deviance interventions on public health to areas such as agriculture, natural resource conservation, and urban planning. Building on four on-going pilot projects in Ecuador, Mexico, Niger and Somalia, we seek to test, evaluate and expand the DPPD method and explore how it can be leveraged for different development issues in different contexts.

This report covers learnings of the first two stages of the DPPD method (Table 1) which are concerned with defining the problem to be tackled and the desired outcome as well as the identification of potential positive deviants (see Figure 1). It is based on lessons learned from the ongoing pilot projects that are introduced in the following section. The six key learnings are then presented and discussed before the report concludes with a reflection on the main insights.

Although there are many calls for the adoption of more data-driven methods in development work, little is documented about the transformative effects of integrating such methods in existing approaches, how these might affect problem definitions, and how they might necessitate a shift in mindset, skills and roles in a team. Much less is known about the potential value of digital data in identifying and scaling local solutions in an attempt to gradually move away from the imposition of external solutions and towards the diffusion of local practices and strategies.

This report attempts to highlight some of the opportunities and challenges associated with the adoption of a method that uses non-traditional data sources in a bottom-up approach to build on a community's inherent assets and capabilities as the starting point in the search for solutions.

[1] http://tiny.cc/PositiveDeviance

[2] http://tiny.cc/Sternin2000

[3] http://tiny.cc/Albanna2018

* Initially introduced as "Big-Data-Based Positive Deviance".

The Five Stages of the DPPD Method

Stage 1: **Assess problem-method fit**	Define the problem and the scope of the intervention. Check if DPPD is a suitable and feasible method by assessing required data and capabilities and by ensuring that potential outcomes are desirable for the target group.
Stage 2: **Determine positive deviants**	Divide the studied population 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 positive deviants.
Stage 3: **Discover factors underlying outperformance**	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 behaviours, practices and other factors that explain the outperformance of positive deviants.
Stage 4: **Design and implement interventions**	Assess the potential of 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 behaviors and outcomes within communities can then be considered for future interventions.
Stage 5: **Monitor and evaluate**	Monitor and evaluate the effectiveness and suitability of the community- and / or policy-interventions

Table 1. The five stages of the DPPD method. Note that the learnings in this report mainly relate to the first two stages (highlighted).

Four Pilot Projects

We are testing the DPPD method with local teams in Ecuador, Mexico, Niger and Somalia to respectively identify grassroots solutions to the interaction between livestock farming and deforestation; gender-based violence and insecurity in dense urban environments; cereal production efficacy and efficiency; and optimized rangeland management (Table 2). At the time of writing this report, the four pilot projects have reached the end of the quantitative analysis phase (Stage 2 of the method—Table 1).

Positive deviants	How they are identified	Data used
Cattle-raising farms in the Ecuadorian Amazon that do not contribute to deforestation.	Farms should have significantly lower deforestation rates than what might be expected for three consecutive years, given their size, land use, soil adaptability and cattle density.	Remote sensing data, vaccination data, official statistics, administrative data, cadastral data and interviews.
Public spaces in Mexico City where women are not subject to gender-based violence.	Public spaces should have significantly lower geneder-based crime reports than what might be expected, given their population density, demographic and socio-economic status.	Geospatial data, interviews, Mexico City's open data portal and official statistics.
Rainfed-cereal-growing communities in Niger that produce healthy crops despite climate change and conflict.	Community boundaries should include a significantly higher soil-adjusted vegetation index (SAVI) than what might be expected, given the level of precipitation and land use.	Remote sensing data, administrative data, geospatial data and interviews.
Pastoral communities in Northern Somalia that preserve healthy rangelands despite recurring droughts.	Pastoral activity boundaries should include a significantly higher SAVI than what might be expected, given the land capability.	Remote sensing data, administrative data and interviews.

Table 2. Overview of the four pilot projects in Ecuador, Mexico, Niger and Somalia.

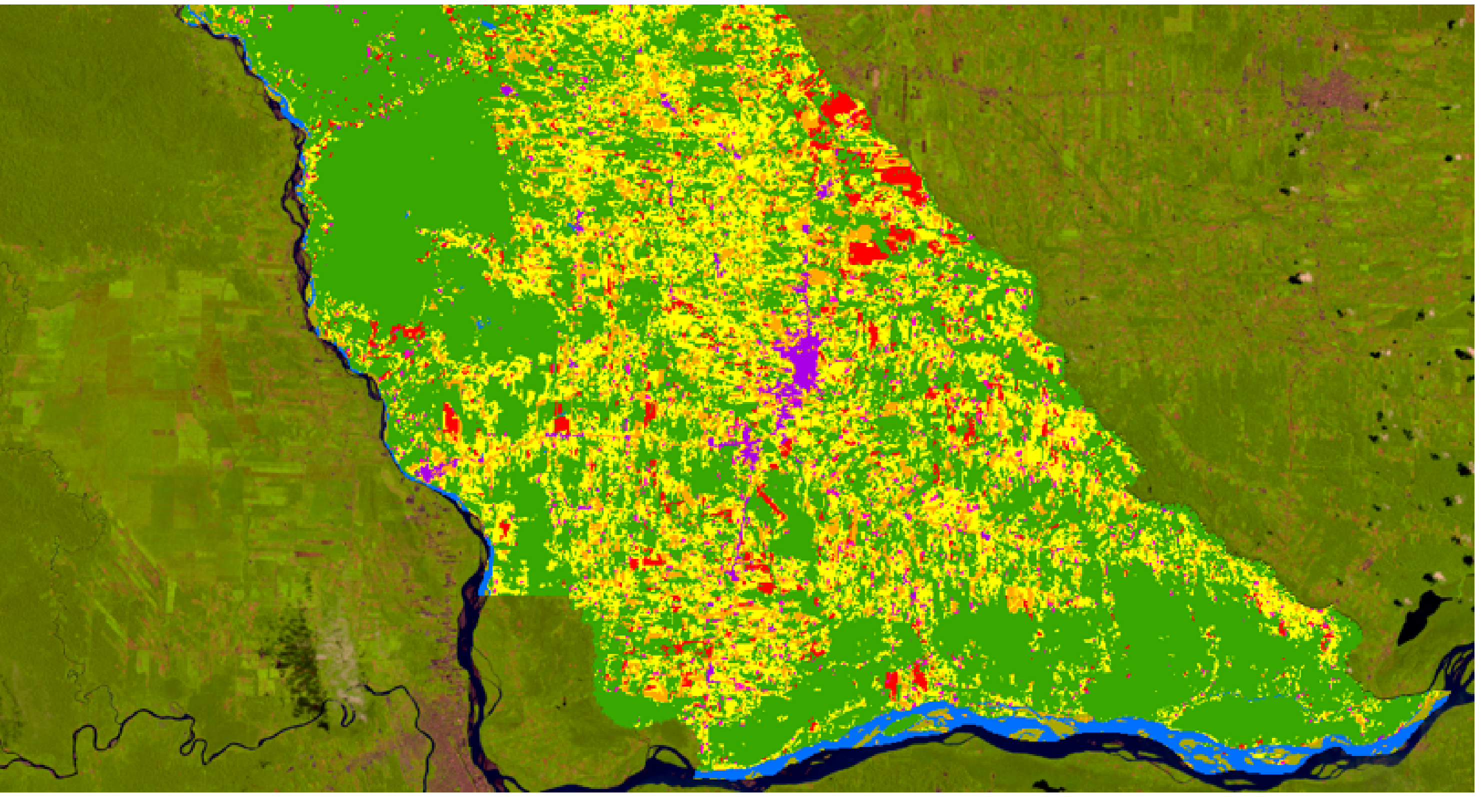


Figure 1. Land cover analysis to identify potential positive deviants in Joya De Los Sachas, Ecuador. Legend: Green: forest; Yellow: no forest; Red: removed forest (deforestation); Blue: water; Purple: urban areas.

Towards tackling deforestation in Ecuador

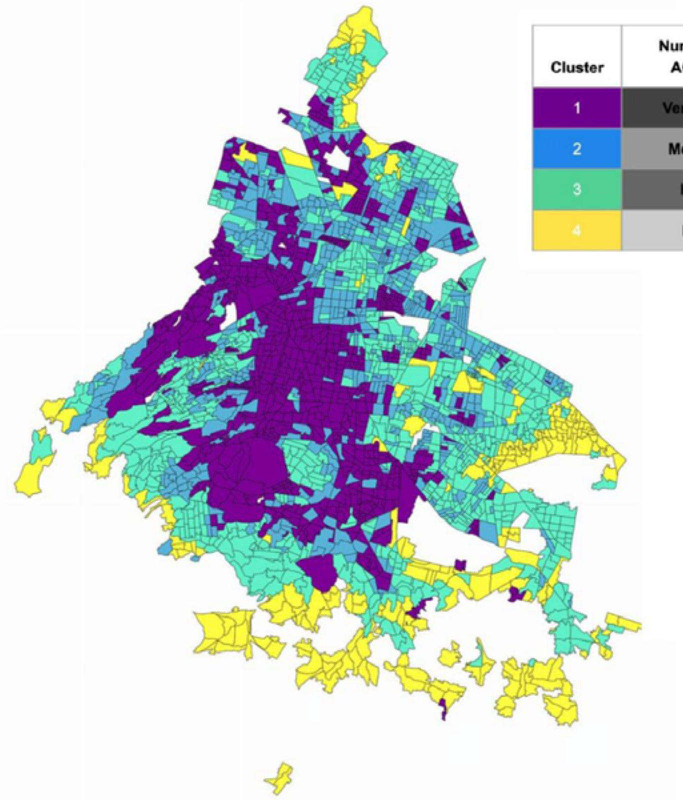
Learn more at:
http://tiny.cc/DPPDEcuador

[4] http://tiny.cc/REDD2016

Agricultural practices are one of the main causes of deforestation in the Ecuadorian Amazon region. Ninety-nine percent of deforested areas are transformed into agricultural land, sixty-four percent of which is used as pasture land for livestock farming and other purposes (REDD+, 2016 [4]).

The pilot project in Ecuador engages with cattle farmers in the Ecuadorian Amazon who operate in areas of potential forest clearance for farming without themselves contributing to deforestation. Positive deviants are defined as cattle-raising farms with deforestation rates that are significantly lower than what might be expected for three consecutive years. Using the DPPD method, our goal is to identify sustainable farming practices and scale them to reduce the negative effects of livestock farming on deforestation.

We identify potential positive deviants using land cover and land use data derived from satellite imagery, as well as climate, soil, socio-economic and cattle vaccination data.



Cluster	Number of AGEBs	Population density	Incoming daily trips	Marginalization index
1	Very high	Medium	Very high	Very low
2	Medium	High	High	Low
3	High	High	Medium	Medium
4	Low	Low	Very low	High

Figure 2. Grouping of AGEBs in Mexico City, based on population density, daily incoming trips and marginalization index.

Towards making public spaces safer for women in Mexico

Learn more at:
http://tiny.cc/DPPDMexico

[5] http://tiny.cc/UNWomen2020

* An Área Geoestadistica Básica (AGEB) is a basic geostatistical area in a Municipality, Town or Delegation Policy in Mexico.

In November 2019, Mexico City's mayor issued a gender-based violence alert, activating a series of measures to reduce violence against women. On average, ten women are murdered per day in Mexico, and in 2019 more than half of those femicides occurred in public spaces (UN Women, 2020 [5]). Violence against women occurs on the street, in parks and, to a lesser extent, on buses, minibuses and the subway.

The pilot project in Mexico looks into factors contributing to lower rates of outdoor violence against women and girls. Positive deviants are defined as geographic areas, so-called AGEBs*, that have lower incidents of outdoor gender-based violence than what might be expected, given their socio-demographic and economic situation. Our goal is to inform local policy makers and grassroots initiatives on what factors might contribute to the design of safer public spaces in Mexico City.

We identify potential positive deviants using geospatial data, official statistics and publicly available crime reports.

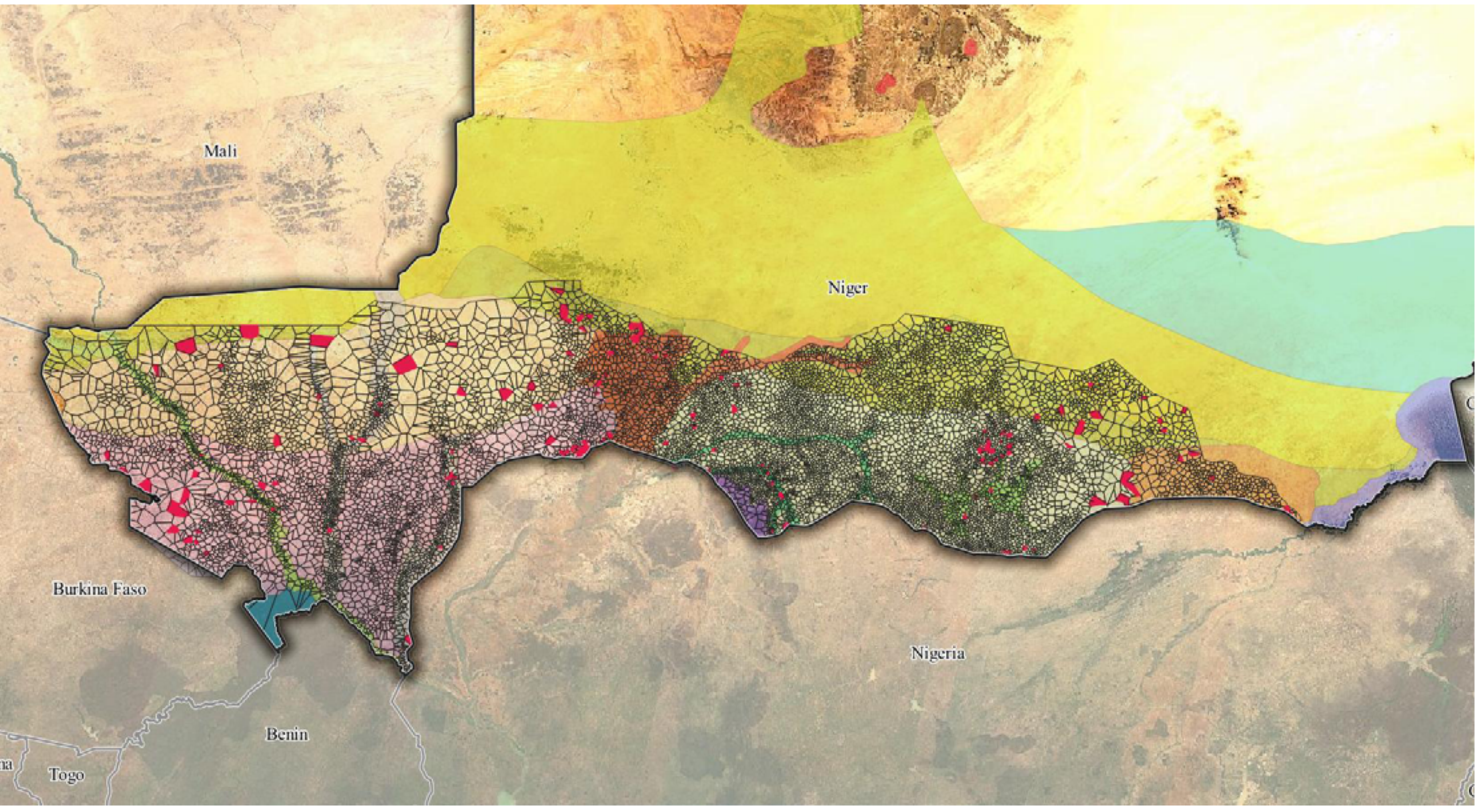


Figure 3. 180 potential positively deviant communities (shown in red) within a sample of 12,093 communities in Southern Niger.

Towards increasing agricultural productivity in Niger

Learn more at:
http://tiny.cc/DPPDNiger

Sustained agriculture in Niger is under tremendous pressure as climate change and the reduction of rainfall affect crop cycles. Crops mature in a drier context, which reduces production quality and aggravates food insecurity. In the broader Sahel region, more than four million people are food insecure, and 80% of lands are at risk of degradation (WFP, 2018 [6]). In Niger, 1.7 million people are estimated to become food insecure in 2021 (WFP, 2021 [7]).

The pilot project in Niger aims to identify and scale practices of positively deviant cereal-growing communities that produce healthy yields of sorghum and pearl millet. Positive deviants are defined as cereal farming villages that have higher quality produce than what might be expected, given their agro-ecological conditions. Our goal is to identify and leverage the local practices and strategies driving this outperformance to inform the design of interventions that can support communities in increasing their agricultural productivity.

We identify potential positive deviants using readily available earth-observation and administrative data.

[6] http://tiny.cc/WFP2018

[7] http://tiny.cc/WFP2021

Figure 4. Communities within 5 km buffer zones (shown as pink circles) in the West Golis pastoral livelihood zone in Somalia.

Towards maintaining pastoral livelihoods in Somalia

Learn more at:
http://tiny.cc/DPPDSomalia

Repeated droughts in Somalia between 2010 and 2017 have caused more than a quarter of a million deaths, and contributed to the displacement of roughly 4.2 million people.

The pilot project in Somalia aims to identify positively deviant pastoral communities that are able to sustain the health of their surrounding rangelands despite the recurring droughts. Positive deviants are defined as community rangelands that are healthier than what might be expected, given local climatic conditions. Our goal is to build on local knowledge to help communities understand how they might preserve their rangelands in order to maintain their pastoral livelihoods.

We identify potential positive deviants using vegetation and climate data along with administrative land cover data.

Six Learnings from the Pilots

Ten months into the pilots, all teams—including the global coordination team—took the time to reflect on the work accomplished and the lessons learned from it. This activity was supported by a number of learning formats, including documentation calls, four online sharing sessions with all teams, and various online surveys. While the global team was responsible for compiling and curating the learnings presented here, they are the result of a collective effort. This section presents the six major learnings that emerged from this process. They are illustrated with examples from the pilot projects and highlight how the learnings might inform future DPPD work.

Learning 1: **Remain flexible in the face of data unavailability**

What we learned

The core idea behind the DPPD method is to leverage digital data to identify potential positive deviants. This assumes that relevant data are available, accessible, reliable and usable. However, in contexts like the ones in which our DPPD pilots are being conducted, digital data is often scarce, and data landscapes are incomplete if not entirely non-existent for specific development issues.

All four teams experienced difficulties in accessing data. Several had to adjust their project designs to be feasible, seeing what data was available, accessible and reliable. In some cases, they had to adapt their way of measuring performance, while in others, they had to simulate units of analysis (hypothetical village boundaries), as no precise data was available. These adjustments however, did not change the domain or regional focus of the pilots.

How we learned it

In Somalia, we initially set out to find pastoral communities in the West Golis region that were able to maintain their livestock size throughout recurring droughts. We hoped that understanding their positively deviant practices could contribute to bending the curve of internal climate migration. We mapped several available data sets, based on our initial understanding of pastoral activities in the region. These included mobile phone data to see mobility patterns, and community-level vaccination data to count livestock.

Unfortunately, setting up the necessary partnership with a local telecom company turned out to be impossible, which meant we could not get access to the mobile phone data. In addition, the high level of aggregation of the accessible vaccination data prevented any meaningful statistical modeling. After looking into several alternatives, we decided to pivot the focus of the project away from counting livestock to measuring the health of community rangelands, a key factor in maintaining livestock. We assumed that rangeland degradation and overgrazing were in fact the main threats to pastoral livelihoods, as these would be among the main drivers of livestock depletion. If it were possible to identify positively deviant communities that preserved their rangelands through severe droughts, we could help scale up these livelihood-sustaining practices. This new direction was mainly driven by data accessibility—in this case remotely sensed vegetation indices that reflect the health of rangelands—but the general focus on pastoral livelihoods was preserved.

What we take away

When applying the DPPD method, the problem to be addressed should drive the project design and provide guidance on the data required. However, necessary data might be unavailable, inaccessible or unreliable. Thus, viewing data as a mere technical aspect that can be accounted for once the project implementation has begun can result in substantial challenges.

There is value in mapping out data landscapes early on as it can help determine whether the DPPD method is applicable. However, there is a tension between the relative rapidity of the mapping exercise, and the lengthy process of securing access to available data, and ensuring its reliability. Put simply, the mapping exercise may indicate that data exist, but it does not guarantee they can be used. The difficulty then is to maintain a focus on the problem, i.e. the development issue to be addressed, as time and resources are spent on efforts to secure access to and validate data, without the guarantee that these will succeed; while properly pondering the temptation to reframe the problem according to whatever data are readily accessible.

We advise to remain flexible and creative in dealing with a potential lack of data, and to be open to adjusting the original project design based on how accessible and reliable certain data sources are. When data is entirely missing for a given issue, it might be possible to look for proxy measures in so-called alternative data sources. However, it may also be that DPPD is not suited for the project at hand.

Learning 2: Leverage existing partnerships for data access

What we learned

The DPPD method requires access to very specific data from a number of different sources at high levels of granularity. While in some cases, data might initially appear to be readily-available and usable, it may well be unsuitable for the identification of positive deviants because of low granularity and/or coverage. Indeed, highly granular data is seldom publicly available and often only accessible through partnerships with holders of the raw data, like government entities or private companies.

While our four pilots used different strategies to get access to adequate data, one common learning emerged: integration of the pilots into pre-existing programmes and partnerships with relevant stakeholders and data holders significantly accelerated the process of identifying and accessing relevant and usable data.

How we learned it

Our pilot in Ecuador is a case in point: it was integrated early on with the UNDP ProAmazonia programme, which already had a multitude of partnerships in place. This made it possible to access granular

cattle vaccination data from the Ecuadorian Ministry of Agriculture, detailed training sets for land cover analysis from the Ministry of Environment and cadaster data from the national agricultural survey.

What's more, preliminary results were repeatedly presented to government partners for feedback. Throughout these interactions, insights emerged on new sources of data that could be used for the pilot, suggesting a strengthened interest of external stakeholders. This led to increased access to quality data for the team. Overall, the pre-existing partnerships accelerated access to a range of different datasets, and facilitated analysis and interpretation.

What we take away

Access to adequate secondary data poses a challenge for all development initiatives. This is particularly true when working with the DPPD method, as it requires both data that can be used to infer individual or collective performance, in addition to data that describes their contextual realities. The data further needs to be sufficiently granular to establish a norm for similar individuals or groups, so that deviance from that norm can be identified.

We stress that the different data sources used must overlap spatially and temporally so that they can be integrated. Spatially, the coverage of the contextual data should be broad enough to include all units of analysis whose performance is being measured. Temporally, the contextual data should be recent and longitudinal enough to enable the identification of "true" positive deviants, i.e. positive deviants that remain so over time, not just at a specific point in time, as the latter could be confounded by random effects.

Learning 3: Map and fill know-how gaps early

What we learned

Access to the right data is essential to the DPPD method. However, simply looking for potential positive deviants in data without understanding the contextual specificities that might make one outperform their peers, or how these might show in digital data can easily lead to wrong conclusions.

While a considerable amount of know-how was already present in each team, additional data analysis skills, local domain expertise and an understanding of the local data ecosystem were necessary for each pilot. All four teams had to consult with or recruit analysts who were able to combine quantitative data skills with an understanding of the specific local context.

How we learned it

The team in Niger included two analysts with domain-specific data expertise. This proved very helpful not only for assessing what data was available, but also for selecting adequate data sources as well as appropriate performance indicators and sampling strategies. With the help of the analysts, we were able to decide on a particular sub-region of Southern Niger where rainfed sorghum and pearl millet were mainly cultivated. We also learned that the soil-adjusted vegetation index (SAVI) was more suitable for arid climates than other, similar indices we had originally focused on. Finally, we realized it was important to narrow our investigation temporally to the local rainy season, in order to reduce possible confounds related to intercropping.

What we take away

Performance and underlying practices of potential positive deviants need to be understood within the country- and domain-specific contexts investigated. This requires a multidimensional analysis that considers socio-economic, cultural and structural factors that might affect the "positive" outcome—and that might do so differently across countries, regions or even communities. Specific know-how is needed for determining the right performance indicators and variables to look into, as well as for analyzing the data in more than a descriptive way.

We advise bringing data analysts and local domain experts on board early to move quickly from brainstorming and abstract conceptualization to iterative, integrated and grounded analyses.

Learning 4: **Scale with caution**

What we learned

The conventional Positive Deviance approach looks at small sample sizes in largely homogeneous contexts. For example, a few dozen families in a single village. This is very powerful for singling out and detailing particular behaviors that might explain successful practices, in part because non-behavioral factors can largely be disregarded since they are more or less the same for the entire observed population.

We initially assumed that by using large amounts of digital data, we would be able to simply transpose the principles of the conventional Positive Deviance approach to larger sample sizes and geographic areas. However, we learned that using such data implied greater heterogeneity and therefore, more possible confounds. For example, more geographic coverage meant a higher level of abstraction and an increased difficulty to determine whether "positive" outcomes were in fact due to certain practices, or whether they were simply due to a combination of contextual factors.

How we learned it

In Ecuador, we used remote sensing data to calculate the deforestation rates of a large sample of cattle raising farms. However, farms were of all shapes and sizes, and we found that cattle density was a likely a driver of deforestation. Indeed, a higher cattle density could put pressure on farmers to expand their pasture. We had to control for this confounding factor to ensure a fair and meaningful comparison between farms, and to be able in the end, to attribute low deforestation rates to sustainable cattle farming practices—not just to an imbalance in the number of animals and the size of the land.

Unfortunately, we identified additional confounding factors that we were not able to control for, as the data did not exist, or because we were not able to access it. For example, the type of farm production system, whether beef or dairy, or the fact that some farmers rented out parts of their land.

What we take away

Contextual factors, from access to roads and other infrastructure, to levels of rainfall and hours of sunlight, differ significantly across places and populations, and these factors are sometimes sufficient for explaining the differences in performance among individuals or communities.

To increase the likelyhood of identifying "true" positive deviants at the scale afforded by digital data, we advise to choose groups with similar contextual factors that might affect performance, while controlling for confounding factors and/or applying intermediary techniques like homogeneous grouping. These strategies can help ensure that identification is based on relative performance, not absolute performance. Doing so increases the chances that the "positive" outcomes are due to individual- or group-level factors that are not structural or contextual, meaning, they can more likely be transferred and amplified within the communities of identified positive deviants.

Learning 5: **Look at deviance over time**

What we learned

Digital data is often collected at regular intervals over long periods of time. This opens the possibility for longitudinal observations of positive deviance, and of variations or consistency in deviance.

In most pilots, we were able to distinguish between individuals or groups who became positive deviants (from low performing to high performing), who remained positive deviants, and who stopped being positive deviants (from high performing to low performing). We learned that consistent positive deviants over several years of data were more likely to be "true" positive deviants.

How we learned it

In Ecuador, Niger and Somalia we conducted longitudinal investigations to identify positive deviants that were able to sustain their relative advantage over time. After engaging with local domain experts in Ecuador, for instance, we found that the average cattle lifecycle (from farm to fork) was three years. It wouldn't be possible to judge whether a farm had reached its maximum cattle density (triggering deforestation), while still preserving forest proportion, without analyzing deforestation rates throughout this lifecycle. "True" positive deviants were thus farms that were able to maintain low deforestation rates for three consecutive years.

What we take away

Being able to identify who is a positive deviant over time, who becomes one, or who stops being one can inform on the determinants of high performance, but also on factors that might reduce positive deviance or evoke any other transition from one deviant state to another. Further, we believe that by observing contextual factors and performance over time, it will be possible to better understand the sustainability of positively deviant practices, as well as their interaction with other extrinsic factors (not indvidual or group practices or behavior).

Learning 6: **Look beyond individual or community practices and behavior**

What we learned

Digital data, and more generally secondary data, can inform on the activities and behaviors of human beings, but also on the environments and settings in which these take place—whether the urban infrastructure, the public service systems or the policy and regulatory mechanisms that govern them. The notion of positive deviance in the DPPD method is therefore not necessarily limited to human practices or behavior.

Throughout the initial design stages of our pilots, we sought to ensure that differences in performance could be traced back to positively deviant practices. We controlled for non-behavioral factors, such as a more suitable climate for agriculture or greater financial endowment. However, we learned that "positive" outcomes often depend on a range of extrinsic factors such as programmatic interventions, infrastructure or government policies. Identifying these factors and understanding how they enable or contribute to positive deviance can provide a suitable basis to design nuanced interventions that take these interactions into account and thus increase their possible effectiveness and contextual fit.

How we learned it

In both our pilots in Somalia and Mexico, we conducted initial data analyses and held conversations with local domain experts to map a range of factors at individual and systemic levels that likely interact and enable specific practices.

In Somalia, we learned that preserving rangelands required good practice at a community level—for example, for soil and water conservation—as well as at a higher, government-level—for land tenure policies and campaigns against private enclosures. Understanding these different factors and how they worked together turned out to be crucial for assessing the conditions in which communities lived—and especially the challenges they faced—the resources they were able to access, as well as other factors that might have inhibited or enabled the emergence of positive deviance.

In Mexico City, we focused directly on positively deviant public spaces that had lower crime rates than others, instead of trying to identify individuals or communities that might have been less prone to act out violently against women and girls. We had to work with the limitations of reported acts of violence—lower reports do not necessarily mean lower incidences of violence—but we controlled for a range of variables that were likely to contribute to the safety of public spaces. Ultimately, we were less concerned with individual behaviours and more with how certain features of the physical environment, for example lighting, the existence of alarm buttons or economic activity, might have led to the emergence of positive deviance.

What we take away

Seeing that the notion of positive deviance in the DPPD method can go beyond the behavior of individuals or communities, we recommend that DPPD interventions focus on more than just adapting practices. For example, an increase in crop yield can be due to applying an existing fertilizer in the "right" way—a practice that can indeed be identified and scaled. However, it can also be due to the introduction of a new fertilizer—a non-behavioral factor that can still be scaled. Even non-behavioral factors that cannot be influenced like the level of rainfall can be compensated, for instance through the usage of irrigation techniques. Further, if positive deviants are non-human entities like public spaces, we advise to focus on how structural attributes like infrastructure or governance mechanisms affect their "performance". Note that this will also affect the selection of data sources and types.

Ultimately, we believe the DPPD method is a good tool for addressing complex development challenges, as it enables the identification of both behavioral and non-behavioral factors that contribute to a "positive" outcome, and thereby provides a suitable basis for the design of nuanced interventions that seek to "influence" multiple factors simultaneously.

Conclusion

This report presents six learnings from the application of the Data Powered Positive Deviance (DPPD) method in four pilots conducted across different countries and domains (Table 3). The learnings revolve mainly around the access to, and use of digital data.

Data access is seldom as straightforward as one might hope, even for large organizations. We advise DPPD practitioners to remain flexible when framing the problem they seek to address if data is unavailable. However, we strongly recommend not to compromise the main purpose of the work (Learning 1). We also encourage practitioners to forge data partnerships at an early stage, and to leverage existing partnerships as much as possible (Learning 2).

In addition, the DPPD method requires a unique set of competencies to be able to deal with the necessary blend of traditional and digital data sources (Learning 3). Practitioners should bring together multi-disciplinary teams, and not shy away from technical problems even in the early stages of project designs. Indeed, while digital data can help reveal performance at a large scale, it requires a lot of technical and contextual knowledge to control for possible confounding factors. Some confounds can be captured remotely—for example, climatic factors—while others require different types of data—possibly more traditional sources—to represent demographic and socio-economic conditions (Learning 4).

An important benefit of the DPPD method is that the large temporal coverage of digital data enables the identification of both sustained positively deviant practices and changes in performance over time—for example, from positive to negative deviance, or vice versa (Learning 5). Finally, we believe DPPD is likely to be useful for investigating both the behavioral and structural factors behind local "solutions" to development problems. Consequently, we expect that the early findings of our pilots will trigger both community-level and policy-level interventions (Learning 6).

That said, our analyses thus far have only provided signals as to where positive deviants are likely to be. All pilot teams are now conducting fieldwork to validate whether deviants identified in the data are indeed "true" deviants, and to further understand the factors underlying the different positive deviances.

Summary of Learnings

	Learning	Illustration	Takeaway
Remain flexible in the face of data unavailability	All teams experienced difficulties in accessing data. They had to adjust their pilots according to what was available, accessible and reliable, while maintaining their original focus.	In Somalia, we pivoted the pilot away from enumerating livestock to measuring the health of community rangelands, while preserving the general focus on pastoral livelihoods.	Remain flexible and creative in dealing with data unavailability, and be open to adjusting the original project design. Also, keep an eye out for proxy measures in data.
Leverage existing partnerships for data access	Applying the DPPD method required access to very specific data from a number of different sources. Integration with pre-existing programmes and data partnerships significantly accelerated the process.	Our pilot in Ecuador was integrated early on within the ProAmazonia programme. This enabled access to more granular and usable data.	When scoping for data partnerships, keep in mind that data sources need to overlap spatially and temporally. They also need to be sufficiently granular to develop valuable insights.
Map and fill know-how gaps early	Identifying potential positive deviants required very specific know-how to select the right performance indicators and contextual variables to avoid false conclusions.	In Niger, including two analysts with data expertise and domain knowledge proved very helpful for assessing data availability, which sources were most adequate, as well as appropriate performance indicators and sampling strategies.	Team up with data analysts and local domain experts as early as possible to advance any DPPD project.

	Learning	Illustration	Takeaway
Scale with caution	Covering large geographic areas made it difficult to determine whether “positive” outcomes were due to practices or to other extrinsic factors (confounds).	In Ecuador, we identified cattle density as an important confounding factor, which we controlled for.	Compare only groups with similar contextual attributes that are likely to affect performance, and control for possible confounding factors.
Look at deviance over time	By using longitudinal data, we identified positive deviance that persisted over time, as well as deviance that shifted from positive to negative, or vice versa.	In Ecuador, we discovered that the average cattle lifecycle was three years. We analyzed deforestation rates for that time interval.	Look for consistent positive deviants in the data over long periods of time, as these are more likely to be “true” positive deviants.
Look beyond individual or community practices and behavior	“Positive” outcomes often depended on a range of factors, beyond practices and behavior.	In Somalia, we realized that preserving rangelands required behavioral and non-behavioral factors. The latter were less related to individual behavior and more to government interventions.	Look for behavioral and non-behavioral factors that impact positive deviance. We believe identifying these factors and understanding how they contribute to positive deviance should help design nuanced development interventions.

Table 3. Summary of learnings, illustrations and takeaways.

About

The DPPD Initiative

The Data Powered Positive Deviance initiative is a global collaboration between the GIZ Data Lab, the UNDP Accelerator Labs Network, the University of Manchester Centre for Digital Development, and UN Global Pulse Lab Jakarta. It is dedicated to finding, understanding and scaling local solutions to development challenges through the use of digitally recorded, readily available quantitative and qualitative data.

Resources

Data Powered Positive Deviance on Medium:
https://dppd.medium.com/
GIZ Data Lab Blog:
https://www.blog-datalab.com/
UNDP Accelerator Labs on Medium:
https://acclabs.medium.com/

Contact

Catherine Vogel (GIZ Data Lab):
catherine.vogel@giz.de
Andreas Glücker (GIZ Data Lab):
andreas.gluecker@giz.de
Jeremy Boy (UNDP Accelerator Labs):
jeremy.boy@undp.org
Basma Albanna (University of Manchester):
basma.albanna@manchester.ac.uk